

Photorealistic Colonoscopy Image Generation

Ángel Villanueva Agudo, Clara Tomasini, O. León Barbed, Ana C. Murillo

Robótica, Visión por Computador e Inteligencia Artificial (RoPeRT)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-876555459, e-mail: 844759@unizar.es

Abstract

This paper explores controlled synthetic colonoscopy image generation with polyps using ControlNet [1] and depth maps. Generated images augment training of a polyp segmentation network, reducing reliance on real clinical data. PolypMini, the best generated dataset, achieves competitive results for standard and small morphologies compared with real-data training.

Motivation and objective

Automatic polyp segmentation in colonoscopy is a clinically relevant task to support early detection of colorectal cancer. In order to train deep learning models for this task, it is required to have large volumes of specialist-annotated images, but in medicine annotated data is scarce, costly to obtain, and subject to ethical and privacy restrictions. This work studies how to reduce this dependency through generation of realistic synthetic data. In particular the generation of images that automatically include the polyp segmentation masks needed for training. The main objective is to design a system that generates image-mask pairs of colonoscopies with polyps, in order to analyse whether they can replace or complement real images in the training of a polyp segmentation model like PraNet [2]. Our work builds on ControlNet [1], a conditional diffusion model that guides image generation using structural inputs, adapted to the endoscopic domain and conditioned by depth maps to control the position, relief, and morphology of the generated polyp.

Methodology

The proposed framework consists of three main stages: control-signal construction, synthetic image generation, and evaluation through segmentation training. In the first stage, real polyp-free backgrounds from the EndoMapper dataset [3] are combined with annotated polyp masks and

examples from Kvasir-SEG [5] and CVC-ClinicDB [6]. Depth maps are estimated from the RGB images using Depth-Anything [4].

To build realistic control signals, the polyp mask is inserted into the selected background using several geometric and structural criteria. Pasting regions are selected using SSIM, the polyp is locally scaled according to the depth map, and its position is randomly sampled within realistic ranges. Small polyps are also considered, and cases partially outside the field of view are analysed using classical computer vision techniques such as Canny, Sobel, and the Hough transform. These operations aim to avoid artificial contours and preserve the geometric coherence of the colonoscopy scene.

In the second stage, the resulting depth-based control signal guides the inference of the ControlNet [1] depth-to-image model, producing a synthetic colonoscopy image with a polyp and its corresponding segmentation mask. Finally, the generated image-mask pairs are used to train PraNet [2] and evaluate whether synthetic data can replace or complement real clinical images.

Experiments and Results

To test the relevance/usefulness of the synthetic generated data for polyp segmentation, we train the polyp segmentation model PraNet [2] on various data combinations consisting of 2900 real and/or synthetic images. Different versions of our fully synthetic training sets have been compared with mixed real-synthetic sets. The evaluation was performed on CVC-300 [7] and CVC-ClinicDB [6] using the standard segmentation metrics Dice and IoU. For all training sets, PraNet [2] was trained for 20 epochs, with learning rate of 10^{-4} and batch size of 16.

Table 1 summarizes the results. Among fully synthetic datasets, PolypMini was the most robust, showing that local scaling, SSIM-based selection,

and small-polyp examples improve the usefulness of generated data for segmentation. Performance on CVC-ClinicDB was lower for synthetic-only training due to its more complex morphologies. In mixed training, RealMiniMax achieved the best overall results, close to real-data training and competitive with ArSDM [8]. This suggests that synthetic data cover standard cases well, while a small amount of real data remains important for rare morphologies.

Conclusions

The results confirm that generation through conditional diffusion can produce useful data for training polyp segmentation models, especially for standard and small morphologies. The developed methodology clearly improves the initial system and makes it possible to reduce the use of real clinical data, promoting privacy and the safe reuse of medical information.

As a limitation, the model still does not reproduce highly irregular polyps, multiple polyps, or complex morphologies with sufficient fidelity. Future work should explore alternative generators or more expressive conditional models, as well as clinical validation strategies and automatic selection of higher-quality synthetic samples.

REFERENCES

- [1]. ZHANG, L. and AGRAWALA, M. Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
- [2]. FAN, D.-P., JI, G.-P., ZHOU, T., et al. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. MICCAI, 2020, pp. 263–273.
- [3]. AZAGRA, P., SOSTRES, C., FERRÁNDEZ, Á., et al. EndoMapper dataset of complete calibrated endoscopy procedures. Scientific Data. 2023, 10, 671.
- [4]. YANG, X., ZHANG, Q., ZHANG, Y., et al. Depth Anything: Unleashing the power of large-scale unlabeled data. In: CVPR, 2024, pp. 28785–28795.
- [5]. JHA, D., SMEDSRUD, P.H., RIEGLER, M.A., et al. Kvasir-SEG: A segmented polyp dataset. In: MultiMedia Modeling. Cham: Springer, 2020, pp. 451–462.
- [6]. BERNAL, J., SÁNCHEZ, F.J., FERNÁNDEZ, G., GIL, D., RODRÍGUEZ, C. and VILARIÑO, F. CVC-ClinicDB: database for polyp segmentation. Computer Vision Center, 2015.
- [7]. COMPUTER VISION CENTER and HOSPITAL CLÍNIC BARCELONA. CVC-300 polyp dataset. Barcelona: CVC, 2012.
- [8]. DU, Y., JIANG, Y., TAN, S., et al. ArSDM: Colonoscopy Images Synthesis with Adaptive Refinement Semantic Diffusion Models. In: MICCAI, 2023.

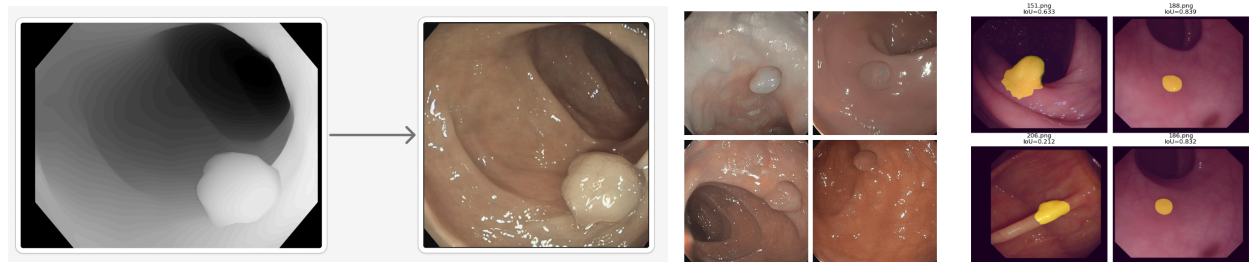


Figure 1. Left: diffusion process guided by a depth map with an inserted polyp. Right: representative synthetic samples generated with PolypMini and segmentation examples.

Table 1. PraNet segmentation results and key synthetic dataset properties. Results are reported as Dice/IoU. S: fully synthetic; M: 50% real, 50% synthetic from PolypMini and PolypMax; R: real only. Check marks indicate properties included during generation.

Dataset	Type	Small	Rare/edge	SSIM	Local scaling	CVC-300	CVC-ClinicDB
PolypMax	S	✓	✓	✗	✗	0.51 / 0.43	0.50 / 0.41
PolypCore	S	✗	✗	✓	✓	0.80 / 0.72	0.45 / 0.39
PolypMini	S	✓	✗	✓	✓	0.84 / 0.75	0.52 / 0.45
RealMiniMax	M	—	—	—	—	0.89 / 0.82	0.91 / 0.85
RealExp	R	—	—	—	—	0.88 / 0.80	0.90 / 0.85