CHAPTER 1

INTRODUCTION

In the last decade, minimally invasive and non–invasive surgical procedures have gained a lot of importance. Digital endoscopy has established itself as a minimally invasive surgery technique. Due to rapid advancement in hardware especially GPU, and training data availability, many research communities with diversified backgrounds proposed methods to analyse innately produced endoscopic images and videos to extract useful information. In the current decade, the world has seen an unanticipated growth in technologies incorporating Artificial Intelligence (AI).

Computer Vision (CV) has given a new definition to digital endoscopy. Most of the research is based on real-time applications such as, the classification or detection of diseases, characterization and estimation of disease severity. Other areas include pre and post-processing techniques. The former embrace de-mosaicking, denoising, detecting and restoring artefacts. Post-processing techniques include management, storage and retrieval of video content.

Research in pre and post-processing techniques is still limited compared to research advancements in real-time processing. This research focusses on a vital pre-processing technique called artefact detection and segmentation. A simple endoscopic imaging pipeline is implemented to stress the need for such pre-processing technique. It includes detection, segmentation and restoration of artefacts followed by classification of Gastro Intestinal (GI) tract diseases after restoration, specifically polyp.

1.1 ENDOSCOPY

A thin, long, firm or flexible tube known as an endoscope is put into the patient's body either via the mouth, any opening, or occasionally by making a small incision. The traditional term "endoscopy" means "look inside". It investigates organs such as, the GI tract, respiratory tract, female reproductive tract, joints and ear. A modified form of endoscope called a laparoscope is used to perform keyhole or buttonhole surgery to make the process less invasive. It can be performed in the joints, abdominal or pelvic cavity.

Endoscopy is also used to take a biopsy to remove tissue for further analysis and look for disease. Other uses of endoscopy include treating digestive tract problems, detecting bleeding, ceasing bleeding, removing polyps and removing foreign bodies. An endoscope is shown in the Figure 1.1.

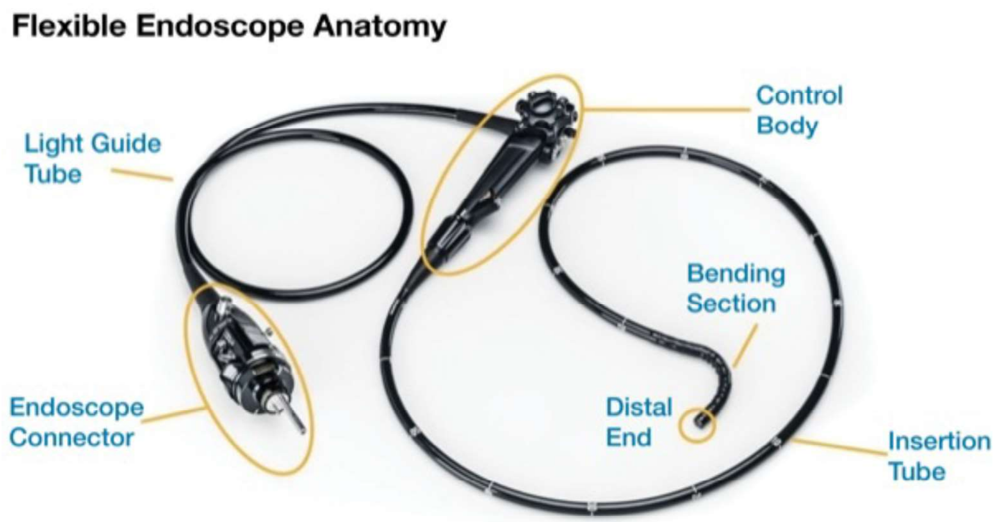


Figure 1.1 External Anatomy of Flexible Endoscope
(https://www.gastroendonews.com)

There are various types of endoscopy. The type varies based on the body's part to be examined. The term endoscopy mainly refers to two methods, namely gastroscopy and colonoscopy. Table 1.1 list the procedure names, tools used, the organ under study and how the endoscope is passed into the human body.

Table 1.1 Types of Endoscopy(https://www.cancer.net/)

Name of Procedure	Instrument Used	Organ under Study	Access
Anoscopy	Anoscope	Anus	Anus
Arthroscopy	Arthroscope	Joints	Small incision
Bronchoscopy	Bronchoscope	Trachea and lungs	Mouth
Colonoscopy	Colonoscope	Large intestine and colon	Anus
Esophagoscopy	Esophagoscope	Esophagus	Mouth
Gastroscopy	Gastroscope	Stomach and duodenum	Mouth
Laparoscopy	Laparoscope	Stomach, liver and other organs of the abdomen	A tiny incision in the abdomen

Advancement in the area includes Wireless Capsule Endoscopy (WCE) (https://www.olympus-europa.com), where humans will swallow a pill-sized camera. It takes photographs of the GI tract and will be recorded in a device attached to the human body (Iddan et al., 2000). It is specially designed for examining the complex structure of the small intestine. Figure 1.2 portrays WCE.

Figure 1.2 Wireless Capsule Endoscopy (https://www.olympus-europa.com)

A camera is attached at the tip of an endoscope. Illuminating the region of interest is done by fibre optic transmission. It aids the clinical team in visualizing and navigating through the internal organ for routine clinical procedures. During the process, the patient's vital signs are be monitored. To avoid gagging and coughing, local anaesthesia is used. With all necessary pre-procedure, the patient is allowed to lie on his left. Later the endoscope is to be passed into the human through any natural opening.

In most cases, the air is passed through the scope to slightly expand the organ, which increases the doctors' viewability of the organ. The organs are then examined, biopsies are taken, and necessary treatments are performed. The procedure may last from a few minutes to an hour. The complete process is recorded in the form of a video signal. Figure 1.3 portrays a typical scenario of an upper GI endoscopy and Figure 1.4 shows the insertion tip of a typical endoscope used by clinicians.

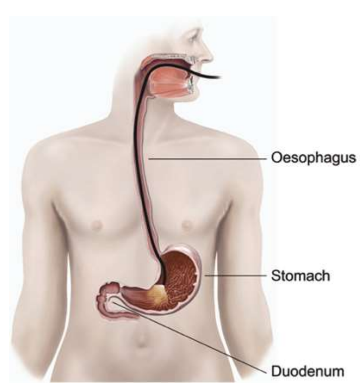


Figure 1.3 Upper GI Endoscopy (https://www.healthdirect.gov.au**)**

Figure 1.4 Insertion Tip of a Typical Endoscope
(https://en.wikipedia.org/wiki/Endoscope)

With the advent of AI technologies, modern digital imaging technology, development in the field of imaging and communication has extend the work of an endoscopist widely available. The advancements in video endoscopy assists clinicians to create texts, images and waveforms as a part of electronic record. The modern day endoscopic systems include inbuild image processors and microcomputers that are capable of capturing, storing, retrieving and delivering the endoscopic images. In a recent research in 2020, advanced endoscopic instruments like endoscopic ultrasound is said have accuracy rate of 85% to 92% (https://www.healthline.com).

Endoscopy has diverse research domains. The clinician must work with an image or video in every field. It is processed at various stages to extract useful information. The endoscopic image analysis has three branches, as indicated in the Figure 1.5.

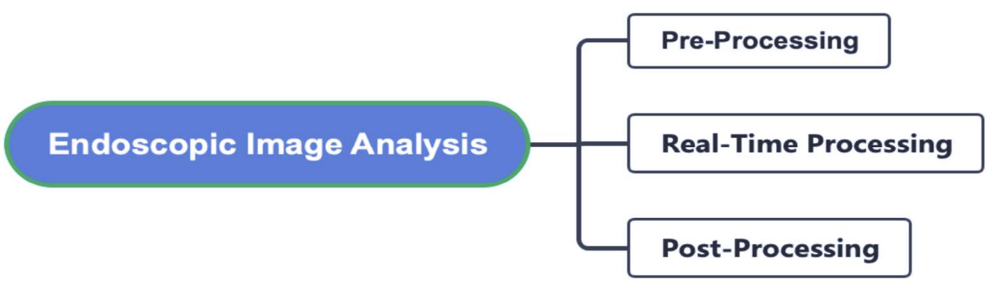


Figure 1.5 Main Branches of Endoscopic Image Analysis

Techniques such as, image enhancement extracts useful information from the frames falls under pre-processing. Real-time processing includes tool tracking for robotic-assisted surgery and augmented reality, to diagnose a disease and assist the surgeon in real-time.

The post-procedure consists of quality assessment, management and storage of medical data. This new paradigm aids various activities as follows:

1. Holds the record of patients for future reference.
2. For further analysis and decisions on treatments.
3. A source of knowledge for novice endoscopists.
4. A good source for researchers to access medical documents.

Thus research in every such area benefits society. From the discussion, it is evident that the research in the endoscopy domain plays around the videos and images that are recorded during the procedure. It can be understood that effective decision-making in real-time and effective pre-processing techniques are necessary for better medical documentation. The pre-processing techniques are considered a preparatory step that helps enhance the image and aid the clinician by providing a better view of the organ of interest. Innately, majority of frames are impacted by artefacts. This could distract the organ's visualisation (Ali et al., 2020). It increases endoscopy procedure time and hence the clinician's fatigue. It also increases missed detection rate by over 15%. Thus the research to detect, segment and restore the artefacts is the need of the hour. The clinicians prefer better-pre-processed images.

1.2 ARTEFACTS

Artefacts are unforeseen pattern appearing in the resultant image. Artefacts in endoscopic images include saturation, specular reflection or specular highlights, blur, contrast, bubbles, instrument, blood, miscellaneous artefacts, comb structures and surgical smoke. Miscellaneous artefacts include chromatic aberration (https://www.adobe.com), floating debris and general miscellaneous artefacts. The artefact surgical smoke (https://pubmed.ncbi.nlm.nih.gov) is produced due to the usage of the energized dissecting and ceasing device. Often, such smoke is immediately evacuated using an advanced suction system such as, intravenous tubing. It is a separate field of research where datasets are also found to be very limited. Comb-structure is found very rare in endoscopic images. One or more of the other artefacts occur randomly in almost all the frames. Artefacts vary in size and do not have a prominent geometrical structure. Every artefact, along with the reason for its presence in an endoscopic image, is discussed in the sub-sections.

1.2.1 Saturation

Saturation in endoscopic images occurs when high-intensity light falls on the tissues. The organ area that is close to the light system is heavily saturated, producing bright pixel areas. Saturation poses several problems for CV algorithms. Saturation occurs due to over-exposure to white light and it saturates all colour channels. Figure 1.6 (a) and (b) highlights the saturated area in endoscopic images.

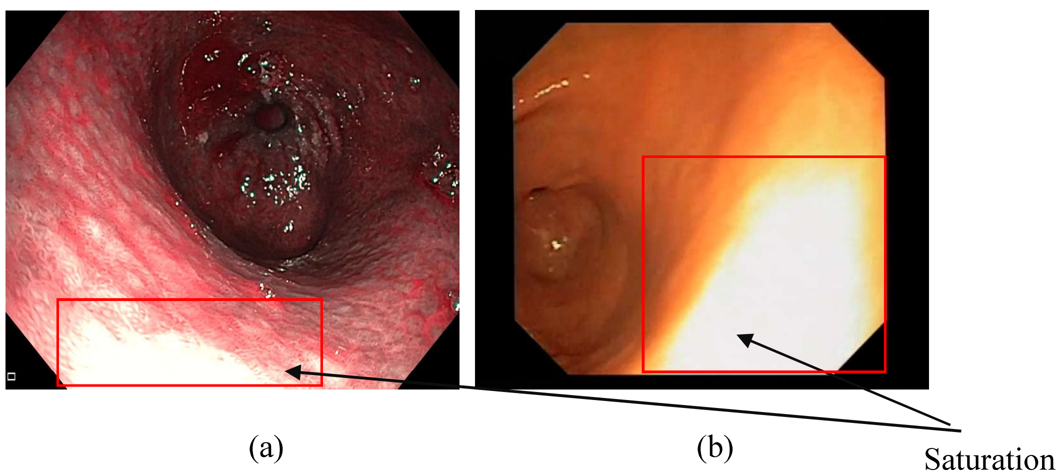
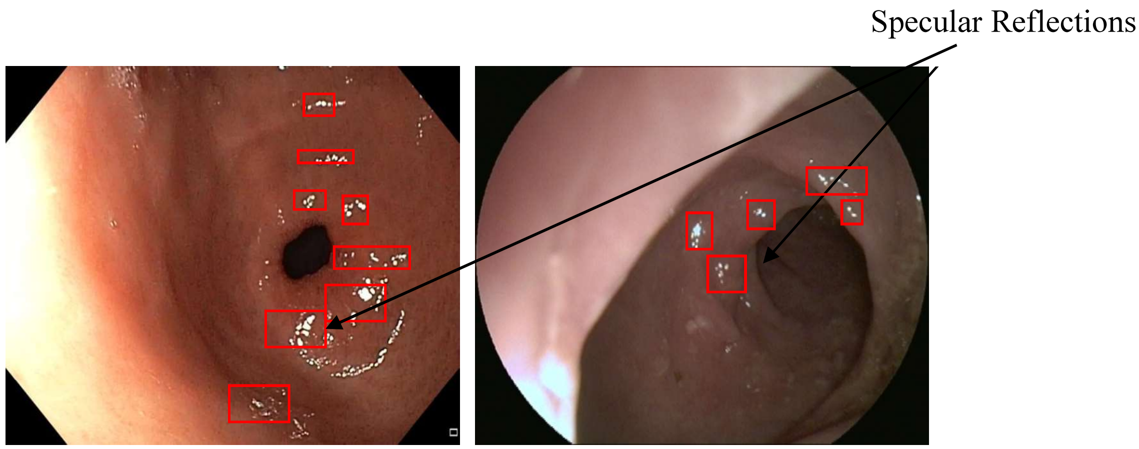


Figure 1.6 (a) and (b) Endoscopic Images Affected by Saturation

1.2.2 Specular Reflections

Specular reflections in an endoscopic image are due to light reflected by a mirror-like tissue surface. It is also called specular highlights. This artefact varies in size and is often found in endoscopic images in clusters or spread across the image randomly. Such bright spots impede the clinician's viewability across the organ leading to increase the endoscopic procedure time. Figure 1.7 (a) and (b) displays endoscopic images affected by specular reflections artefact. A few of the specular highlights are highlighted for better viewability



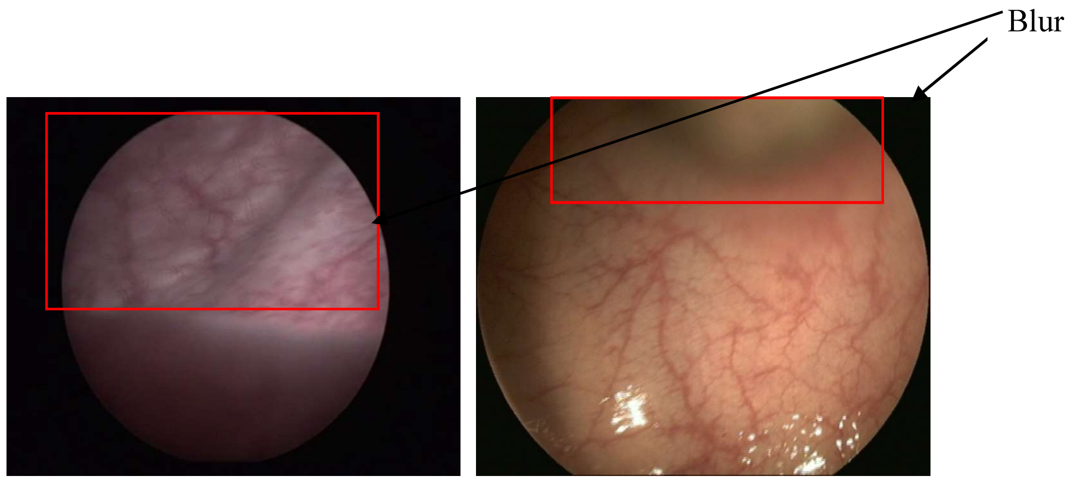
(a) (b)

Figure 1.7 (a) and (b) Endoscopic Images Affected by Specular Reflections

The specular reflections impede the analysis procedure. It proposes wrong pixel values and adds additional edges during automated analysis.

1.2.3 Blur

The endoscope is attached with a miniaturized camera to guide it through the internal organ. The imaged organ will be displayed on a monitor attached to the external imaging system setup. The movement of the endoscope is controlled by small hand motions outside. A small or fast hand motion causes severe blur in the resulting image. Blur due to such motion is termed as motion blur. An endoscope also allows zoom in or zoom out, which aids a closer look at the organ. During zooming in or zooming out, blurring may occur. Blur due to defocus is called defocus blur. Both motion and defocus blur affect the resultant image. Figure 1.8 (a) and (b) exhibits a partially blurred endoscopic images. The result of a study says that 25% of frames of a colonoscopy video are blurry.

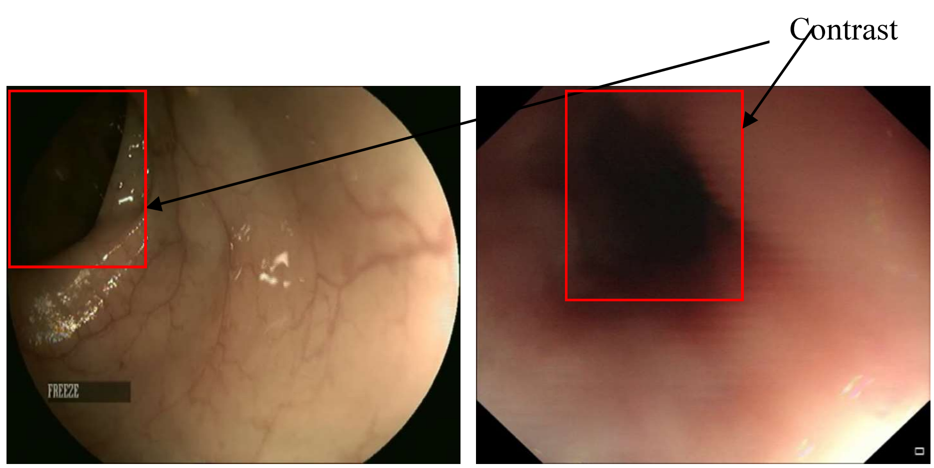


(a) (b)

Figure 1.8 (a) and (b) Endoscopic Images Affected by Blur

1.2.4 Contrast

Contrast is considered an artefact that occurs due to underexposure or occlusion. The low-contrast areas appear dark, which may impede a clinical abnormality. So to eliminate the possibility of the presence of a clinical abnormality in low contrast areas, the clinician moves the endoscope to validate the same; in turn, it consumes much of the procedure time. Figure 1.9 (a) and (b) shows endoscopic images affected by contrast artefact.

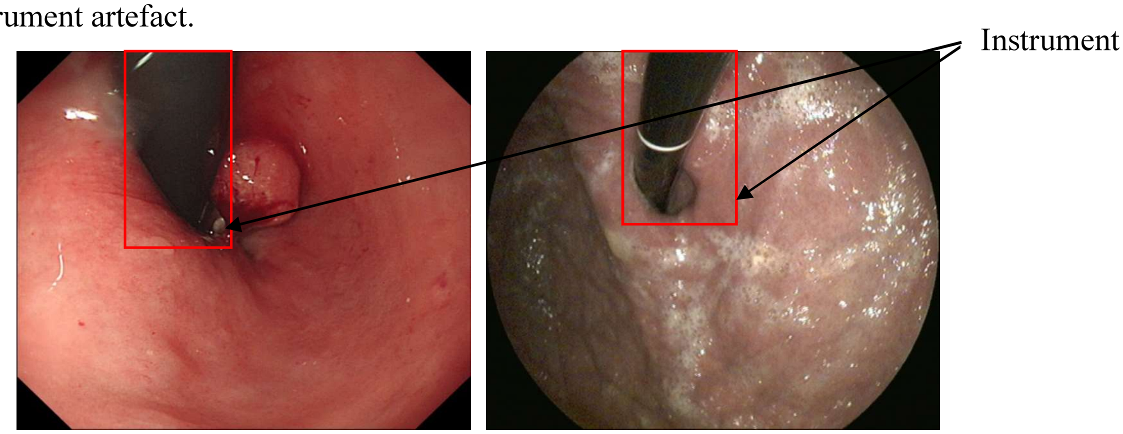


(a) (b)

Figure 1.9 (a) and (b) Endoscopic Images Affected by Contrast

1.2.5 Instrument

Professionals commonly use endoscopic instruments during simple laparoscopic surgeries to colonoscopies. Apart from anatomical objects, instruments such as, biopsy forceps, tissue scissors and staplers gain much importance during the procedure. A light delivery system is also attached to the scope, along with a camera. When the clinician performs the procedure, the instruments falling within the camera's scope hide the underlying tissue. The doctor cannot see the underlying tissue since the internal organ being examined is displayed on a monitor that is fixed outside. Thus the instrument falling within the scope is also considered as an artefact. Figure 1.10 (a) and (b) shows endoscopic images with instrument artefact.

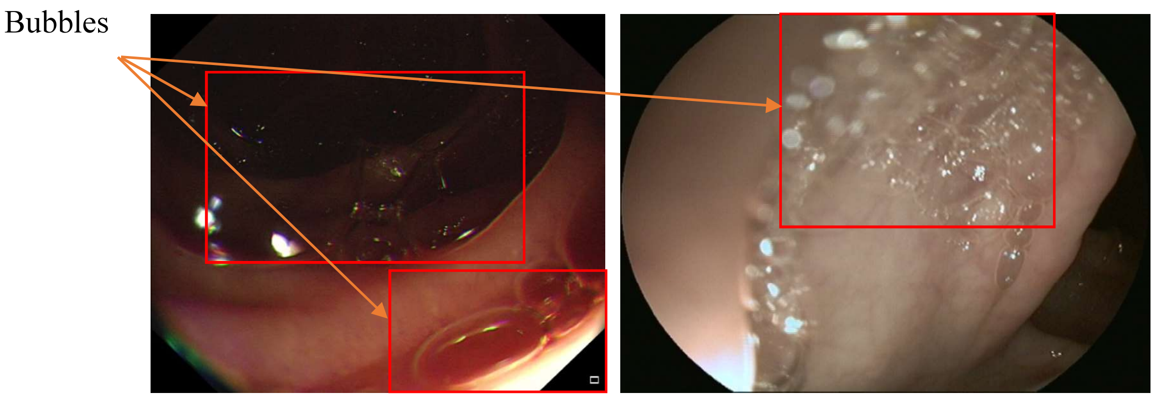


(a) (b)

Figure 1.10 Endoscopic Images Affected by Instrument

1.2.6 Bubbles

Gastrointestinal juice is found to be present inside the tract. A thin layer of intestinal juice envelopes air sacs forming bubbles. It distorts the appearance of the tissue and occludes the visualization field. According to a study, 23% of such images are discarded referring to increase the procedure and the processing time of clinical experts and automatic diagnostic systems. The reflective surface of the water bubbles aids other artefacts, such as, specular reflections and saturation. In such cases, specular reflections and bubbles overlap and impede the tissue underneath. Figure 1.11 (a) and (b) shows endoscopic images affected by bubbles artefact.

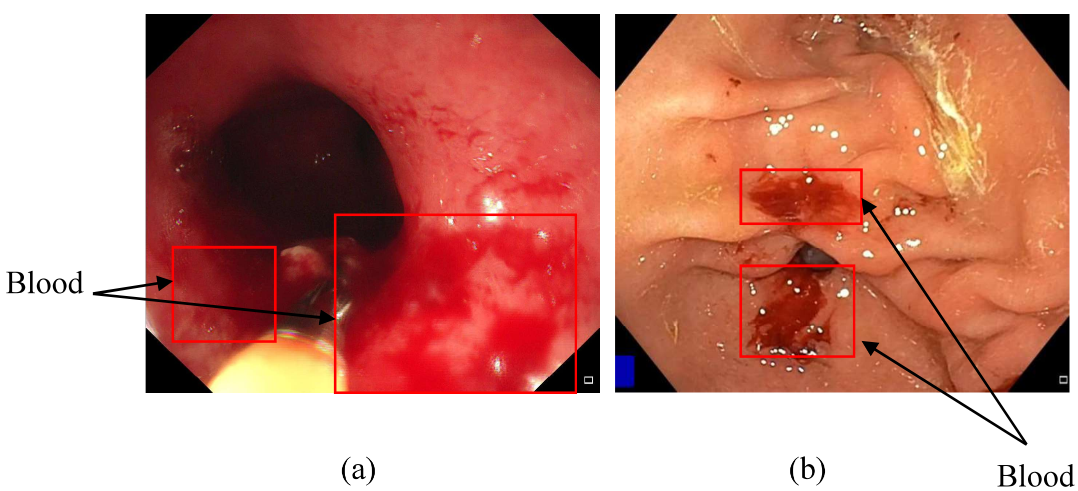


(a) (b)

Figure 1.11 (a) and (b) Endoscopic Images Affected by Bubbles

1.2.7 Blood

Blood is an artefact. It may be a symptom of an abnormality in the upper or lower GI tract. Various abnormalities include peptic ulcers, tears in walls, foreign body ingestion, tumours and colon cancer. Spotting the bleeding site becomes a time-consuming procedure amidst the artefact. Blood may also be due to clinical procedures on the affected site. Blood exuding from the site may occlude the surrounding area. Figure 1.12 (a) and (b) shows endoscopic images affected by the artefact blood.

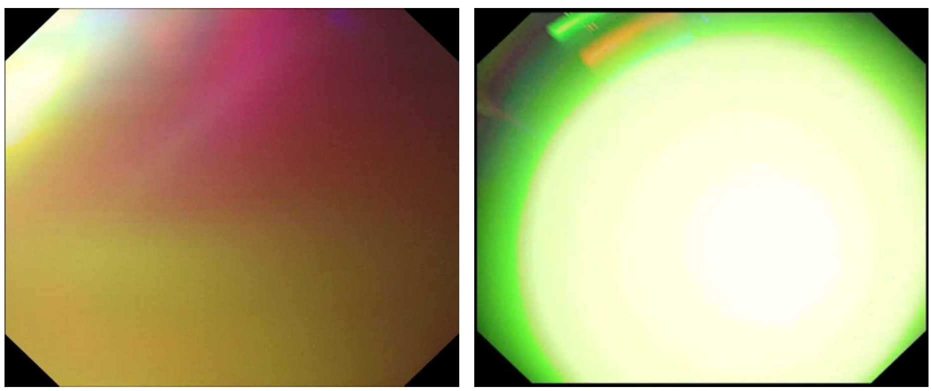


(a) (b)

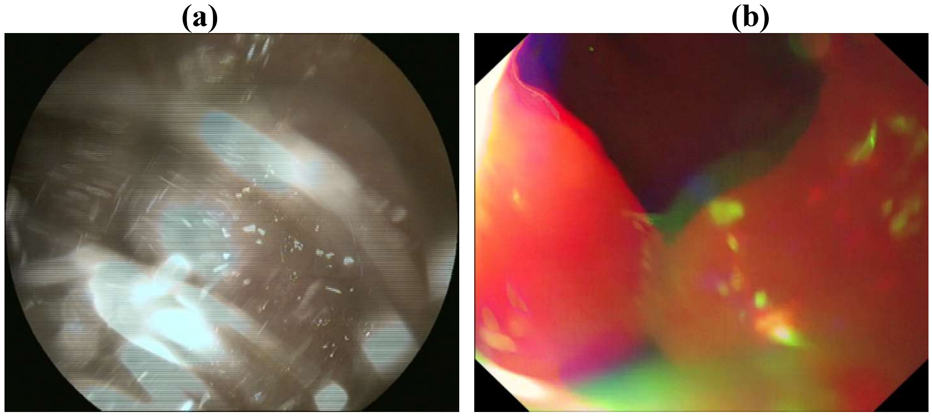
Figure 1.12 (a) and (b) Endoscopic Images with Blood

1.2.8 Miscellaneous Artefacts

Miscellaneous artefacts include chromatic aberration (the lens in the camera does not focus all colours at the same point), debris and imaging artefacts. Due to miscellaneous artefacts, the viewability of the organ is affected, and images affected by such an artefact will not help to extract useful information. Figure 1.13 (a) to (d) displays the endoscopic images affected by miscellaneous artefacts. Figure 1.13 (a), (b) and (d) denotes images affected by chromatic aberration. No useful information can extracted from these images. Figure 1.13 (c) is due to the motion along with scattering of light. Such images are generally discarded.



(a) **(b)**



(c) **(d)**

Figure 1.13 (a) - (d) Endoscopic Images Affected by Miscellaneous Artefacts

Endoscopic images are often said to contain artefacts. These artefacts occur due to various phenomena. Artefacts generally distort the clinician's observation. Often, the clinician prefers to restore artefacts for better and faster decision-making. The artefact blights image analysis procedure, and hence a lot of importance is given to this area of research. An efficient algorithm for efficient detection, segmentation and restoration could eventually help endoscopists to have a better view of the organ, which can be attained through deploying AI into endoscopy.

1.3 ARTIFICIAL INTELLIGENCE

In the present era, digital medicine has gained a lot of significance. Medical images are generated from clinical procedures such as, ultrasound , pathology, endoscopy and radiology. Every day thousands of images are generated. There is a great demand for smart and intelligent equipment to manage the situation. With the advent of AI, algorithms based on Convolutional Neural Networks (CNN) (https://www.ibm.com) advanced rapidly. It plays a key task in medical imaging. Such technologies aid doctors towards diagnosis in a shorter time with improved accuracy.

AI is a tributary of computer science where the computers are trained to execute tasks like human beings. Machine Learning (ML) is a subclass of AI where labelled or the unlabelled data are used to train the algorithm, which in turn, is expected to perform as humans carry out the specified job after training. The ML technique is classified into four groups, as shown in the Figure 1.14.

Figure 1.14 Classification of ML Algorithms

Supervised Learning: The ML algorithm will be trained with the help of labelled medical images. Labelling is a process where the raw medical images will be taken, and informative labels will be added to provide the algorithm with an environment to learn meaningful information from the images.

***Unsupervised Learning**:* The ML algorithm will be trained with the help of unlabelled data.

***Semi-supervised Learning**:* A chunk of data will be labelled, and others will be left unlabelled.

Reinforcement Learning: In reinforcement learning, the algorithm periodically receives positive feedback if the action is good and vice versa.

The disadvantage of ML algorithms is that feature extraction must be done manually. In the case of Deep Learning (DL) algorithms, that algorithm extracts features by itself. DL models resemble the human brain in many ways. It simulates the neural network architecture of the human brain. DL algorithms are often preferred for high-dimensional data.

The DL algorithms are classified into CNNs, Long Short Term Memory networks (LSTMs) (Hochreiter & Schmidhuber, 1997), Recurrent Neural Networks (RNNs) (Sherstinsky, 2020), Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), Radial Basis Function Networks (RBFNs) (Bors, 2001), Multi-Layer Perceptron (MLP) (Marius-Constantin et al., 2009), Self-Organizing Maps (SOMs) (Kohonen, 1998), Deep Belief Networks (DBNs) and Restricted Boltzmann Machines (RBMs) (Cueto et al., 2010). This research mainly focuses on CNNs. CNN is otherwise called as ConvNets. The network has multiple layers. The first network LeNet (LeCun et al., 1998), came into existence in 1998, proposed by the author Yann LeCun. After which CNN incurred a drastic growth in two decades. Such CNN based networks are mostly preferred for applications like image processing.

CNN has several layers. A convolutional layer, a pooling layer and a Fully Connected Network (FCN) layer. The network process the data by passing it through multiple layers of the network. It extracts image features to exhibit convolutional operations. A pooling layer is used to lessen the dimension of the extracted feature map. The output from the preceding layer is transformed into a single, continuous linear vector by a flattening layer. Finally FCN classifies the image.

1.3.1 Popular Object Detection Algorithms

An object detection algorithm along with labelled or unlabelled data, can be rigorously trained using a Graphical Processing Unit (GPU). The training of an object detection algorithm is computationally very expensive. Once trained, the algorithm can predict meaningful information from the test images. The healthcare industry assists doctors with such algorithms in detecting lesions, cancer stages, fractures and simulation-based surgical platforms for skill development.

The most popular object detection algorithms are Region-based CNN (R-CNN) (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015), Single Shot Detector (SSD) (Liu et al., 2016), You Only Look Once (YOLO) (Redmon & Farhadi, 2018), Feature Pyramid Network (FPN) (Lin et al., 2017) and CenterNet (Duan et al., 2019). Object detection framework can be split into four major types as shown in Figure 1.15.

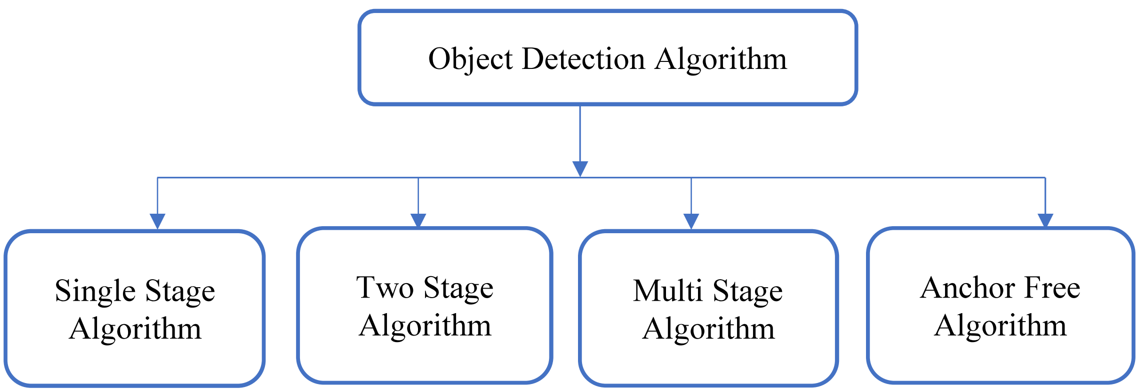


Figure 1.15 Classification of Object Detection Algorithms

- **Single-stage object detection algorithm:** Single-stage object detection algorithm runs over the entire image in one shot. Popular single-stage algorithms include SSD, YOLO and RetinaNet. These algorithms are said to be faster but less accurate than two-stage detectors.

- **Two-stage object detection algorithm:** This object detection algorithm has four modules: The first module produces region proposals. Next module extricates feature vectors for each region proposed by first module; the third module classifies the object using a Support Vector Machine (SVM) classifier, and the fourth module is draws a precise bounding box. Basic R-CNN is a popular among it. Updated versions of R-CNN includes Fast R-CNN and Faster R-CNN.

- **Multi-stage object detection algorithm:** Multi-stage detectors achieve better accuracy than one-stage and speed than two-stage detectors. It is said to overcome the problem of overfitting during training and inference time mismatch between Intersection over Unions (IoUs). The network is a muti stage extension of R-CNN. Every stage is trained consecutively. That is turnouts of first stage acts as input to the next stage. Cascade R-CNN (Cai & Vasconcelos, 2018), an extended version of Faster R-CNN.

These DL algorithms depend on anchors for foreseeing an image's semantic objects. The anchors have various scales and aspect ratios. It is based on the object to be identified in an image. The accuracy and speed of the object detector are based on the anchors; fewer anchors are faster than the detectors, but it may reduce accuracy. At the same time, it involves a lot of hyper-parameters which directly affect the IoU score.

- **Anchor free object detection algorithm:** It works on the principle of key-point detection. It generates a CNN heatmap and relies on Non-Max Suppression (NMS) to suppress unnecessary bounding boxes. It is said that anchor-free object detectors find it difficult to detect a dense and overlapping objects. Present-day anchor-free object detectors include CenterNet (Duan et al., 2019), and CornerNet (Law & Deng, 2020).

1.3.2 Image Segmentation

Image segmentation is another prime area of CV, where the whole image is grouped into segments based on some commonality like texture, colour or intensity. It is used to identify the object present in every grouped region for some simplified post-analysis. Each group will be labelled with a class name. Image segmentation is also considered an add-on to image classification tasks. This technique works based on one of the two principles, either discontinuity or similarity.

The discontinuity approach establishes image segmentation by considering edges. Edges are usually found between two different regions. The abrupt changes are identified using the intensity values. It can be detected by identifying an images' points, lines and edges. Likewise, the similarity-based approach works on partitioning images based on similarity. The similarity follows some predefined criteria. A common approach to identify discontinuities is to run a mask over an image. A mask, the same as the Laplacian operator, could be chosen to detect a point in an image. It can be achieved by adaptive gradient, Prewitt (Prewitt, 1970) and Sobel operators (Sobel & Feldman, 2015).

1.3.3 Classification of Segmentation Algorithms

Segmentation algorithms are performed manually by feeding images and other required inputs to segment an image. It all started when digital image processing is clubbed with an optimization algorithm. Figure 1.16 portrays the types of the segmentation algorithm.

Figure 1.16 Classification of Image Segmentation Algorithms

Thresholding-based segmentation: It's the oldest method used to segment an image. The whole image will be divided into two classes in threshold-based segmentation based on the threshold value set. Pixels holding values greater than the threshold are set to '1', and pixel holding values lesser than the threshold is grouped and assigned a value of '0'. This results in a binary image. This process is also known as binarization. This method is well suited when the difference between both classes are vast. This process acts as an initial step for further processes such as, contour detection. Some known thresholding-based techniques include simple thresholding, Otsu binarization (Otsu, 1979) and adaptive thresholding (J. D. Yang et al., 1994).

Edge-based segmentation: This process is also called as edge detection and is carried out using edge detection filters. Popular filters to detect edges included canny, Sobel and Prewitt operators. The output of a simple edge detection algorithm would consist both the background and the edges. After edge detection, some processing must be done to segment the image. The output of edge detection operators is not the fully segmented image. Common edge detection algorithm includes search-based, zero-crossing-based edge detection algorithms (Canny, 1986; X. Zhang et al., 2018).

Region-based segmentation: It works based on similarities detected between adjacent pixels. Based on the similarity, detected pixels will be grouped. This process of segmentation starts with a seed pixel. The immediate boundaries of the seed pixel are classified and grouped based on whether the pixel is found to be similar or dissimilar. In the next cycle of operation, the next set of neighbours are considered and again grouped whether the pixels are similar to the seed or different. The same steps are repeated until the complete image is segmented. Region-based segmentation can be used for two different purposes. With the seed pixel, either more pixels could be added to them to form segments or shrink to merge with the other seeds. Region growing, splitting and merging are some of the segmentation techniques.

Clustering-based segmentation: Clustering-based algorithms are popular as they perform far better than all previously described algorithms. It is comparatively faster when compared to all. It groups pixels based on common attributes found between them. Fuzzy C-means (Bezdek et al., 1984), K-means clustering (Hartigan & Wong, 1979), and improved K-means algorithm (Huang & Su, 2014) are some of the popular examples.

Graph-based segmentation: It characterize the problem using a graph. Assume that the graph is represented as G where G = (V,E), V represents vertices and E represents edges. For segmentation of an image, the elements present in the set V are the pixels. The difference in local attributes like intensity, colour or motion are measured as dissimilarity. The two pixels connected by the edge E are taken into consideration for the dissimilarity assessment. The final result of segmentation is a graph G' where G'= (V',E'). This final result contains distinct regions where v' and E' represents the vertices and edges of the resultant image.

CNN-based segmentation: CNN-based algorithms are popular in image segmentation. It can be divided into three major types as follows:

1. Semantic segmentation
2. Instance segmentation
3. Panoptic segmentation

Semantic Segmentation: This algorithm classifies pixels in an image into semantic classes. In an image, if there are three objects of the same class, then all three objects would be grouped into one group or the same class. For example, if an image of a crowded place is considered, there would be many pedestrians; all pedestrians would be grouped into one class called "pedestrians". This kind of segmentation results in providing less information from the image.

Instance segmentation: In semantic segmentation, the segmentation is done based on the classes, but in contrast, in instance segmentation it is performed based on the number of instances. Consider the same example used for semantic segmentation. If an image of a crowded place is considered, there will be many pedestrians. Instead of grouping all of them into one instance called pedestrians, in instance segmentation, every instance of the pedestrian will be individually segmented. Thus the depth of information extracted from instance segmentation is more than semantic segmentation.

***Panoptic segmentation*:** Panoptic segmentation is considered as a blend of instance and semantic segmentation, where every pixel is separately segmented and assigned a separate label along with an identifier for every instance. Few or more objects may be overlapping. In such cases, the discrepancy is managed by favouring the instance of the object. Panoptic segmentation is considered an emerging area of research.

1.4 ARTIFICIAL INTELLIGENCE IN THE FIELD OF ENDOSCOPY

Pacing towards AI in endoscopy, various researchers proposed several research outcomes for scope guidance during colonoscopy (Yarze, 2022), polyp detection (Ijspeert et al., 2017; Kim et al., 2021; Sikka et al., 2008), early cancer identification and identifying the stage of cancer (Davidson et al., 2021; Gao et al., 2019; Hassan et al., 2010; Ijspeert et al.,

2017) and polyp characterization (Bhandari et al., 2018; Korbar et al., 2017; Nazarian et al., 2021).

In every single instance, the system generates either an image or a sequence of the images. However, there are possibilities that endoscopists may make inappropriate interpretations during the procedure due to fatigue, or a novice endoscopist may make incorrect decisions. To assist the endoscopist, many imaging modalities in endoscopy have emerged, such as, Narrow Band Imaging (NBI), Auto Fluorescence Imaging (AFI) and Magnifying Endoscopy (ME). 3D imaging in the endoscopy field has recently improved CV systems' performance. Amidst all the technological advancements and improved diagnostic capabilities, endoscopists must be trained to handle the advanced systems effectively.

On the other hand, artefacts are said to occur due to recent advancements in miniaturizing the hardware components, mishandling of imagers and some natural phenomena. These artefacts occlude the underlying tissue and increase the possibilities of false detection rates in automated Computer Aided Diagnostic (CAD) systems, and increase manual procedure time. Thus the detection and segmentation of these artefacts emerged as a new area of research. Especially in endoscopic images, such artefacts are said to occur in clusters and distinct.

AI is frequently used by CAD systems to recognise aberrant structures. Features are extracted and fed into a classifier after going through a number of pre-processing processes; the classifier then delivers the diagnosis.

1.5 MOTIVATION OF THE RESEARCH

The primary challenges identified behind implementing such an AI-powered artefact detection and segmentation system are as follows:

- The tissue appears different in different modalities of imaging, namely NBI, White light and AFI.
- Most of the artefacts overlap each other, which leads to over-estimation of the bounding box.
- Occurrence of different artefacts are due to various physical phenomena.

- Frequent occurrence of artefacts in clusters in every frame.
- The size, location, and nature of artefacts vary, lacking prominent geometrical structures.

1.6 OBJECTIVES OF THE RESEARCH

The following are the objectives set:

- Set benchmarks by tuning hyperparameters and training existing object detection algorithms such as, YOLOv3, YOLOv4 and Faster R-CNN on endoscopic images affected by artefacts.
- To design an ensembled architecture for endoscopic artefact detection for outstanding results.
- To design an ensembled architecture for endoscopic artefact segmentation.
- To implement a simple restoration and classification pipeline to project the necessity of artefact detection and segmentation.

1.7 CONTRIBUTION OF THE THESIS

The proposed research aims to detect and segment artefacts present in endoscopic images using images from a custom dataset and Endoscopic Artefact Detection (EAD) dataset. It is divided into three stages,

a. **Custom dataset preparation**: Images from hospitals are collected, annotated and a custom dataset is prepared. The images available in the EAD public datasets are not just sufficient to train an model and propose effective results. Hence a new dataset is created and combined with images from EAD to improve the efficiency of artefact detectors in terms of mAP and IoU.

b. **Artefact detection**: Recently, deep learning based techniques give a better detection results. In this work, artefact detection is performed using deep learning based object detection algorithms. The pre-trained deep learning based object detection algorithms are retrained to detect artefacts like saturation, specular reflections, contrast, blur, bubbles, blood, instrument and miscellaneous artefacts. Deep learning based object detection algorithms like YOLOv3, YOLOv4 and

Faster R-CNN are adopted for this research. Images from the dataset are divided into train and test dataset. The training dataset is used to train endoscopic artefact detectors. The test set is used for evaluating model performance. Predictions of all three artefact detectors are combined using ensemble technique. Performance metrics like mean Average Precision (mAP) and Intersection over Union (IoU) and inference time are considered for evaluating the model.

c. **Artefact segmentation**: Deep learning based segmentation algorithms shows improved performance in the recent areas especially in the field of healthcare. Deep learning based segmentation algorithms like U-Net with Efficient Net B3, U-Net with SEREsNeXt101 and Link-Net with Efficient-net B3 are adopted for artefact segmentation. Endoscopic artefacts like saturation, specular reflections, contrast, bubbles and instrument are considered for segmentation. Images from EAD dataset is adopted for training and testing. Performance evaluation is achieved through metrics like F2 score and Jaccard Score.

To insist the need for artefact detection and segmentation in endoscopic imaging pipeline a simple polyp classifier is designed using a simple CNN architecture. The polyp classifier is trained with images with and without polyp collected from kvasir-SEG and EAD dataset. Artefacts are restored in few images using traditional fast marching algorithm and the same set of images without restoration is considered for training the polyp classifier with and without artefacts. The classifier's performance is evaluated using metrics like accuracy.

1.8 ORGANIZATION OF THE REPORT

The report comprises the following seven chapters:

- Introduction
- Literature review
- Dataset
- Endoscopic artefact detection
- Endoscopic artefact segmentation
- Artefact restoration and polyp classification
- Conclusion and future direction

Chapter 1 explains the endoscopic process, various artefacts that can be seen in endoscopic images, an introduction to object detection, and the necessity of detecting and segmenting artefacts.

Chapter 2 elucidates various research works carried out in the field of multiple artefact detection and segmentation with a short note on every architecture proposed by the researchers.

Chapter 3 explains the public datasets available for multi-class endoscopic artefact detection, its advantages and shortcomings. This chapter also deliberates the need to curate a new dataset, annotation protocols and software used with sample images.

Chapter 4 ponders the architecture proposed for endoscopic artefact detection, its training procedure and various hyperparameters set for training, the details of the proposed ensemble model along with its results.

Chapter 5 discusses the various image segmentation algorithms simulated with and the proposed DL-based artefact segmentation model with results.

Chapter 6 explains the need for artefact restoration through a simple application and explains the classification model trained to classify polyps before and after restoration along with its performance.

Chapter 7 concludes with a summarization of the findings and discusses potential enhancements to be made in the future.