Synthetic data for unsupervised polyp segmentation *

Enric Moreu[1,2][0000−0003−1336−6477], Kevin McGuinness[1,2][0000−0003−1336−6477], and Noel E. O'Connor[1,2][0000−0002−4033−9135]

[1] Insight SFI Centre for Data Analytics, Ireland
[2] Dublin City University, Ireland

Abstract. Deep learning has shown excellent performance in analysing medical images. However, datasets are difficult to obtain due privacy issues, standardization problems, and lack of annotations. We address these problems by producing realistic synthetic images using a combination of 3D technologies and generative adversarial networks. We use zero annotations from medical professionals in our pipeline. Our fully unsupervised method achieves promising results on five real polyp segmentation datasets. As a part of this study we release Synth-Colon, an entirely synthetic dataset that includes 20 000 realistic colon images and additional details about depth and 3D geometry: https://enric1994.github.io/synth-colon

Keywords: Computer Vision · Synthetic Data · Polyp Segmentation · Unsupervised Learning

1 Introduction

Colorectal cancer is one of the most commonly diagnosed cancer types. It can be treated with an early intervention, which consists of detecting and removing polyps in the colon. The accuracy of the procedure strongly depends on the medical professionals experience and hand-eye coordination during the procedure, which can last up to 60 minutes. Computer vision can provide real-time support for doctors to ensure a reliable examination by double-checking all the tissues during the colonoscopy.

The data obtained during a colonoscopy is accompanied by a set of issues that prevent creating datasets for computer vision applications. First, there are privacy issues because it is considered personal data that can not be used without the consent of the patients. Second, there are a wide range of cameras and lights used to perform colonoscopies. Every device has its own focal length, aperture,

* This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765140. This publication has emanated from research supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2, co-funded by the European Regional Development Fund.

Fig. 1. Synth-Colon dataset samples include: synthetic image, annotation, realistic image, depth map, and 3D mesh (from left to right).

and resolution. There are no large datasets with standardized parameters. Finally, polyp segmentation datasets are expensive because they depend on the annotations of qualified professionals with limited available time.

We propose an unsupervised method to detect polyps that does not require annotations by combining 3D rendering and a CycleGAN [23]. First, we produce artificial colons and polyps based on a set of parameters. Annotations of the location of the polyps are automatically generated by the 3D engine. Second, the synthetic images are used alongside real images to train a CycleGAN. The CycleGAN is used to make the synthetic images appear more realistic. Finally, we train a HarDNeT-based model [3], a state-of-the-art polyp segmentation architecture, with the realistic synthetic data and our self-generated synthetic labels.

The contributions of this paper are as follows:

- To the best of our knowledge, we are the first to train a polyp segmentation model with zero annotations from the real world.
- We propose a pipeline that preserves the self-generated annotations when shifting the domain from synthetic to real.
- We release Synth-Colon (see Figure 1), the largest synthetic dataset for polyp segmentation including additional data such as depth and 3D mesh.

The remainder of the paper is structured as follows: Section 2 reviews relevant work, Section 3 explains our method, Section 4 presents the Synth-Colon dataset, Section 5 describes our experiments, and Section 6 concludes the paper."

2 Related work

Here we briefly review some relevant works related to polyp segmentation and synthetic data.

2.1 Polyp segmentation

Early polyp segmentation was based in the texture and shape of the polyps. For example, Hwang et al. [8] used ellipse fitting techniques based on shape. However, some corectal polyps can be small (5mm) and are not detected by these

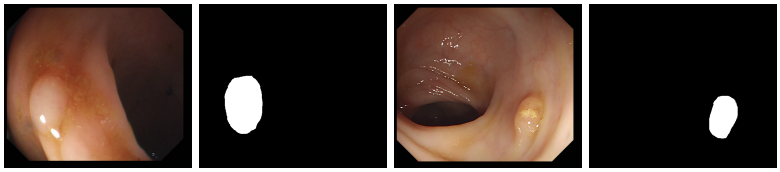

Fig. 2. Real samples from CVC-ColonDB with the corresponding annotation made by a medical professionals indicating the location of cancerous polyps.

techniques. In addition, the texture is easily confused with other tissues in the colon as can be seen in Figure 2.

With the rise of convolutional neural networks (CNNs) [10] the problem of the texture and shape of the polyps was solved and the accuracy was substantially increased. Several authors have applied deep convolutional networks to the polyp segmentation problem. Brandao et al. [2] proposed to use a fully convolutional neural network based on the VGG [16] architecture to identify and segment polyps. Unfortunately, the small datasets available and the large number of parameters make these large networks prone to overfitting. Zhou et al. [22] used an encoder-decoder network with dense skip pathways between layers that prevented the vanishing gradient problem of VGG networks. They also significantly reduced the number of parameters, reducing the amount of overfitting. More recently, Chao et al. [3] reduced the number of shortcut connections in the network to speed-up inference time, a critical issue when performing real-time colonoscopies in high-resolution. They focused on reducing the memory traffic to access intermediate features, reducing the latency. Finally, Huang et al. [7] improved the performance and inference time by combining HarDNet [3] with a cascaded partial decoder [21] that discards larger resolution features of shallower layers to reduce latency.

2.2 Synthetic data

The limitation of using large neural networks is that they often require large amounts of annotated data. This problem is particularly acute in medical imaging due to problems in privacy, standardization, and the lack of professional annotators. Table 1 shows the limited size and resolution of the datasets used to train and evaluate existing polyp segmentation models. The lack of large datasets for polyp segmentation can be addressed by generating synthetic data.

Thambawita et al. [18] used a generative adversarial network (GAN) to produce new colonoscopy images and annotations. They added a fourth channel to SinGAN [14] to generate annotations that are consistent with the colon image. They then used style transfer to improve the realism of the textures. Their results are excellent considering the small quantity of real images and professional annotations that are used. Gao et al. [6] used a CycleGAN to translate colonoscopy images to polyp masks. In their work, the generator learns how to segment polyps by trying to fool a discriminator.

Table 1. Real polyp segmentation datasets size and resolution.

Dataset	#Images	Resolution
CVC-T [19]	912	574 x 500
CVC-ClinicDB [1]	612	384 x 288
CVC-ColonDB [17]	380	574 x 500
ETIS-LaribPolypDB [15]	196	1225 x 966
Kvasir [9]	1000	Variable

Synthetic images combined with generative networks have also been widely used in the depth prediction task [11,12]. This task helps doctors to verify that all the surfaces in the colon have been analyzed. Synthetic data is essential for this task because of the difficulties to obtain depth information in a real colonoscopy.

Unlike previous works, our method is entirely unsupervised and does not require any human annotations. We automatically generate the annotations by defining the structure of the colon and polyps and transferring the location of the polyps to a 2D mask. The key difference between our approach and other state-of-the-art is that we combine 3D rendering and generative networks. First, the 3D engine defines the structure of the image and generates the annotations. Second, the adversarial network makes the images realistic.

Similar unsupervised methods have also been successfully applied in other domains like crowd counting. For example, Wang et al. [20] render crowd images from a video game and then use a CycleGAN to increase the realism.

3 Method

Our approach is composed of three steps: first, we procedurally generate colon images and annotations using a 3D engine; second, we feed a CycleGAN with images from real colonoscopies and our synthetic images; finally, we use the realistic images created by CycleGAN to train an image segmentation model.

3.1 3D colon generation

The 3D colon and polyps are procedurally generated using Blender, a 3D engine that can be automated via scripting.

Our 3D colons structure is a cone composed by 2454 faces. Vertices are randomly displaced following a normal distribution in order to simulate the tissues in the colon. Additionally, the colon structure is modified by displacing 7 segments as in Figure 3. For the textures we used a base color [0.80, 0.13, 0.18] (RGB). For each sample we shift the color to other tones of brown, orange and pink. One single polyp is used on every image, which is placed inside the colon. It can be either in the colon's walls or in the middle. Polyps are distorted spheres with 16384 faces. Samples with polyps occupying less than 20,000 pixels are removed.

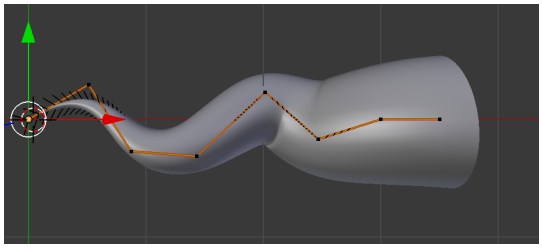

Fig. 3. The structure of the colon is composed by 7 segments to simulate the curvature of the intestinal tract.

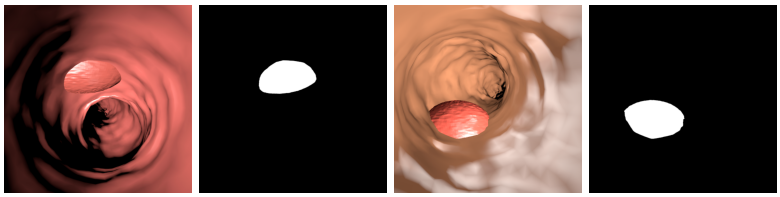

Fig. 4. Synthetic colons with corresponding annotations rendered using a 3D engine.

Lighting is composed by a white ambient light, two white dynamic lights that project glare into the walls, and three negative lights that project black light at the end of the colon. We found that having a dark area at the end helps CycleGAN to understand the structure of the colon. The 3D scene must be similar to real colon images because otherwise, the CycleGAN will not translate properly the images to the real-world domain. Figure 4 illustrates the images and ground truth generated by the 3D engine.

3.2 CycleGAN

A standard CycleGAN composed by two generators and two discriminators is trained using real images from colonoscopies and synthetic images generated using the 3D engine as depicted in Figure 6. We train a CycleGAN for 200 epochs and then we infer real images in the "Generator Synth to Real" model, producing realistic colon images.

Figure 5 displays synthetic images before and after the CycleGAN domain adaptation. Note that the position of the polyps is not altered. Hence, the ground truth information generated by the 3D engine is preserved.

3.3 Polyp segmentation

After creating a synthetic dataset that has been adapted to the real colon textures, we train an image segmentation model. We used the HarDNeT-MSEG [7] model architecture because of its real-time performance and high accuracy. We use the same hyperparameter configuration as in the original paper.

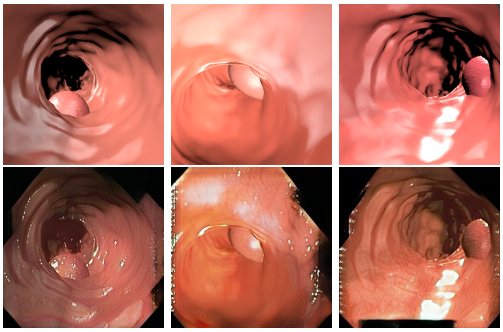

Fig. 5. Synthetic images (first row) and realistic images generated by our CycleGAN (second row).

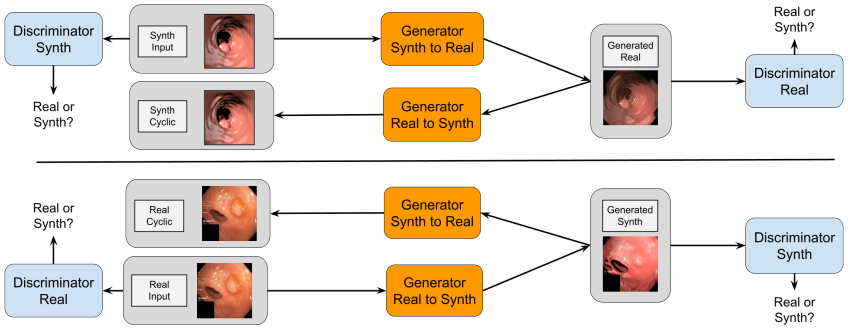

Fig. 6. Our CycleGAN architecture. We train two generator models that try to fool two discriminator models by changing the domain of the images.

4 Synth-Colon

We publicly release Synth-Colon, a synthetic dataset for polyp segmentation. It is the first dataset generated using zero annotations from medical professionals. The dataset is composed of 20 000 images with a resolution of 500×500. Synth-Colon additionally includes realistic colon images generated with our CycleGAN and the Kvasir training set images. Synth-Colon can also be used for the colon depth estimation task [12] because we provide depth and 3D information for each image. Figure 1 shows some examples from the dataset. In summary, Synth-Colon includes:

- Synthetic images of the colon and one polyp.
- Masks indicating the location of the polyp.
- Realistic images of the colon and polyps. Generated using CycleGAN and the Kvasir dataset.
- Depth images of the colon and polyp.
- 3D meshes of the colon and polyp in OBJ format.

5 Experiments

5.1 Metrics

We use two common metrics for evaluation. The mean Dice score, given by:

$$\text{mDice} = \frac{2 \times tp}{2 \times tp + fp + fn}, \tag{1}$$

and the mean intersection over union (IoU):

$$\text{mIoU} = \frac{tp}{tp + fp + fn}, \tag{2}$$

where in both forumlae, tp is the number of true positives, fp the number of false positives, and fn the number of false negatives.

5.2 Evaluation on real polyp segmentation datasets

We evaluate our approach on five real polyp segmentation datasets. Table 2 shows the results obtained when training HarDNeT-MSEG [7] using our synthetic data. Note that our method is not using any annotations. Results are satisfactory considering the fact that labels have been generated automatically. We found that training the CycleGAN with only the images from the target dataset performs better than training the CycleGAN with all the datasets combined, indicating a domain gap among the real-world datasets.

Table 2. Evaluation of our synthetic approach on real-world datasets. The metrics used are mean Dice similarity index (mDice) and mean Intersection over Union (mIoU).

	CVC-T		ColonDB		ClinicDB		ETIS		Kvasir	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
U-Net [13]	0.710	0.627	0.512	0.444	0.823	0.755	0.398	0.335	0.818	0.746
SFA [5]	0.467	0.329	0.469	0.347	0.700	0.607	0.297	0.217	0.723	0.611
PraNet [4]	0.871	0.797	0.709	0.640	0.899	0.849	0.628	0.567	0.898	0.840
HarDNet-MSEG [7]	0.887	0.821	0.731	0.660	0.932	0.882	0.677	0.613	0.912	0.857
Synth-Colon (ours)	0.703	0.635	0.521	0.452	0.551	0.475	0.257	0.214	0.759	0.527

5.3 Study with limited real data

In this section we evaluate how our approach based on synthetic imagery and domain adaptation compares with the fully supervised state-of-the-art HarDNeT-MSEG network when there are fewer training examples available. We train the CycleGAN used in the proposed approach, without ground truth segmentation labels, on progressively larger sets of imagery, and compare this with the supervised method trained on the same amount of labelled imagery. Table 3 shows the results of the experiment, which demonstrates that synthetic data is extremely useful for domains where annotations are very scarce. While our CycleGAN can produce realistic images with a small sample of only five real images, supervised methods require many images and annotations to achieve good performance. Table 3 shows that our unsupervised approach is useful when there are less than 50 real images and annotations. Note that zero images here means there is no domain adaptation via the CycleGAN.

Table 3. Evaluation of the proposed approach on the Kvasir dataset when few real images are available. The performance is measured using the mean Dice metric.

	Synth-Colon (ours)	HarDNeT-MSEG [7]
0 images	0.356	-
5 images	0.642	0.361
10 images	0.681	0.512
25 images	0.721	0.718
50 images	0.735	0.781
900 (all) images	0.759	0.912

6 Conclusions

We successfully trained a polyp segmentation model without annotations from doctors. We used 3D rendering to generate the structure of the colon and generative adversarial networks to make the images realistic, and demonstrated

that it can perform quite reasonably in several datasets, even outperforming some fully supervised methods in some cases. We hope this study can help aligning synthetic data and medical imaging in future. As future work, we will explore how to include our synthetic annotations in the CycleGAN.

References

1. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics **43**, 99–111 (2015)
2. Brandao, P., Mazomenos, E., Ciuti, G., Caliò, R., Bianchi, F., Menciassi, A., Dario, P., Koulaouzidis, A., Arezzo, A., Stoyanov, D.: Fully convolutional neural networks for polyp segmentation in colonoscopy. In: Medical Imaging 2017: Computer-Aided Diagnosis. vol. 10134, p. 101340F. International Society for Optics and Photonics (2017)
3. Chao, P., Kao, C.Y., Ruan, Y.S., Huang, C.H., Lin, Y.L.: Hardnet: A low memory traffic network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3552–3561 (2019)
4. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 263–273. Springer (2020)
5. Fang, Y., Chen, C., Yuan, Y., Tong, K.y.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 302–310. Springer (2019)
6. Gao, H., Ogawara, K.: Adaptive data generation and bidirectional mapping for polyp images. In: 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). pp. 1–6. IEEE (2020)
7. Huang, C.H., Wu, H.Y., Lin, Y.L.: Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps (2021)
8. Hwang, S., Oh, J., Tavanapong, W., Wong, J., De Groen, P.C.: Polyp detection in colonoscopy video using elliptical shape feature. In: 2007 IEEE International Conference on Image Processing. vol. 2, pp. II–465. IEEE (2007)
9. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: International Conference on Multimedia Modeling. pp. 451–462. Springer (2020)
10. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)
11. Mahmood, F., Chen, R., Durr, N.J.: Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. IEEE transactions on medical imaging **37**(12), 2572–2581 (2018)
12. Rau, A., Edwards, P.E., Ahmad, O.F., Riordan, P., Janatka, M., Lovat, L.B., Stoyanov, D.: Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy. International journal of computer assisted radiology and surgery **14**(7), 1167–1176 (2019)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

14. Rott Shaham, T., Dekel, T., Michaeli, T.: Singan: Learning a generative model from a single natural image. In: Computer Vision (ICCV), IEEE International Conference on (2019)
15. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. International journal of computer assisted radiology and surgery **9**(2), 283–293 (2014)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
17. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE transactions on medical imaging **35**(2), 630–644 (2015)
18. Thambawita, V., Salehi, P., Sheshkal, S.A., Hicks, S.A., Hammer, H.L., Parasa, S., de Lange, T., Halvorsen, P., Riegler, M.A.: Singan-seg: Synthetic training data generation for medical image segmentation. arXiv preprint arXiv:2107.00471 (2021)
19. Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdzal, M., Courville, A.: A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of healthcare engineering **2017** (2017)
20. Wang, Q., Gao, J., Lin, W., Yuan, Y.: Learning from synthetic data for crowd counting in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8198–8207 (2019)
21. Wu, Z., Su, L., Huang, Q.: Cascaded partial decoder for fast and accurate salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3907–3916 (2019)
22. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 3–11. Springer (2018)
23. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on (2017)