Report

BKAI-IGH NeoPolyp Segmentation

Vũ Trung Thành - 20220066

Abstract

The purpose of this experiment was to train a model for semantic segmentation of neoplastic and non-neoplastic polyps using the BKAI-IGH NeoPolyp-Small dataset. The TransResU-Net architecture was selected, and its hyperparameters were fine-tuned to achieve optimal performance. The model was trained and validated using a combination of Focal Tversky loss and BCE loss, achieving a test score of 0.78798 on the public test set.

1 Model

The model used in this experiment is TransResU-Net [6], an encoder-decoder-based architecture built upon residual blocks that leverages the transformer self-attention mechanism and dilated convolution. While the original architecture only supported binary classification (polyp and non-polyp), the final output convolution layer was modified to enable classification of non-polyp, neoplastic polyp and non-neoplastic polyp.
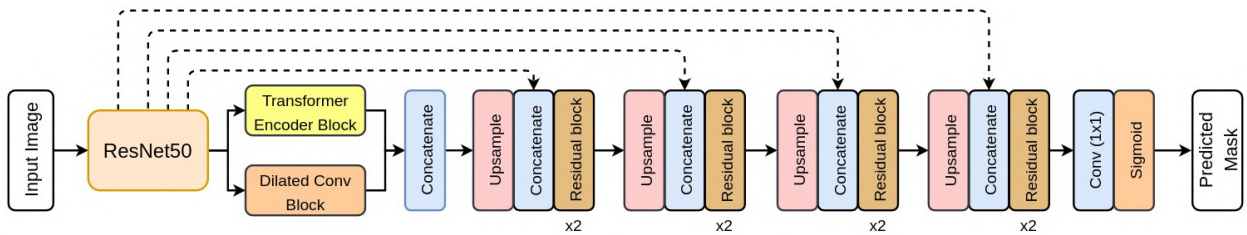

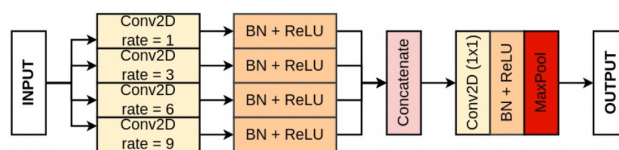Figure 1: *Block diagram of the original TransResU-Net architecture*

Figure 2: *Block diagram of the Dilated Convolution Pooling Block*

2 Experiments

2.1 Dataset

The model was trained on BKAI-IGH NeoPolyp-Small [4, 2, 3, 5] dataset provided in the Kaggle competition, which contains 1000 images on training set (with ground-truth masks) and 200 images on test set. All polyps are classified into neoplastic or non-neoplastic classes denoted by red and green colors, respectively.
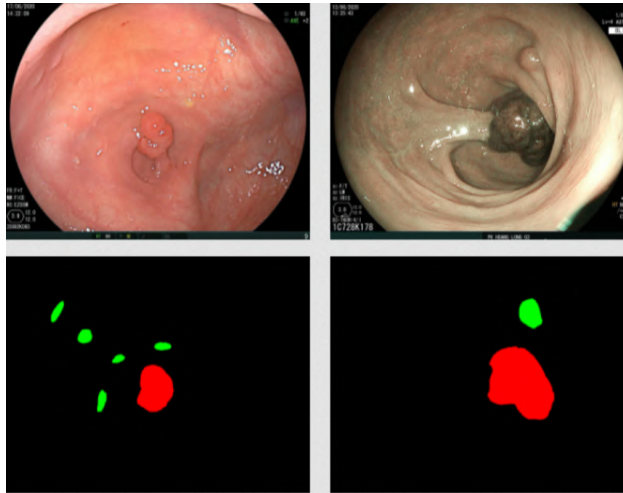

Figure 3: *Some data from the BKAI-IGH NeoPolyp dataset*

2.2 Loss function

For the experiment, Focal Tversky loss [1] was combined with Binary Cross-Entropy (BCE) loss as the evaluation metric.

Due to the imbalance in the dataset, where the number and size of non-neoplastic polyps are relatively smaller than those of neoplastic polyps, Focal Tversky loss was employed with $\alpha = 1 - \beta = 0.3$ to place greater emphasis on false negatives, ultimately improving the recall rate. The hyperparameter γ was set to $4/3$, as suggested in [4] for this specific dataset.

The loss was calculated pixel-wise for each class and then averaged across all classes. By combining Focal Tversky and BCE loss, the model effectively detects polyps while accurately classifying them as neoplastic or non-neoplastic.

2.3 Data augmentation and transformation

Data augmentation and transformation were used to increase the size of the training data and enhance the model's generalization capability. Specifically, rotations (with a maximum angle of 45º)

and flips were applied with a probability of 0.4, while random brightness, contrast, and RGB shifts were applied with a probability of 0.7.

To address class imbalance, oversampling of the non-neoplastic class was performed to ensure that the model received approximately the same number of pixels for both classes. Oversampling was achieved by randomly sampling images containing non-neoplastic polyps from the original 1,000 images, and adding them to a copy of the original dataset until the classes in the copy were balanced. This process increased the dataset size to 4,183 images, which significantly enhanced the model's ability to learn features from the underrepresented class.

	Images	N_{non}	N_{neo}	Ratio
Before	1,000	8,581,706	58,717,102	0.1462
After	4,183	96,618,347	96,618,049	1.0000

Table 1: *Dataset statistics, before and after oversampling. N_{non} and N_{neo} is the number of pixels in non-neoplastic and neoplastic class, respectively.*
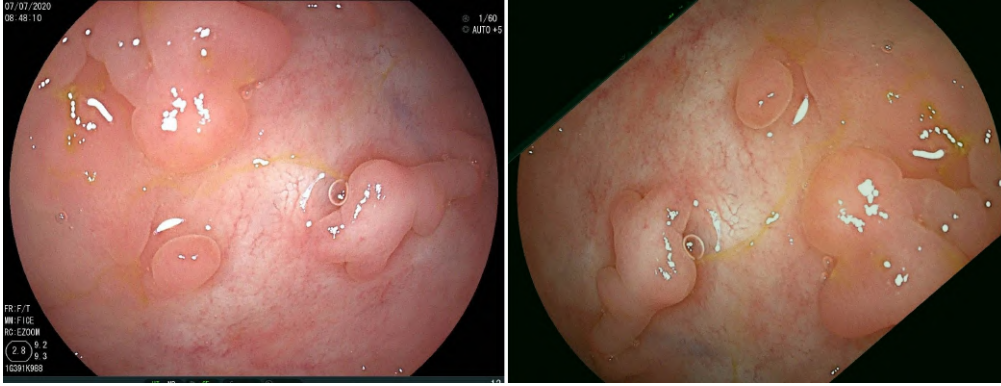

Figure 4: *Before and after data transformation*

2.4 Experimental setup

First, the images and masks were read and split into training and validation sets with a ratio of 80% training and 20% validation. Before being fed into the model, the images were scaled down to a size of 256×256 pixels. The training images were then processed in batches of 8.

The model was trained for 250 epochs, with early stopping applied if the validation loss did not improve for 25 consecutive epochs. An Adam optimizer was used with an initial learning rate of 10^{-4}, a weight decay of 10^{-5} and a learning rate scheduler that reduced the learning rate by a factor of 0.5 after 5 epochs of plateau.

The training was done on a machine equipped with an Intel i7-12700H processor, an NVIDIA GeForce RTX 3050 GPU and 16 GB of RAM.

2.5   Result

Figure 6 shows the losses during training. The model achieves the test score of 0.78798 on the test set, using submission on dataset owner's Kaggle competition, which evaluate results using mean Dice coefficient. Some segmentation results can be seen in figure 7.
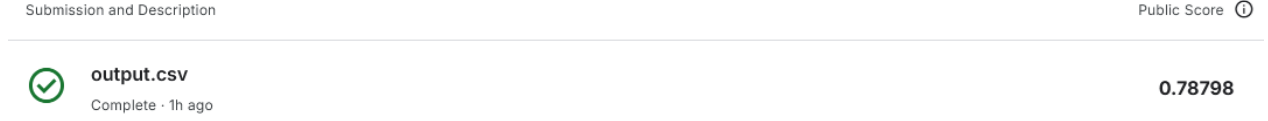

Figure 5: *Submission result on Kaggle*

Figure 6: *Plots of losses during training. Plots taken from Weights & Biases*
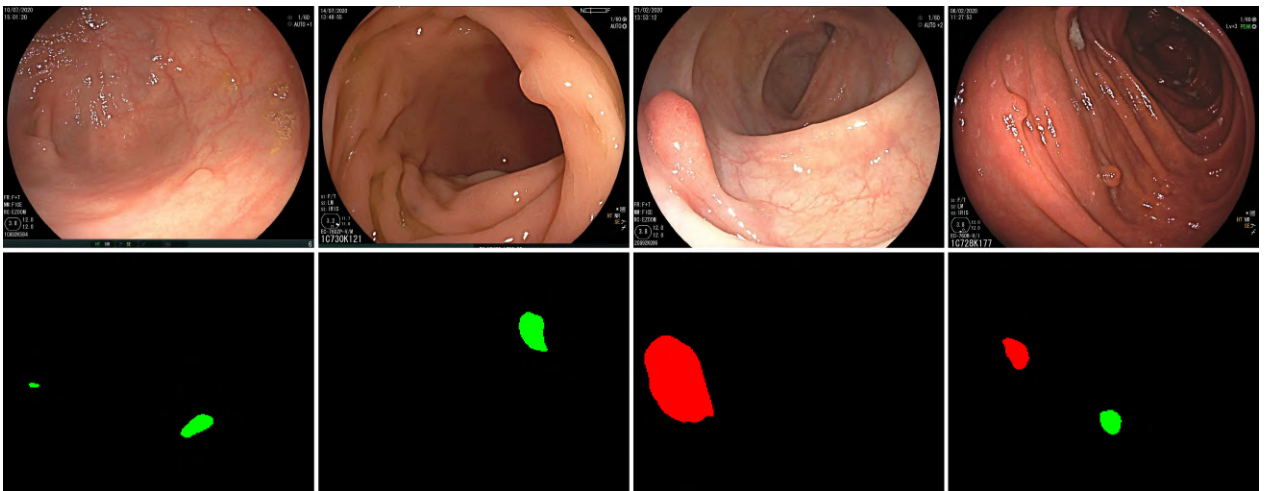

Figure 7: *Segmentation result*

2.6 Conclusion

In this experiment, a model for semantic segmentation of neoplastic and non-neoplastic polyps was successfully trained using the BKAI-IGH NeoPolyp-Small dataset. The TransResU-Net architecture, adapted for multi-class segmentation, achieved a test Dice score of 0.78798, demonstrating its ability to identify and classify polyps effectively. Data augmentation, oversampling, and hyperparameter tuning contributed to this result by improving the model's robustness and performance.

In conclusion, while the current approach showed promising results, further optimization and experimentation will be necessary to improve the model's generalization and ensure reliable segmentation in practical applications.

References

[1] N. Abraham and N. M. Khan. A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 683–687, 2019. doi: 10.1109/ISBI.2019.8759329.

[2] N. S. An, P. N. Lan, D. V. Hang, D. V. Long, T. Q. Trung, N. T. Thuy, and D. V. Sang. BlazeNeo: Blazing Fast Polyp Segmentation and Neoplasm Detection. *IEEE Access*, 10:43669–43684, 2022. doi: 10.1109/ACCESS.2022.3168693.

[3] N. T. Duc, N. T. Oanh, N. T. Thuy, T. M. Triet, and V. S. Dinh. ColonFormer: An Efficient Transformer Based Method for Colon Polyp Segmentation. *IEEE Access*, 10:80575–80586, 2022. doi: 10.1109/ACCESS.2022.3195241.

[4] P. Ngoc Lan, N. S. An, D. V. Hang, D. V. Long, T. Q. Trung, N. T. Thuy, and D. V. Sang. Neounet : Towards accurate colon polyp segmentation and neoplasm detection. In G. Bebis, V. Athitsos, T. Yan, M. Lau, F. Li, C. Shi, X. Yuan, C. Mousas, and G. Bruder, editors, *Advances in Visual Computing*, Cham, 2021. Springer International Publishing. ISBN 978-3-030-90436-4.

[5] N. H. Thuan, N. T. Oanh, N. T. Thuy, S. Perry, and D. V. Sang. RaBiT: An Efficient Transformer using Bidirectional Feature Pyramid Network with Reverse Attention for Colon Polyp Segmentation, 2023. URL https://arxiv.org/abs/2307.06420.

[6] N. K. Tomar, A. Shergill, B. Rieders, U. Bagci, and D. Jha. TransResU-Net: Transformer based ResU-Net for Real-Time Colonoscopy Polyp Segmentation, 2022. URL https://arxiv.org/abs/2206.08985.

Appendix

The code and model for implementing the architecture can be found at the link below:

https://github.com/thanh309/polyp-segmentation.