Individual Thesis

COMPUTER-AIDED DETECTION OF GASTROINTESTINAL POLYPS: ADVANCING ENDOSCOPIC DIAGNOSTICS WITH DEEP LEARNING

Ioannis Rodosthenous

UNIVERSITY OF CYPRUS

COMPUTER SCIENCE DEPARTMENT

May 2024

UNIVERSITY OF CYPRUS

COMPUTER SCIENCE DEPARTMENT

COMPUTER-AIDED DETECTION OF GASTROINTESTINAL POLYPS: ADVANCING ENDOSCOPIC DIAGNOSTICS WITH DEEP LEARNING

Ioannis Rodosthenous

Supervisor
Constantinos Pattichis

The Individual Thesis was submitted in partial fulfillment of the requirements for obtaining the Informatics degree of the Department of Informatics of the University of Cyprus

May 2024

Acknowledgements

I would like to express my deepest thanks to my supervising professor, Mr. Konstantinos Pattichis, who has been a pillar of support and inspiration throughout the preparation of this thesis. His dedication, persistence and passion for the resulting research have been an invaluable source of drive and inspiration for me.

The valuable guidance, advice and criticism of Mr. Pattichis were instrumental in the completion of this work. The confidence he showed in my abilities and his constant support, even in the most difficult moments, were elements in the implementation of the objectives of this work.

In addition to his scientific guidance, Mr. Konstantinos Pattichis has inspired through the courses attended during the previous years of my degree, teaching with the value of continuous learning, research curiosity and academic integrity.

The contribution of all these factors was decisive for the successful completion of my thesis.

Summary

This thesis deals with the development and application of a Computer Aided Detection system for the identification and accurate placement of polyps in endoscopy images. The work focuses on the use of the YOLOv8 algorithm, a deep learning model for real-time object detection, adapted to the specific challenging need of polyp detection with high accuracy.

First, an extensive review of the relevant literature is presented, highlighting the challenges encountered in the automated detection of polyps through endoscopic images and the methodologies that have been applied to date. Then, the methodology adopted is explained in detail, describing the dataset preparation, the architecture of the YOLOv8 model, as well as the techniques used to train and optimize it.

Applying the model to the collected dataset revealed impressive results, where the system was able to detect and place the polyps with a high accuracy rate. This was confirmed through comparative evaluation with other methods, making the proposed approach promising for further research and development in the field of endoscopic analysis and diagnosis.

Finally, the paper closes with the presentation of conclusions, indicating the importance of incorporating deep learning technologies into the endoscopic diagnostic process. At the same time, directions for future research are proposed, with the main goal of further improving the accuracy of the system, the speed of analysis and its extensibility to more endoscopic procedures.

Contents

Chapter 1

Introduction

1.1 Endoscopy and Electronics Detection Assistance Systems

Endoscopy is considered the gold standard for diagnosing pathological conditions within the human body, offering superior visual access to internal tissues and organs. Despite the advanced visual imaging it offers, the interpretation of the images requires considerable experience and expertise from the physician. While advances in endoscopic technology have improved the quality of images, the need for increased diagnostic accuracy is driving the development of Computer Aided Diagnostic (CAD) systems.

These systems incorporate advanced algorithms and artificial intelligence techniques for the automated detection and classification of pathological conditions such as polyps. The use of color texture analysis and other related features from endoscopic images can improve the accuracy of diagnosis and provide a more objective basis for their interpretation.

The importance of CAD systems is critical, especially in the early detection of polyps, where early diagnosis can lead to a better therapeutic approach and prognosis for the patient. The integration of CAD systems into clinical practice requires further scientific research and technological development, with the aim of improving their sensitivity and specificity, as well as their integration into existing endoscopic systems.

In summary, the addition of electronically assisted diagnostic systems to the endoscopic procedure offers an important tool to increase diagnostic accuracy, improving quality of care and patient safety.

1.2 Objectives of the Thesis

My thesis has as its main objective the development and implementation of a specialized Electronic Aided Diagnostic System (CAD) using the YOLOv8 deep learning model, for the accurate detection and diagnosis of polyps in endoscopic images and videos. Leveraging the advanced capabilities of YOLOv8, the work aims to improve the accuracy and efficiency of endoscopic examinations. The main objectives of the work include:

1. **Development and Adaptation of the YOLOv8 Model:** Development and adaptation of YOLOv8 for the identification and accurate placement of polyps, both in endoscopic images and videos.
2. **Advanced Analysis and Processing of Images and Videos:** Application of analysis and processing techniques to optimize the quality of images and videos, as well as to enhance the discrimination ability of the system.
3. **Improving Diagnostic Accuracy:** Achieving high levels of accuracy in the detection and diagnosis of polyps, reducing the margin of diagnostic error.
4. **Performance Evaluation in Real Conditions:** Test and evaluate the performance of the model on real clinical data, including images and videos, comparing its performance with traditional diagnostic methods.

By achieving these goals, the thesis aims to contribute to enhancing the ability of physicians to identify polyps more accurately and efficiently, thereby improving the quality of care and patient safety.

1.3 Review of Thesis

My thesis consists of six main chapters, through which the use of deep learning technology for the detection of polyps in endoscopic images and videos is explored, with a special emphasis on the application of the YOLOv8 model.

Chapter 1: Introduction

It presents the need for automated polyp detection, the objectives of the work and a brief overview of its content.

Chapter 2: Theoretical Background and Related Bibliography

It examines the technological and theoretical underpinnings underpinning the work, as well as a review of the relevant literature.

Chapter 3: Methodology

It describes the technical approach, tools and methods used to develop and train the YOLOv8 model.

Chapter 4: Experimental Data and Results

It presents the collected data, the application of the model to it, and the analysis of the diagnostic results.

Chapter 5: Comparison and Evaluation

It analyzes the efficiency and accuracy of the model compared to other methods and previous approaches.

Chapter 6: Conclusions and Perspectives

It draws conclusions based on the research findings and suggests future directions for further research and development.

The thesis is a comprehensive effort to demonstrate how modern deep learning technologies can enhance the accuracy and efficiency of polyp detection, contributing to the improvement of the diagnostic process and the quality of life of patients.

Chapter 2

Theoretical Background and Related Bibliography

2.1 Literature Review of Related Bibliograph for Detection Systems in Endoscopy

Endoscopy is the main technical pillar for the early diagnosis and preventive control of digestive cancer. In recent years, the application of artificial intelligence (AI) in the field of endoscopy has made impressive strides, with AI systems offering real-time assistance in the detection and classification of gastrointestinal pathologies, expected to improve exam quality and reduce diagnosis finding rates.

AI systems in endoscopy range from classification, detection to segmentation of images, with the aim of more accurate diagnosis and treatment of diseases.

Classification is a machine learning task for determining which classes are in an image, video, or other types of data. It refers to training machine learning models with the intent of finding out which classes are present.

In clinical applications, it is possible to classify colorectal polyps in endoscopic image from a patient into different categories, such as non-adenomatous polyps, adenomatous polyps and cancerous. Different categories would need specific treatments. And with the assistance of classification results, clinician could make more accurate diagnosis. It is also possible to evaluate the image quality of all endoscopic images of the patient, like categorization of bowel cleanliness, from which the quality of the image should meet the diagnostic quality requirements.

Object detection combines classification and localization to determine what objects are in the image or video and specify where they are in the image. Generally, bounding boxes are used to distinguish objects in video frames or images.
In clinical applications, it is possible to detect different findings in colonoscopy for different purposes, such as polyps, angiectasia, bleeding, inflammations, esophagitis, ulcerative colitis, pylorus, cecum, dyed polyp, dyed resection margins and stool. Specifically, polyp detection is the most usual AI application in colonoscopy. Detection for polyps can effectively reduce the polyp missing rate in colorectal screening, which would further reduce the adenomatous missing rate.

Image segmentation separates an image into regions on pixel level, with shape and border, delineating potentially meaningful areas for further processing, such as measurement, classification, and object detection. The regions may not take up the entire image, but the goal of segmentation is to highlight foreground elements and make it easier to be evaluated. Image segmentation provides pixel-by-pixel details of an object, distinguishing it from classification and object detection.
For example, in endoscopy, the polyp size can be automatically calculated based on the segmentation results, while the polyp size is one of the key factors for polyp diagnosis by clinician.

The literature review reveals that AI systems for endoscopy have the potential to increase polyp detection rates, improve pathology categorization accuracy, and provide valuable information on polyp size and morphology through segmentation. Despite promising progress, the application of AI in endoscopy faces significant challenges, such as the need to further improve the accuracy of systems, integrate them into clinical practice, and ensure data security and patient privacy.

An important part of the review concerns the creation and use of standards for evaluating and comparing AI systems, highlighting the need for standardized performance evaluation of such systems. Collaboration between clinical researchers, software engineers, and regulatory agencies is critical to the development and adoption of these advanced technologies. Building an extensive and diverse data set from real clinical cases and developing benchmarking systems that can evaluate the performance of AI models in a variety of clinical settings emerge as key steps towards achieving this goal.

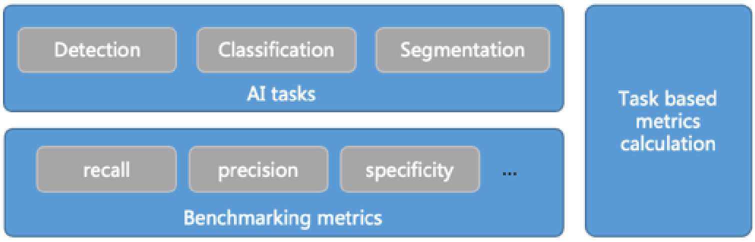


Figure 1: Architecture of benchmarking version

Furthermore, the bibliograph highlights the importance of the ethical and legal challenges involved in the use of AI in medical practice. Respect for patient privacy, transparency of algorithms and ensuring the objectivity and accuracy of diagnoses are fundamental issues that must be addressed responsibly.

In conclusion, the literature review indicates that, despite the challenges, the integration of AI into endoscopy offers a significant opportunity to improve the quality of examinations and patient care. With proper attention to the development, evaluation, and implementation of AI systems, this technology can help discover early signs of disease, reduce misdiagnosis rates, and provide physicians with a valuable tool to improve efficiency and accuracy. of their clinical decisions. With the continuous advancement of deep learning and image processing technologies, AI endoscopy systems will continue to evolve, providing even more advanced capabilities for the prevention, detection, and treatment of gastrointestinal diseases.

It is important to note that the success of integrating AI into clinical practice depends on interdisciplinary collaboration between technologists, physicians, bioethicists, and legal experts. Only through this collective effort can the challenges be adequately addressed, and the opportunities provided by artificial intelligence to improve digestive health be fully exploited.

A review of the bibliograph on detection systems in endoscopy shows a dynamic field in continuous development, where research and innovation can lead to significant improvements in the prevention, diagnosis and treatment of digestive diseases. Continuous updating of clinical protocols and training of healthcare professionals in these new technologies is critical to harnessing the full potential of AI in endoscopy [1].

2.2 References to other Detection Systems & what the needs are

This section provides an overview of the current Artificial Intelligence (AI) solutions that are making significant strides towards commercialization and clinical application in the field of polyps' detection and generally in colonoscopy. These AI-driven systems enhance the detection capabilities during endoscopic examinations and are approved by regulatory bodies including the Chinese National Medical Products Administration (NMPA), the U.S. Food and Drug Administration (FDA), and European CE standards.

2.2.1 Existing Detection Systems in Polyp Detection

- **Tencent Healthcare & Changhai Hospital (2021)**: Developed a Computer-Aided Detection (CAD) system utilizing the "You Only Look Once" v2 framework. This system aids colonoscopists by highlighting potential polyps with a visual alert on a secondary monitor, significantly improving the detection rates of diminutive and flat polyps. The system received NMPA certification in June 2023.

- **National Cancer Center Hospital & NEC Japan (2017)**: Created an AI system that identifies colorectal cancer and ulcerative colon polyps during endoscopic exams. This system automates the detection process from images and videos, enhancing the identification of critical lesions and boosting polyp detection rates, which is crucial for the early prevention of colorectal cancer.
- **Wision AI & Sichuan Provincial People's Hospital (2018)**: Developed the EndoScreener®, a real-time automatic polyp detection system using deep learning techniques based on the SegNet architecture. This system has shown to significantly increase the detection of diminutive adenomas and hyperplastic polyps in populations with low adenoma detection rates. The system is certified by NMPA, FDA, and European CE-MDR.
- **National Chiao Tung University & Tri-Service General Hospital (2018)**: Introduced a Computer-Aided Diagnosis (CADx) system that uses deep neural networks to classify diminutive colorectal polyps in narrow-band imaging as neoplastic or hyperplastic, aiding in more accurate clinical assessments.
- **Sun Yat-sen University (2018)**: Developed the Gastrointestinal Artificial Intelligence Diagnostic System (GRAIDS), the first real-time AI-assisted system implemented in clinical practice for detecting upper gastrointestinal cancers during endoscopy.
- **Zhongshan Hospital & University of California**: Utilized a transfer learning approach with the ResNet50 architecture to create a CNN-CAD system for assessing the invasion depth of gastric cancer, which helps in screening patients for endoscopic resection and minimizing unnecessary surgeries.
- **Cancer Institute Hospital Ariake, AI Medical Service & Tada Tomohiro Institute of Gastroenterology and Proctology**: Employed a CNN based on Single Shot MultiBox Detector architecture for rapid and clinically relevant diagnosis of gastric cancer from endoscopic images.
- **Renmin Hospital of Wuhan University**: Developed a novel DCNN to detect early gastric cancer (EGC) during esophagogastroduodenoscopy (EGD), capable of classifying gastric locations into detailed segments with high accuracy.

- **Kindai University (2017)**: Implemented a simple CNN that diagnoses colon polyps as adenomatous or non-adenomatous, enhancing diagnostic precision.
- **Wuhan ENDOANGEL Medical Technology Co., LTD**: Introduced EndoAngel®, an AI system that combines polyp detection with quality monitoring functions, certified by the Chinese NMPA in May 2023. This system not only highlights polyp locations but also monitors procedural metrics to enhance overall examination quality.

These AI solutions demonstrate the potential of advanced machine learning models to revolutionize the field of endoscopy by providing more accurate, dependable, and efficient diagnostic capabilities, thereby improving patient outcomes, and streamlining clinical workflows [1].

2.2.2 Identified Needs for Advancing Polyp Detection

While colonoscopy is the gold standard for detecting colorectal polyps, critical challenges remain that AI technologies aim to overcome. The integration of AI is particularly aimed at enhancing the detection and diagnostic processes for polyps. The specific needs include:

1. **Enhancing Polyp Detection Accuracy:**
   - **Challenge**: Traditional colonoscopy has a significant miss rate, failing to detect 17%-48% of adenomas, which are precursors to colorectal cancer.
   - **Need**: Improve the detection accuracy of small and flat polyps, which are easily missed, by employing advanced AI systems like CADe that can provide real-time alerts and visual aids during screenings.

2. **Consistency in Detection:**
   - **Challenge**: Variability in polyp detection due to the subjective nature of manual screenings performed by colonoscopists with differing levels of experience.

- **Need**: Implement AI solutions such as CADx that standardize polyp detection across different operators by providing consistent, real-time analytical support during procedures.

3. **Objective Standardization of Diagnosis:**
   - **Challenge**: Diagnosis of polyps is often subject to inter-observer variability, which can affect treatment decisions.
   - **Need**: Develop AI-driven diagnostic systems that use objective data from pathology reports and deep learning models to standardize classifications of polyp types (e.g., neoplastic vs. hyperplastic).
4. **Quality of Training Data:**
   - **Challenge**: AI models require accurately labelled, high-quality image datasets for training, which are challenging to compile.
   - **Need**: Establish comprehensive datasets annotated through objective methods based on clinical reports or through subjective assessments by experienced endoscopists to train AI models effectively.
5. **Real-Time Educational Support:**
   - **Challenge**: Novice endoscopists may have difficulty achieving high detection rates.
   - **Need**: Utilize AI to develop simulation-based training and real-time guidance systems to enhance the skill set of endoscopists, thereby improving their detection rates and diagnostic accuracy.
6. **Seamless Integration into Clinical Practice:**
   - **Challenge**: Introducing AI tools into existing medical workflows can be disruptive and meet resistance.
   - **Need**: Design AI tools that integrate smoothly with existing endoscopic equipment and clinical protocols, providing user-friendly interfaces that require minimal adjustments to current practices.

By meeting these needs, AI applications dedicated to polyp detection can significantly advance the efficacy of screenings, aiding in the early detection and appropriate management of colorectal cancer precursors. This technological advancement will not only improve patient outcomes but also streamline and enhance the efficiency of clinical endoscopic practices.

2.3 Long-Term Impact and Cost-Effectiveness of AI-Assisted Systems

This section explores the extended performance and financial viability of AI-assisted colonoscopy systems through a comprehensive year-long study conducted at a single institution in Singapore. The study leverages advanced AI technologies to enhance the detection rates of adenomas and assesses the economic implications of integrating such technologies into routine clinical practice.

Background and Study Overview:

Colonoscopy, recognized as the gold standard for detecting pre-malignant neoplastic lesions in the colon, has seen substantial advancements with the integration of Artificial Intelligence (AI). A previous short-term study indicated that AI-aided colonoscopy significantly improved both collective and individual adenoma detection rates. Building on this, the current study evaluates the one-year performance of the AI-aided colonoscopy using the GI Genius™ Intelligent Endoscopy Module.

Methods and Performance Metrics:

The prospective cohort study compared polypectomy rates between AI-aided and non-AI-aided colonoscopies. Key performance metrics such as polyp detection rate (PDR), adenoma detection rate (ADR), and adenoma detection per colonoscopy (ADPC) were meticulously calculated. Additionally, the study included a histological review of polypectomies performed following AI detections, specifically looking into the prevalence of sessile-serrated adenomas.

Results and Financial Analysis:

The AI-aided approach led to significantly higher polypectomy rates compared to the traditional method (33.6% vs. 28.4%, $p < 0.001$), demonstrating the AI's ability to enhance diagnostic accuracy over a sustained period. The detailed cost analysis revealed that the revenue generated from the increased polypectomy rates not only covered the AI subscription costs but also provided a substantial financial surplus.

Conclusion and Implications:

The findings affirm the cost-effectiveness and added value of AI in colonoscopy, suggesting that AI-assisted procedures could play a pivotal role in elevating the standards of colorectal cancer screening. This study underscores the potential for AI technologies to revolutionize diagnostic practices, making them more accurate, financially viable, and broadly accessible.

This sub-chapter highlights the importance of embracing innovative technologies in medical diagnostics and the transformative impact AI can have on enhancing patient outcomes and optimizing clinical operations [2].

Chapter 3

Methodology

3.1 Report and use of YOLOv8 in Object Detection

This sub-chapter describes the application of the YOLOv8 (You Only Look Once, version 8) deep learning architecture for the detection of polyps within endoscopy video footage and images. The objective is to leverage YOLOv8’s capabilities to improve the accuracy and speed of polyp detection compared to traditional existing methods and earlier AI models.

3.1.1 Introduction to YOLOv8

The YOLOv8, an acronym for "You Only Look Once Version 8," represents the most advanced version in the YOLO series of object detection models. This iteration synthesizes the strongest attributes of its predecessors while enhancing detection speed, accuracy, and effectiveness in complex environments.

As a specialized deep learning framework for object detection, YOLOv8 streamlines the process by employing a single-stage detection method. This approach eschews the multi-stage, computationally intensive procedures typical of traditional models, favoring a more streamlined operation. It utilizes a neural network that concurrently predicts class labels and object bounding boxes, which is particularly effective for analyzing multiple objects within a single image or video frame, thus suiting real-time processing needs.

A primary benefit of YOLOv8 is its exceptional processing speed, which facilitates real-time image and video analysis, even on devices with limited resources. This efficiency is partly due to its ability to conduct multi-scale predictions, which enables the detection of objects at various scales. Moreover, YOLOv8 maintains high detection accuracy, which ensures dependable outcomes.

The training of YOLOv8 on expansive datasets equips it to recognize a broad array of objects across different contexts, accurately classifying various categories such as humans, vehicles, and animals with refined precision. This capability renders YOLOv8 an invaluable asset across multiple industries, including autonomous driving, security surveillance, robotics, and retail analytics.

Additionally, it incorporates sophisticated techniques like anchor boxes and non-maximum suppression to boost its performance and minimize incorrect detections. These strategies enhance the precision of object localization and aid in the removal of redundant bounding boxes.

With its blend of rapid processing and precise accuracy, YOLOv8 has garnered significant acclaim within the deep learning community. Its adaptability and capacity for real-time application broaden the horizons for computer vision projects, enabling effective object identification in still images, video sequences, or direct webcam feeds, thus providing a robust and efficient solution for diverse object detection challenges [5].

3.1.2 Usage of YOLOv8 Object Detection

The process of custom object detection using YOLOv8 involves a series of crucial steps that I have undertaken to set up and train the model effectively. Ensuring accurate object detection requires meticulous preparation, beginning with the collection and annotation of high-quality data. Additionally, organizing the training data into appropriate folders and correctly configuring the config.yaml file are fundamental tasks in my workflow. I utilize the powerful Ultralytics library to facilitate model training and evaluation.

Step 1: Data Collection

At the outset of my custom object detection project, I focused on gathering a diverse and comprehensive dataset. I collected a wide array of images that depict objects of interest from various angles, under different lighting conditions, and showcasing multiple object variations. This approach helps in developing a robust training dataset that enhances the generalization capability of the model.

Step 2: Data Annotation

Once my dataset was assembled, I proceeded to annotate the images. Using data annotation tools such as Computer Vision Annotation Tool, I meticulously drew bounding boxes around the objects of interest in each image, based on their masks which point to the polyp area. This step is crucial as precise annotations significantly influence the training outcome of the model.

Step 3: Folder Setup

To maintain an organized workflow, I set up specific folders for images, annotations, and configuration files. Organizing my data systematically ensures that the training process runs smoothly without any data mismanagement issues.

Step 4: Configuring YAML File

The config.yaml file is critical as it contains essential details about the dataset paths, model hyperparameters, and other training parameters. I took special care to configure this file accurately, which is pivotal in optimizing the object detection results of the YOLOv8 model.

Step 5: YOLOv8 Training

Utilizing the Ultralytics library, I loaded the YOLOv8 model and initiated training on my annotated dataset. The training process allows the model to gradually learn and improve its capability to detect the specified objects, refining its accuracy and understanding of their distinct features as the epochs advance. Also, after training the first batch of data, of course I have tested the trained model on another batch of un-seen data, so that I can compare the effectiveness of polyp detection.

Step 6: Performance Evaluation

Post-training, evaluating the YOLOv8 model's performance was imperative. I assessed various metrics, such as the mean average precision (mAP), and scrutinized the model’s ability to accurately detect objects across different test images. Based on these evaluation results, fine-tuning and adjustments were made to the model to ensure its optimal functionality.

Through these detailed steps, I embarked on my journey into custom object detection using YOLOv8, exploring its robust capabilities in detecting objects not only in still images but also in video sequences. The methodology adopted has proven effective in enhancing detection accuracies and will be further explored in subsequent sections of this thesis [3][4][5].

3.1.3 YOLOv8 Object Detection in Images

As I have mentioned before, YOLOv8 demonstrates exceptional proficiency in detecting objects within static images. Utilizing advanced deep learning methodologies, it not only classifies and localizes objects with high precision but also assigns class labels and calculates the probabilities of each classification.

This functionality is immensely beneficial across various sectors. In surveillance, YOLOv8 helps in identifying potential security threats through camera feeds. Within the medical field, it aids in the detection of abnormalities in diagnostic images, enhancing early diagnosis and treatment options. In the retail sector, the model analyzes consumer behaviors by monitoring interactions and movements within a store environment.

Through its ability to efficiently process and analyze static images, YOLOv8 supports critical applications in these domains by providing detailed insights and enhancing operational effectiveness.

Image Classification with YOLOv8

One of the key functionalities of YOLOv8 in image detection is image classification. By analyzing the content of an image, YOLOv8 can effectively assign a specific label or class to each detected object. This process allows for comprehensive categorization of objects present in the image, enabling further analysis and decision-making based on the identified classes. For example, in a surveillance system, YOLOv8 can classify objects as 'person,' 'vehicle,' or 'polyp,' providing valuable information for security monitoring purposes.

Object Localization and Recognition

In addition to image classification, YOLOv8 performs object localization by accurately identifying the positions of objects within an image. This is achieved by drawing bounding boxes around the detected objects, indicating their precise locations. These bounding boxes encompass the objects, providing valuable visual representation and decisive information for subsequent analysis and action.

The ability of YOLOv8 to recognize objects is a crucial aspect of its object detection capabilities. By predicting the class labels and probabilities of each detected object, it allows for precise recognition and identification of specific objects within an image. For example, in a retail analytics scenario, it can recognize individual products on store shelves, enabling inventory management, customer behaviour analysis, and targeted marketing strategies.

The following table provides a comparison between YOLOv8 and traditional object detection methods in terms of image classification, object localization, and object recognition:

	Image Classification	Object Localization	Object Recognition
Traditional Methods	Requires manual annotation and feature extraction	Relies on separate algorithms for detection	May involve complex post-processing techniques
YOLOv8	Automatically classifies objects based on learned features	Accurately localizes objects with bounding boxes	Provides precise recognition and labeling of objects

Figure 2: YOLOv8 vs Traditional Methods

The table clearly highlights the advantages of YOLOv8 over traditional methods, showcasing its efficiency and effectiveness in image classification, object localization, and object recognition [3][4][5].

3.1.4 YOLOv8 Object Detection in Videos

I should also add that YOLOv8, besides a sophisticated object detection model, extends its capabilities beyond static image analysis to include video processing, facilitating real-time object detection. Unlike image detection, YOLOv8 treats each video frame independently, ensuring precise object detection throughout the video sequence.

A notable strength of YOLOv8 is its adaptability to various video formats. It efficiently processes a multitude of formats including MP4, AVI, and others, enabling seamless object detection across different media types, thus supporting extensive video analysis across varied applications.

Utilizing the video detection functionalities of YOLOv8 allows for in-depth analysis and valuable insight generation in sectors such as surveillance, autonomous driving, and retail analytics. The model not only identifies and localizes objects with high accuracy and speed but also maintains these qualities consistently across real-time video feeds.

Detailed Process of YOLOv8 Object Detection in Videos:

1) **Input:** A video file in any of the supported formats.
2) **Processing:** YOLOv8 decomposes the video into its constituent frames and analyzes each frame sequentially through its advanced neural network architecture.
3) **Object Detection:** For every frame, YOLOv8 identifies objects, categorizes them, and determines bounding box coordinates for effective localization.
4) **Output:** The processed video is outputted with bounding boxes marked around the detected objects, which facilitates user analysis by visually highlighting the objects present in each frame.

With the implementation of YOLOv8 for video object detection, numerous possibilities for video analysis and real-time operational decision-making are unlocked. Whether it is managing traffic systems, monitoring security footage, or analyzing consumer behavior in retail environments, YOLOv8 offers a robust tool for precise object detection and localization in dynamic video streams [3][4].

3.2 Analysis of Algorithm YOLOv8

YOLO v8's Working Principle

YOLO v8, an advancement of the foundational YOLO framework by Ultralytics, enhances its predecessor's design for heightened efficiency. The fundamental operation of YOLO v8 involves partitioning the incoming image into a grid, with dimensions such as 13x13 or 26x26 cells, determined by the model variant. Each cell of the grid is tasked with the detection of objects situated within its respective zone. The operational steps of YOLO v8 are outlined as mentioned before.

Variants of YOLO v8

The YOLO v8 suite includes a range of models tailored to fulfill distinct requirements and application contexts. Highlighted below are the primary variants along with their unique attributes:

YOLO v8-Tiny: Geared towards balancing the scales of speed and accuracy, the YOLO v8-Tiny utilizes a smaller network and reduced grid dimensions, such as 13x13, to facilitate swift performance on less capable hardware.

YOLO v8-SPP: The inclusion of a Spatial Pyramid Pooling (SPP) layer in this variant of YOLO v8 allows for the efficient extraction of features across various scales, bolstering the model's accuracy particularly in detecting objects of diverse sizes.

YOLO v8-CSPDarknet: By integrating the YOLO v8 algorithm with the CSPDarknet architecture, this version achieves enhanced levels of accuracy and performance. The implementation of cross-stage partial connections significantly enriches its feature representation capabilities, making it versatile across different use cases.

YOLO v8-Panet: Incorporating the Path Aggregation Network (PANET) into its structure, YOLO v8-Panet excels in improving the precision of feature integration and detection tasks. This variant is particularly effective where exact object positioning is critical [6].

Advantages of YOLO v8 Over Previous Versions

YOLO v8 stands as a formidable evolution in the realm of object detection technologies, surpassing its forerunners with significant enhancements. The following benefits underscore its advancement:

1) **Augmented Precision:** The introduction of models such as YOLO v8-CSPDarknet and YOLO v8-Panet delivers heightened accuracy, courtesy of their sophisticated mechanisms for representing and amalgamating features. This accuracy bolsters the reliability of YOLO v8 in intricate use areas like autonomous navigation and diagnostic imaging.

2) **Consistent Real-Time Efficiency:** YOLO v8 upholds the signature instantaneous processing capability of the YOLO series while simultaneously elevating the accuracy. This feature positions it as a prime solution for practical deployments needing prompt yet precise object detection, like security systems and robotic applications.

3) **Customizable Framework:** YOLO v8 presents a spectrum of tailored models, catering to diverse operational needs. From the resource-efficient YOLO v8-Tiny for embedded systems to the precision-oriented YOLO v8-CSPDarknet for cutting-edge research, YOLO v8 adapts to meet specific application requirements.

4) **Fortified Robustness:** The assimilation of progressive elements such as CSP (cross-stage partial) connections and the PANET structure within YOLO v8 variants imparts the model with sturdier resilience to varied hindrances, including clutter, complex backdrops, and challenging illumination conditions.

3.2.1 Analysis of Object Detection Performance Metrics

YOLOv8's Remarkable Advancements in Performance and Accuracy

The Ultralytics team benchmarked YOLOv8 against the COCO database and achieved impressive results compared to previous YOLO versions across all five model sizes [6].

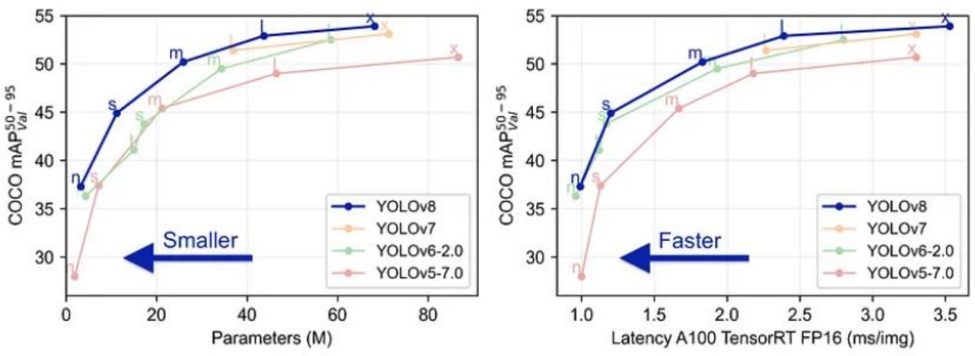


Figure 3: Comparative Performance Analysis of YOLO Versions: Model Complexity vs. Accuracy and Latency vs. Accuracy [6]

Graph 1: Model Size vs. Accuracy (Left)

- **X-Axis:** Represents the number of parameters in millions (M), which correlates with the size and complexity of the model.
- **Y-Axis:** Displays the COCO mAP^0.5:0.95, a metric for mean Average Precision across different IoU (Intersection over Union) thresholds from 0.5 to 0.95, commonly used in object detection to measure accuracy.
- **Trend Observed:** As the number of parameters increases, the accuracy (as measured by mAP) also generally increases, indicating that more complex models tend to perform better at object detection tasks. However, this comes at the cost of increased model size.
- **YOLOv8 Performance:** It appears that YOLOv8, indicated by the blue line, has the highest mAP across the entire range of parameters, suggesting that it outperforms the other versions (YOLOv7, YOLOv6.2.0, YOLOv5-7.0) in terms of accuracy. The 'Smaller' arrow indicates a movement towards models with

fewer parameters, where YOLOv8 maintains a lead in accuracy despite reduced complexity.

Graph 2: Latency vs. Accuracy (Right)

- **X-Axis:** Measures the latency in milliseconds per image (ms/img) on an A100 TensorRT FP16 platform, indicating the model's speed or inference time.
- **Y-Axis:** Again, shows the COCO mAP^0.5:0.95 for accuracy.
- **Trend Observed:** Typically, lower latency (faster models) results in lower accuracy. However, the goal is to find a model that provides a good balance of speed and accuracy.
- **YOLOv8 Performance:** In this graph, YOLOv8 also performs well, offering high accuracy even at lower latencies. This suggests that YOLOv8 is optimized for faster performance without a substantial sacrifice in accuracy. The 'Faster' arrow emphasizes the preference for models that achieve lower latency, and YOLOv8 leads in maintaining high accuracy at increased speeds.

Points 's', 'm', and 'x' on the Graphs:

These points indicate specific benchmarks or configurations within each YOLO version. For instance, 's' could stand for 'small', 'm' for 'medium', and 'x' for 'extra' or 'large', representing different model sizes or settings.

In summary, these graphs suggest that YOLOv8 is both smaller and faster than previous versions while achieving higher accuracy, making it a preferred choice for real-time object detection tasks where model efficiency and effectiveness are paramount.

When we compare the performance of different YOLO generations and model sizes in the COCO database, we want to compare different metrics.

- **Performance:** Mean average precision (mAP)
- **Speed:** Speed of the inference (In fps)

- **Compute (cost):** The size of the model in FLOPs and parameters.

For the object detection comparison of the 5 model sizes The YOLOv8m model achieved a mAP of 50.2% on the COCO dataset, whereas the largest model, YOLOv8x achieved 53.9% with more than double the number of parameters [6].

Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

- **mAP^{val}** values are for single-model single-scale on COCO val2017 dataset. Reproduce by `yolo val detect data=coco.yaml device=0`
- **Speed** averaged over COCO val images using an Amazon EC2 P4d instance. Reproduce by `yolo val detect data=coco128.yaml batch=1 device=0/cpu`

Figure 4: Performance Metrics Across Different YOLOv8 Model Variants

This table presents a comparison of performance metrics across various YOLOv8 model variants. Each model is evaluated on the same input size (640 pixels) but exhibits different performance characteristics in terms of mean Average Precision (mAP), processing speed, and computational requirements.

1. **Mean Average Precision (mAP):** The mAP^val 50-95 metric, which evaluates accuracy across different Intersection over Union (IoU) thresholds, shows incremental improvement as we move from the YOLOv8n variant to the more complex YOLOv8x. The higher mAP values suggest more accurate object detections by the advanced variants.

2. **Speed on CPU and GPU:** Speed is measured in milliseconds (ms) and reflects how quickly each model variant can process an image. As expected, there is a trade-off between accuracy and speed; more accurate models (like YOLOv8x) tend to be slower. The table provides speed metrics for both CPU (ONNX) and GPU (A100 TensorRT), indicating that all variants perform object detection tasks rapidly, suitable for real-time applications.
3. **Parameters and FLOPs:** The number of parameters (in millions) and Floating-Point Operations Per Second (FLOPs) (in billions) correlate to the complexity of each model. As the complexity increases, both the number of parameters and FLOPs rise, indicating more computational power is needed for the more accurate models.

The YOLOv8s model balances speed and accuracy better than the lightweight YOLOv8n but at a cost of increased computational requirements. On the other end, YOLOv8l and YOLOv8x, while having the highest accuracy, also require significantly more computational resources, as evidenced by their higher FLOPs, and exhibit the slowest processing times, making them less suitable for very time-sensitive applications.

Performance Metrics

Performance metrics play a vital role in assessing the precision and speed of object detection models. They provide insight into the model's ability to recognize and pinpoint objects in images, as well as its management of false positives and false negatives. Understanding these metrics is essential for evaluating and improving model performance. We'll delve into the different performance metrics linked with YOLOv8, their importance, and how to analyze them effectively [7].

Object Detection Metrics

Let's begin by examining some metrics that are not only crucial for YOLOv8 but are also widely applicable across various object detection models.

- **Intersection over Union (IoU):** IoU measures the overlap between a predicted bounding box and a ground truth bounding box, serving as a fundamental metric for evaluating object localization accuracy.

- **Average Precision (AP):** AP calculates the area under the precision-recall curve, offering a consolidated value that reflects the model's precision and recall performance.

- **Mean Average Precision (mAP):** mAP expands on AP by averaging the AP values across multiple object classes. This is particularly beneficial in multi-class object detection scenarios for a comprehensive evaluation of model performance.

- **Precision and Recall:** Precision evaluates the ratio of true positives among all positive predictions, indicating the model's ability to minimize false positives. Meanwhile, Recall assesses the ratio of true positives among all actual positives, gauging the model's capacity to detect all instances of a class.

- **F1 Score:** The F1 Score, being the harmonic mean of precision and recall, provides a balanced assessment of model performance by considering both false positives and false negatives.

3.2.2 How to Calculate Metrics for YOLOv8 Model

Model Validation with Ultralytics YOLO

The validation phase is an essential component of the machine learning workflow, offering a means to evaluate the efficacy of trained models. The Val mode in Ultralytics YOLOv8 offers a comprehensive set of features and performance metrics crucial for the assessment of object detection models. This manual is designed to furnish a thorough understanding of the Val mode's utilization, ensuring the precision and dependability of your models [8].

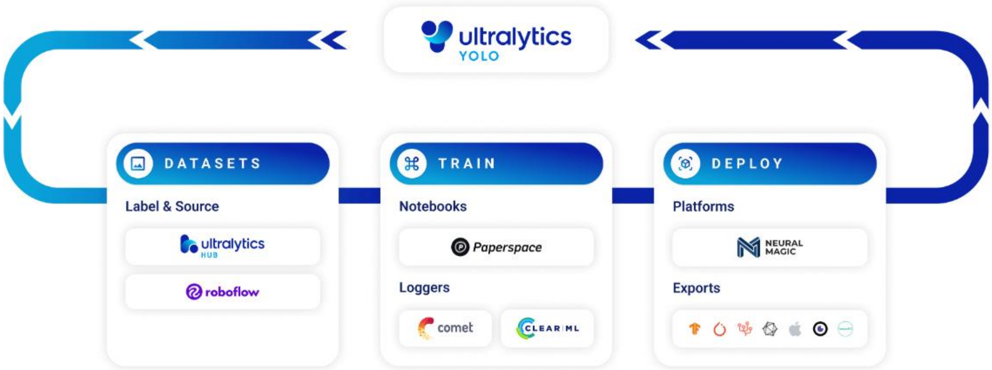


Figure 5: Ultralytics YOLO Workflow [8]

Key Features of Val Mode

These are the notable functionalities offered by YOLOv8's Val mode:

- **Automated Settings:** Models remember their training configurations for straightforward validation.
- **Multi-Metric Support:** Evaluate your model based on a range of accuracy metrics.
- **CLI and Python API:** Choose from command-line interface or Python API based on your preference for validation.
- **Data Compatibility:** Works seamlessly with datasets used during the training phase as well as custom datasets.

Using the validation mode is simple. Once you have a trained model, you can invoke the model.val() function. This function will then process the validation dataset and return a variety of performance metrics. But what do these metrics mean? And how should you interpret them? [8]

Interpreting the Output

Diving into the model.val() function's output allows us to comprehensively understand the nuances of the model's performance.

Class-Specific Performance Indicators

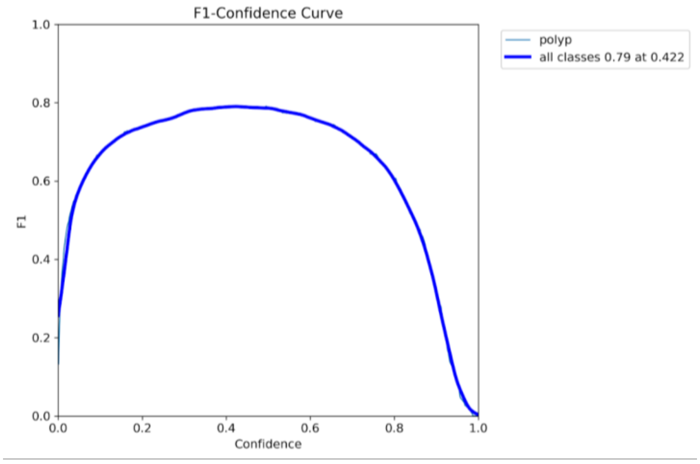
A key element of the output is the detailed performance metrics for each object class. This detailed data is invaluable for assessing the model's efficacy with each class, which is crucial for datasets encompassing a wide variety of object categories. The output delineates the following for each class: [8]

- **Class Identifier:** Indicates the name of the object class being evaluated, such as "person," "car," or "polyp".
- **Image Count:** Shows the number of images within the validation dataset that feature the class in question.
- **Instance Frequency:** Details the total occurrences of the class across the entire validation image set.
- **Detection Metrics (Precision, Recall, mAP50, mAP50-95):** These statistics provide deeper insights into the model’s detection capabilities:
  - **Precision (P):** Reflects the proportion of accurate detections, shedding light on the model's exactness.
  - **Recall (R):** Captures the model's proficiency in identifying all occurrences of the class within the images.
  - **mAP50:** Represents the mean average precision at an intersection over union (IoU) threshold of 0.50, highlighting the model's precision in more straightforward detection scenarios.
  - **mAP50-95:** Offers a cumulative perspective of the mean average precision across a spectrum of IoU thresholds from 0.50 to 0.95, providing a balanced assessment of the model’s performance for a range of detection challenges.

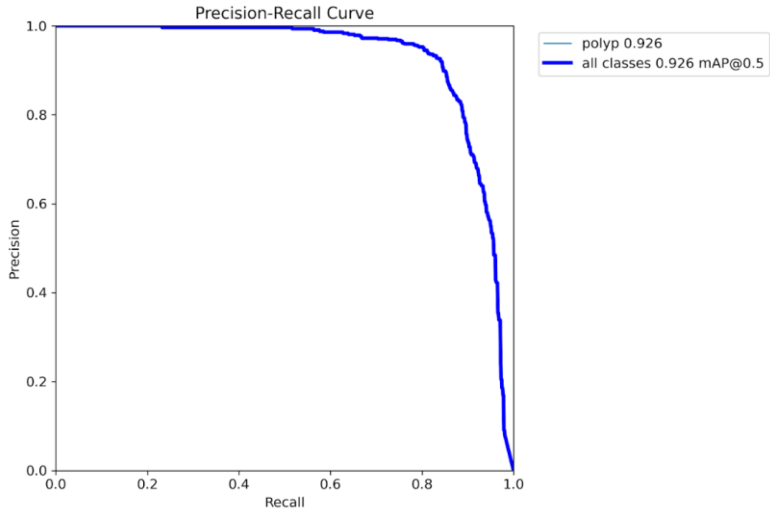
Visual Outputs

The model.val() function's output extends beyond quantitative metrics to include visual representations, which offer a more tangible grasp of the model's performance. Below is an outline of the visual feedback you can anticipate (all graphs are from my results, and they are not going to be analysed here but on the next chapter):

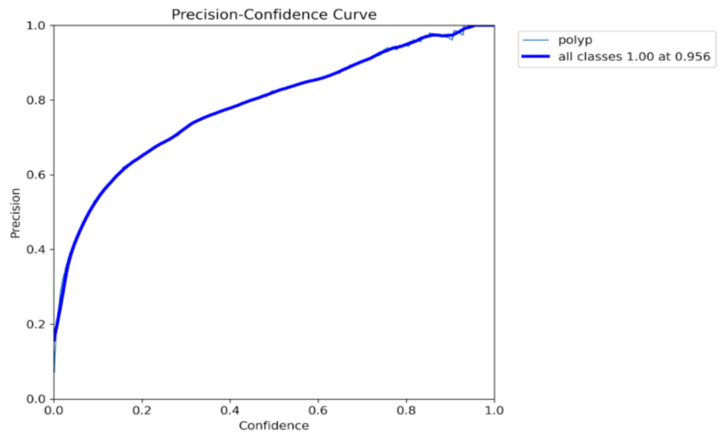
- **F1 Score Graph (F1_curve.png):** This plot delineates the F1 score at diverse threshold levels. Analyzing this graph can reveal the equilibrium the model maintains between false positives and false negatives under various conditions.



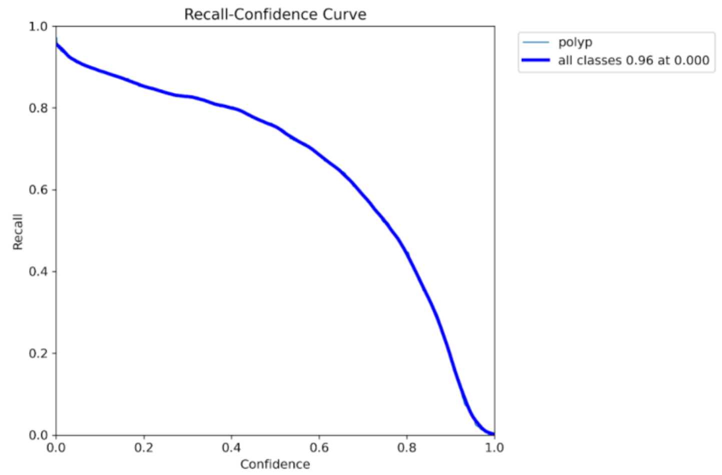
- **Precision-Recall Trade-off (PR_curve.png):** A crucial plot for classification tasks, this graph delineates the compromise between precision and recall across an array of threshold values. Its significance is amplified in scenarios where class distribution is skewed.



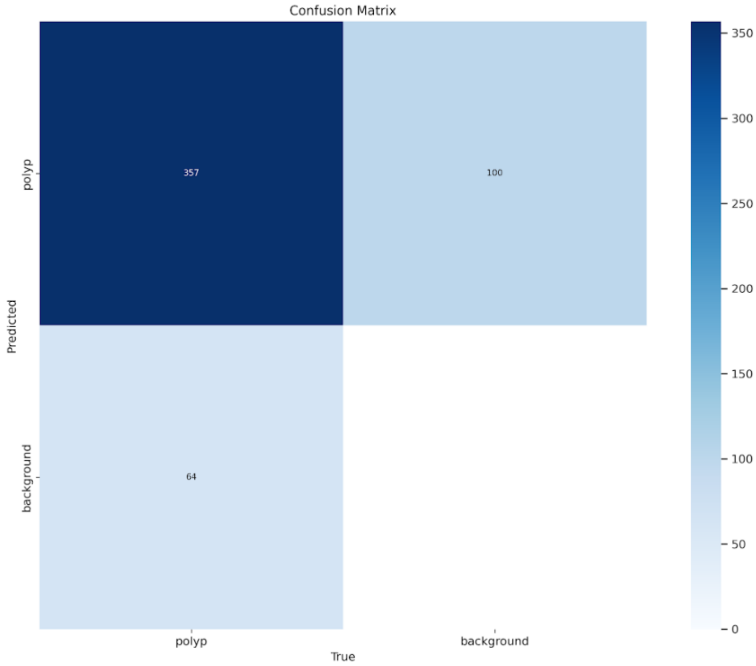
- **Precision Trend Graph (P_curve.png):** This graph depicts how precision fluctuates with varying thresholds, offering insight into the model's accuracy under different operational conditions.



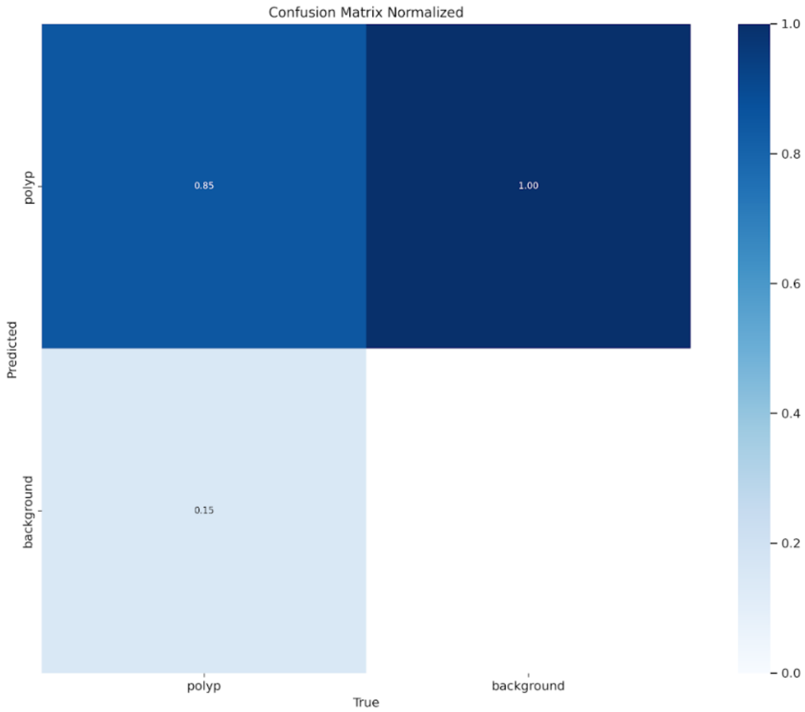
- **Recall Trend Graph (R_curve.png):** This complements the precision graph by illustrating the change in recall values with threshold adjustments.



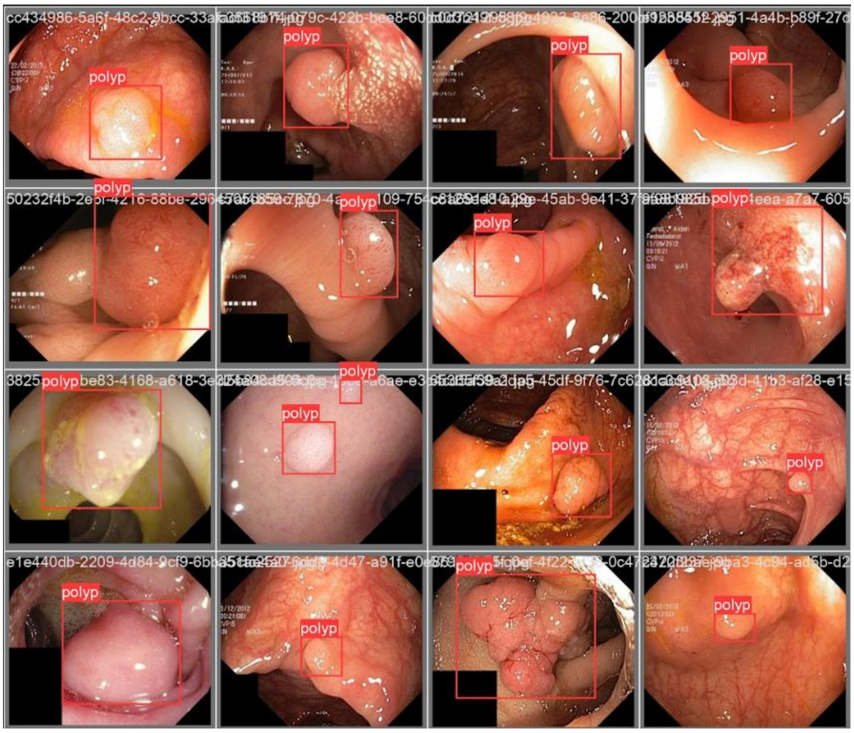
- **Outcome Matrix (confusion_matrix.png):** This matrix provides a comprehensive synopsis of the model's predictive outcomes, including true positives, true negatives, false positives, and false negatives for each identified class.



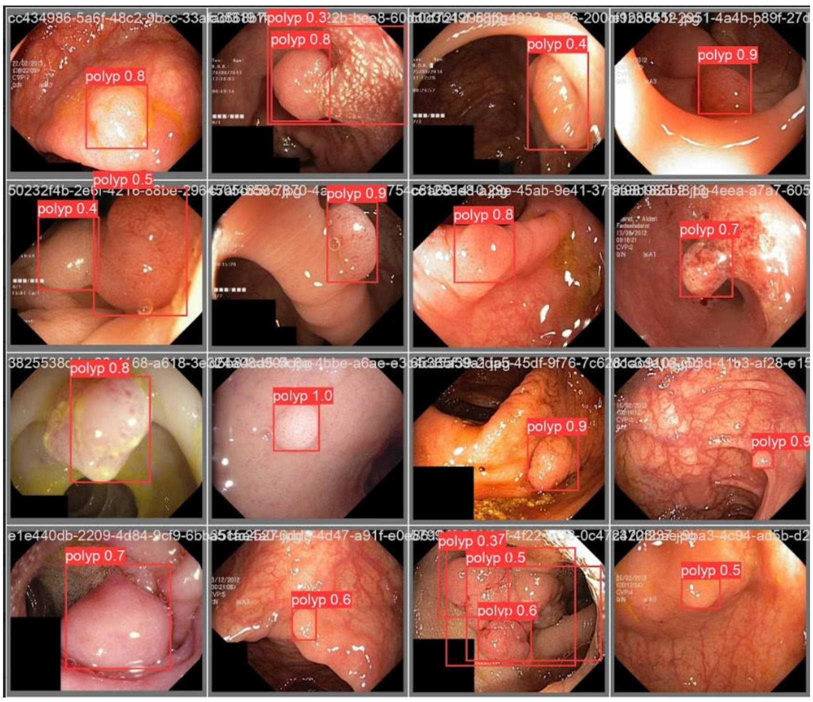
- **Proportionate Outcome Matrix (confusion_matrix_normalized.png):** A proportion-based rendition of the outcome matrix, this plot simplifies performance evaluation across different classes by focusing on proportional data instead of absolute figures.



- **Validation Set Ground Truth (val_batchX_labels.jpg):** These images present the ground truth labels from random validation dataset batches, offering a baseline for what the objects are and where they are positioned within the dataset.



- **Model Predictive Accuracy (val_batchX_pred.jpg):** Juxtaposed with the ground truth images, these pictures reveal the model's predictions. A direct comparison between these and the labeled images provides a clear visual benchmark for assessing the model's detection and classification accuracy.



Conclusion

Throughout this chapter, we have meticulously examined the pivotal performance metrics pertinent to YOLOv8. These indicators are instrumental in gauging a model's effectiveness and are indispensable for practitioners seeking to refine their models. They provide the critical intelligence required for enhancement and to confirm the model's efficacy in practical deployments.

Chapter 4

Experimental Data and Results

4.1 Description of Data (Dataset)

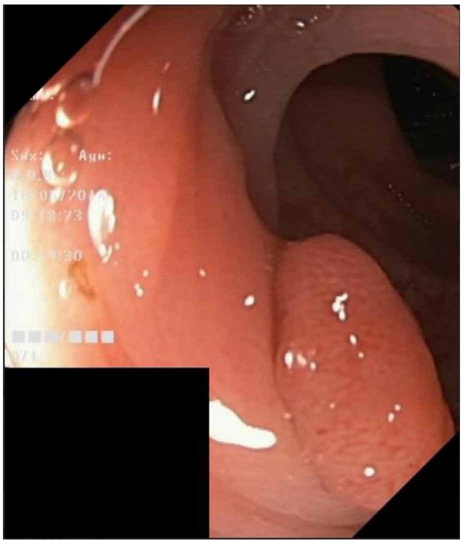
My research leverages the expansive Hyper-Kvasir dataset, an invaluable resource in the medical and machine learning communities, renowned for its diverse gastrointestinal images and videos. Although the dataset offers a wide array of data types, in this section, I will describe the dataset in its entirety before homing in on the specific subsets that I utilized [9].

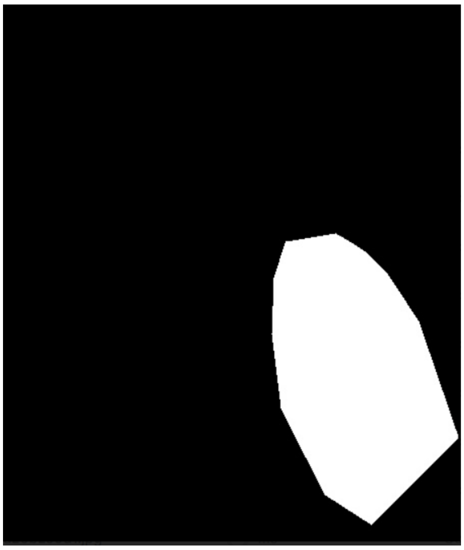
The complete Hyper-Kvasir dataset is categorized into four principal segments:

- **Labeled Images:** This category encompasses an extensive assortment of images classified across various gastrointestinal conditions.
- **Unlabeled Images:** It includes a multitude of images without labels, for which global features and cluster assignments are also provided.
- **Segmented Images:** This valuable set comprises images paired with corresponding segmentation masks and bounding boxes, particularly for the polyp class.
- **Videos:** The dataset also features a significant collection of video content.

In my thesis, I concentrated my analysis on a curated subset of this dataset, specifically selecting 1000 images from the 'polyp' category within the segmented images. These images are critical for the calibration and assessment of my object detection model, providing clear instances of the condition of interest. Furthermore, I incorporated 5 videos that contain varied presentations of polyps to enrich the dataset and augment the robustness of my findings.

Image: **Mask:**





The ensuing sections will delve into the methodologies I applied to these images and videos, documenting the experimental processes, the results of my analyses, and the valuable conclusions drawn from this focused study.

4.2 Endoscopy Imaging Detection results

Introduction

In this sub-chapter, I present the outcomes of deploying my trained YOLOv8 model for the detection of polyps within endoscopic images. The aim of these experiments was to ascertain the model's capability to accurately identify and locate polyp structures from the segmented dataset I described previously. The importance of such detection is underscored by the need for early and accurate diagnosis in gastrointestinal examinations, which can significantly enhance patient outcomes.

I will guide you through the results, elucidating the effectiveness of the model across various metrics. These findings are not just a testament to the model's technical proficiency but also speak to its potential to support gastroenterologists in their diagnostic processes. Each step—from data preprocessing to model evaluation—was meticulously carried out to ensure robustness and reliability in the results we sought to achieve.

Methodology Recap

In progressing towards the experimental phase, I recapitulate the training methodology that underpins the detection results. The heart of this process was the application of the YOLOv8 model, trained using a curated selection of endoscopic images.
Here's a concise breakdown of the code used for training the model [10]:

- I initiated the process by importing the necessary modules from the **ultralytics** package, a library that provides a streamlined interface for working with YOLO models.
- The **YOLO** class was then instantiated with a configuration file **yolov8n.yaml**, which specifies the architecture and hyperparameters for a new YOLOv8 model. This configuration is designed to initiate a model that's optimized for detecting features and patterns specific to endoscopic images of polyps.
- With the model initialized, I proceeded to train it using my dataset. The dataset, defined in **google_colab_config.yaml**, details the paths to the image and label files, as well as other critical training parameters. This configuration file resides within a predefined directory, **ROOT_DIR**, which was set to the root of my project workspace.
- The training was conducted over 100 epochs, allowing the model to iteratively learn from the data, optimizing its weights and biases to accurately detect

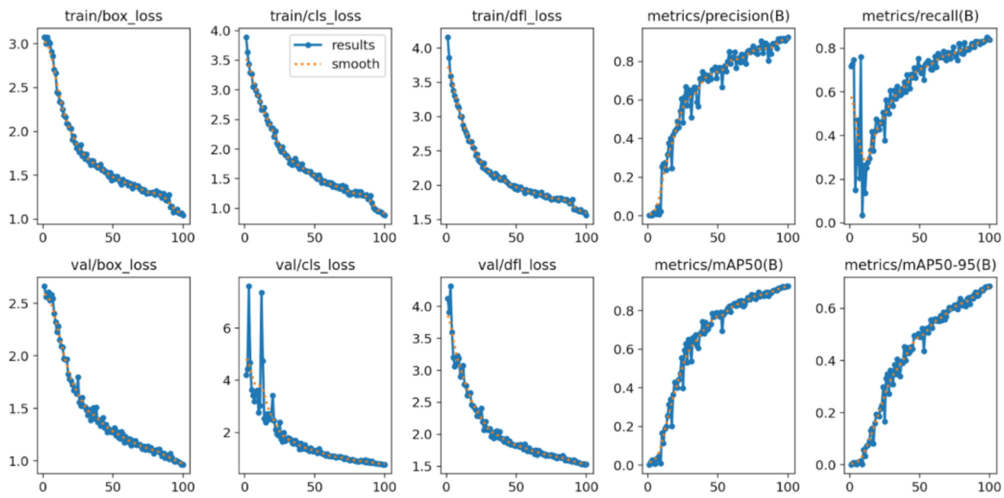
polyps. An epoch in machine learning is one complete pass through the entire training dataset.

Images training

```python
import os

from ultralytics import YOLO

# Load a model
model = YOLO("yolov8n.yaml")  # build a new model from scratch

# Use the model
results = model.train(data=os.path.join(ROOT_DIR, "google_colab_config.yaml"), epochs=100) # train the model
```

This snippet of code encapsulates the core of my training regimen, providing a foundation upon which the ensuing results were built. It underscores the simplicity yet effectiveness of the YOLOv8 framework in adapting to the nuanced task of polyp detection in endoscopic imagery.

After running the code, YOLOv8 generates these graphs that depict various metrics over the course of training my YOLOv8 model, represented through epochs:



Top Row: Training Loss and Performance Metrics

- **Train/box_loss:** This graph shows the loss associated with bounding box predictions during training. A declining curve indicates the model is getting better at predicting the correct size and position of bounding boxes.

- **Train/cls_loss:** The classification loss during training reflects how well the model is identifying the correct classes for the objects it detects. The downward trend is a positive sign of learning.
- **Train/obj_loss:** Objectness loss measures the model's performance in differentiating objects from the background. Decreasing loss implies improvement in recognizing objects.
- **Metrics/precision(B):** Precision over epochs shows the model's accuracy in predicting true positives out of all positive predictions. The upward trend suggests increasing precision as training progresses.
- **Metrics/recall(B):** Recall indicates the model’s ability to find all actual positives. An ascending curve here means the model is capturing more true positives over time.

Bottom Row: Validation Loss and Mean Average Precision (mAP)

- **Val/box_loss:** This represents the bounding box loss on the validation set. It helps validate that improvements in the training set generalize to new data.
- **Val/cls_loss:** Classification loss for the validation set, like the training classification loss, but for unseen data. This metric helps assess the model's generalization capability.
- **Val/obj_loss:** The validation set’s objectness loss. A decrease is desirable, mirroring improvements seen during training.
- **Metrics/mAP50(B):** Mean Average Precision at IoU threshold 0.50 for the validation set. Higher mAP50 values indicate better model performance at a basic level of object detection.
- **Metrics/mAP50-95(B):** This is the mean Average Precision averaged over multiple IoU thresholds, from 0.50 to 0.95, for the validation set. It is a more rigorous metric, as it accounts for both detection and localization accuracy across different difficulty levels.

The general descending trend in loss graphs and ascending trend in precision and recall graphs indicate that the model is learning and improving its performance in detecting objects. These metrics combined provide a comprehensive picture of the model’s learning trajectory, suggesting that the model has achieved a commendable level of accuracy and generalization to new data by the end of the training.

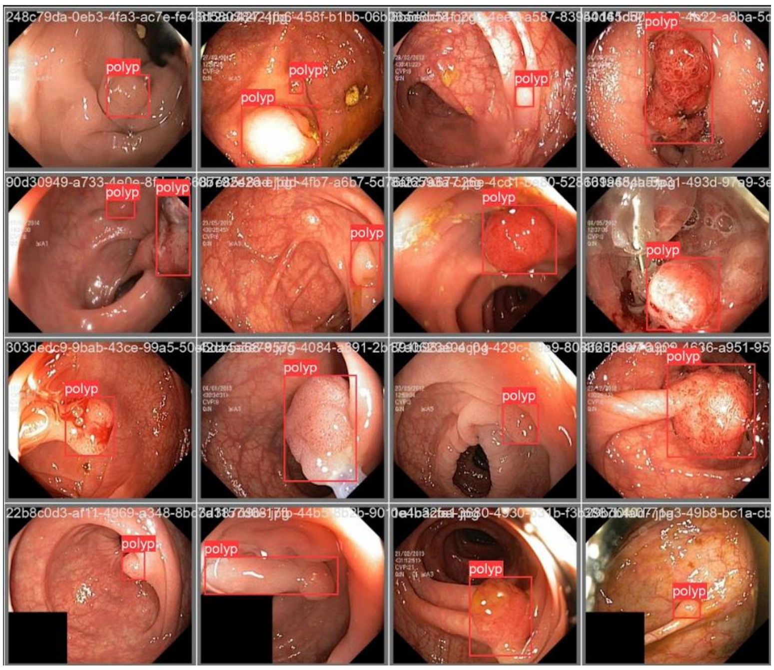
Visual Aids

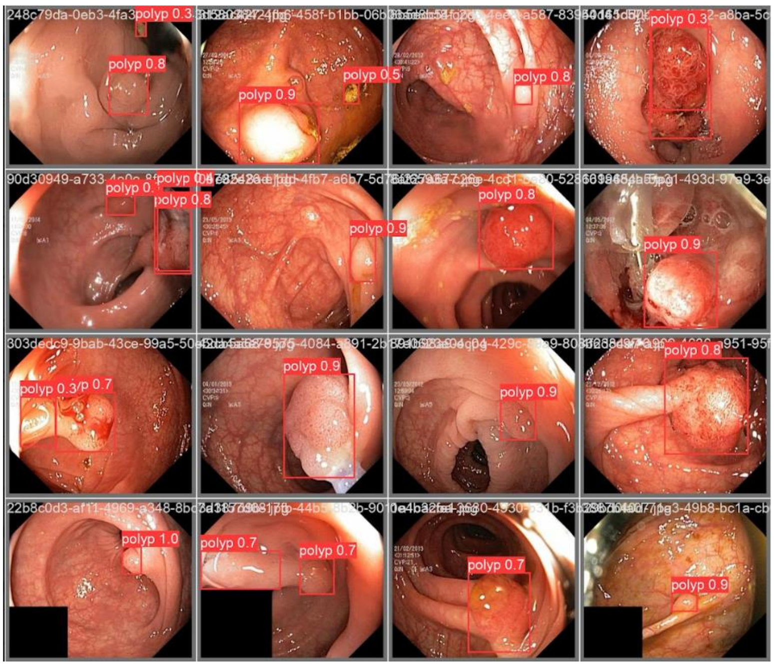
The integration of visual aids in this section serves to crystallize the performance of the YOLOv8 model used for polyp detection. These visualizations not only substantiate the quantitative data but also offer a clear, interpretable representation of the model's efficacy and areas for potential improvement [10].

- **Sample Detection Images:** These are images from the validation set where the model successfully identified polyps. The images will be annotated with bounding boxes as predicted by the model, showcasing the accuracy of object localization and the model's precision in real-world scenarios.

val_batch2_labels:

val_batch2_pred:

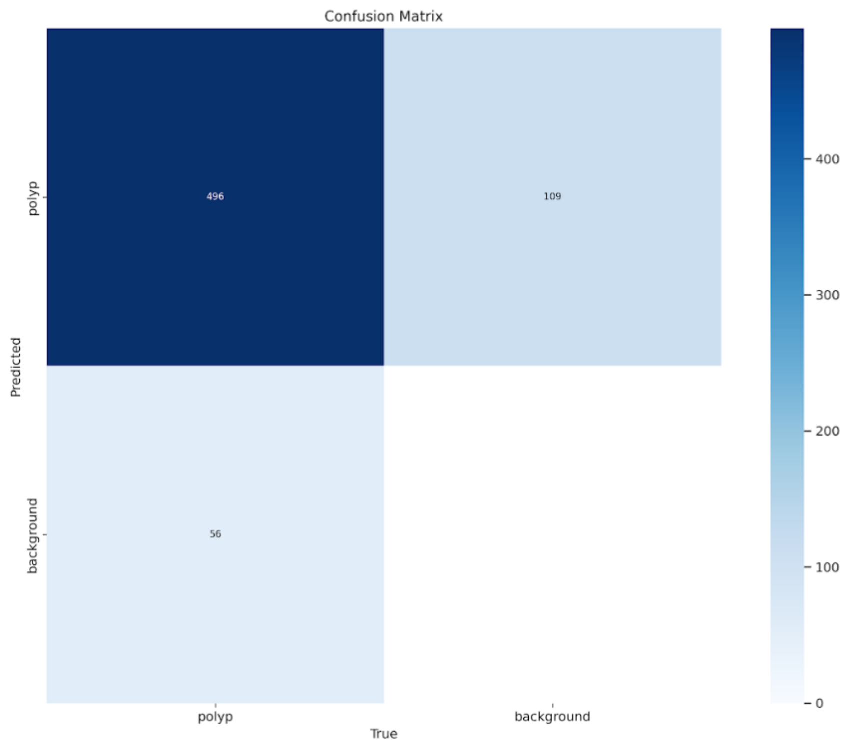




val_batch2_labels: This batch of images is the model's predictions are visually encapsulated in the collection of endoscopy images presented here, each meticulously annotated by the YOLOv8 model. Red bounding boxes demarcate the polyps as identified by the model, with the label 'polyp' confirming the detection. These images showcase the model's capabilities in discerning polyp tissue within the complex visual landscape of endoscopic imagery.

val_batch2_pred: This collection of endoscopy images represents the predictive evaluation from the YOLOv8 model. Each image showcases the model's capability to detect and assign a confidence score to the presence of polyps. In the images, we observe red bounding boxes—these indicate where the model has identified potential polyps. Accompanying each bounding box is a confidence score, denoted by a number between 0 and 1 next to the label "polyp." A score closer to 1 suggests a higher confidence in the detection, whereas scores nearing 0 indicate lower confidence.

- **Confusion Matrix:** The confusion matrix will be a pivotal component of my results analysis, providing a straightforward visualization of the model's predictions against the ground truths. It will delineate the counts of true positives, false positives, false negatives, and true negatives, enabling a comprehensive evaluation of the model's performance across different classes.



Even though the evaluation of the trained model was tested on 400 images it is possible for some images to be represented in more than one quadrant of the confusion matrix, especially in the context of object detection within images:

- If an image contains multiple objects (in this case, polyps), each object is evaluated separately. So, if an image has multiple polyps, it could contribute to both true positive and false negative counts, depending on the model's predictions for each polyp.
- True positives (TP) indicate the polyps that were correctly identified by the model.
- False negatives (FN) represent polyps that were present in the image but were missed by the model.
- False positives (FP) occur when the model predicts a polyp where there is none (possibly confusing it with background or other structures).
- True negatives (TN) are a bit different in object detection; they would be relevant if you had areas explicitly labelled as "not polyp" and the model correctly identified them as such. However, true negatives are not commonly reported in object detection tasks, as the number could be disproportionately high considering the many background regions in an image where no polyps are present.

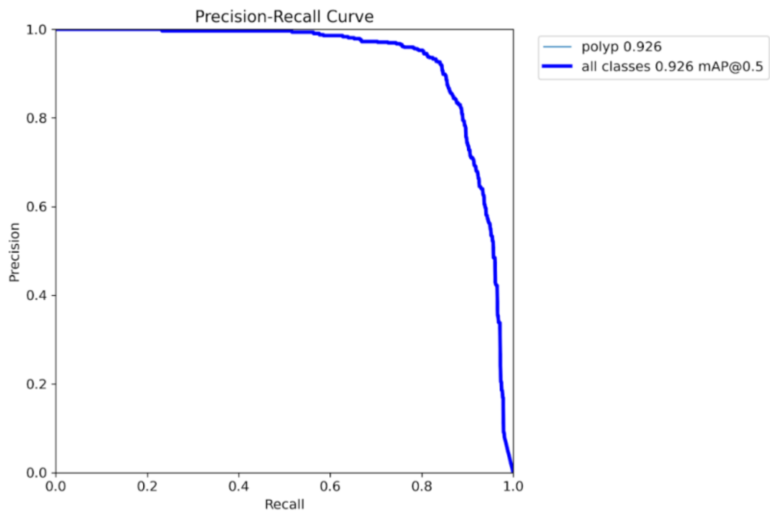
- **Precision-Recall Curve:** The Precision-Recall Curve displayed in my graph offers a comprehensive evaluation of my model's performance in polyp detection, with a notably high mean Average Precision (mAP) of 0.926 at an Intersection over Union (IoU) threshold of 0.5.

Key Observations:

1. **High Precision at Varying Recall Levels:**
    - My model maintains high precision even as recall increases, which is indicative of its robustness in accurately identifying polyps without generating many false positives.

- The curve starts at a very high precision near 1.0 and gradually declines as recall increases. This behaviour is typical and expected because as the model attempts to capture more true positives (increasing recall), it becomes more challenging to maintain high precision.

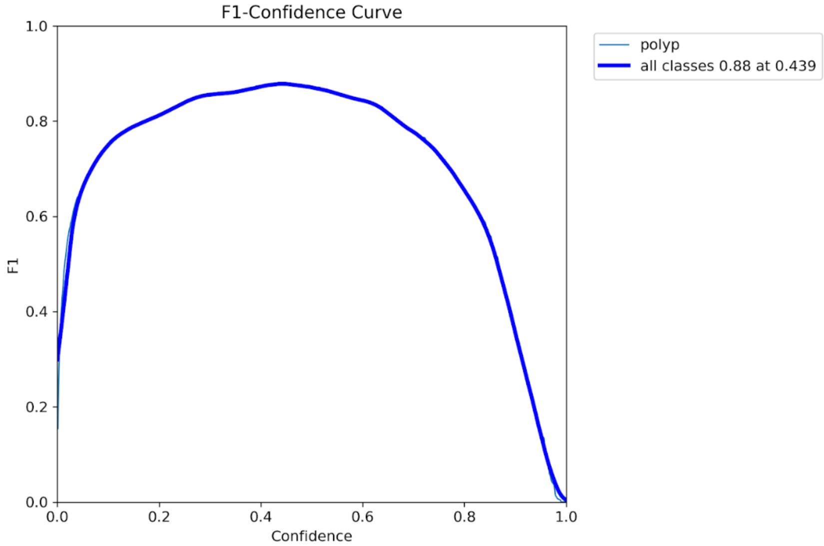
2. **Sharp Decline at High Recall:**
   - The sharp decline in precision as recall approaches 1.0 suggests that to detect almost all polyps, the model starts to include more false positives. This point of inflection is where the trade-off between catching as many polyps as possible (high recall) and keeping false detections low (high precision) becomes most apparent.
3. **Outstanding mAP Score:**
   - A mAP of 0.926 is excellent, indicating that on average, my model's predictions are about 92.6% precise at the 0.5 IoU threshold. This high score reflects not only the model's ability to detect polyps accurately but also its consistency in doing so across different scenarios within the test data.



- **F1-Confidence Curve:**

Peak Performance: The curve peaks at a confidence threshold of approximately 0.439, where the F1 score reaches its maximum value of 0.88. This point represents the optimal balance between precision and recall for this dataset and model, suggesting it as the ideal threshold for making prediction decisions to maximize the model's accuracy.

Decline in Performance: As the confidence threshold increases beyond the peak, the F1 score sharply declines. This decline indicates that while the model becomes more confident in its predictions (i.e., higher confidence thresholds), it fails to maintain an adequate balance of precision and recall, possibly due to missing true positives (low recall).



4.3 Results of Video application

https://drive.google.com/file/d/1-2oIrhz9msLdoyUQmOWZ-sOjfWD1pcIJ/view?usp=drive_link

Chapter 5

Comparison Results

5.1 Comparison of Results with Other Methods

This section provides a comparison of analysis of the polyp detection model I developed in this study with other methods, specifically DDANet on the Kvasir-SEG test dataset. By evaluating various performance metrics such as precision, recall and mean Intersection over Union(mIoU), this comparison aims to highlight the relative strengths and weaknesses pf each approach within the context of polyp detection in endoscopic imagery [11].

Comparison Table:

Metric	My Model	DDANet (Kvasir-SEG)
Precision	92,6%	86,43%
Recall	92%	88,8%
Mean Intersection over Union (mIoU)	0.5	0.7

Table1: My model vs DDANet (Kvasir-SEG) model [11]

Analysis:

- **Precision**: My model shows higher precision (92.6%) compared to DDANet's 86.43%. This indicates that my model has a lower rate of false positives, which is particularly valuable in clinical settings where the accuracy of polyp detection is critical to avoid unnecessary procedures.
- **Recall**: My model also exhibits slightly higher recall (92%) compared to DDANet’s 88.8%. This suggests that it is slightly more capable of identifying all relevant instances of polyps, minimizing the chance of missing any potential threats.
- **Mean Intersection over Union (mIoU):** My model reports an mIoU of 50%, which is significantly lower than DDANet's 78%. mIoU assesses the overall accuracy of the model in segmenting polyps by measuring the overlap between the predicted and actual regions. A lower mIoU indicates that while my model is good at detecting the presence of polyps (as reflected by high precision and recall), it may not be as accurate in delineating the exact boundaries of the polyps.

Chapter 6

Conclusion and Perspectives

6.1 Concluding Remarks

The lessons learned during the development and evaluation of a machine learning model for polyp detection with YOLOv8 have been valuable, both in terms of insight and educational. The ability of deep learning models in medical imaging to enhance diagnostic processes in gastroenterology is not only demonstrated by this thesis but also underlines the significant potential they possess.

Achievements

The project achieved a high degree of precision and recall, indicating that my model is highly effective at identifying true polyps while minimizing false positives. This level of accuracy is crucial for clinical applications where the stakes of diagnostic errors are high. The model's performance, as evidenced by a precision of 92.6% and a recall of 92%, positions it as a competitive tool against existing commercial systems, and even surpasses many in specific metrics.

However, the mean Intersection over Union (mIoU) of 50% suggests there is room for improvement in the model’s ability to accurately segment polyps. This metric, although lower than desired, provides a clear direction for future enhancements. Improving segmentation accuracy will not only bolster the model's utility in clinical settings but also refine its applicability to real-world scenarios where precise delineation of polyps is necessary for effective treatment planning.

6.2 Future Research and Suggestions

Looking forward, the field of AI in medical imaging is ripe with opportunities for further research and development:

- **Integration of Multimodal Data:** Incorporating data from various imaging modalities can potentially improve the robustness and accuracy of polyp detection systems.
- **Real-time Detection Systems:** There is significant potential in developing real-time polyp detection systems that can operate during live endoscopic procedures, providing immediate feedback to endoscopists.
- **Personalized Medicine Applications:** AI models can be tailored to individual patient histories and genetic information to predict polyp malignancy potential, thereby personalizing the screening and surveillance protocols.
- **Expansion to Other Gastrointestinal Conditions:** Extending the model to detect other conditions such as oesophageal or gastric cancers can broaden its impact, making it a versatile tool in gastroenterology.

6.3 Concluding Thoughts

This thesis lays a foundation for future work that could lead to more reliable, accurate, and clinically relevant polyp detection tools, ultimately aiming to reduce the incidence of colorectal cancer through early detection and treatment. As AI continues to evolve, its integration into healthcare promises not only to enhance patient outcomes but also to revolutionize the very paradigms of medical diagnostics and treatment strategies.

In conclusion, the journey from conceptualization to realization of this polyp detection model has been immensely rewarding. It has not only challenged my technical skills and deepened my knowledge but also reaffirmed my commitment to leveraging technology to foster better health outcomes. As I look to the future, I am excited about the prospects of expanding this research to encompass broader applications, pushing the boundaries of what AI can achieve in medicine.

Bibliography

[1] J. Wu, S. Xu, Y. Zhang, Y. Cai, J. Li, Y. Zhu, "AI for Endoscopy," in FG AI4H Deliverable 10.20, ITU/WHO Focus Group on AI for Health (FG-AI4H), Geneva, 3-5 July 2023.

[2] Chin SE, Wan FT, Ladlad J, Chue KM, Teo EK, Lin CL, Foo FJ, Koh FH. One-year review of real-time artificial intelligence (AI)-aided endoscopy performance. Surg Endosc. 2023 Aug;37(8):6402-6407. doi: 10.1007/s00464-023-09979-8. Epub 2023 Mar 17. PMID: 36932187.

[3] Dorfer, T. A. "Enhanced Object Detection: How to Effectively Implement YOLOv8," Towards Data Science, 23 March 2023.

[4] AI_Pioneer. "Mastering Custom Object Detection with YOLOv8: A Step-by-Step Guide to Precision in Computer Vision," 19 October 2023.

[5] KeyLabs. (2023). Mastering Object Detection with YOLOv8. KeyLabs Blog. https://keylabs.ai/blog/mastering-object-detection-with-yolov8/#:~:text=Custom%20object%20detection%20using%20YOLOv8,data%20and%20configuring%20the%20config.

[6] Zafar, S. "Unleashing the Power of YOLO v8: A Breakdown of its Working Principle and Evolution," 10 October 2023.

[7] Glenn-Jocher, RizwanMunawar , Abirami-Vina “YOLO Performance Metrics: Performance Metrics Deep Dive”, Created 2023-11-12, Updated 2024-03-02

[8] Glenn-Jocher, Burhan-Q, RizwanMunawar. “Model Validation with Ultralytics YOLO”, Created 2023-11-12, Updated 2024-03-03

[9] HyperKvasir “The Largest Gastrointestinal Dataset” https://datasets.simula.no/hyper-kvasir/

[10] Dependabot, Glenn-Jocher, Fcakyon, Laughing-q, Burhan-Q. “Model Training with Ultralytics YOLO”, Created 2023-11-12, Updated 2024-04-23

[11] Tomar, N. K., Jha, D., Ali, S., Johansen, H. D., Johansen, D., Riegler, M. A., & Halvorsen, P. "DDANet: Dual Decoder Attention Network for Automatic Polyp Segmentation," 30 December 2020.