Improving CCE video review time with a model based on frame similarity

Pere Gilabert[1] PERE.GILABERT@UB.EDU
Santi Seguí[1] SANTI.SEGUI@UB.EDU
[1] *Facultat de Matemàtiques i Informàtica, Universitat de Barcelona.*

Editors: Under Review for MIDL 2022

Abstract

Many advances have been made in the detection of pathologies through the use of the Colon Capsule Endoscopy (CCE), a non-invasive procedure that allows physicians to view the entire interior of a patient's digestive system without the need for sedation. In this article we focus on the subsequent process, assuming that we already have a good model to detect a pathology (polyps in this case) and see how to improve the video review process by re-sorting the high score frames. With a simple sorting method, we obtain a 7% improvement compared to the linear method where the frames are reviewed in decreasing order of score. This accuracy boost occurs in the first 100 frames, which allows the videos to be reviewed more quickly and efficiently.

1. Introduction

Colon Capsule Endoscopy (CCE) is a relatively new technology that allows physicians to see the inside of the colon with no need to perform a colonoscopy. It consists of a capsule-shaped device with two cameras that records the entire digestive system from the time it is swallowed until it is naturally expelled. Later, experienced personnel review the video recorded by the capsule and can identify injuries, diseases or health problems. One of the most important advantages of CCE is that patients do not need to be sedated as it is a minimally invasive procedure.

The videos are about eight hours long, the reviewing process is tedious and time consuming and it requires experienced personnel. Lesions can appear anywhere in the video and a high level of focus is required when reviewing them.

Recently, different studies have been proposed to detect Crohn's disease, bleeding, tumours, polyps, etc. in images obtained with CCE. However, all these applications focus on improving a binary metric of pathologic / non-pathologic and do not pay attention on how these decisions impact the review of the videos by the medical staff in terms of both time and accuracy.

In this paper we present an initial idea to solve this problem. Based on the super-expert metric presented in (Gilabert et al., 2022), we developed an algorithm capable of reordering videos obtained with CCE for polyp detection.

© 2022 P. Gilabert & S. Seguí.

2. Methods

The data we have used are from a retrospective study. We used a total of 18 videos obtained with PillCam COLON 2 capsule (Medtronic). Following the procedure of (Gilabert et al., 2022), medical experts labeled all images containing a polyp. They also assigned a unique identifier to images of the same polyp. The experts reported a total of 52 different polyps of varying sizes from 1cm to 10cm. There are polyps appearing in a very small number of frames (1-3 frames) to polyps appearing over many consecutive images (more than 100 images). Details can be found in (Gilabert et al., 2022).

Let V be a video with n frames: $f_*^1;:::;f_*^n$ temporally ordered. Let M be a polyp detector model that assigns to each frame a score, $M(f_*^i)$. M induces a new ordering of the video, $\hat{f}_*^1;:::;\hat{f}_*^n$ where each frame has the same or less score of being a polyp than the previous one, i.e., $M(\hat{f}_*^i) \geq M(\hat{f}_*^{i+1})\ \forall i = 1;:::;n-1$.

Moreover, let $P_V = \{p_1;:::;p_k\}$ be the set of different polyps in video V. Each frame of the video has a P_V label if it contains a polyp, or a non-polyp label label, $p_\emptyset$. This label is indicated in the sub-index, e.g., $f_{p_\emptyset}^1$. Let $\hat{P_V} = P_V \cup \{p_\emptyset\}$ be the set of all possible labels of a frame.

The goal is to find a new ordering $\bar{f}_{q_1}^1;:::;\bar{f}_{q_n}^n;\ q_i \in \hat{P_V}\ \forall i = 1;:::;n$ such that the number of frames required to display all polyp labels is the minimum possible, i.e, we want to find an ordering such that the value m is minimal:

$$\bar{f}_{q_1}^1;:::;\bar{f}_{q_m}^m \ \Big|\ \bigcup_{i=1}^{m} q_i \supset P_V;\ m \leq n \tag{1}$$

To transform the initial ordering $\hat{f}_*^1;:::;\hat{f}_*^n$ to the new ordering $\bar{f}_*^1;:::;\bar{f}_*^n$ we use a **similarity distance metric** between frames computed using the image content. Let S be a model that extracts an embedding from each image, $S(f^i) = e_{f^i}$, we compute the similarity between two frames f^i and f^j as:

$$d_s(f^i; f^j) = ||S(f^i) - S(f^j)||_2 = ||e_{f^i} - e_{f^j}||_2 \tag{2}$$

Then, to reorder the sequence of frames $\hat{f}_*^1;:::;\hat{f}_*^n$ we compute the similarity distance between each frame and all the next ones and we modify its score if it is below a threshold, $_S$, i.e.,

$$\begin{aligned} &\bar{f}_*^1;:::;\bar{f}_*^n \leftarrow \hat{f}_*^1;:::;\hat{f}_*^n \\ &\text{score}(\bar{f}_*^j) \leftarrow \text{score}(\bar{f}_*^j) d_s(\bar{f}_*^i; \bar{f}_*^j) \quad \forall i \geq 1 \quad \forall j > i \quad (\text{ if } \quad d_s(\bar{f}_*^i; \bar{f}_*^j) < \ _S\ ) \end{aligned} \tag{3}$$

At each step i we reorder the sequence of frames $\bar{f}_*^{i+1};:::;\bar{f}_*^n$ according to this new score. To avoid underflow problems we use logarithms in Equation 3.

3. Results and Conclusions

Figure 1 shows the result of applying this process using three different similarity models, S, pretrained using ImageNet: ResNet50(0.3), EfficientNetB3(0.4), ViT-B/16(0.4). Inside the parenthesis we indicate the value of obtained after a gridsearch process. We used the