

# 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

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

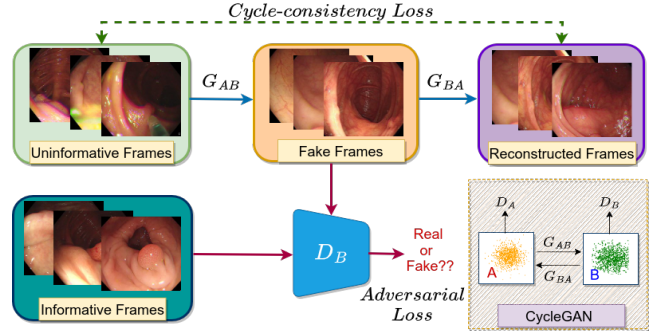


Figure 1: The proposed framework contains two generators  $G_{AB}$  and  $G_{BA}$  and two discriminators  $D_A$  and  $D_B$ .

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:

$$L_{adv}(G_{AB}, D_B) = \mathbb{E}_{b \sim p_{data}(b)} [(D_B(b) - 1)^2] + \mathbb{E}_{a \sim p_{data}(a)} [(D_B(G_{AB}(a)))^2] \quad (1)$$

$G_{AB}$  aims to translate uninformative frames such that they appear similar to the informative frames, while  $D_B$  tries to distinguish the translated frames from the high-quality, informative frames of domain B. In other words,  $D_B$  is trained to minimize  $L_{adv}(G_{AB}, D_B)$  and  $G_{AB}$  is trained to minimize  $\mathbb{E}_{a \sim p_{data}(a)} [(D_B(G_{AB}(a)) - 1)^2]$ .

To ensure cycle-consistency and to reduce randomness in mapping, a cycle-consistency loss is used, which is given by:

$$L_{cyc}(G_{AB}, G_{BA}) = \mathbb{E}_{a \sim p_{data}(a)} [\|G_{BA}(G_{AB}(a)) - a\|_1] + \mathbb{E}_{b \sim p_{data}(b)} [\|G_{AB}(G_{BA}(b)) - b\|_1] \quad (2)$$

An identity mapping loss is also added to help preserve color in translated images. With this model, we intend to determine the clinically relevant details obscured by the artifacts. Furthermore, we carried out the following investigations:

1. Polyp detection is performed using YOLOv5 (Ultralytics 2020) to determine the impact of GAN-translated frames.
2. Qualitative analysis is done to identify the artifacts successfully handled by the CycleGAN and analyze the ones that still persist in the translated frames.

## Experiments

To assess the effectiveness of the adversarial approach in mitigating the impact of artifacts, we conducted experiments using a publicly available SUN database (Misawa et al. 2021) consisting of 1, 09, 554 non-polyp and 49, 136 polyp frames. In addition to the localization information, the polyp frames are manually annotated by experts as informative or uninformative. We used only the polyp frames with a patient-wise split. The translation is done on a Titan Xp GPU at 14 frames per second. We report the results based on visual perception and consider feature space representation by evaluating the polyp detection outcomes using YOLOv5. We conducted training and testing in two scenarios using: a) Raw frames comprising both high and low-quality frames and b) Translated frames along with high-quality frames. The results in Table 1 show that the translated frames complement the detection ability of YOLOv5 in terms of precision, recall, F1-score, and mAP@0.5. The detector correctly identified more polyps with lower deviations, presenting a more robust model. However, this is achieved with slightly less precise bounding boxes, as indicated by a minor decrease in mAP@0.5:0.95. Fecal depositions, ghost colors, and low-illumination are significantly reduced using CycleGAN, as shown in Fig. 2. However, motion blur and interlacing are not handled adequately in the process. This could be overcome by adopting blur removal approaches.

## Conclusion and Future Work

In this work, we propose a GAN-based framework to translate uninformative colonoscopy frames into clinically significant frames. We showed that the translated frames improve polyp detection F1-score and mAP@0.5, with negligible reduction in mAP@0.5:0.95. We analyzed the types of artifacts where the CycleGAN performed well and identified the scope of improvements. Since the artifacts in colonoscopy

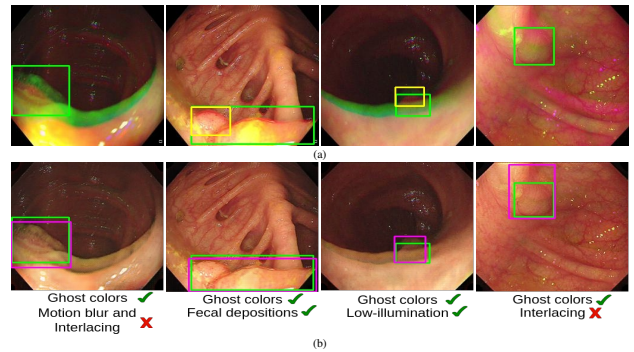


Figure 2: Detection performance using: (a) Raw frames and (b) Translated frames. Green bounding boxes denote the ground truth. Ticks and cross marks represent the successful and unsuccessful artifacts translations, respectively.

Metrics	SUN Database	
	Raw Frames	Translated Frames
<b>Precision (%)</b>	92.03±0.60	<b>93±0.87</b>
<b>Recall (%)</b>	88.9±3.12	<b>90.2±1.3</b>
<b>F1-score (%)</b>	90.4±1.51	<b>91.57±0.38</b>
<b>mAP@0.5 (%)</b>	95.37±0.95	<b>95.6±0.21</b>
<b>mAP@0.5:0.95 (%)</b>	<b>57.53±0.32</b>	57.07±0.31

Table 1: Comparative analysis of polyp detection results

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

- Almalioglu, Y.; Ozyoruk, K. B.; Gokce, A.; Incetan, K.; Gokceler, G. I.; Simsek, M. A.; Ararat, K.; Chen, R. J.; Durr, N. J.; Mahmood, F.; et al. 2020. EndoL2H: deep super-resolution for capsule endoscopy. *IEEE Transactions on Medical Imaging*, 39(12): 4297–4309.
- Ma, M.; Mei, S.; Wan, S.; Wang, Z.; Ge, Z.; Lam, V.; and Feng, D. 2020. Keyframe extraction from laparoscopic videos via diverse and weighted dictionary selection. *IEEE Journal of Biomedical and Health Informatics*, 25(5): 1686–1698.
- Misawa, M.; Kudo, S.-e.; Mori, Y.; Hotta, K.; Ohtsuka, K.; Matsuda, T.; Saito, S.; Kudo, T.; Baba, T.; Ishida, F.; et al. 2021. Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). *Gastrointestinal Endoscopy*, 93(4): 960–967.
- Ultralytics. 2020. YOLOv5. <https://github.com/ultralytics/yolov5>. *git/*. Accessed: 25 Feb 2022.
- Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.