

# Multi-Center Polyp Segmentation with Double Encoder-Decoder Networks

Adrian Galdran<sup>a</sup>, Gustavo Carneiro<sup>b</sup> and Miguel A. González Ballester<sup>c,d</sup>

<sup>a</sup>Bournemouth University, Bournemouth, United Kingdom

<sup>b</sup>University of Adelaide, Adelaide, Australia

<sup>c</sup>BCN Medtect, Dept. of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

<sup>d</sup>Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

## Abstract

Polyps are among the earliest sign of Colorectal Cancer, with their detection and segmentation representing a key milestone for automatic colonoscopy analysis. This work describes our solution to the EndoCV 2021 challenge, within the sub-track of polyp segmentation. We build on our recently developed framework of pretrained double encoder-decoder networks, which has achieved state-of-the-art results for this task, but we enhance the training process to account for the high variability and heterogeneity of the data provided in this competition. Specifically, since the available data comes from six different centers, it contains highly variable resolutions and image appearances. Therefore, we introduce a center-sampling training procedure by which the origin of each image is taken into account for deciding which images should be sampled for training. We also increase the representation capability of the encoder in our architecture, in order to provide a more powerful encoding step that can better capture the more complex information present in the data. Experimental results are promising and validate our approach for the segmentation of polyps in a highly heterogeneous data scenarios.

## Keywords

Polyp segmentation, multi-center data, double encoder-decoders

## 1. Introduction

Colorectal Cancer (CRC) is among the most concerning diseases affecting the human gastrointestinal tract, representing the second most common cancer type in women and third most common for men [1]. CRC treatment begins with colorectal lesion detection, which is typically performed during colonoscopic screenings. In these procedures, a flexible tube equipped with a camera is introduced through the rectum to look for such lesions throughout the colon. Early detection of CRC is known to substantially increase survival rates. Unfortunately, it is estimated that around 6-27% of pathologies are missed during a colonoscopic examination [2]. Colonoscopic image analysis and decision support systems have shown great promise in improving examination effectiveness and decreasing the amount of missed lesions [3].


---

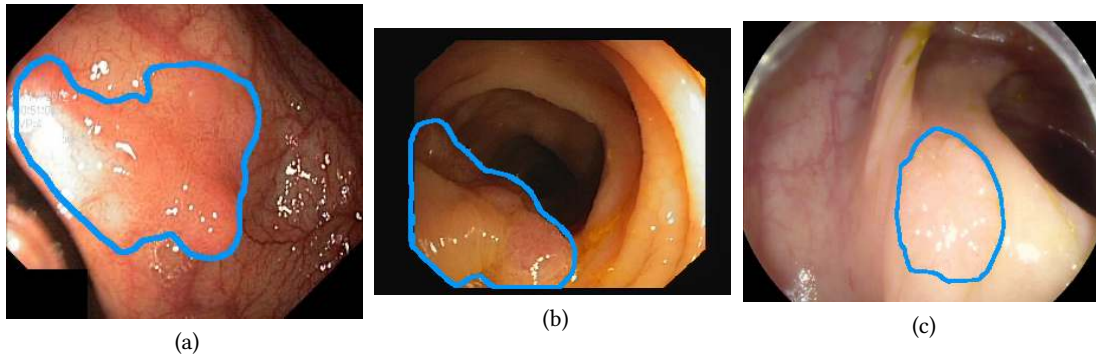
*3rd International Workshop and Challenge on Computer Vision in Endoscopy (EndoCV2021) in conjunction with the 18th IEEE International Symposium on Biomedical Imaging ISBI2021, April 13th, 2021, Nice, France*

✉ agaldran@bournemouth.ac.uk (A. Galdran); gustavo.carneiro@adelaide.edu (G. Carneiro); ma.gonzalez@upf.edu (M. A. González Ballester)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

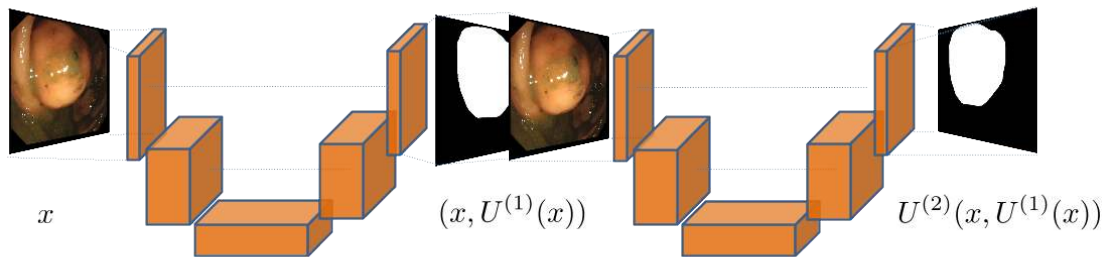


**Figure 1: Polyp visual aspects have a wide variability in terms of shape, appearances, and boundaries.** In this figure, polyps sampled from different databases, acquired in different centers and with different equipment: (a) Kvasir-Seg [11], (b) CVC-ColonDB [12], (c) ETIS [13].

Gastrointestinal polyps are among the most relevant pathologies to be found in colonoscopies, since they are one of the main early signs of CRC [4]. However, their correct identification and accurate segmentation are challenging tasks for both clinicians and computational techniques, due to their wide and highly variable range of shapes and visual appearances, as illustrated in Fig. 1. For this reason, computer-aided polyp detection has been extensively explored in recent years as a supplementary tool for colonoscopic procedures to improve detection rates, enable early treatment, and increase survival rates.

Polyp segmentation is often approached by means of encoder-decoder convolutional neural networks. In [5] an encoder-decoder network containing multi-resolution, multi-classification, and fusion sub-networks was introduced, whereas [6] explored several combinations of different encoder and decoder architectures. In [7] an architecture with a shared encoder and two inter-dependent decoders was proposed to model polyp areas and boundaries respectively, and in [8] ensembles of instance-segmentation architectures were studied. More recently, in [9] the authors proposed parallel reverse attention layers to model the relationship between polyp areas and their boundaries. A recent review of different approaches to polyp segmentation (and detection) on gastroendoscopic images can be found in [10].

This work describes our solution to the EndoCV 2021 challenge on the polyp segmentation track [14]. The proposed approach is based on our recently introduced solution for polyp segmentation [15], consisting of a cascaded double encoder-decoder Convolutional Neural Network, which achieved the first position on the EndoTect Challenge [16]. We improve upon our previous approach by 1) increasing the representation capability of the pre-trained encoder, and 2) adopting a multi-site sampling scheme to better capture the varying nature of endoscopic data during training. Our approach is straightforward to implement, yet it delivers outstanding performance for the task of polyp segmentation. Our experimental analysis, even if limited due to the final results of the competition not being released at the time of writing, demonstrate that the proposed technique is highly effective and can reliably generate accurate polyp segmentations on endoscopic images of a highly varying visual aspect.



**Figure 2: Double encoder-decoder networks for polyp segmentation.** The first encoder-decoder  $U^{(1)}$  generates an initial attempt to segment the polyp, which is supplied to the second encoder-decoder  $U^{(2)}$  together with the original image  $x$ . This guides the learning of  $U^{(2)}$  towards more challenging regions within  $x$ .

## 2. Methodology

### 2.1. Double Encoder-Decoder Networks

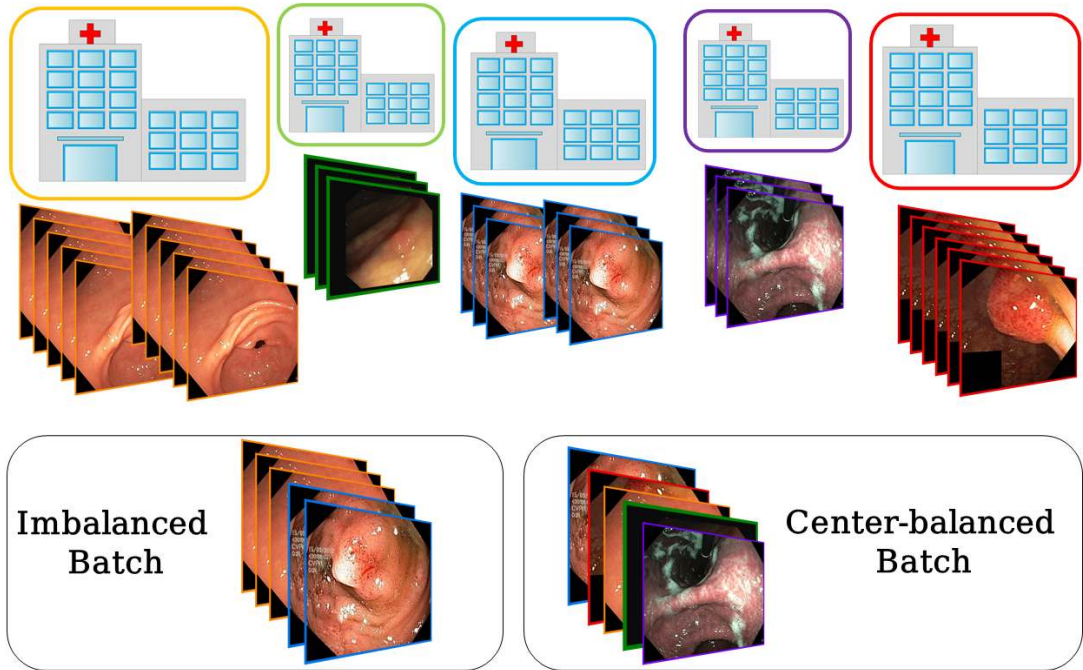
Dense semantic segmentation tasks are nowadays typically approached with encoder-decoder networks [17] equipped with skip connections that produce pixel-wise probabilities. The encoder acts as a feature extractor downsampling spatial resolutions while increasing the number of channels by learning convolutional filters. The decoder then upsamples this representation back to the original input size. Double encoder-decoders are a direct extension of encoder-decoder architectures in which two encoder-decoder networks are sequentially combined [18]. Being  $x$  an input RGB image,  $U^{(1)}$  the first network, and  $U^{(2)}$  the second network, in a double encoder-decoder, the output  $U^{(1)}(x)$  of the first network is fed to the second network together with  $x$ , behaving like an attention map that allows  $U^{(2)}$  to focus on the most interesting areas of the image:

$$U(x) = U^{(2)}(x, U^{(1)}(x)), \quad (1)$$

where  $x$  and  $U^{(1)}(x)$  are stacked so that the input to  $U^{(2)}$  has four channels, as illustrated in Fig. 2. In this work we employ the same structure in both  $U^{(1)}$  and  $U^{(2)}$ : we select a Feature Pyramid Network architecture as the decoder [19]. In addition, in order to take into account the more complex data characteristics in this challenge, we increase the encoder capability (as compared to [15]) by leveraging the more powerful DPN92 architecture instead of the DPN68 CNN [20]. Note also that during training we apply pixel-wise supervision on both  $U^{(1)}$  and  $U^{(2)}$  by computing the Cross-Entropy loss between  $U^{(2)}(x)$  and the corresponding manual segmentation  $y$ , but also between  $U^{(1)}(x)$  and  $y$ .

### 2.2. Multi-Center Sampling

The nature of the provided database of segmented polyps for this competition is highly heterogeneous, with samples collected from 6 different centers. This leads to a widely variable training set containing images of varying resolutions, visual quality, and even diverse annotation styles.



**Figure 3: Multi-Center training data sampling strategy.** Instead of a regular data sampling in which each image in the training set is shown to the model based on a uniform probability distribution (resulting in center-imbalanced batches, bottom left), in this work we implement a center-based sampling technique in which each mini-batch of images contains a proportionate representation of images from the different centers on which the data was collected (center-balanced batches, bottom right).

In this work, we attempt to facilitate the training of the CNN described in the previous section on such irregular dataset by considering the origin of each sample (its center) when designing our training sampling approach.

Modified sampling strategies are typically used in classification problems when there is a high class imbalance during training, the most typical scheme being oversampling under-represented categories. In our case, we consider the set of different centers provided by the organization,  $\mathcal{C}_1, \dots, \mathcal{C}_6$  as our categories. We denote our training set as  $\mathcal{D} = \{(x_i, y_i, c_i), i = 1, \dots, N\}$ , where  $x_i$  is an image containing a polyp,  $y_i$  its manual segmentation, and  $c_i \in \{1, \dots, 6\}$  its original center. In our case, each class/center  $c$  contains  $n_c$  examples, so that  $\sum_{c=1}^6 n_c = N$ . With this notation, most data sampling strategies can be described with a single equation as follows:

$$p_j = \frac{n_j^q}{\sum_{c=1}^6 n_c^q}, \quad (2)$$

where  $p_j$  is the probability of sampling an image from center  $j$  during training. By specifying  $q = 1$ , we are defining a sampling scheme akin to selecting examples with a probability equal to the frequency of their center in the training set (conventional sampling), while setting  $q = 0$  leads to a uniform probability  $p_j = 1/6$  of sampling from each center, this is, oversampling of minority centers in order to supply our CNN with mini-batches containing representative

**Table 1**

Performance comparison for polyp segmentation of different double encoder-decoder networks on the hidden validation set in terms of mean and standard deviation of Dice score

	FPN/ResNet34	FPN/DPN68	FPN/DPN92
<b>Dice (Mean <math>\pm</math> Std)</b>	79.12 $\pm$ 4.32	81.69 $\pm$ 2.34	<b>81.81 <math>\pm</math> 1.19</b>

images from all sites. The latter is the sampling strategy implemented in this work.

### 2.3. Training Details

Our models are trained on training data (five different center data) provided by EndoCV2021 challenge organisers [14]. Here, we minimized the cross-entropy loss using Stochastic Gradient Descent with a batch-size of 4 and a learning rate of  $l = 0.01$ , which is cyclically decayed following a cosine law from its initial value to  $l = 1e - 8$  during 25 epochs, which defines a training cycle. We repeat this process for 20 cycles, resetting the learning rate at the start of each cycle. Images are re-sampled to  $640 \times 512$  and during training they are augmented with standard operations (random rotations, vertical/horizontal flipping, contrast/saturation/brightness changes). The mean Dice score is monitored on a separate validation set and the best model is kept for testing purposes. In test time, we generate four different versions of each image by horizontal/vertical flipping, predict on each of them, and average the results.

## 3. Experimental Results

At the time of writing, final results on this challenge have not been released yet. An offline validation phase on unseen images (a subset of the final test set) was run by the organization for the participants to be able to perform model selection. This allows us to compare internally the performance of different versions of our approach<sup>1</sup>. Table 1 shows the performance of the system described in the previous sections when using three different double encoder-decoder networks, all of them trained with the center-sampling approach. It can be appreciated that increasing the complexity of the encoder correlates with a greater performance in terms of average Dice score. In addition, we can also observe a substantial decrease in standard deviation measured across centers when the more powerful DPN92 encoder architecture is employed, as compared to the smaller DPN68 or a more simple ResNet34, highlighting the relevance of this design decision.

## 4. Conclusions

In this work, we have detailed our solution for the EndoCV 2021 challenge on the polyp segmentation track. The proposed approach consists of a double encoder-network, enhanced with an improved encoder architecture and a training data sampling strategy specifically designed to deal with the multi-site nature of the data associated to this competition. The

<sup>1</sup>Details on performance analysis metrics for this problem can be found in [21].

limited experimental results show that our method achieves a consistently high Dice score with a remarkably low standard deviation, which indicates that it is suitable for polyp segmentation on endoscopic images, and it has enough generalization capability to perform well on images collected from different centers.

## Acknowledgments

Adrian Galdran was funded by a Marie Skłodowska-Curie Global Fellowship (No 892297).

## References

- [1] F. A. Haggar, R. P. Boushey, Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors, *Clinics in Colon and Rectal Surgery* 22 (2009) 191–197.
- [2] S. B. Ahn, D. S. Han, J. H. Bae, T. J. Byun, J. P. Kim, C. S. Eun, The Miss Rate for Colorectal Adenoma Determined by Quality-Adjusted, Back-to-Back Colonoscopies, *Gut and Liver* 6 (2012) 64–70.
- [3] T. K. Lui, C. K. Hui, V. W. Tsui, K. S. Cheung, M. K. Ko, D. aCC Foo, L. Y. Mak, C. K. Yeung, T. H. Lui, S. Y. Wong, W. K. Leung, New insights on missed colonic lesions during colonoscopy through artificial intelligence–assisted real-time detection (with video), *Gastrointestinal Endoscopy* (2020).
- [4] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdal, A. Courville, A benchmark for endoluminal scene segmentation of colonoscopy images, *Journal of healthcare engineering* 2017 (2017).
- [5] Y. Guo, J. Bernal, B. J. Matuszewski, Polyp Segmentation with Fully Convolutional Deep Neural Networks—Extended Evaluation Study, *Journal of Imaging* 6 (2020) 69.
- [6] L. F. Sánchez-Peralta, A. Picón, J. A. Antequera-Barroso, J. F. Ortega-Morán, F. M. Sánchez-Margallo, J. B. Pagador, Eigenloss: Combined PCA-Based Loss Function for Polyp Segmentation, *Mathematics* 8 (2020) 1316.
- [7] Y. Fang, C. Chen, Y. Yuan, K.-y. Tong, Selective Feature Aggregation Network with Area-Boundary Constraints for Polyp Segmentation, in: *Medical Image Computing and Computer Assisted Intervention - MICCAI 2019, Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2019, pp. 302–310.
- [8] J. Kang, J. Gwak, Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images, *IEEE Access* 7 (2019) 26440–26447.
- [9] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, PraNet: Parallel Reverse Attention Network for Polyp Segmentation, in: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020*, pp. 263–273.
- [10] S. Ali, M. Dmitrieva, N. Ghatwary, S. Bano, G. Polat, A. Temizel, A. Krenzer, A. Hekalo, Y. B. Guo, B. Matuszewski, M. Gridach, I. Voiculescu, V. Yoganand, A. Chavan, A. Raj, N. T. Nguyen, D. Q. Tran, L. D. Huynh, N. Boutry, S. Rezvy, H. Chen, Y. H. Choi, A. Subramanian, V. Balasubramanian, X. W. Gao, H. Hu, Y. Liao, D. Stoyanov, C. Daul, S. Realdon, R. Cannizzaro, D. Lamarque, T. Tran-Nguyen, A. Bailey, B. Braden, J. E. East, J. Rittscher, Deep learning for detection and segmentation of artefact and dis-

- ease instances in gastrointestinal endoscopy, *Medical Image Analysis* 70 (2021) 102002. doi:10.1016/j.media.2021.102002.
- [11] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H. D. Johansen, Kvasir-seg: A segmented polyp dataset, in: *International Conference on Multimedia Modeling*, 2020, pp. 451–462.
- [12] J. Bernal, J. Sánchez, F. Vilariño, Towards automatic polyp detection with a polyp appearance model, *Pattern Recognition* 45 (2012) 3166–3182.
- [13] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer, *International Journal of Computer Assisted Radiology and Surgery* 9 (2014) 283–293.
- [14] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, M. A. Riegler, P. Halvorsen, C. Daul, J. Rittscher, O. E. Salem, D. Lamarque, T. de Lange, J. E. East, Polypgen: A multi-center polyp detection and segmentation dataset for generalisability assessment, *arXiv* (2021).
- [15] A. Galdran, G. Carneiro, M. A. González Ballester, Double Encoder-Decoder Networks for Gastrointestinal Polyp Segmentation, in: *Pattern Recognition. ICPR International Workshops and Challenges, Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2021, pp. 293–307. doi:10.1007/978-3-030-68763-2\_22.
- [16] S. A. Hicks, D. Jha, V. Thambawita, P. Halvorsen, H. L. Hammer, M. A. Riegler, The EndoTect 2020 Challenge: Evaluation and Comparison of Classification, Segmentation and Inference Time for Endoscopy, in: *Pattern Recognition. ICPR International Workshops and Challenges, Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2021, pp. 263–274. doi:10.1007/978-3-030-68793-9\_18.
- [17] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2015, pp. 234–241.
- [18] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, I. B. Ayed, The Little W-Net That Could: State-of-the-Art Retinal Vessel Segmentation with Minimalistic Models, *arXiv:2009.01907* (2020).
- [19] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature Pyramid Networks for Object Detection, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–944. ISSN: 1063-6919.
- [20] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, J. Feng, Dual path networks, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 4470–4478.
- [21] S. Ali, F. Zhou, B. Braden, A. Bailey, S. Yang, G. Cheng, P. Zhang, X. Li, M. Kayser, R. D. Soberanis-Mukul, S. Albarqouni, X. Wang, C. Wang, S. Watanabe, I. Oksuz, Q. Ning, S. Yang, M. A. Khan, X. W. Gao, S. Realdon, M. Loshchenov, J. A. Schnabel, J. E. East, G. Wagnieres, V. B. Loschenov, E. Grisan, C. Daul, W. Blondel, J. Rittscher, An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy, *Scientific Reports* 10 (2020) 2748. doi:10.1038/s41598-020-59413-5.