



# Automatic Bleeding Risk Rating System of Gastric Varices

Yicheng Jiang<sup>1</sup>, Luyue Shi<sup>1</sup>, Wei Qi<sup>2</sup>, Lei Chen<sup>2</sup>, Guanbin Li<sup>5</sup>,  
Xiaoguang Han<sup>3,4</sup>, Xiang Wan<sup>1</sup>, and Siqi Liu<sup>1</sup>(✉)

<sup>1</sup> Shenzhen Research Institute of Big Data, Shenzhen, China  
siqiliu@sribd.cn

<sup>2</sup> Department of Gastroenterology, The Second Hospital of Hebei Medical University,  
Hebei Key Laboratory of Gastroenterology, Hebei Institute of Gastroenterology,  
Hebei Clinical Research Center for Digestive Diseases, Shijiazhuang, China

<sup>3</sup> FNii, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China

<sup>4</sup> School of Science and Engineering, The Chinese University of Hong Kong,  
Shenzhen, Shenzhen, China

<sup>5</sup> School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou,  
China

**Abstract.** An automated bleeding risk rating system of gastric varices (GV) aims to predict the bleeding risk and severity of GV, in order to assist endoscopists in diagnosis and decrease the mortality rate of patients with liver cirrhosis and portal hypertension. However, since the lack of commonly accepted quantification standards, the risk rating highly relies on the endoscopists' experience and may vary a lot in different application scenarios. In this work, we aim to build an automatic GV bleeding risk rating method that can learn from experienced endoscopists and provide stable and accurate predictions. Due to the complexity of GV structures with large intra-class variation and small inter-class variations, we found that existing models perform poorly on this task and tend to lose focus on the important varices regions. To solve this issue, we constructively introduce the segmentation of GV into the classification framework and propose the region-constraint module and cross-region attention module for better feature localization and to learn the correlation of context information. We also collect a GV bleeding risks rating dataset (**GVbleed**) with 1678 gastroscopy images from 411 patients that are jointly annotated in three levels of risks by senior clinical endoscopists. The experiments on our collected dataset show that our method can improve the rating accuracy by nearly 5% compared to the baseline. Codes and dataset will be available at <https://github.com/LuyueShi/gastric-varices>.

---

Y. Jiang, L. Shi and W. Qi—Contributed equally to this work.

---

**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-43904-9\\_1](https://doi.org/10.1007/978-3-031-43904-9_1).

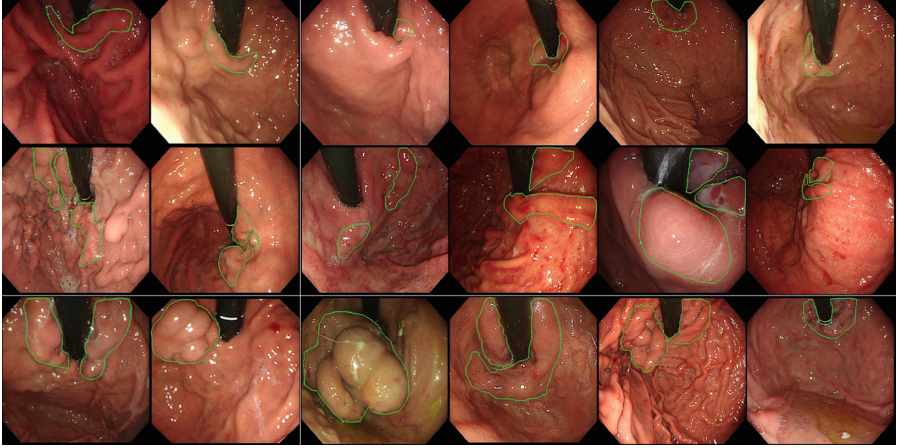
© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023  
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 3–12, 2023.  
[https://doi.org/10.1007/978-3-031-43904-9\\_1](https://doi.org/10.1007/978-3-031-43904-9_1)

**Keywords:** Gastric Varices · Bleeding Risk Rating · Cross-region Attention

## 1 Introduction

Esophagogastric varices are one of the common manifestations in patients with liver cirrhosis and portal hypertension and occur in about 50 percent of patients with liver cirrhosis [3, 6]. The occurrence of esophagogastric variceal bleeding is the most serious adverse event in patients with cirrhosis, with a 6-week acute bleeding mortality rate as high as 15%–20% percent [14]. It is crucial to identify high-risk patients and offer prophylactic treatment at the appropriate time. Regular endoscopy examinations have been proven an effective clinical approach to promptly detect esophagogastric varices with a high risk of bleeding [7]. Different from the grading of esophageal varices (EV) that is relatively complete [1], the bleeding risk grading of gastric varices (GV) involves complex variables including the diameter, shapes, colors, and locations. Several rating systems have been proposed to describe GV based on the anatomical area. Sarin et al. [16] described and divided GV into 2 groups according to their locations and extensions. Hashizume et al. [10] published a more detailed examination describing the form, location, and color. Although the existing rating systems tried to identify the risk from different perspectives, they still lack clear quantification standard and heavily rely on the endoscopists’ subjective judgment. This may cause inconsistency or even misdiagnosis due to the variant experience of endoscopists in different hospitals. Therefore, we aim to build an automatic GV bleeding risk rating method that can learn a stable and robust standard from multiple experienced endoscopists.

Recent works have proven the effectiveness and superiority of deep learning (DL) technologies in handling esophagogastroduodenoscopy (EGD) tasks, such as the detection of gastric cancer and neoplasia [4]. It is even demonstrated that AI can detect neoplasia in Barrett’s esophagus at a higher accuracy than endoscopists [8]. Intuitively we may regard the GV bleeding risk rating as an image classification task and apply typical classification architectures (e.g., ResNet [12]) or state-of-the-art gastric lesion classification methods to it. However, they may raise poor performance due to the large intra-class variation between GV with the same bleeding risk and small inter-class variation between GV and normal tissue or GV with different bleeding risks. First, the GV area may look like regular stomach rugae as it is caused by the blood vessels bulging and crumpling up the stomach (see Fig. 1). Also, since the GV images are taken from different distances and angles, the number of pixels of the GV area may not reflect its actual size. Consequently, the model may fail to focus on the important GV areas for prediction as shown in Fig. 3. To encourage the model to learn more robust representations, we constructively introduce segmentation into the classification framework. With the segmentation information, we further propose a region-constraint module (RCM) and a cross-region attention module (CRAM) for better feature localization and utilization. Specifically, in RCM, we utilize

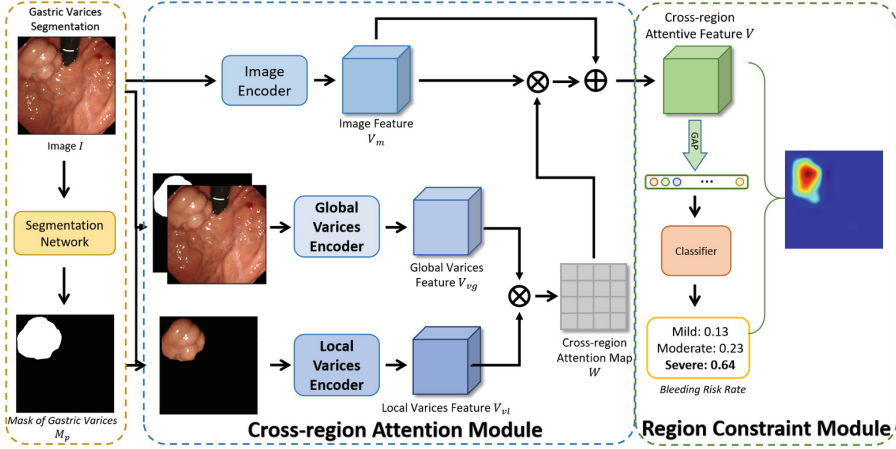


**Fig. 1. Visualization of our GVbleed dataset.** The three rows from up to down show the typical samples of gastric varices with mild, moderate, and severe bleeding risk, respectively. The green line represents the contour of the target gastric varices regions, from which we can sense the complex patterns involving the diameter, shapes, and colors.

the segmentation results to constrain the CAM heatmaps of the feature maps extracted by the classification backbone, avoiding the model making predictions based on incorrect areas. In CRAM, the varices features are extracted using the segmentation results and combined with an attention mechanism to learn the intra-class correlation and cross-region correlation between the target area and the context.

To learn from experienced endoscopists, GV datasets with bleeding risks annotation is needed. While most works and public datasets focus on colonoscopy [13, 15] and esophagus [5, 9], with a lack of study on gastroscopy images. In the public dataset of EndoCV challenge [2], the majority are colonoscopies while only few are gastroscopy images. In this work, we collect a GV bleeding risks rating dataset (**GVbleed**) that contains 1678 gastroscopy images from 411 patients with different levels of GV bleeding risks. Three senior clinical endoscopists are invited to grade the bleeding risk of the retrospective data in three levels and annotated the corresponding segmentation masks of GV areas.

In sum, the contributions of this paper are: 1) a novel GV bleeding risk rating framework that constructively introduces segmentation to enhance the robustness of representation learning; 2) a region-constraint module for better feature localization and a cross-region attention module to learn the correlation of target GV with its context; 3) a GV bleeding risk rating dataset (**GVbleed**) with high-quality annotation from multiple experienced endoscopists. Baseline methods have been evaluated on the newly collected GVbleed dataset. Experimental results demonstrate the effectiveness of our proposed framework and modules, where we improve the accuracy by nearly 5% compared to the baseline model.



**Fig. 2. Pipeline of the proposed automatic bleeding risk rating framework.** The framework consists of a segmentation module, a cross-region attention module, and a region constraint module.

## 2 Methodology

The architecture of the proposed framework is depicted in Fig. 2, which consists of a segmentation module (SM), a region constraint module (RCM), and a cross-region attention module (CRAM). Given a gastroscopy image, the SM is first applied to generate the varices mask of the image. Then, the image together with the mask are fed into the CRAM to extract the cross-region attentive feature map, and a class activation map (CAM) is calculated to represent the concentrated regions through RCM. Finally, a simple classifier is used to predict the bleeding risk using the extracted feature map.

### 2.1 Segmentation Module

Due to the large intra-class variation between GV with the same bleeding risk and small inter-class variation between GV and normal tissue or GV with different bleeding risks, existing classification models exhibit poor perform and tend to lose focus on the GV areas. To solve this issue, we first embed a segmentation network into the classification framework. The predict the varices mask is then used to assist the GV feature to obtain the final bleeding risk rate. Specifically, we use SwinUNet [11] as the segmentation network, considering its great performance, and calculate the DiceLoss between the segmentaion result  $M_p$  and ground truth mask of vaices region  $M_{gt}$  for optimizing the network:

$$l_{se} = 1 - \frac{2\Sigma M_p * M_{gt}}{\Sigma M_p^2 + \Sigma M_{gt}^2 + \epsilon}, \quad (1)$$

where  $\epsilon$  is a smooth constant equals to  $10^{-5}$ .

A straightforward strategy to utilize the segmentation mask is directly using it as an input of the classification model, such as concatenating the image with the mask as the input. Although such strategy can improve the classification performance, it may still lose focus in some hard cases where the GV area can hardly be distinguished. To further regularize the attention and fully utilize the context information around the GV area, on top of the segmentation framework we proposed the cross-region attention module and the region-constraint module.

## 2.2 Cross-Region Attention Module

Inspired by the self-attention mechanism [17, 18], we propose a cross-region attention module (CRAM) to learn the correlation of context information. The CRAM consists of an image encoder  $f_{im}$ , a varices local encoder  $f_{vl}$  and a varices global encoder  $f_{vg}$ . Given the image  $I$  and the predicted varices mask  $M_p$ , a feature extraction step is first performed to generate the image feature  $V_m$ , the local varices feature  $V_{vl}$  and global varices feature  $V_{vg}$ :

$$V_m = f_{im}(I), \quad V_{vl} = f_{vl}(I * M_p), \quad V_{vg} = f_{vg}(\text{concat}[I, M_p]), \quad (2)$$

Then, through similarity measuring, we can compute the attention with

$$A = (V_{vl})^T V_{vg}, \quad W_{ij} = \frac{\exp(A_{ij})}{\sum_p (\exp(A_{pj}))}, \quad (3)$$

which composes of two correlations: self-attention over varices regions and cross-region attention between varices and background regions. Finally, the output feature is calculated as:

$$V = \gamma V_m W + V_m, \quad (4)$$

where  $\gamma$  is a learnable parameter. Then the cross-region attentive feature  $V$  is fed into a classifier to predict the bleeding risk.

## 2.3 Region Constraint Module

To improve the focus ability of the model, we propose the region constraint module (RCM) to add a constraint on the class activation map (CAM) of the classification model. Specifically, we use the feature map after the last convolutional layer to calculate the CAM [19], which computes the weighted sum of feature maps from the convolutional layer using the weights of the FC layer. After getting the CAM, we regularize CAM by calculating the dice loss between the CAM and ground truth mask of varices region  $l_{co}$ .

## 2.4 Network Training

In our framework, we use the cross entropy loss as the classification loss:

$$l_{cl} = - \sum_{c=1}^C \log \frac{\exp(p_c)}{\sum_{i=1}^C \exp(p_i)} y_c \quad (5)$$

**Table 1.** Distribution of the three-level GV bleeding risks in the GVBleed dataset.

	Mild	Moderate	Severe	Total
Train	398	484	455	1337
Test	94	145	102	341
Total	462	629	557	1678

where  $p$  is the prediction of the classifier and  $y$  is the ground-truth label. And the total loss of our framework can be summarized as:

$$L_{total} = \frac{1}{N} \sum (\omega_s l_{se} + \omega_{co} l_{co} + \omega_{cl} l_{cl}), \quad (6)$$

where  $N$  is the total number of samples,  $\omega_s$ ,  $\omega_{co}$  and  $\omega_{cl}$  are weights of the three losses, respectively.

The training process of the proposed network consists of three steps: 1) The segmentation network is trained first; 2) The ground-truth segmentation masks and images are used as the inputs of the CRAM, the classification network, including CRAM and RCM, are jointly trained; 3) The whole framework is jointly fine-tuned.

### 3 GVBleed Dataset

**Data Collection and Annotation.** The GVBleed dataset contains 1678 endoscopic images with gastric varices from 527 cases. All of these cases are collected from 411 patients in a **Grade-III Class-A hospital** during the period from 2017 to 2022. In the current version, **images from patients with ages elder than 18 are retained**<sup>1</sup>. The images are selected from the raw endoscopic videos and frames. To maximize the variations, non-consecutive frames with larger angle differences are selected. To ensure the quality of our dataset, senior endoscopists are invited to remove duplicates, blurs, active bleeding, chromoendoscopy, and NBI pictures.

**Criterion of GV Bleeding Risk Level Rating.** Based on the clinical experience in practice, the GV bleeding risks in our dataset are rated into three levels, i.e., mild, moderate, and severe. The detailed rating standard is as follows: 1) Mild: low risk of bleeding, and regular follow-up is sufficient (usually with a diameter less than or equal to 5 mm). 2) Moderate: moderate risk of bleeding, and endoscopic treatment is necessary, with relatively low endoscopic treatment difficulty (usually with a diameter between 5 mm and 10 mm). 3) Severe: high risk of bleeding and endoscopic treatment is necessary, with high endoscopic treatment difficulty. The varices are thicker (usually with a diameter greater than 10 mm) or less than 10mm but with positive red signs. Note that the diameter is only one reference for the final risk rating since the GV is with

<sup>1</sup> Please refer to the supplementary material for more detailed information about our dataset.

**Table 2.** Quantitative results of the proposed method and modules.

Models	Accuracy(%)	F1-Score(%)		
		Mild	Moderate	Severe
ResNet18	65.40	60.17	60.38	<b>73.25</b>
ResNet18 + SegMask (Concat)	66.86	69.30	66.46	65.98
ResNet18 + SegMask + Attention	67.16	69.21	65.50	66.62
ResNet18 + SegMask + Attention + Constraint	<b>70.97</b>	<b>73.33</b>	<b>68.32</b>	72.88

various 3D shapes and locations. The other facts are more subjectively evaluated based on the experience of endoscopists. To ensure the accuracy of our annotation, three senior endoscopists with more than 10 years of clinical experience are invited to jointly label each sample in our dataset. If three endoscopists have inconsistent ratings for a sample, the final decision is judged by voting. A sample is selected and labeled with a specific bleeding risk level only when two or more endoscopists reach a consensus on it.

The GVBleed dataset is partitioned into training and testing sets for evaluation, where the training set contains 1337 images and the testing set has 341 images. The detailed statistics of the three levels of GV bleeding risk in each set are shown in Table 1. The dataset is planned to be released in the future.

## 4 Experiments

### 4.1 Implementation Details

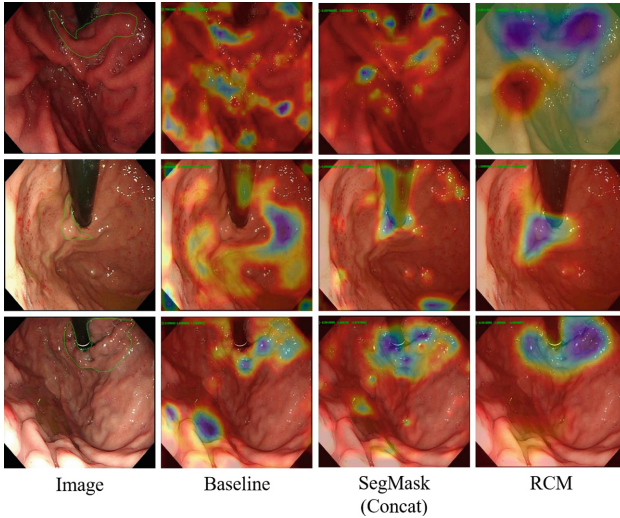
In experiments, the weights  $\omega_s$ ,  $\omega_{co}$ , and  $\omega_{cl}$  of the segmentation loss, region constraint loss, and classification loss are set to 0.2, 1, and 1, respectively. The details of the three-step training are as follows: 1) **Segmentation module**: We trained the segmentation network for 600 epochs, using Adam as the optimizer, and the learning rate is initialized as  $1e-3$  and drops to  $1e-4$  after 300 epochs. 2) **Cross-region attention module and region constraint module**: We used the ground-truth varices masks and images as the inputs of the CRAM, and jointly trained the CRAM and RCM for 100 epochs. Adam is used as the optimizer, the learning rate is set to  $1e-3$ ; 3) **Jointly fine-tuning**: The whole framework is jointly fine-tuned for 100 epochs with Adam as optimizer and the learning rate set to  $1e-3$ . In addition, common data augmentation techniques such as rotation and flipping were adopted here.

### 4.2 Results Analysis

Table 2 reports the quantitative results of different models and Fig. 3 shows the CAM visualizations. We tested several baseline models, including both simple CNN models and state-of-the-art transformer-based models. However, the transformer-based models achieves much worse performances since they always require more training data, which is not available in our task. Thus, we selected



the simple CNN models as baselines since they achieve better performances. As shown in the figure, the baseline model cannot focus on the varices regions due to the complexity of GV structures with large intra-class variations and small inter-class variations. By introducing the segmentation of GV into the framework, concatenating the image with its segmentation mask as the inputs of the classifier can improve the classification accuracy by 1.2%. And the focus ability of the classifier is stronger than the baseline model. With the help of CRAM, the performance of the model can be further improved. Although the model can extract more important context information at the varices regions, the performance improvement is not very large since the focus ability is not the best and the model may still make predictions based on the incorrect regions for some hard images. By adding the RCM to the CRAM, the focus ability of the model can be further improved, and thus the model has a significant improvement in performance by 5% compared to the baseline model, this proves the effectiveness of our proposed modules. Note that, the baseline model tends to predict the images as severe, thus the f1-score of severe is high but the f1-scores of mild and moderate are significantly lower than other models. More quantitative and visualization results are shown in supplementary material. In addition, given the input image with resolution  $512 \times 512$ , the parameters and computational cost of our framework are 40.2M, and 52.4G MACs, and 29 ms inference time for a single image on GPU RTX2080. For comparison, a single ResNet152 model has 60.19M parameters with 60.62 G MACs.



**Fig. 3. Visualization of CAM.** Green lines represent the contours of varices regions. The blue areas represent the primary areas of concern. With the proposed SegMask and RCM module, the network tends to focus on the important gastric varices areas as endoscopists do.



## 5 Conclusions

In this paper, we propose a novel bleeding risk rating framework for gastric varices. Due to the large intra-class variation between GV with the same bleeding risk and small inter-class variation between GV and normal tissue or GV with different bleeding risks, existing classification models cannot correctly focus on the varices regions and always raise poor performance. To solve this issue, we constructively introduce segmentation to enhance the robustness of representation learning. Besides, we further design a region-constraint module for better feature localization and a cross-region attention module to learn the correlation of target GV with its context. In addition, we collected the GVBleed dataset with high-quality annotation of three-level of GV bleeding risks. The experiments on our dataset demonstrated the effectiveness and superiority of our framework.

**Acknowledgements.** This work is supported by Chinese Key-Area Research and Development Program of Guangdong Province (2020B0101350001), and the Shenzhen Science and Technology Program (JCYJ20220818103001002), and the Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen. It was also supported by the Shenzhen Sustainable Development Project (KCXFZ20201221173008022), the National Natural Science Foundation of China (NO. 61976250), the Guangdong Basic and Applied Basic Research Foundation (NO. 2020B1515020048).

## References

1. Abby Philips, C., Sahney, A.: Oesophageal and gastric varices: historical aspects, classification and grading: everything in one place. *Gastroenterol. Rep.* **4**(3), 186–195 (2016)
2. Ali, S., Ghatwary, N., Jha, D., Halvorsen, P.: EndoCV 2021 3rd International Workshop and Challenge on Computer Vision in Endoscopy
3. Brocchi, E., et al.: Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices: a prospective multicenter study. *New England J. Med.* **319**(15), 983–989 (1988)
4. Chen, H., Sung, J.J.: Potentials of AI in medical image analysis in gastroenterology and hepatology. *J. Gastroenterol. Hepatol.* **36**(1), 31–38 (2021)
5. Du, W., et al.: Improving the classification performance of esophageal disease on small dataset by semi-supervised efficient contrastive learning. *J. Med. Syst.* **46**, 1–13 (2022)
6. Garcia-Tsao, G., Abraldes, J.G., Berzigotti, A., Bosch, J.: Portal hypertensive bleeding in cirrhosis: risk stratification, diagnosis, and management: 2016 practice guidance by the American association for the study of liver diseases. *Hepatology* **65**(1), 310–335 (2017)
7. Gralnek, I.M., et al.: Endoscopic diagnosis and management of esophagogastric variceal hemorrhage: European society of gastrointestinal endoscopy (ESGE) guideline. *Endoscopy* **54**(11), 1094–1120 (2022)
8. de Groof, A.J., et al.: Deep-learning system detects neoplasia in patients with Barrett’s esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. *Gastroenterology* **158**(4), 915–929 (2020)

9. Guo, L., et al.: Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest. Endosc.* **91**(1), 41–51 (2020)
10. Hashizume, M., Kitano, S., Yamaga, H., Koyanagi, N., Sugimachi, K.: Endoscopic classification of gastric varices. *Gastrointest. Endosc.* **36**(3), 276–280 (1990)
11. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images. In: Crimi, A., Bakas, S. (eds.) *BrainLes 2021*. LNCS, vol. 12962, pp. 272–284. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-08999-2\\_22](https://doi.org/10.1007/978-3-031-08999-2_22)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
13. Kang, J., Gwak, J.: Ensemble of instance segmentation models for polyp segmentation in colonoscopy images. *IEEE Access* **7**, 26440–26447 (2019)
14. Lv, Y., et al.: Covered tips versus endoscopic band ligation plus propranolol for the prevention of variceal rebleeding in cirrhotic patients with portal vein thrombosis: a randomised controlled trial. *Gut* **67**(12), 2156–2168 (2018)
15. Ma, Y., Chen, X., Cheng, K., Li, Y., Sun, B.: LDPolypVideo benchmark: a large-scale colonoscopy video dataset of diverse polyps. In: de Bruijne, M., et al. (eds.) *MICCAI 2021*. LNCS, vol. 12905, pp. 387–396. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87240-3\\_37](https://doi.org/10.1007/978-3-030-87240-3_37)
16. Sarin, S., Kumar, A.: Gastric varices: profile, classification, and management. *Am. J. Gastroenterol.* (Springer Nature) **84**(10) (1989)
17. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803 (2018)
18. Ye, L., Rochan, M., Liu, Z., Wang, Y.: Cross-modal self-attention network for referring image segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10502–10511 (2019)
19. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929 (2016)