A Real-time Polyp Detection Method Based on GhostAtt-YOLOv8

Phuong Thao Nguyen
Waseda University

Hiroshi Watanabe
Waseda University

Abstract

Convolutional Neural Network (CNN) in medical image processing has lately received a lot of interest. Computer-aided polyp detection in gastrointestinal endoscopy has been the subject of research over the past few decades. However, despite significant advances, automatic polyp detection in real-time is still an unsolved problem. In this paper, we propose a Deep Learning method for reliable real-time polyp detection on endoscopic images and videos. We improve the performance of YOLOv8 model by modifying YOLOv8 model architecture with Ghost Convolution and Spatial and Channel Attention mechanisms (GhostAtt-YOLOv8). These techniques are integrated into the backbone network to enhance detection result. The proposed method is applied on Showa University and Nagoya University polyp database (SUN) dataset. Experimental results show that a better performance is archived with mAP@50 of 80.13% compared to the original YOLOv8, and FPS of our proposed model is 294, faster than original YOLOv8.

Keywords: Polyp Detection, Endoscopy, Deep Learning, YOLOV8, Ghost Convolution, Attention

1. Introduction

Colorectal cancer (CRC) ranks third in terms of cancer-related mortality. While colon cancer has a five-year survival rate of approximately 68%, the rate for cancer stomach is only 44% [1]. To mitigate CRC-related fatalities, one of the most effective strategies is identifying and removing precancerous lesions, such as colon polyps, which have the potential to progress into CRC at a later stage. Therefore, early detection of CRC is essential for improving survival rates. Colonoscopy is an invasive medical procedure in which an endoscopist uses a flexible endoscope to inspect and treat the colon. It is thought to be the best diagnostic tool for a colonoscopy for the early detection and treatment of polyps. Therefore, gastroenterologists often choose colonoscopic screening over other methods.

Polyps are abnormal tissue growths that protrude from the mucous membrane. They can occur anywhere in the gastrointestinal (GI) tract, but are most frequently detected in the colorectal area, and are often considered to be a predecessor of CRC. Larger polyps can usually be detected and removed. Some polyps, however, may be overlooked due to their small size, low image quality due to the colonoscopy device, or the gastroenterologists' skills. Other challenges include the patient variability and presence of different sizes, shapes, textures, colors, and orientations of these polyps. For the aforementioned reasons, computer-based detection methods come to the aid of physicians for a more accurate diagnosis. In this paper, we propose an improved model, GhostAtt-YOLOv8 based on YOLOv8 [2] to solve the polyp detection task in a real-time manner. Ghost Convolution [3] combined with Spatial and Channel Attention mechanism (CBAM) [4] is introduced to the backbone network.

2. Related Works

Over the last two decades, there has been a lot of effort put into developing effective methods and algorithms for automated polyp detection. Earlier research concentrated on polyp color and texture, employing handmade descriptors-based feature learning. Methods based on CNNs have recently gained a lot of attention and have become the most popular option for both production and those competing in public competitions.

Shin et al. [5] applied Inception ResNet as a transfer learning approach and implemented post-processing techniques to enhance the polyp detection result during colonoscopy. Liu et al. [6] presented a YOLOv3-based approach by fusing a two-dimensional CNN-based real-time object detection network with spatio-temporal data to address the issue of missed polyp detection and enhance accuracy. In [7], the authors proposed a method for real-time polyp detection utilizing YOLO-based models for small datasets. Nogueira-Rodríguez et al. [8] conducted an extensive analysis of all accessible data to evaluate the effectiveness of CNN on polyp detection task.

3. Proposed Method

3.1. Overview of YOLOv8

YOLOv8 [2] is the latest version of the YOLO family of detection models. Much like its predecessor, YOLOv5 [9], its architecture is comprised of a backbone, neck and detection head. YOLOv8 introduces a new architectural de-

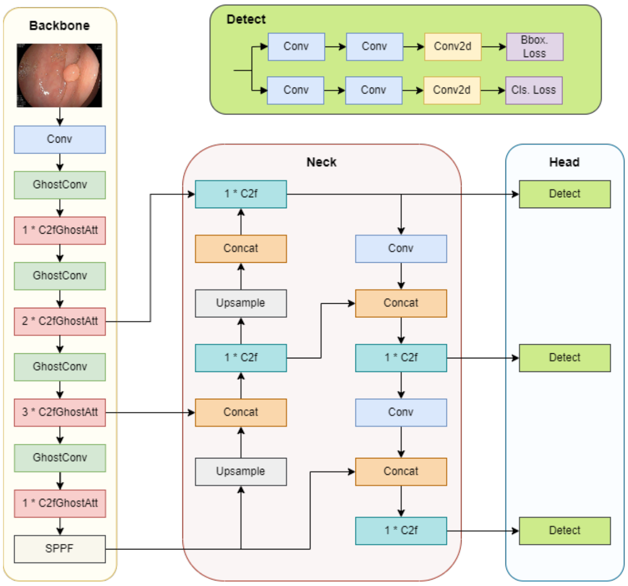

Figure 1: The architecture of GhostAtt-YOLOv8.

sign, enhanced convolutional layers in the backbone, and an upgraded detection head, positioning it as the top choice for real-time object detection. The model uses CSPDarknet53 with an SPPF layer as the backbone network, same as YOLOv5, but the C3 module has been replaced by the C2f module. The C2f module (cross-stage partial bottleneck with two convolutions) merges high-level features with contextual information, thereby enhancing detection accuracy. YOLOv8 adopts an anchor-free model with a decoupled head, enabling independent processing of objectness, classification, and regression tasks. This design allocates each branch to its specific task, ultimately boosting the overall accuracy of the model.

3.2. GhostAtt-YOLOv8

The framework of GhostAtt-YOLOv8 is illustrated in Fig. 1. The detailed architecture is presented below.

3.2.1. Ghost Attention Convolution

Recently, many CNN architectures have been designed to reduce the number of parameters and required resources. Inspired by Ghost Convolution (GhostConv) [3], (shown in Fig. 2a), we propose a newly efficient Ghost Attention Convolution (GhostAttConv), which its architecture is illustrated in Fig. 2b. The Ghost Convolution reduces computational complexity by using cheap operations to extract feature maps. However, it does not increase other useful information because cheap operations only copy inherent features from the primary convolution. For the above reason, the Convolution Block Attention Module (CBAM) [4]

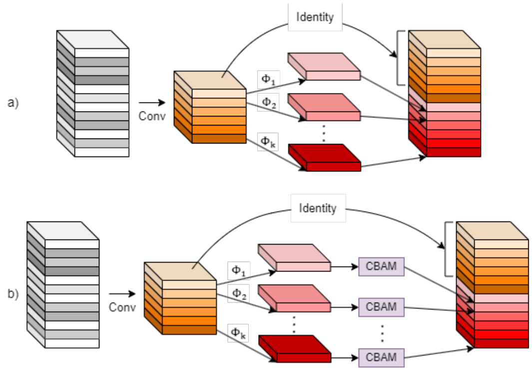

Figure 2: **a)** Ghost Convolution. **b)** Ghost Attention Convolution. ϕ represents the cheap operations.

is added after each cheap operation in the Ghost Convolution to elevate feature information. CBAM is a simple yet effective attention module and can be integrated into many CNN architectures, which consists of Channel and Spatial Attention mechanism. Integrating CBAM after the cheap operations enables the model to focus more on important areas in the input image, emphasizing essential polyp features while reducing the influence of unimportant or noisy sections. This module also efficiently manages computational resources, resulting in enhancement in both accuracy and efficiency. Ghost Attention Convolution is utilized in the backbone network of GhostAtt-YOLOv8 to extract crucial features of the polyps in the image.

3.2.2. Ghost Attention Bottleneck

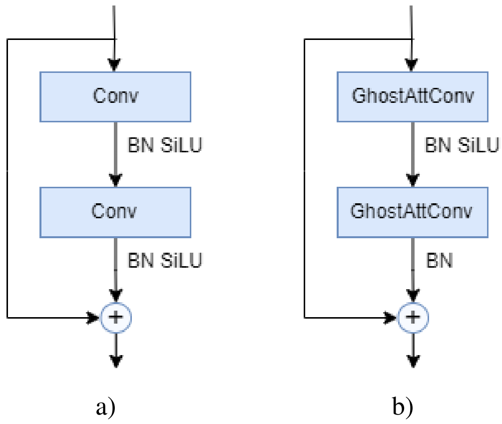

Figure 3: **a)** Bottleneck. **b)** Ghost Attention Bottleneck.

We propose Ghost Attention Bottleneck (GhostAttBottleneck), as shown in Fig. 3b. Compared to the original Bottleneck (Fig. 3a), Ghost Attention Bottleneck replaces the Convolution layer with Ghost Attention Convolution layer.

Batch Normalization and SiLU activation function is used after the first Ghost Attention Convolution, while only Bach Normalization is performed after the second Ghost Attention Convolution to achieve the linearity. The first Ghost Attention Convolution increases the number of channels after extraction, while the number of channels is reduced by the second, then is connected with the input features by the residual connection. Bottleneck and Ghost Attention Bottleneck is the main block in C2f and C2fGhostAtt module, respectively. The two modules are introduced in section 3.2.3.

3.2.3. C2fGhostAtt Module

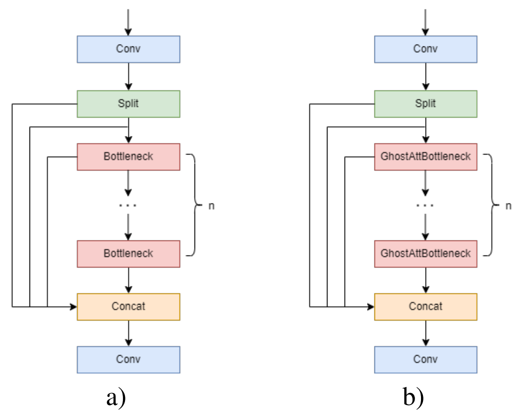

Figure 4: **a)** C2f module. **b)** C2fGhostAtt module.

The C2fGhostAtt module (Fig. 4b) replaces the Bottleneck used in C2f module (Fig. 4a) with Ghost Attention Bottleneck. The C2fGhostAtt module has two parallel gradient flow branches, thereby the model could obtain richer gradient flow information while reducing parameters with Ghost Attention Bottleneck. The C2fGhostAtt module is used in the backbone, and the C2f module is applied in the neck of the proposed GhostAtt-YOLOv8 model.

4. Experiments

4.1. Datasets

We have conducted experiments on five datasets for polyp detection: Kvasir-SEG [10], NeoPolyp-Small [11], PolypsSet [12], LDPolypVideo [13], and SUN [14]. The first four datasets were used to evaluate the learning capability of the model, while SUN dataset was used to analyze the generalization capability of the model.

4.2. Learning Capability

In this part, four datasets were used to evaluate the learning capability of our proposed model, Kvasir-SEG [10], NeoPolyp-Small [11], PolypsSet [12], and LDPolypVideo [13]. We deleted some images with no polyps in PolypsSet [12]. Since LDPolypVideo [13] includes images extracted from videos, we also eliminated images that have identical viewpoint and distance of the same polyp and images that are too blurry and contain too many artifacts. Totally, there are 30,918 images. We split the dataset to the ratio 80:10:10 for training, validation, and testing. Table 1 shows the comparison results between GhostAtt-YOLOv8 and the original YOLOv8s model (hereinafter referred as YOLOv8) on the test set. GhostAtt-YOLOv8 has slightly higher mAP scores compared to YOLOV8, with mAP@50 of 98.94% and mAP@50-95 of 86.01%.

4.3. Generalization Capability

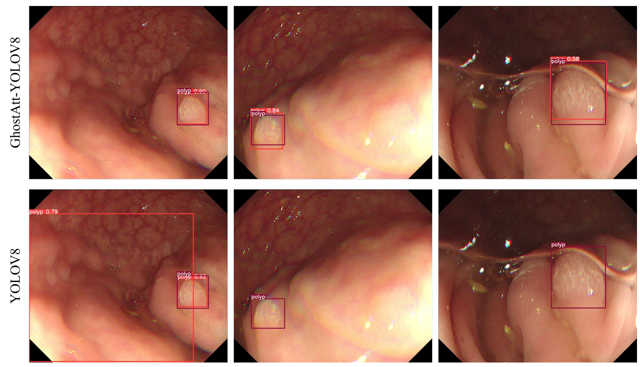

Figure 5: Qualitative results comparison on some images in SUN [14] dataset (Purple bounding box: Groundtruth. Red bounding box: Detection result). GhostAtt-YOLOv8 produces better bounding boxes than YOLOv8.

The trained models were applied to SUN [14] dataset to evaluate their performance on unseen data. Table 2 describes the comparison results of this experiment. GhostAtt-YOLOv8 outperforms YOLOv8 in all metrics, with about 3.4% higher mAP and Recall scores, and 1.5% higher Precision score. We also analyzed their performance in a real-time manner. GhostAtt-YOLOv8 has higher FPS and lower number of Parameters and GFLOPs compared to YOLOv8. Fig. 5 illustrates some qualitative results of GhostAtt-YOLOv8 and YOLOv8 model. This paper is the first one to combine Ghost Convolution with CBAM after the cheap operations and integrate it to YOLOV8 backbone, enhancing both detection performance and computational efficiency.

5. Conclusion

This paper proposes GhostAtt-YOLOv8 model for reliable real-time polyp detection on endoscopic images and videos. GhostAtt-YOLOv8 integrates Ghost Convolution and Spatial and Channel Attention mechanisms into the backbone

Table 1: Performance comparison on test dataset.

Method	mAP@50 (%)	mAP@50-95 (%)	Recall (%)	Precision (%)
GhostAtt-YOLOv8	**98.94**	**86.01**	96.94	99.03
YOLOv8	98.91	85.92	**97.42**	**99.14**

Table 2: Performance comparison on SUN [14] dataset.

Method	mAP@50 (%)	mAP@50-95 (%)	Recall (%)	Precision (%)	FPS	Params (M)	GFLOPs
GhostAtt-YOLOv8	**80.13**	**44.21**	**67.24**	**85.34**	**294**	**8.5**	**21.2**
YOLOv8	76.64	40.91	64.03	83.82	285	11.1	28.4

of the original YOLOv8 model, not only outperforming YOLOv8 in the generalization capability but also being faster and having lower number of Parameters and GFLOPs, showing a promising result in real-world applications. In the future, we will work on modifying the neck and detection head to further improve the detection result of our model.

Reference

[1] J. Asplund, J. H. Kauppila, F. Mattsson, and J. Lagergren, "Survival trends in gastric adenocarcinoma: a population-based study in sweden," *Annals of surgical oncology*, vol. 25, pp. 2693–2702, 2018.

[2] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[3] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1580–1589.

[4] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[5] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, and I. Balasingham, "Automatic colon polyp detection using region based deep cnn and post learning approaches," *IEEE Access*, vol. 6, pp. 40 950–40 962, 2018.

[6] X. Liu, X. Guo, Y. Liu, and Y. Yuan, "Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images," *Medical image analysis*, vol. 71, p. 102052, 2021.

[7] I. Pacal and D. Karaboga, "A robust real-time deep learning based automatic polyp detection system," *Computers in Biology and Medicine*, vol. 134, p. 104519, 2021.

[8] A. Nogueira-Rodríguez, M. Reboiro-Jato, D. Glez-Peña, and H. López-Fernández, "Performance of convolutional neural networks for polyp localization on public colonoscopy image datasets," *Diagnostics*, vol. 12, no. 4, p. 898, 2022.

[9] G. Jocher, "YOLOv5 by Ultralytics," May 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[10] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," in *International Conference on Multimedia Modeling*. Springer, 2020, pp. 451–462.

[11] P. Ngoc Lan, N. S. An, D. V. Hang, D. V. Long, T. Q. Trung, N. T. Thuy, and D. V. Sang, "Neounet: Towards accurate colon polyp segmentation and neoplasm detection," in *Advances in Visual Computing: 16th International Symposium, ISVC 2021, Virtual Event, October 4-6, 2021, Proceedings, Part II*. Springer, 2021, pp. 15–28.

[12] K. Li, M. I. Fathan, K. Patel, T. Zhang, C. Zhong, A. Bansal, A. Rastogi, J. S. Wang, and G. Wang, "Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations," *Plos one*, vol. 16, no. 8, p. e0255809, 2021.

[13] Y. Ma, X. Chen, K. Cheng, Y. Li, and B. Sun, "Ldpolypvideo benchmark: a large-scale colonoscopy video dataset of diverse polyps," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*. Springer, 2021, pp. 387–396.

[14] M. Misawa, S.-e. Kudo, Y. Mori, K. Hotta, K. Ohtsuka, T. Matsuda, S. Saito, T. Kudo, T. Baba, F. Ishida *et al.*, "Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video)," *Gastrointestinal endoscopy*, vol. 93, no. 4, pp. 960–967, 2021.