Ensemble U-Net Model for Efficient Polyp Segmentation

Shruti Shrestha[1], Bishesh Khanal[1], Sharib Ali[2]
[1]NepAL Applied Mathematics and Informatics Institute for Research (NAAMII), Kathmandu, Nepal
[2]Institute of Biomedical Engineering, Department of Engineering Science, Oxford, UK

ABSTRACT

This paper presents our approach developed for the Medico automatic polyp segmentation challenge 2020 [1]. We used a U-Net model with two different encoder backbones: ResNet-34 and EfficientNet-B2. The two models were trained separately, and trained for ensembling using Tversky loss. We performed CutMix and standard augmentations for data pre-processing. For ensembling, we chose the hyperparameter of the loss function in the range that makes individual models have high recall while relaxing the precision. We evaluated the individual models and the ensemble model on validation data. ResNet-34 backbone model and the ensemble model were submitted to the challenge website for further evaluation on the test data. Our ensemble model improved performance on metrics compared to the single networks by achieving a Dice Coefficient of 0.8316, Intersection Over Union of 0.7550, Precision of 0.8851, and Overall Accuracy of 0.9583.

1 INTRODUCTION

Colorectal cancers are one of the leading causes of death worldwide. Colonoscopy is preferred for detecting and removing the colorectal polyps, which are the predecessors of Colorectal Cancers(CRC) [3]. Polyps generally occur as a protrusion of the mucosa looking like a bumpy structure. However, wide variation in shape, size, intensity of polyps, and specular reflection in colonoscopy images can make polyps very difficult to detect by endoscopists that can have a severe impact on CRC patients and often are contributor to higher mortality rate in CRC [3]. In recent years, several computer-aided polyp detection and segmentation methods has been developed [2, 4, 7]. While the detection methods provide image level presence or absence of polyps or locate them with a rectangular box, semantic segmentation provides pixel-wise classification targeting finer polyp boundaries. In this paper, we focus on semantic segmentation for automated delineation of polyps.

2 RELATED WORK

The state-of-the-art polyp segmentation methods use Convolutional Neural Networks (CNN). Akbari et al. [1] used FCN-8S [15] network to get region of probable polyps followed by Otsu thresholding to select the largest connected component to segment polyp regions, resulting in 81% accuracy in the CVC- ColonDB database[2]. Sanchez et al. [16] first proposed a polyp detection system using texture to find potential polyps windows, which were further segmented to produce masks for polyp location and extension. Kang et al. [10] used a transfer learning-based ensemble method. They ensembled Mask R-CNN [5] models, one with ResNet-50 backbone and another with ResNet-101 [6], and then performed bit wise combination of two predicted masks. CNN based polyp segmentation method must have uncertainty in predictions. [18] studied uncertainty estimation and model interpretability for polyp segmentation task. It also provided the advancements on two methods, firstly in FCN-8 [15] by keeping batch normalization after each layer, and secondly in SegNet by including dropouts. Their best performance method on EndoScene dataset used Monte Carlo Dropout model and had far fewer parameters.

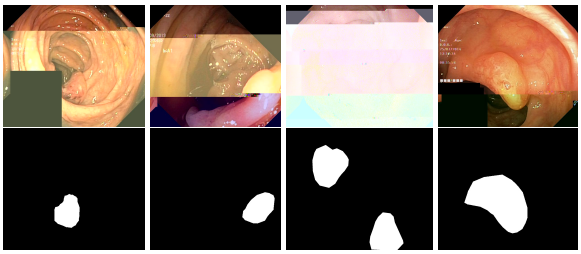

Figure 1: Original colonoscopy images and corresponding ground truth polyp masks for Kvasir-SEG training dataset [9]

3 DATASET

We use publicly available Kvasir-SEG dataset [9] that consists of 1000 images of gastrointestinal polyp images and corresponding manually annotated segmentation masks verified by an experienced gastroenterologist. The sample-images of this data set are shown in Figure 1. We performed a random split of the dataset into 80% and 20% train-validation split resulting into 880 training set and 120 validation set. 160 test images were provided by the organisers during the [8] challenge for which no ground truth masks were provided.

4 METHOD

An encoder decoder architecture with transfer learning was used for computing the predicted mask on the provided polyp dataset [9]. In addition to this, we have also exploited different data augmentation techniques and used Tversky loss function [14] to tune the precision and recall of the individual models for effective ensembling.

4.1 Encoder-decoder architecture

The encoder-decoder architecture is one of the widely used architectures for medical image segmentation. The encoder takes the input and downscales it by computing feature representations at various resolution scales and outputs feature maps that hold encoded information of the input image. In the decoder part these

[1]https://multimediaeval.github.io/editions/2020/tasks/medico/
[2]http://mv.cvc.uab.es/projects/colon-qa/cvccolondb/

MediaEval'20, 14-15 December 2020, Online

feature maps are up sampled and restored to the full segmentation map. Here we use a U-Net architecture developed by Ronnerberger et al. [13]. In this model, the authors include a skip-connection to propagate the original resolution information from encoder to the decoder layers. In this work, we have exploited ResNet-34 [6] and E□cientNet-B2 [17] backbones in the U-Net architecture.

Single model. We used ResNet-34 as our □rst model. The weights saved after the training phase were loaded in the network and test data were fed to get the predicted polyp masks.

Ensemble model. We used two models, ResNet-34 and E□cientNet-B2, to predict our masks. Then we ensembled the predictions by using bit wise multiplication between the two predicted masks.

4.2 Data Augmentation

We used random angles for rotations, contrast, gaussian noise, zoom, elastic deformation, resize, □ips, a□ne, and scaling to overcome over□tting. We also used CutMix regularization [19] in the data augmentation process which chooses a patch from another random image of the same batch and appends the patch in the current training image. We observed that using CutMix regularizer increased the accuracy by up to 3% in the validation set.

4.3 Loss function

Tversky loss [14] $\mathcal{L}_{Tv}$ is a generalisation of Dice similarity coe□cient and F_β scores. This loss is used for an imbalance dataset. By adjusting the hyperparameters as in [12], we used random beta values from 0.9 to 1. Random values of beta were used to create variation between the two models, ResNet-34 and E□cientNet-B2. By using beta in this range, it focuses more on the false negatives and decreases them.

$$\mathcal{L}_{Tv} = 1 - \frac{\sum_{j=1}^{N} y^j f^j}{\sum_{j=1}^{N} [y^j f^j + \beta y^j (1 - f^j) + (1 - \beta)(1 - y^j) f^j]} \quad (1)$$

where, y^j is 1 if the pixel j is a ground truth polyp mask and 0 if it is a non polyp mask. Also, f^j is the probability of pixel j to be a polyp and (1-f^j) is the probability of a pixel j to be a non-polyp. $\beta \in [0.9,1)$ is a hyperparameter. This loss function penalizes false negatives when β is kept in this range. N is the number of pixels.

5 EXPERIMENTS

5.1 Implementation Details

We used ResNet-34 as backbone for our □rst model (model-I), and a combined ensemble model with E□cientNet-B2 as backbone for our second model (model-II). Transfer learning based approach with a pre-trained mechanism using the ImageNet dataset was implemented. Adam optimiser [11] was used with a learning rate of $1e^{-3}$, and default beta values of $\beta_1 = 0.9$, $\beta_2 = 0.99$.

5.2 Evaluation metrics

We have used dice coe□cient (DSC), Jaccard or intersection-over-union (IoU), precision (Prec.), recall (Rec.), overall accuracy (Acc.) and frames-per-second (FPS) to evaluate our approach.

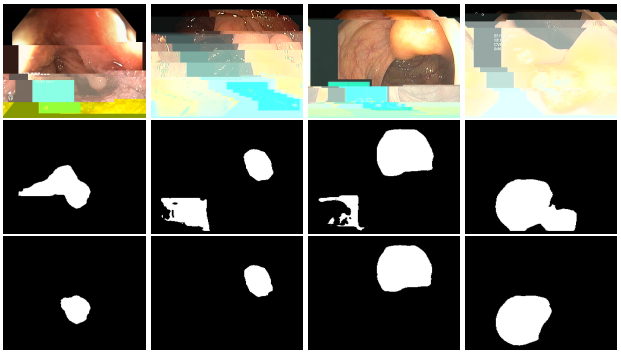

Figure 2: Original colonoscopy images (top row), predicted masks from model-I (middle row) and model-II (bottom row) for the provided test dataset of this competition.

5.3 Results and Discussion

Quantitative results for both of our model on validation set are shown in Table 1. It can be observed that our ensemble model (model-II) outperformed our single method (model-I). However, the FPS is reduced to half for the model-II. Similar observation can be seen from Table 2 where model-II has nearly 2% improved DSC and IoU metric scores compared to the model-I. This better outcome with model-II was obtained as the multiplied outputs between the two models was considered. Qualitative results for both the models on unseen test data provided by the challenge organisers are shown in Figure 2.

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

Model	DSC	IoU	Recall	Prec.	Acc.	FPS
model-I	0.8212	0.7393	0.8748	0.8460	0.9423	60
model-II	0.8379	0.7603	0.8417	0.9001	0.9451	30

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

Model	DSC	IoU	Recall	Prec.	Acc.	F2	FPS
model-I	0.8148	0.7342	0.8764	0.8145	0.9452	0.8354	27
model-II	0.8316	0.7550	0.8316	0.8851	0.9583	0.8249	16

6 CONCLUSION

We have proposed to use an ensemble model that performs a bit-wise operation to output the □nal mask between two backbone architectures. Additionally, we have performed several data augmentation techniques and weighted loss that provided us with improved results on both validation and unseen test set. In future, we aim to apply dilated convolutions and attention networks to exploit the strength of the encoder-decoder architecture.

REFERENCES

[1] Mojtaba Akbari, Majid Mohrekesh, Ebrahim Nasr-Esfahani, S. M. Reza Soroushmehr, Nader Karimi, Shadrokh Samavi, and Kayvan Najarian. 2018. Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network. (2018). arXiv:eess.IV/1802.00368

[2] 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

[3] 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

[4] Jorge Bernal and others. 2018. Polyp detection benchmark in colonoscopy videos using gtcreator: A novel fully con□gurable tool for easy and fast annotation of image databases. In *Proc. Comput. Assist. Radiol. Surg. (CARS)*.

[5] K. He, G. Gkioxari, P. Dollár, and R. Girshick. 2017. Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 2980–2988. https://doi.org/10.1109/ICCV.2017.322

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. (2015). arXiv:cs.CV/1512.03385

[7] 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

[8] Debesh Jha, Steven A. Hicks, Krister Emanuelsen, Håvard D. 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 MediaEval 2020 CEUR Workshop*.

[9] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D. Johansen. 2019. Kvasir-SEG: A Segmented Polyp Dataset. (2019). arXiv:eess.IV/1911.07069

[10] J. Kang and J. Gwak. 2019. Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images. *IEEE Access* 7 (2019), 26440–26447. https://doi.org/10.1109/ACCESS.2019.2900672

[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] Tianyu Ma, Hang Zhang, Hanley Ong, Amar Vora, Thanh D. Nguyen, Ajay Gupta, Yi Wang, and Mert Sabuncu. 2020. Ensembling Low Precision Models for Binary Biomedical Image Segmentation. (2020). arXiv:eess.IV/2010.08648

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. (2015). arXiv:cs.CV/1505.04597

[14] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. 2017. Tversky loss function for image segmentation using 3D fully convolutional deep networks. (2017). arXiv:cs.CV/1706.05721

[15] E. Shelhamer, J. Long, and T. Darrell. 2017. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 4 (2017), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

[16] A. Sánchez-González, B. Garcia-Zapirain, D. Sierra-Sosa, and A. Elmaghraby. 2018. Colon Polyp Segmentation Using Texture Analysis. In *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. 579–588. https://doi.org/10.1109/ISSPIT.2018.8642748

[17] Mingxing Tan and Quoc V. Le. 2020. E□cientNet: Rethinking Model Scaling for Convolutional Neural Networks. (2020). arXiv:cs.LG/1905.11946

[18] K. Wickstrøm, M. Kamp□meyer, and R. Jenssen. 2018. UNCERTAINTY MODELING AND INTERPRETABILITY IN CONVOLUTIONAL NEURAL NETWORKS FOR POLYP SEGMENTATION. In *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*. 1–6. https://doi.org/10.1109/MLSP.2018.8516998

[19] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. 2019. CutMix: Regularization Strategy to Train Strong Classi□ers with Localizable Features. (2019). arXiv:cs.CV/1905.04899