Poster ID: 05

Study on Gastric Cancer Detection from Gastroendoscopic Videos Based on Deep Learning

Youshen Zhao, 2nd Master Student, Supervised by Xin Zhu, Bioinformation Engineering Lab, The University of Aizu

Abstract

Gastric cancer is a global health problem and the 3^{th} cause of cancer death among males and females. Patients with early stage gastric cancer have 5-year survival rate greater than 90%. Therefore, early gastric cancer is a significant research in medicine field. In this study, we collaborate with endoscopists from Kyoto Prefectural University of Medicine, Kyoto, Japan, and apply YOLOv5 object detector to diagnose gastric cancer. The results shows that sensitivity based on positive videos achieved 90.6%, and 92.8% of the specificity based on negative frames.

Keywords: Gastric cancer detection, deep learning, object detection, medical imaging

Introduction

Gastric cancer (also known as stomach cancer), is a global health problem, which is the 5^{th} most commonly diagnosed cancer and the 3^{rd} cause of cancer death among males and females, equating to 1 in every 12 deaths.

Over the last decade, AI has been successfully applied in biological data analysis, including medical signal processing, medical imaging, surgical assistance, etc. It is essential to develop gastric cancer detector based on deep learning framework to help endoscopists screening tumor in the process of diagnosis.

In this study, we use the dataset provided from endoscopists applied YOLOv5 object detector to develop a gastric cancer framework.

Approach

In object detection, we are given a set of labeled data $\mathcal{D}_l = \{(x_i^l, y_i^l)\}_{i=1}^{n_l}$ where x and y denote image and ground-truth (classification labels and bounding box coordinates) respectively. The goal of gastric cancer object detector is to train a model on labeled data which can detect tumors on unlabeled images and videos.

The output of classification probability vector after softmax operation is denoted as $f_{cls}^{p,r,c,d}(A,\theta)$ which corresponding to the $p-$th pyramid, $r-$th row, $c-$th column and $d-$th default box, A is augmentation of dataset. The localization result for the $k-$th bounding box $f_{loc}^{k}(A,\theta)$ consists of $[t_x, t_y, t_w, t_h]$, where t_x and t_y represent the center coordinate of an object in image, and t_w and t_h are scale coefficients of an object bounding box. The final loss is shown as follow:

$$\mathcal{L} = \mathcal{L}_{loc} + \mathcal{L}_{cls}$$

where

$$\mathrm{L}_{cls} = f_{cls}^{p,r,c,d} = -\omega_n[y_n \cdot \log\sigma(x_n) + (1-y_n)\cdot\log\sigma(x_n)],$$
$$\mathrm{L}_{loc} = 1 - IoU(box_p, box_t)$$

The overview of gastric cancer detector is shown as Figure 1.

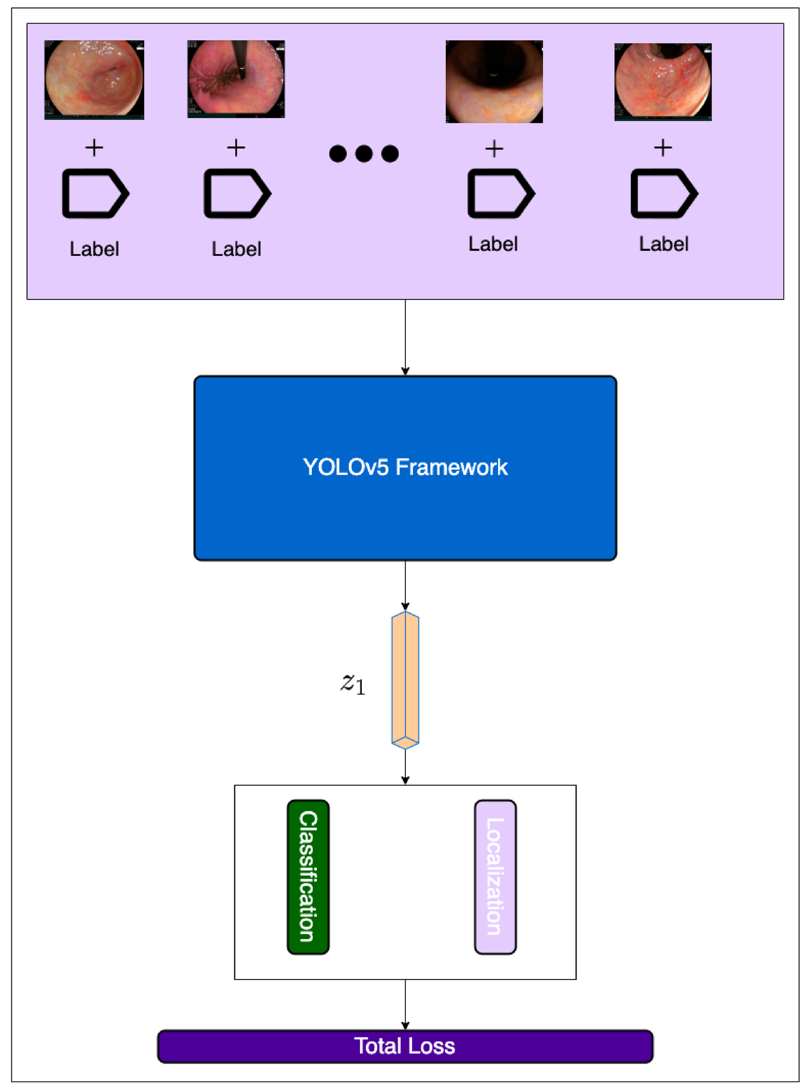

Figure 1: Overview of gastric cancer detector

Dataset Preparation

The training dataset includes in total 341 patients, were obtained from Kyoto Prefectural University of Medicine, Kyoto, Japan, from May 2017 to Nov 2020. All videos are consist of Narrow Band Imaging (NBI), Flexible spectral Image Color Enhancement (FICE), Blue Laser Imaging (BLI) and White Light Imaging (WLI), as shown in Fig 2. After selection, 8,697 images were collected as a training image dataset from 279 cases.

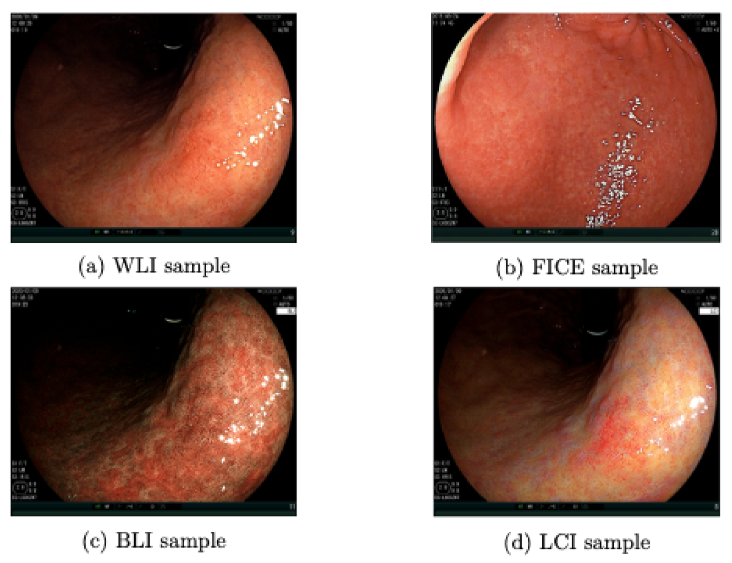

Figure 2: Examples of different endoscopic images

Results

We set thresholds to 0.4, 0.5 and 0.6 respectively, the model are evaluated with 32 LCI videos. The sensitivity is calculated with 32 cases without patients overlapped, the specificity is calculated from 50001 frames which are extracted from negative clips based on 32 LCI videos above. The formulas are show as follows:

$$sensitivity = \frac{TP}{TP+FP}$$

$$specificity = \frac{TN}{TN+FN}$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

The results are shown in Table follow:

Threshold	Sensitivity	Specificity
0.4	90.6%	89.3%
0.5	90.6%	92.8%
0.6	81.3%	95.7%

Conclusion

- Analysis of performance on videos: Although the sensitivity of models on 32 test videos is 90.6%, we still need to focus on the false positive rate, or specificity. We divide all negative frames from videos to evaluate specificity, from negative frames which consist of 50,001 frames, from the specificity the experiment results we can find it is extremely difficult to improve both sensitivity and specificity simultaneously and essential to find a balance. The video-based evaluation can reflect the progress of clinical use.
- Frame linked detection: In this study, detection in videos is the independent detection on frames, object tracking algorithms in videos need to be considered such as sequence model construction among frames, vision in transformer (ViT) can be considered to apply in this task.

References

[1] Alexey Bochkovskiy, etc. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

[2] Peng Jin, etc. Artificial intelligence in gastric cancer: a systematic review. *Journal of cancer research and clinical oncology*, 146(9):2339–2350, 2020.

Future Work

- Considering that the false detection rate of gastric cancer detection is higher than other detection tasks, we will try to use bimodal to improve the classification ability of the model.
- In future, we will consider combining image and un-iamge data to train model, in gastric cancer detection task, we read some medicine history from some patients and found that a great number of patients had history of colorectal cancer. It can help us to detect gastric cancer more accurately.

Contact Information

Web http://bitlab.u-aizu.ac.jp/

Email zhuxin@u-aizu.ac.jp, m5242118@u-aizu.ac.jp