INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	Vanshali Sharma
Roll Number	:	186101103
Programme of Study	:	Ph.D.

Thesis Title: Anomaly Detection in Endoscopic Videos: Keyframe Extraction to Designing Clinical and Synthetic Datasets

Name of Thesis Supervisor(s)	:	Prof. Pradip K. Das and Prof. M.K. Bhuyan
Thesis Submitted to the Department/ Center	:	Department of Computer Science and Engineering
Date of completion of Thesis Viva-Voce Exam	:	27/08/2024
Key words for description of Thesis Work	:	Medical image analysis, deep learning, endoscopic videos, colorectal cancer, diffusion models

SHORT ABSTRACT

Gastrointestinal (GI) cancers, specifically colorectal cancers (CRC), are prevalent and significant contributors to global cancer-related deaths. CRC originates from pre-malignant polyps, which can be detected through a colonoscopy procedure, during which videos of a patient's colon are captured. However, analyzing screening videos for related diagnosis and treatment faces challenges due to a large proportion of low-quality data, risking human review errors. Further, the low-quality data and the limited availability of large-scale annotated datasets pose significant hurdles in building automated computer-aided diagnostic systems. This thesis addresses these challenges while aligning with standard clinical procedures. To maintain this uniformity, we mimic these manual procedures in our proposed automated pipeline and present solutions to problems encountered at different stages.

A standard clinical analysis of colonoscopy videos generally begins with manually reviewing recordings and gradually confines the analysis to keyframes for retrospective treatments. Hence, initially, this thesis focuses on automating this task to reduce the clinicians' burden. However, the keyframe count could be significantly low in some video recordings captured under extremely unfavorable conditions. Therefore, techniques to extract obscured details of uninformative frames are proposed. Following the pre-processing stage, the thesis addresses issues related to automated diagnostic systems, enhancing lesion detection, localization, segmentation, and classification outcomes. One of the crucial concerns in the literature is the lack of reproducibility and fair comparison across different segmentation techniques due to inconsistent evaluation datasets, as revealed in our case study. Motivated by this, our focus is on resolving dataset availability issues, as a good-quality, diverse dataset enhances lesion detection performance and promotes reproducibility. Thus, this thesis incorporates effective keyframe selection and other pre-processing techniques, meticulous dataset curation, and synthetic image generation. The four significant contributions of the thesis are highlighted below.

First, a multi-stage framework is presented that focuses on *keyframe extraction* to select good-quality, non-redundant frames and enforce diversity in the final frames for analysis. The framework enhances polyp

detection and polyp localization outcomes while reducing processing time. Our novel multi-scale attention-based localization model, *YcOLOn*, further improves the localization task within the framework.

Second, we propose two approaches to overcome the limitations of our keyframe extraction framework. These techniques focus on obtaining obscured clinical details from uninformative frames with artifacts. One method is an *adversarial-based* approach that focuses on translating uninformative frames into clinically significant frames. This helps in improved polyp localization. With a similar aim to deal with artifacts, a *DWT-based encoder-decoder* architecture is designed to segment specular highlights while overcoming the issue of overexposed regions in the colonoscopy images.

Third, a *case study* is presented that analyzes different polyp and instrument segmentation algorithms involved in two competitions conducted in the years 2020 and 2021. Such analysis provides an opportunity to compare different state-of-the-art techniques on the same dataset for transparency and reproducibility. Additionally, we release an open-access multi-class dataset, *GastroVision*, for computer-aided diagnosis of GI cancer. It comprises 8000 images from 27 classes covering pathological and normal findings, anatomical landmarks, and cases from therapeutic interventions.

Fourth, two frameworks are proposed to generate synthetic medical images using diffusion models, overcoming the lengthy procedures to acquire real medical datasets. The first framework, *ControlPolypNet*, leverages easily accessible non-polyp frames and converts them into hard-to-find polyp images. The generated polyp images are utilized to augment a real dataset to perform a downstream task of polyp segmentation. The second framework, *PathoPolyp-Diff*, is a text-controlled model to generate diverse polyp images covering different pathologies, imaging modalities, and quality. The generated images are used to augment real datasets to enhance pathology based polyp classification. Further, cross-class label learning is introduced, which learns features from other classes without additional annotations.