Can Adversarial Networks Make Uninformative Colonoscopy Video Frames Clinically Informative? (Student Abstract)

Vanshali Sharma, M.K. Bhuyan, Pradip K. Das
Indian Institute of Technology Guwahati, Assam, India-781039
{vanshalisharma, mkb, pkdas}@iitg.ac.in

Abstract
Various artifacts, such as ghost colors, interlacing, and motion blur, hinder diagnosing colorectal cancer (CRC) from videos acquired during colonoscopy. The frames containing these artifacts are called uninformative frames and are present in large proportions in colonoscopy videos. To alleviate the impact of artifacts, we propose an adversarial network based framework to convert uninformative frames to clinically relevant frames. We examine the effectiveness of the proposed approach by evaluating the translated frames for polyp detection using YOLOv5. Preliminary results present improved detection performance along with elegant qualitative outcomes. We also examine the failure cases to determine the directions for future work.

Introduction
Colonoscopy is a minimally invasive procedure widely adopted for polyp detection to diagnose colorectal cancer (CRC). In a colonoscopy, diagnostic accuracy relies on the correct analysis of the acquired recordings. However, the traditional assessment approaches by physicians suffer from inter-observer variations and demand extensive manual efforts. In recent years, accessibility to several colonoscopy datasets has paved the way for many machine learning based research works for automated CRC detection. However, the well-trained models proposed in the existing works still report limited diagnostic success. This limited success of automated methods is attributed to low-quality frames in the video samples, which contain artifacts, namely, ghost colors, low-illumination, interlacing due to camera motion, and fecal depositions due to inadequate patient preparation.

To overcome the low-quality frames, some related fields of laparoscopy and endoscopy followed keyframe selection (Ma et al. 2020) or performed super-resolution (Almalioglu et al. 2020), but no work in the colonoscopy domain explored the idea of extracting obscured clinical details from such low-quality uninformative video frames. Therefore, our work investigates whether GANs can convert uninformative frames to informative frames. In this direction, we propose a GAN-based image-to-image translation approach to generate informative frames from the degraded frames of the colonoscopy videos. We highlight the cases where GANs fail and where it helps, which gives us directions for future work. The main contributions are summarized below:

1. To the best of our knowledge, this is the first framework to address the issue of uninformative colonoscopy frames using adversarial networks.
2. We investigate the impact of translating uninformative frames on polyp detection performance and discuss future directions in this context.

Methodology
The overview of the proposed framework is shown in Fig. 1. Given the uninformative colonoscopy frames \( \{a_i\}_{i=1}^{N} \) from domain A, the aim is to learn a mapping function \( G_{AB} : A \rightarrow B \) to generate frames such that the data distribution of obtained frames is indistinguishable from that of informative colonoscopy frames \( \{b_j\}_{j=1}^{M} \) of domain B. Due to the unavailability of paired data, our work is inspired by the unpaired translation approach of CycleGAN (Zhu et al. 2017). Hence, another mapping function \( G_{BA} : B \rightarrow A \) is also introduced. Our implementation involves ResNet-based generators and PatchGAN discriminators \( D_A \) and \( D_B \). The CycleGAN objective integrates adversarial loss and cycle-consistency loss. The adversarial loss can be expressed as:

\[
L_{adv}(G_{AB}, D_B) = \mathbb{E}_{b \sim p_{data}(b)}[(D_B(b) - 1)^2] + \mathbb{E}_{a \sim p_{data}(a)}[(D_B(G_{AB}(a)))^2] \tag{1}
\]
implement image-to-image translation using cycle-consistent adversarial networks. Since the artifacts in colonoscopy video frames alter the various aspects of images, such as structure, texture, and color, this work lays the foundation for a more interesting future work of developing a standalone model to address all the artifacts in one go.

Acknowledgments
Vanshali Sharma would like to thank the Department of Science and Technology, Government of India, for providing the INSPIRE fellowship (IF190362).

References


