

Real-Time Polyp Segmentation Using U-Net with IoU Loss

George Batchkala¹, Sharib Ali²

¹Department of Computer Science, University of Oxford, Oxford, UK

²Institute of Biomedical Engineering, Department of Engineering Science, Oxford, UK
george.batchkala@gmail.com

sharib.ali@eng.ox.ac.uk

ABSTRACT

Colonoscopy is the third leading cause of cancer deaths worldwide. While automated segmentation methods can help detect polyps and consequently improve their surgical removal, the clinical usability of these methods requires a trade-off between accuracy and speed. In this work, we exploit the traditional U-Net methods and compare different segmentation-loss functions. Our results demonstrate that IoU loss results in an improved segmentation performance (nearly 3% improvement on Dice) for real-time polyp segmentation.

1 INTRODUCTION

Colorectal cancer (CRC) is the commonly diagnosed malignancy and the third leading cause of cancer-related deaths worldwide [4]. Colorectal polyps are abnormal protrusions from the mucosa that are usually identified during standard medical procedure referred to as colonoscopy; the associated malignancy is classified through histopathological examinations [13]. Patients with conventional adenomas or serrated polyps are advised to undergo polypectomy, which is a non-invasive surgical procedure usually done during colonoscopy surveillance to prevent CRC [7]. While detection and segmentation of polyps are critical, missed detection and inaccurate removal of polyps can lead to subsequent risk of CRC. Due to advancements in hardware and algorithmic revolutions such as deep learning, building accurate real-time systems is now possible. However, a trade-off between accuracy and speed is still vital for the use of automated systems during CRC surveillance and surgical removal of polyps.

Medico automatic polyp segmentation challenge¹ held in 2020 aims to address the automated delineation of polyps and evaluate the capability of built models for real-time performance that directly implicates clinical utility of the methods. We participated in both polyp segmentation and algorithm efficiency sub-tasks in this challenge. To this end, we have investigated the successful and widely used for semantic segmentation U-Net architecture [12]. In this paper, we propose and shed light on U-Net-based deep learning architecture and evaluate it using different loss functions and data augmentation strategies for polyp segmentation.

2 RELATED WORK

In the past, several biomedical challenges related to the endoscopy data have been accomplished [1, 2, 5, 6]. These challenges curate endoscopy video image frames and provide to the computational

scientists to benchmark their methods. Among these challenges, the very first challenge on polyp segmentation² was introduced in 2015 with comprehensive single images and video data. This dataset has been widely used by the researchers. GIANA dataset³ was introduced in 2017 with the added detection task [6].

Kvasir-SEG dataset [10], released in 2020, contains 1000 pairs of colonoscopy images and their ground-truth segmentation masks⁴. Similarly, multi-class endoscopy disease detection and segmentation challenge [3] includes polyps as one of its five disease categories. A comprehensive comparison of deep learning methods on this dataset can be found in [1]. Likewise, [8] provides an extensive comparison of the state-of-the-art methods for Kvasir-SEG dataset.

3 APPROACH

U-Net [12] is an established encoder-decoder architecture with skip-connections. Classically, binary cross-entropy (BCE) is used for binary segmentation tasks [8, 12]. While preserving the standard U-Net design, we used intersection-over-union loss \mathcal{L}_{IoU} and experimented with a combination of BCE and IoU losses. To boost the performance on this dataset, we have also added augmentation techniques that include random rotations (up to 180 degrees in each direction) and random horizontal flips (with probability 0.5) followed by cropping to return the rotated images to their original sizes. Here, we have directly used negative of IoU instead of classically used $1 - IoU$ as shown in Eq. 1, where M^P and M^{GT} are the predicted and ground-truth masks, respectively.

$$\mathcal{L}_{IoU} = - \frac{|M^P \cap M^{GT}|}{|M^P \cup M^{GT}|} \quad (1)$$

During the training stage, the IoU loss computation showed convergence already at 55 epochs providing validation IoU value over 70% (refer Figure 1).

4 EXPERIMENTS

4.1 Dataset and set-up

We split our training data into 88% training and 12% for validation on the 1000 training images provided by the organisers [9]. The resolution of images varies from 332×487 to 1920×1072 pixels, so we resized all the images to 256×256 pixels for training purposes. A hidden test dataset that included additional 160 images was provided.

We used Adam optimiser [11] for minimisation of our loss function with a learning rate of $1e^{-4}$ and default weight decay of $1e^{-8}$. For each experiment, we trained our network for 100 epochs with

¹<https://multimediaeval.github.io/editions/2020/tasks/medico/>

Copyright 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

MediaEval'20, December 14-15 2020, Online

²<https://polyp.grand-challenge.org/Home/>

³<https://giana.grand-challenge.org>

⁴<https://datasets.simula.no/kvasir-seg/>

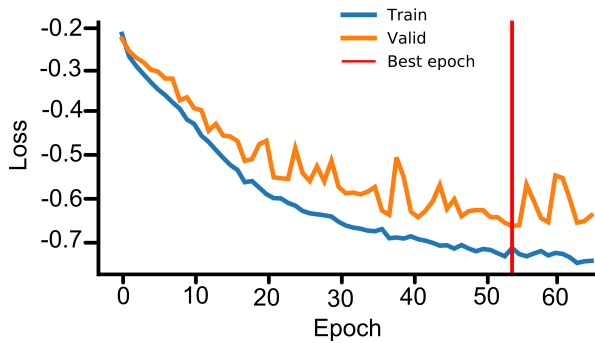


Figure 1: IoU loss computation for training and validation. Red line shows the achieved stopping criteria reached on the 54th epoch (starting from 0) with validation IoU of 0.703.

Table 1: Results on the validation subset of the provided Kvasir-SEG training dataset

Model	IoU	DSC	Rec.	Prec.	Acc.	F2	FPS
U-Net							
+ IoU loss	0.6761	0.7703	0.8360	0.7967	0.9346	0.7772	241
+ BCE loss	0.6639	0.7556	0.8373	0.7769	0.9304	0.7714	221
+ BCE + IoU loss	0.6497	0.7415	0.8275	0.7745	0.9298	0.7509	223
+ IoU loss, aug	0.7005	0.7868	0.8307	0.8435	0.9391	0.7820	252
+ IoU loss, subm	0.6928	0.7821	0.8686	0.7895	0.9391	0.8026	243

early stopping (patience 10) and a batch size of 20. The implemented code is available at <https://github.com/GeorgeBatch/kvasir-seg>. We implemented the network architecture in PyTorch (1.7.0) and ran the computations on Tesla V100 32GB GPU.

4.2 Evaluation metrics

We used standard computer vision metrics for evaluating semantic segmentation of polyps: intersection-of-union (IoU), Dice similarity coefficient (DSC), recall (Rec.), precision (Prec.), overall accuracy (Acc.) and F2-error (F2). Additionally, we demonstrated the real-time application of our approach using frames-per-second (FPS) measurement.

4.3 Results

Table 1 shows the quantitative results of the U-Net model with different loss functions and augmentation. It can be observed that using IoU loss as a minimization objective is better than using the BCE loss or the combined (IoU + BCE) loss. Furthermore, using IoU loss and data augmentation results in the best DSC of 0.7868, the best IoU of 0.7005, and the best trade-off between precision (0.8435) and recall (0.8307). It is worth noting that our method with IoU loss has the highest FPS on our hardware of over 240.

Table 2 presents the results of our method on the unseen test dataset provided by the challenge organisers. We have achieved the DSC of 0.7328 and precision of 0.8229. Again, it can be observed that our method has an FPS of 197, which is sufficient to be used in clinical practice. In general, with available high-definition colonoscopy equipment, the required rate is below 100 FPS.

Table 2: Results on the previously unseen test dataset (provided by the organisers)

Model	mIoU	DSC	Rec.	Prec.	Acc.	F2	FPS
U-Net + IoU loss	0.6351	0.7328	0.7500	0.8229	0.9422	0.7361	197

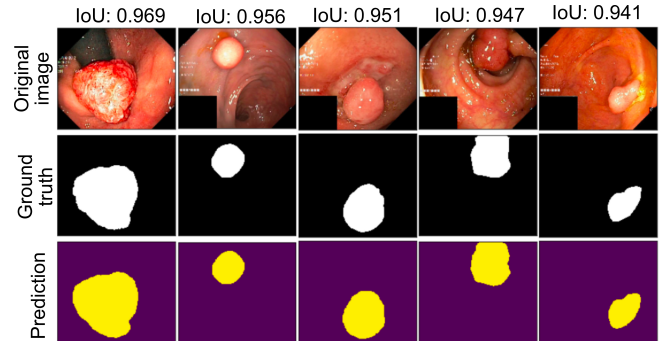


Figure 2: Quantitative results on validation set

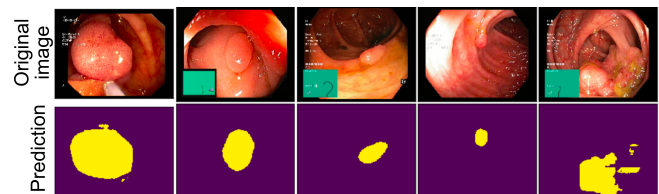


Figure 3: Qualitative results on (unseen) test set

Figure 2 shows polyps with different size present at various locations in the colon. Additionally, there are different textures present on these protrusions. We see that our method is able to segment these polyps accurately with IoU of nearly over 0.95. Figure 3 presents predicted masks from our trained network on the unseen test dataset. For this data, no ground truth was provided. Visual inspection suggests that our method is able to segment the most protruded polyps accurately. However, the method confuses the large polyp structure with the colon folds.

5 CONCLUSION

We have presented different loss combinations and showed that using widely used U-Net with IoU loss results in a descent segmentation performance on the Kvasir-SEG dataset. Additionally, our method provides strong clinical applicability due to its real-time capability. In future, we will work on improving the segmentation accuracy using attention mechanism and apply shape context information to boost performance.

ACKNOWLEDGMENTS

G.B. is funded by a full Health Data Science Studentship through Professor Fergus Gleeson's A2 research funds and S.A. is supported by Oxford NIHR BRC.

REFERENCES

- [1] Sharib Ali, Mariia Dmitrieva, Noha M. Ghatwary, Sophia Bano, Gorkem Polat, Alptekin Temizel, and others. 2020. A translational pathway of deep learning methods in GastroIntestinal Endoscopy. *CoRR abs/2010.06034* (2020). <https://arxiv.org/abs/2010.06034>
- [2] Sharib Ali, Felix Zhou, Barbara Braden, and others. 2020. An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. *Scientific Reports* 10 (2020), 2748.
- [3] Sharib Ali *et al.* 2020. Endoscopy disease detection challenge 2020. *arXiv preprint arXiv:2003.03376* (2020). <https://arxiv.org/abs/2003.03376>
- [4] Melina Arnold, Mónica S Sierra, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 2017. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 66, 4 (2017), 683–691. <https://doi.org/10.1136/gutjnl-2015-310912>
- [5] Jorge Bernal and others. 2017. Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. *IEEE Trans. Med. Imag* 36, 6 (2017), 1231–1249.
- [6] Jorge Bernal and others. 2018. Polyp detection benchmark in colonoscopy videos using gtcreator: A novel fully configurable tool for easy and fast annotation of image databases. In *Proc. Comput. Assist. Radiol. Surg. (CARS)*.
- [7] Xiaosheng He, Dong Hang, Kana Wu, Jennifer Nayor, David A Drew, Edward L Giovannucci, Shuji Ogino, Andrew T Chan, and Mingyang Song. 2020. Long-term Risk of Colorectal Cancer After Removal of Conventional Adenomas and Serrated Polyps. *Gastroenterology* 158, 4 (March 2020), 852–861.e4. <https://doi.org/10.1053/j.gastro.2019.06.039>
- [8] Debesh Jha, Sharib Ali, Håvard D. Johansen, Dag D. Johansen, Jens Rittscher, Michael A. Riegler, and Pål Halvorsen. 2020. Real-Time Polyp Detection, Localisation and Segmentation in Colonoscopy Using Deep Learning. *CoRR abs/2011.07631* (2020). <https://arxiv.org/abs/2011.07631>
- [9] Debesh Jha, Steven A. Hicks, Krister Emanuelsen, Håvard Johansen, Dag Johansen, Thomas de Lange, Michael A. Riegler, and Pål Halvorsen. 2020. Medico Multimedia Task at MediaEval 2020: Automatic Polyp Segmentation. In *Proc. of the MediaEval 2020 Workshop*.
- [10] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. 2020. Kvasir-seg: A segmented polyp dataset. In *International Conference on Multimedia Modeling*. Springer, 451–462.
- [11] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- [12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*. 234–241.
- [13] J. G. Williams, R. D. Pullan, J. Hill, P. G. Horgan, E. Salmo, G. N. Buchanan, S. Rasheed, S. G. McGee, and N. Haboubi. 2013. Management of the malignant colorectal polyp: ACPGBI position statement. *Colorectal Disease* 15, s2 (2013), 1–38. <https://doi.org/10.1111/codi.12262>