Enhancing Generalization and Reducing Bias in Polyp Segmentation with Diffusion-Based Inpainting Models

Jiajian Ma
Student ID: 222041049
School of Data Science
The Chinese University of Hong Kong, Shenzhen
222041049@link.cuhk.edu.cn

Abstract

Enhancing generalization and reducing bias are crucial challenges for polyp segmentation models in clinical applications. From a causal inference perspective, these challenges share a common resolution pathway: effectively learning the causal factors of current segmentation outcomes amidst confounding variables. Addressing this, we propose that inpainting polyps within different normal backgrounds acts as a do-operator for disentangling causal and non-causal factors, thereby potentially solving issues related to generalization and bias simultaneously. To validate this approach, we developed an inpainting model based on pre-trained Stable Diffusion Inpaint and ControlNet. To address their limitation of imprecise boundaries, we utilized the prior that synthetic polyps are confined to the inpainted region, establishing an inpainted region-guided pseudo-mask refinement network that significantly improves the quality of pseudo-labels. Experiments demonstrate that our inpainting-based data augmentation method significantly enhances the generalization of polyp segmentation models across datasets (overall mDice improved by +2.3%), and effectively reduces bias in sensitive metrics such as Polyp Diagnosis (ES-Dice improved by +5.4%). Our code and data are available at https://github.com/497662892/Ai-Security-Final-Project.

1 Introduction

Polyp segmentation models can effectively help doctor to identify polyps during endoscopic examinations, which is crucial for colon cancer screening [6]. However, these models are often trained on small scale labeled data from limited sources, constrained by data privacy and high costs of annotation [6]. These datasets frequently fail to capture the diversity of real-world clinical environments and exhibit biases in their distribution [2, 22, 8], leading to generalization problems and performance disparity across different clinical scenarios.

Improving model generalization and reducing bias are essential for enhancing polyp segmentation models in real-world applications [9]. As these objectives differ, previous studies typically focus on either enhancing generalization [25, 4, 3, 16] or reducing bias [9], with few addressing both simultaneously. However, from a causal inference perspective [10], we can identify a shared solution pathway for these challenges: accurately capturing the causal factors related to actual segmentation (Figure 1a). This suggests that methods enhancing the acquisition of causal factors have the potential to simultaneously improve generalization and reduce bias.

In practice, the causal factors corresponding to actual segmentation outcomes are not directly observable; instead, we must train a model to implicitly extract their representation from data[10, 23]. However, since samples and annotations are obtained through specific distributional sampling, external factors can induce correlations between causal and non-causal factors (Figure 1a). This coupling

Preprint. Under review.

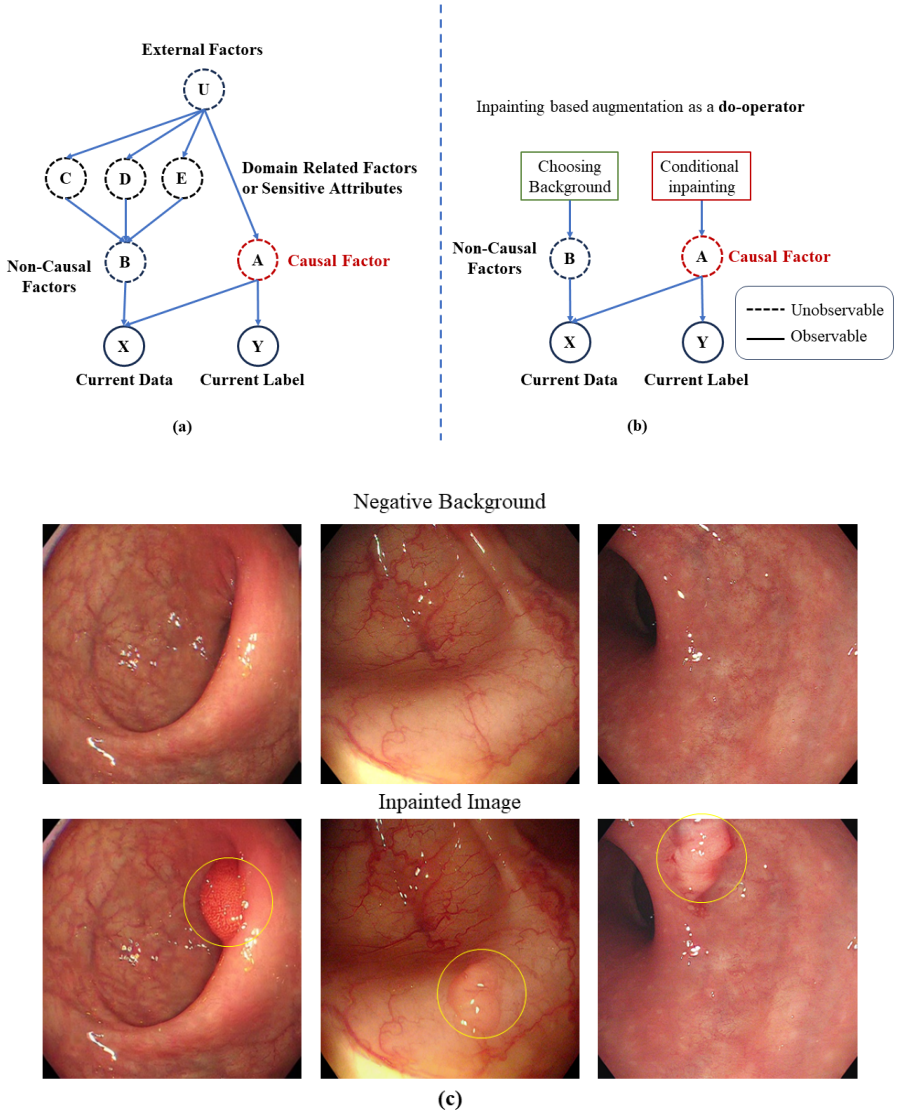


Figure 1: (a) Causal graph for segmentation model, how to learn a good representation of causal factor would be critical for addressing generalization and bias problem. (b) Inpainting polyps into different background can be view as a do-operator for the non-causal and causal factors. (c) Examples of inpainting polyps into different background.

can be embedded into the learned representations, leading to predictions that are correlated with domain-specific features or sensitive attributes, consequently diminishing generalization performance and introducing model bias [10].

To address this issue, we need to break the correlation between causal and non-causal factors. Given that negative background images contain extensive information related to non-causal factors, realistically inpainting polyp into these normal backgrounds (Figure 1c) can change the causal factors while keeping the non-causal factors unchanged. In other word, inpainting based data augmentation method provides a means to perform a do-operation on both causal and non-causal factors (Figure 1b). By selectively using negative backgrounds and strategically planning the inpainting of polyps, we can decouple their correlation in the synthetic samples.

However, challenges remain in realistically inpainting polyps into diverse backgrounds and providing high-quality pseudo-labels for synthetic images. We introduce the Polyp Inpainter, a diffusion-

based polyp inpainting model [17] that incorporates fine-grained boundary and surface features via ControlNet [26] modules. This model can realistically introduce polyps into various backgrounds. We also utilize the prior that synthetic polyps are confined to the inpainted region and develop an inpainted region-guided segmentation network to generate high-quality pseudo-masks. Additionally, we implement a sample selection strategy prioritizing high-quality synthetic images for further training of the segmentation model. Our experiments demonstrate that this data augmentation approach effectively enhances model generalization and reduces bias simultaneous.

In summary, our contributions are twofold:

- Analyze generalization and bias problems from a causal inference perspective and propose using inpainting polyps into different backgrounds as a do-operator to enhance causal factor learning.
- Propose an inpainting based data augmentation method, which can generate high-quality synthetic images with corresponding pseudo-masks and effectively improve generalization and reduce bias in polyp segmentation.

2 Related Work

2.1 Data augmentation in medical images

Data augmentation is a common technique in the field of medical image analysis. In early stage, researchers focus on data augmentation method that can change the image style across different domains. These methods can be considered as do-operators for domain-related information, which help enhance model generalization. For example, Zhang [25] proposed a deep stacked transformation, suggesting that domain shifts can be simulated through extensive augmentation within a single domain. Similarly, Fick [5] utilized Cycle-GAN to transform images across different domains for augmentation purposes. Additionally, Wei [22] proposed enhancing performance by swapping the color of polyp images from different devices. However, merely adjusting the style information of images does not fully decouple the causal factors from non-causal factors, as biases may relate to structural information.

In addition, numerous works on synthetic data from scratch have emerged recently. Du [3] introduced ArSDM, which enables the generation of endoscopic images under controlled polyp mask conditions. Ktena [9] proposed a diffusion model that synthesizes data conditioned by sensitivity indicators, which significantly improved fairness in medical image classification tasks across three different modalities: pathology, chest X-rays, and skin lesions. From a causal inference perspective, Ktena's work essentially performs an explicit do-operator on sensitivity indicators, and its outstanding results are expected. Nevertheless, these models face multiple challenges, including the need for extensive labeled data, substantial training costs, and limited flexibility of sensitivity indicators, which hinder their practical application.

2.2 Controllable inpainting methods

Controllable inpainting refers to the technique of inserting specific semantic content into designated areas of a target image, a task vital for applications like medical imaging and digital restoration [17, 24, 27]. Methods for achieving controllable inpainting can be primarily categorized into three types: image blending-based methods, Generative Adversarial Network (GAN)-based methods, and diffusion-based methods.

Image blending-based methods utilize techniques such as copy-paste [29], Poisson blending [15], and deep image blending [1]. These methods excel at preserving the original boundaries and details of the image. However, their performance diminishes in complex backgrounds where they often require manual adjustments to align with realistic distributions. Furthermore, the diversity of the inpainted content is limited as it closely resembles the source material.

GAN-based controllable inpainting methods have been applied in areas like chest CT [18] and polyp segmentation [4]. However, these studies have primarily focused on model performance within internal test datasets, with a lack of assessment of model generalization capabilities. Due to the absence of open-source foundational models for these tasks, GAN-based methods require retraining from scratch, presenting substantial computational resource challenges [18, 4].

Diffusion-based methods, exemplified by Stable Diffusion, have gained prominence due to their ability to generate realistic and diverse images [17]. Previous works such as the text-controlled Stable Diffusion Inpaint[17], image-controlled Paint-by-Example[24], and PHD[27], which supports both text and image control, represent recent advances in this area. By integrating additional modules like Controlnet[26] and T2I-Adaptor[14], they offer refined control over generated image boundaries, contours, and textures. Despite these advances, diffusion-based methods still struggle to precisely adhere to specified boundary conditions, often producing synthesized images that do not possess high-quality pseudo-masks, a critical requirement for applications in medical imaging and other precision-dependent fields.

3 Method

3.1 Overall pipeline for data augmentation

The overall pipeline of our proposed data augmentation method comprises five stages, as illustrated in Figure 2: training the inpainting model, generating inpainted images, refining pseudo-masks, selecting samples, and training the polyp segmentation model.

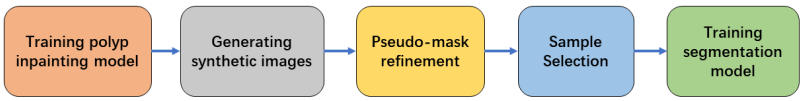
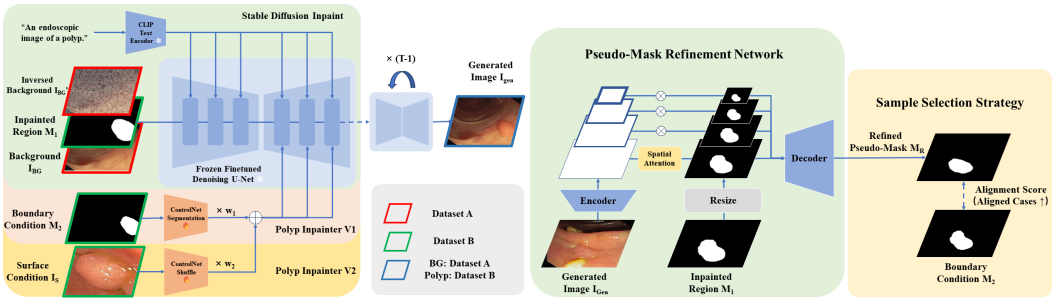


Figure 2: The overall pipeline of our proposed data augmentation method.

3.2 Diffusion-based polyp inpainting model

Figure 3a displays the architecture of our proposed polyp inpainting model. We employ the Stable Diffusion Inpaint v1.5 as its backbone [17], which will be frozen, after fine-tuning on a single endoscopic dataset (Figure 3a, green), with the training loss as depicted in [17]. This backbone is capable of generating polyps within the inpainted region M_1, without modifying the background.

To enhance the inpainting quality, we further add and fine-tune 2 ControlNet modules, which are initialized by segmentation v1.1 and shuffle v1.1 [26], to inject the fine-grained boundary feature and the surface feature into the Stalbe Diffusion Inpaint decoder, respectively. These modules can utilize real mask M_2 as boundary conditions (Figure 3a, red) or cropped polyp images I_S as surface conditions (Figure 3a, yellow).



(a) The structure of our polyp inpainting model (b) Pseudo-mask refinement and case selection strategy

Figure 3: The framework of our data augmentation method

3.3 Generating inpainted images

The inference process of our polyp inpainting model is detailed in Algorithm 1. To enhance the naturalness of the generated images I_{Gen}, we implemented three inference techniques.

First, we divide the background I_{BG} into different patches and reposition the center of the inpainted region M_1 and the boundary condition M_2 to the patch that best matches the color of the reference polyp.

Second, inspired by [12], instead of sampling from standard Gaussian noise, we try to sample from an inversed noisy background $I_{BG'}$, which will help us get more natural transition in the boundary.

Lastly, to ensure accurate generation of smaller polyps, which occupy less than 2.5% of the area, we preprocess the inpainting background with a copy-paste technique. This method allows the sampling process towards generating lesions rather than normal mucosa.

Algorithm 1 Inference of Polyp Inpainting Model

Require: Polyp Inpainter (*PI*), Stable Diffusion Inpainter (*SD*), Surface control module (*SF*), Boundary control module (*BD*), Background (I_{BG}), Reference (I_{Ref}), Inpainted region (M_1), Boundary Condition (M_2), Surface Condition (I_S)

Prepare: ▷ choose suitible location for inpainting
$center \leftarrow getcenter(I_{BG}, M_2, I_{Ref})$
$M_1 \leftarrow movecenter(M_1, center)$
$M_2 \leftarrow movecenter(M_2, center)$
$I_{Ref} \leftarrow movecenter(I_{Ref}, center)$

if $sum(M_2)/(M_2.H * M_2.W) \leq 0.025$ **then** ▷ update background for small polyps
    $I_{BG} \leftarrow copypaste(I_{BG}, M_2, I_{Ref})$
end if

$k = inverse\_strength \times total\_timesteps$
$I'_{BG} = \sqrt{\overline{\alpha}_k} \cdot I_{BG} + \sqrt{1 - \overline{\alpha}_k} \cdot \epsilon$ ▷ sampling from inversed background

Input:
SD input: I'_{BG} , M_1, I_{BG}
SF input: I_S
BD input: M_2
k
while $k > 0$ **do** ▷ sampling
    $\epsilon \leftarrow PI(SD(I'_{BG}, M_1, I_{BG}), BD(M_2), SF(I_S))$
    $I'_{BG} \leftarrow Sampler(I'_{BG}, \epsilon, k)$
    $k \leftarrow k - 1$
end while
$I_{gen} \leftarrow I'_{BG}$
return I_{gen} ▷ return the generated image

3.4 Pseudo-mask refinement network

In order to obtain high quality pseudo-mask of the synthetic images, we leverage the fact that polyps are exclusively present within the inpainted region M_1 to design our pseudo-mask refinement network (Figure 3b, green block). By applying either element-wise multiplication or spatial attention with M_1 to gate the multi-scale features of encoder, We can direct the model's attention to this region. Consequently, it can filter out distractions from the complex endoscopic background, and produce high-quality refined pseudo-mask M_R in a semi-supervised manner.

3.5 Sample selection strategy

After obtaining the pseudo-mask M_R, we implement a strategy to select high-quality samples, as illustrated in Figure 3b and detailed in Algorithm 2. We employ the alignment score, defined as the Dice Coefficient between M_R and the conditional mask M_2, to assess the generation quality indirectly. Samples that achieve an alignment score of ≥ 0.8 are classified as well-aligned and are subsequently added to the training set for further model training.

Algorithm 2 Sample Selection Strategy

Require: Generated Image (I_{gen}), Pseudo-Mask (M_R), Conditional Mask (M_2), Training Set (S_{train}), Synthetic Data Set (S_{Syn})

for each I_{gen}, M_R in S_{Syn} **do**
  $align\_score \leftarrow Dice(M_R, M_2)$
  if $align\_score \geq 0.8$ **then**
    $S_{train} \leftarrow S_{train} \cup \{(I_{gen}, M_R)\}$
  end if
end for

4 Experiment

4.1 Datasets

Our study utilized five publicly available datasets for polyp segmentation: SUN-SEG [13, 8] (49,136 images), Kvasir-SEG [7] (1,000 images), CVC-ClinicDB [21] (612 images), CVC-ColonDB [21] (300 images), and ETIS-LaribPolypDB [19] (196 images). To expedite the validation of our work's efficacy, we sampled every 10th frame from each video in the SUN-SEG dataset and maintained the original partitioning scheme to derive the training set (1,993 images), validation set (1,442 images), and test set (1,574 images). Additionally, we employed negative images from SUN-SEG (781 images) and LD-PolypVideo[11] (1,220 images) as negative backgrounds for inpainting.

4.2 Implementation details

All experiments were performed on a single RTX 3090 GPU, utilizing the PyTorch framework. Our inpainting models were exclusively fine-tuned on the SUN-SEG training set with the default AdamW optimizer, using hyperparameters listed in table 1.

Table 1: Training parameters for inpainting models

Model	Steps	LR	Batch Size
SD Inpaint Backbone	5,000	1×10^{-5}	4
ControlNet Modules	16,000	5×10^{-5}	4

During the synthetic data generation, we utilized data from the SUN-SEG training set (1,993 images) to serve as the boundary and surface conditions. We randomly selected one negative image from SUN-SEG (721 images) and LD-PolypVideo (1,220 images) at each time as the negative background. We employed 0.8 as the inverse strength [12], 50 as the sampling steps and UniPCMultistepScheduler as the sampler [28] in the sampling process. In total, 1,993 synthetic images were generated in this experiment.

We adapted the polyp-PVT model, a robust polyp segmentation baseline [2], as the backbone for our pseudo-mask refinement network. This model was then trained on the SUN-SEG training set, adhering to the protocols outlined by Dong et al. [2]. The model demonstrating the highest performance during the validation was used to refine the pseudo-masks of synthetic images.

After applying our sample selection strategy, we obtain a total of 1,568 qualified cases, which will be combined with the SUN-SEG training set to train the Polyp-PVT segmentation model under the same protocols [2]. The best segmentation model in validation was advanced to the testing phase.

To evaluate the model's generalization performance, we trained the model on the training set and then tested it across five datasets, including one internal (SUN-SEG test set) and four external test sets (Kvasir-SEG, CVC-ClinicDB, CVC-ColonDB, ETIS-LaribPolypDB). The Dice coefficient was employed as the metric to assess the segmentation model's performance.

$$Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

To evaluate model bias, we used diagnosis, location, size, and shape of polyps, provided as sensitive attributes in the SUN-SEG dataset. We employed the Equity-Scaled Segmentation Performance

(ESSP) to measure the extent of model bias [20]. ESSP is defined as follows, where I is the performance metric for segmentation (Dice coefficient in this project) and A represents a specific subgroup.

$$\Delta = \sum_{A \in \mathcal{A}} |I((z', y)) - I((z', a, y)|a = A)|$$

$$ESSP = \frac{I((z', y))}{1 + \Delta}$$

4.3 Qualitative evaluation of polyp inpainting quality

The visualization results of our inpainting model are shown in Figure 4. It can be observed that our model is capable of inpainting polyps very realistically across various negative backgrounds. Furthermore, through our pseudo-mask refinement network, high-quality pseudo-labels are generated. For more visualization results, please check the synthetic dataset we uploaded.

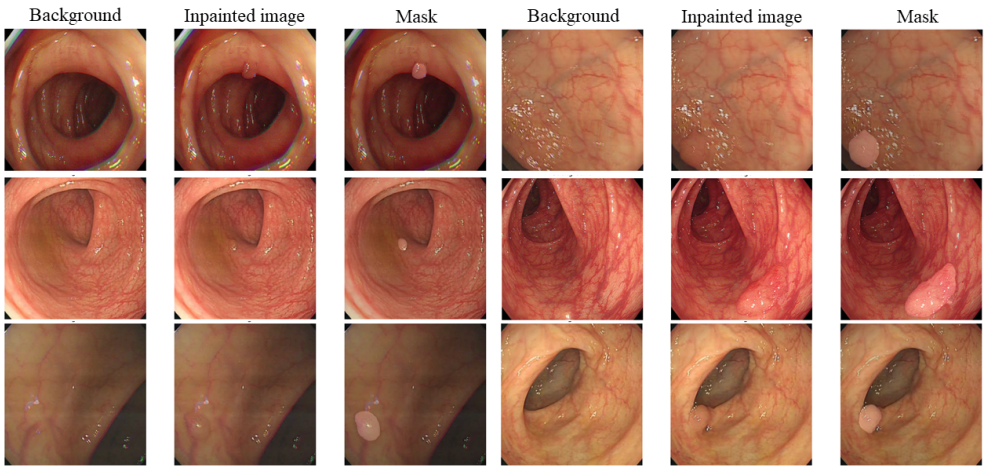


Figure 4: The visualization of our synthetic images.

4.4 Generalization in polyp segmentation

Table 2 displays the polyp-PVT's performance across diverse datasets with or without augmentation strategies. The models trained on the SUN-SEG training set show a performance reduction on external datasets (row 1). This drop is pronounced in the ETIS-LaribPolypDB (mDice: 0.8254 v.s. 0.7781), highlighting a potential generalization problem for models trained solely on SUN-SEG.

Implementing our data augmentation method results in significant performance enhancements on both the SUN-SEG and external datasets. As indicated in row 2 of Table 2, augmentation with polyp inpainting improved the mDice by 2.3% on SUN-SEG (0.8484 v.s. 0.8254) and exhibited increments of 1.9%, 1.5%, 0.3%, and 5.3% on the Kvasir-SEG, CVC-ClinicDB, CVC-ColonDB and ETIS-LaribPolypDB, respectively. These findings affirm that our method can considerably boost model generalization capabilities, with additional usage of small amount easily accessible negative backgrounds.

Table 2: The mDice of polyp-PVT across different datasets with/without data augmentation

Dataset	**SUN-SEG**	**Kvasir-SEG**	**ClinicDB**	**ColonDB**	**LaribPolypDB**	**Overall**
Baseline	0.8254	0.8143	0.8358	0.8498	0.7781	0.8207
Ours	**0.8484**	**0.8334**	**0.8506**	**0.8532**	**0.8313**	**0.8434**

4.5 Reducing Bias in polyp segmentation

The following tables illustrate the performance of the Polyp-PVT model under four sensitive attributes: Diagnosis, Size, Shape, and Location, both with and without the use of data augmentation. The comparison reveals consistent improvements in the bias metric ES-Dice, with increments of 5.4%, 1.7%, 0.4%, and 0.01% for Diagnosis (Table 3), Size (Table 4), Shape (Table 5), and Location (Table 6) respectively. Across nearly every subgroup within these sensitive attributes, the model enhanced with data augmentation demonstrated superior performance. These results suggest that our method can effectively mitigate model bias.

Table 3: Performance with **Diagnosis** as Sensitive Attribute in SUN-SEG

Metric	ES Dice	Overall Dice	Hyperplastic	Low-grade	High-grade
Baseline	0.7547	0.8254	**0.8548**	0.8318	0.7677
Ours	**0.8094**	**0.8484**	0.8384	**0.8443**	**0.8825**

Table 4: Performance with **Size** as Sensitive Attribute in SUN-SEG

Metric	ES Dice	Overall Dice	Large Dice	Small Dice
Baseline	0.8141	0.8254	0.8200	0.8339
Ours	**0.8312**	**0.8484**	**0.8564**	**0.8357**

Table 5: Performance with **Shape** as Sensitive Attribute in SUN-SEG

Metric	ES Dice	Overall	IIa Dice	Ip Dice	Is Dice	Isp Dice
Baseline	0.7174	0.8254	0.7679	0.8040	0.8747	0.8477
Ours	**0.7213**	**0.8484**	**0.7784**	**0.8431**	**0.8980**	**0.8996**

Table 6: Performance with **Location** as Sensitive Attribute in SUN-SEG

Metric	ES Dice	Dice	Ascending	Descending	Rectum	Sigmoid	Transverse
Baseline	0.7183	0.8254	0.7754	0.8026	0.8417	0.8361	0.8745
Ours	**0.7184**	**0.8484**	**0.7868**	**0.8170**	**0.9060**	**0.8612**	**0.8659**

5 Conclusion

In this project, we analyze generalization and bias problems from a causal inference perspective and propose using inpainting polyps into different backgrounds as a do-operator to enhance causal factor learning. In our experiments, we validated that our inpainting-based data augmentation strategy is successful in improving generalization and reducing bias simultaneously in polyp segmentation models.

However, our work still has some limitations. It has only been validated in the polyp-PVT segmentation model and on five relatively small datasets. Additionally, due to limited time, we lack comparisons with other inpainting-based or data augmentation methods. Moreover, our random augmentation strategy during synthetic data generation did not focus on improving poorly performing subgroups, which might not fully leverage the potential to reduce bias. These areas will be the focus of our future work.

References

[1] Junyan Cao, Yan Hong, and Li Niu. Painterly image harmonization in dual domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 268–276, 2023.

[2] Bo Dong, Wenhai Wang, Deng-Ping Fan, Jinpeng Li, Huazhu Fu, and Ling Shao. Polyp-pvt: Polyp segmentation with pyramid vision transformers. *arXiv preprint arXiv:2108.06932*, 2021.

[3] Yuhao Du, Yuncheng Jiang, Shuangyi Tan, Xusheng Wu, Qi Dou, Zhen Li, Guanbin Li, and Xiang Wan. Arsdm: Colonoscopy images synthesis with adaptive refinement semantic diffusion models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 339–349. Springer, 2023.

[4] Jan Andre Fagereng, Vajira Thambawita, Andrea M Storås, Sravanthi Parasa, Thomas De Lange, Pål Halvorsen, and Michael A Riegler. Polypconnect: Image inpainting for generating realistic gastrointestinal tract images with polyps. In *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 66–71. IEEE, 2022.

[5] Rutger HJ Fick, Alireza Moshayedi, Gauthier Roy, Jules Dedieu, Stéphanie Petit, and Saima Ben Hadj. Domain-specific cycle-gan augmentation improves domain generalizability for mitosis detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 40–47. Springer, 2021.

[6] Dexin Gong, Lianlian Wu, Jun Zhang, Ganggang Mu, Lei Shen, Jun Liu, Zhengqiang Wang, Wei Zhou, Ping An, Xu Huang, et al. Detection of colorectal adenomas with a real-time computer-aided system (endoangel): a randomised controlled study. *The lancet Gastroenterology & hepatology*, 5(4):352–361, 2020.

[7] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26*, pages 451–462. Springer, 2020.

[8] Ge-Peng Ji, Guobao Xiao, Yu-Cheng Chou, Deng-Ping Fan, Kai Zhao, Geng Chen, and Luc Van Gool. Video polyp segmentation: A deep learning perspective. *Machine Intelligence Research*, 19(6):531–549, 2022.

[9] Ira Ktena, Olivia Wiles, Isabela Albuquerque, Sylvestre-Alvise Rebuffi, Ryutaro Tanno, Abhijit Guha Roy, Shekoofeh Azizi, Danielle Belgrave, Pushmeet Kohli, Taylan Cemgil, et al. Generative models improve fairness of medical classifiers under distribution shifts. *Nature Medicine*, pages 1–8, 2024.

[10] Fangrui Lv, Jian Liang, Shuang Li, Bin Zang, Chi Harold Liu, Ziteng Wang, and Di Liu. Causality inspired representation learning for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8046–8056, 2022.

[11] Yiting Ma, Xuejin Chen, Kai Cheng, Yang Li, and Bin 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*, pages 387–396. Springer, 2021.

[12] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

[13] Masashi Misawa, Shin-ei Kudo, Yuichi Mori, Kinichi Hotta, Kazuo Ohtsuka, Takahisa Matsuda, Shoichi Saito, Toyoki Kudo, Toshiyuki Baba, Fumio Ishida, et al. 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, 2021.

[14] Chong Mou, Xintao Wang, Liangbin Xie, Jian Zhang, Zhongang Qi, Ying Shan, and Xiaohu Qie. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. *arXiv preprint arXiv:2302.08453*, 2023.

[15] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 577–582. 2023.

[16] Alexander K Pishva, Vajira Thambawita, Jim Torresen, and Steven A Hicks. Repolyp: A framework for generating realistic colon polyps with corresponding segmentation masks using diffusion models. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 47–52. IEEE, 2023.

[17] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[18] Zhenrong Shen, Xi Ouyang, Bin Xiao, Jie-Zhi Cheng, Dinggang Shen, and Qian Wang. Image synthesis with disentangled attributes for chest x-ray nodule augmentation and detection. *Medical Image Analysis*, 84:102708, 2023.

[19] Juan Silva, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International journal of computer assisted radiology and surgery*, 9:283–293, 2014.

[20] Yu Tian, Min Shi, Yan Luo, Ava Kouhana, Tobias Elze, and Mengyu Wang. Fairseg: A large-scale medical image segmentation dataset for fairness learning with fair error-bound scaling. *arXiv preprint arXiv:2311.02189*, 2023.

[21] David Vázquez, Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Antonio M López, Adriana Romero, Michal Drozdzal, Aaron Courville, et al. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of healthcare engineering*, 2017, 2017.

[22] Jun Wei, Yiwen Hu, Ruimao Zhang, Zhen Li, S Kevin Zhou, and Shuguang Cui. Shallow attention network for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 699–708. Springer, 2021.

[23] Chenwang Wu, Xiting Wang, Defu Lian, Xing Xie, and Enhong Chen. A causality inspired framework for model interpretation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2731–2741, 2023.

[24] Binxin Yang, Shuyang Gu, Bo Zhang, Ting Zhang, Xuejin Chen, Xiaoyan Sun, Dong Chen, and Fang Wen. Paint by example: Exemplar-based image editing with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18381–18391, 2023.

[25] Ling Zhang, Xiaosong Wang, Dong Yang, Thomas Sanford, Stephanie Harmon, Baris Turkbey, Bradford J Wood, Holger Roth, Andriy Myronenko, Daguang Xu, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE transactions on medical imaging*, 39(7):2531–2540, 2020.

[26] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.

[27] Xin Zhang, Jiaxian Guo, Paul Yoo, Yutaka Matsuo, and Yusuke Iwasawa. Paste, inpaint and harmonize via denoising: Subject-driven image editing with pre-trained diffusion model. *arXiv preprint arXiv:2306.07596*, 2023.

[28] Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. *arXiv preprint arXiv:2302.04867*, 2023.

[29] Lei Zhou. Spatially exclusive pasting: A general data augmentation for the polyp segmentation. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 01–07. IEEE, 2023.