A GAN Model with Controllable Lesion Generation for Synthetic Capsule Endoscopy Datasets

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

In this paper, we will address a novel approach to create a synthetic capsule endoscopy dataset. In the medical area, research using deep learning has been actively conducted. It is important to secure a large amount of high-quality datasets to develop a deep learning model. However, medical data have privacy concerns or data bias issues. For this reason, medical data for learning can be noisy and incomplete. Also, it is difficult to obtain qualitative and quantitative medical data. To overcome these limitations, one of the studies that has recently been in the spotlight is synthetic data research. If we use synthetic data to learn deep learning models, we can maintain a more uniform data format and label. In this study, we want to solve the problem of lack of data by creating enough endoscopic datasets by naturally synthesizing the desired lesions in the desired location. We applied the crop and paste method and CycleGAN to the capsule endoscopy dataset for the first time. After placing the desired lesion at the desired coordinates using the crop and paste method, a widely used Data Augmentation Technique, we achieve natural synthesis using the CycleGAN model. We propose an Image-to-Image model that adjusts the type of location and lesion of the generated synthetic data. Through high-quality synthetic data generated in this way, we aim to realize the potential of deep learning in the medical field.

1 Introduction

Studies using synthetic data for deep learning have recently gained popularity for several reasons: First, in many areas, such as healthcare or autonomous driving, where data for learning is scarce, collecting and labeling large amounts of real-world data can be difficult and time-consuming. Synthetic data can be generated to supplement the insufficient real data or to create a required dataset from scratch. Second, synthetic data can reduce the data imbalance that is common in deep learning. Synthetic data allows complete control of the data distribution, which can help reduce bias and generalization in the learning process. Finally, synthetic data can be used to protect sensitive information and maintain privacy, especially when processing medical or financial data. If we replace a dataset that contains a lot of sensitive data with synthetic data, we are free to use it for research.

Based on these advantages, synthetic data is used in many research areas, and synthetic medical data research is also being actively conducted in the medical area. Deep learning has the potential to innovate healthcare by improving diagnosis, treatment, and outcomes for patients. However, there are some problems that need to be addressed before deep learning fully integrate into the healthcare sector. This paper covers the quality and availability of data among those problems.

This paper is a study to create a synthetic capsule endoscopic dataset with the aim of generating high-quality synthetic data for algorithm training and performance improvement. Capsule endoscopy is a medical procedure that involves swallowing a pill-sized camera to capture images of the digestive

Submitted to the Workshop on Deep Generative Models for Health at NeurIPS 2023. Do not distribute.

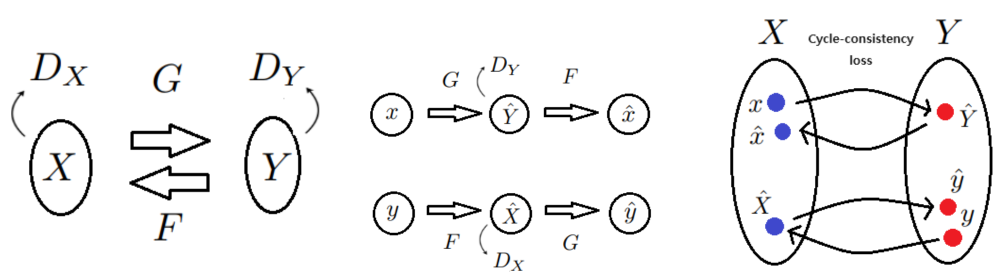
tract. It provides comprehensive data that can be used to diagnose and monitor various gastrointestinal disorders, including Crohn's disease, ulcers, and tumors. However, limited availability is a problem. Since capsule endoscopy is a relatively new technology, the availability of datasets is limited. As a result, there is a challenge of getting enough data to develop and train powerful deep learning models. However, capsule endoscopy datasets are promising tools for improving patient outcomes and advancing research in the medical field, requiring research to increase usability.

In this paper, we conduct research to synthesize data and improve image quality using image-to-image deep learning models. In addition, when synthesizing, we aim to place the desired lesion in the desired location. We create synthetic images that are indistinguishable from the actual dataset with lesions. The generated synthetic capsule endoscopy images can supplement the dataset to address the imbalance in the existing dataset, particularly in terms of lesion images that may be lacking. Also we can increase the total amount of datasets that are sufficient to increase the performance of detection or segmentation. With these studies, synthetic data is expected to improve the quality and diversity, making it a much more valuable tool for deep learning tasks.

2 CycleGAN with Crop and Paste Method

2.1 Model Baseline

The baseline model, as a CycleGAN[11] and depicted in Figure 1, is designed to learn the translation of images between x-domain and y-domain datasets without the need for paired images. We benefit from this approach as it eliminates the necessity for paired images with distinct attributes for the same image. Each domain constitutes a dataset containing images of non-natural lesions created using the crop and paste method, as well as a normal dataset without lesions. In this model, images of non-natural lesions, generated through the crop and paste method, are employed for training, alongside a dataset comprising natural, lesion-free images. Subsequently, the model synthesizes lesions into images in a natural manner. Other existing models[2, 4, 9, 12] have the advantage of achieving precise image conversion because they utilize paired data. Nevertheless, it's essential to acknowledge that the availability of paired data may not always be practical or feasible in many real-world scenarios. This is where the strength of CycleGAN, with its capability to work with unpaired data, becomes particularly valuable. Furthermore, we do not employ generative models that primarily focus on unconditional image generation, such as DCGAN[5] or StyleGAN[3]. While these models excel at generating realistic and high-quality images from random noise, our objective is to modify the content or style of images based on specific input conditions.



(a) (b) (c)

Figure 1: Overview of CycleGAN's translation task. CycleGAN allows the mapping function G(x) to reconstruct the original image x. (a) illustrates the connection between the two mapping functions, G and F, which are inversely functional, and includes discriminators Dy and Dx. Additionally, cycle consistency loss is introduced to enhance the ability of the translators G and F to restore the original image. (b) and (c) are graphs representing forward cycle-consistency loss and backward cycle-consistency loss.

2.2 Dataset

In this study, we employed the Kvasir capsule dataset[6], originally comprising 14 different classes. However, for our research, we focused on a subset of 5 classes, including normal images. To create our training dataset using the crop and paste method, we excluded lesions that covered more than half of an existing normal image due to excessively large bounding boxes. Specifically, lesions with bounding box sizes nearly equivalent to the image size, such as polyps, were omitted. We chose to work with blood, ulcer, erosion, and erythema lesions, as they could be naturally integrated into existing normal images. For the lesions primarily used in our learning process, we included 446 blood-fresh, 506 erosion, 117 erythema, and 854 ulcer images. It's important to note that the dataset we used in this study exhibited data imbalance, which we address as part of our research.

2.3 Implementation Details

Data Preparation: The initial step in this study involved creating the required dataset for CycleGAN using the crop and paste method. To achieve this, we first edited the csv file to select the desired classes for use in the crop and paste method. If the csv file contained only lesion information, we proceeded with the crop and paste method.

Crop and Paste Options: For the crop and paste method in this study, we had four options:

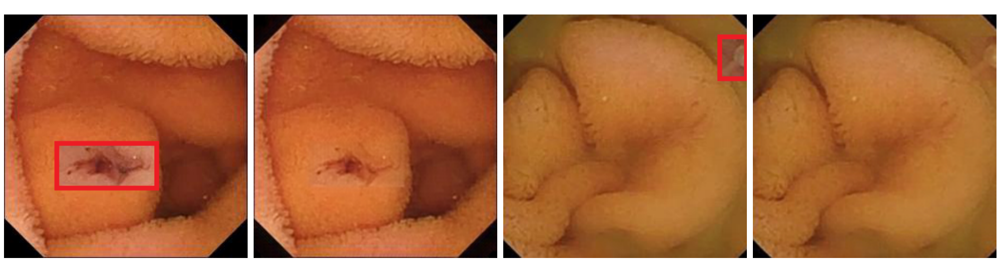
1. Apply the crop and paste method to all lesions in the csv file without position change.
2. Apply the crop and paste method to all lesions in the csv file with adjusting position.
3. Apply the crop and paste method to the desired lesion in the csv file without position change.
4. Apply the crop and paste method to the desired lesion in the csv file with adjusting position.

In our experiments, we primarily used methods 1 and 3. When applying the crop and paste method to the desired location, we made position adjustments as needed to ensure that the lesions fit within the image without altering their position significantly.

Training Details: For training, we utilized a CycleGAN model that was not pre-trained. We conducted training for 200 epochs with a learning rate set to 0.0002. Learning rate decay was linear, with the rate maintained at the same level during the first 100 epochs and then linearly decaying to zero over the next 100 epochs. We initialized weights from a Gaussian distribution $\mathcal{N}(0, 0.02)$. The input training data were pre-processed as 256×256 images.

2.4 Experiments and Results

In this study, lesions such as blood and ulcers were used to produce synthetic capsule endoscopic images.



(a) (b) (c) (d)

Figure 2: Original and result image of crop and paste method: (a) is cropped blood image, and (b) is the generated synthetic image result of (a). (c) is cropped ulcer image, and (d) is the generated synthetic image result of (c).

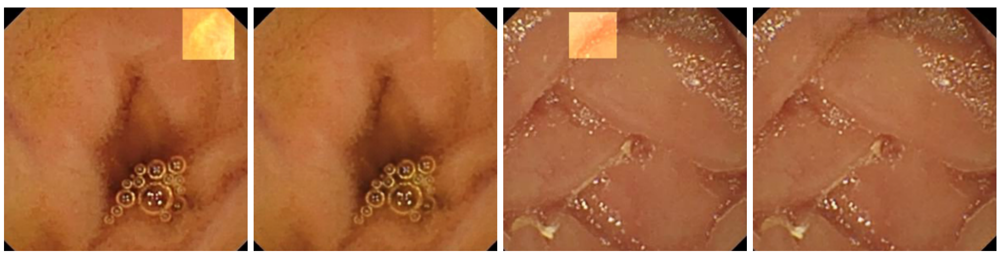
As shown in Figure 2, these images exemplify the crop and paste method applied to cropped blood and ulcer images, along with their resulting synthetic images.

In Figure 3, we observe that our model can produce more natural synthetic images when the background color or view closely resembles that of the original image. These findings highlight the

(a) (b) (c)

Figure 3: Original and result image of crop and paste method with similar background. (a) is cropped erosion image, and (b) is the generated synthetic image result of (a). (c) is the original image of the lesion and looks similar to (a), and the color is also similar.

importance of background similarity in generating more natural synthetic images, although such pre-processing tasks can be time-consuming.



(a) (b) (c) (d)

Figure 4: The result of crop and paste transparency method. Ulcer in (a) and erosion in (c) were synthesized, but lesions are hard to find in (b) and (d).

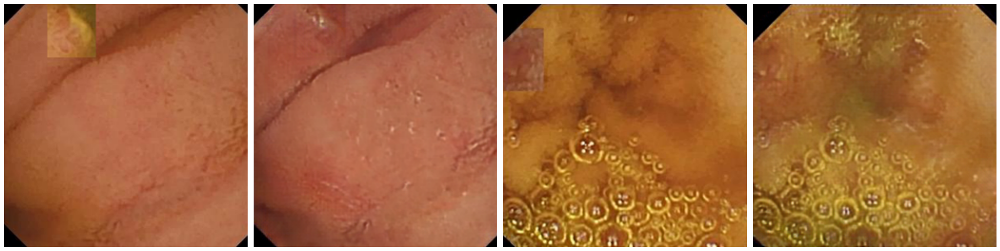
Additionally, we employed the crop and paste transparency method to compare its results with the crop and paste method. In the case of the crop and paste transparency method in Figure 4, the results indicate that the image of the cropped lesion was not adequately preserved. The characteristics of the original image remain strong, and the lesion image becomes lighter, making it difficult to discern the lesion properly.

(a) (b) (c)

Figure 5: Another lesion crop and paste method with the same normal image. (a) is the original normal image, and (b) and (c) synthesized each ulcer and blood.

Also, we can see that it's a model in Figure 5 that can easily create a image that represents a lesion or a image that we need by synthesizing another lesion with the same normal image.

We conducted an experiment in which we filled the normal image domain with the original lesion image for training. The goal was to investigate whether learning the image with the original lesion as a control group would result in more natural synthesis of the cropped lesions. However, this experiment revealed that when learning from the original lesion image, our model recognized a



(a) (b) (c) (d)

Figure 6: Training with the lesion image domain. In (a), we wanted to synthesize erosion, but the quality of (b) synthesis is low and overall red. We analyze under the influence of the large blood dataset of the lesion. In (c), erythema was attempted to be synthesized, but the lesion disappeared and the result (d) was produced with a lower resolution.

relatively wide range of lesions that were not present in the normal image. As seen in Figure 6, the overall color of the image was significantly affected, with some images turning red due to the influence of bleeding. Additionally, the generated images exhibited awkward lesions overall. Consequently, these experiments confirmed that, as initially intended, it is better to train the model on images without lesions and then synthesize the cropped lesion images naturally.

2.5 Limitations

In this section, we will address some of the limitations and potential solutions. Firstly, there is a resolution issue when synthesizing lesions using the crop and paste method, which leads to overall image smoothing and a reduction in resolution compared to the original. To mitigate this quality concern, we can employ super-resolution techniques[8, 10, 7], such as EndoL2H[1] in the endoscopy field, to restore the image to or above the original quality. Additionally, another challenge is the relatively lower ratio of natural outcomes compared to unnatural outcomes. This challenge can be addressed by adjusting the position and attaching the lesion to a natural location (Crop and Paste Options No. 2 or No. 4) instead of using methods that maintain the original position (Crop and Paste Options No. 1 or 3). Alternatively, better results can be achieved by introducing rotation or flip operations in addition to the crop and paste method, allowing for more natural input values.

3 Conclusion

In conclusion, this study is significant for its ability to naturally synthesize multiple lesion images, which addresses the data imbalance problem. In this study, CycleGAN was applied to capsule endoscopy for the first time, offering the advantage of customizing the desired results by freely adjusting both lesion type and location. The utilization of the Kvasir capsule dataset, a real-world medical dataset, enhances the credibility and relevance of the research, demonstrating the practical applicability of the proposed methodology. However, it has limitations, such as minor image cropping and limited diversity in the lesions synthesized. Future research can focus on refining data imbalance mitigation techniques and incorporating a more diverse range of lesions in the training dataset, and improving the ratio of favorable results through resolution enhancement methods. Research aimed at improving data quality and addressing data imbalance issues in healthcare not only enhances healthcare applications but also has significant implications for synthetic data research in other domains, such as robotics and autonomous driving.

References

[1] Y. Almalioglu, K. B. Ozyoruk, A. Gokce, K. Incetan, G. I. Gokceler, M. A. Simsek, K. Ararat, R. J. Chen, N. J. Durr, F. Mahmood, et al. Endol2h: deep super-resolution for capsule endoscopy. *IEEE Transactions on Medical Imaging*, 39(12):4297–4309, 2020.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2017.

[3] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks, 2019.

[4] M.-Y. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks, 2018.

[5] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2016.

[6] P. H. Smedsrud, V. Thambawita, S. A. Hicks, H. Gjestang, O. O. Nedrejord, E. Næss, H. Borgli, D. Jha, T. J. D. Berstad, S. L. Eskeland, M. Lux, H. Espeland, A. Petlund, D. T. D. Nguyen, E. Garcia-Ceja, D. Johansen, P. T. Schmidt, E. Toth, H. L. Hammer, T. de Lange, M. A. Riegler, and P. Halvorsen. Kvasir-Capsule, a video capsule endoscopy dataset. *Scientific Data*, 8(1):142, 2021.

[7] X. Song, H. Tang, C. Yang, G. Zhou, Y. Wang, X. Huang, J. Hua, G. Coatrieux, X. He, and Y. Chen. Deformable transformer for endoscopic video super-resolution. *Biomedical Signal Processing and Control*, 77:103827, 2022.

[8] M. Turan. A generative adversarial network based super-resolution approach for capsule endoscopy images. *Medicine Science*, 10(3):1002–1007, 2021.

[9] Z. Yi, H. Zhang, P. Tan, and M. Gong. Dualgan: Unsupervised dual learning for image-to-image translation, 2018.

[10] C. Zhou, K. Qiu, C. Chen, D. Zhang, and Y. Guo. Video super-resolution for wireless capsule endoscopy imaging sensor. *IEEE Sensors Journal*, 22(17):17283–17290, 2022.

[11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[12] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*, 2017.