GLoFF-Net: Global-Local Feature Fusion Network for Polyp Segmentation

Jiabao Sean Xiao

September 2024

Abstract

Colonoscopy is recognized as a primary standard for detecting colorectal cancer and its precursory symptoms. With hardware development in recent years, contemporary methods using deep learning models have already gained great progress, but holding limitations of relatively high miss rates of abnormalities. Some current models are also highly biased toward local features and fail to detect the global aspects of their input. In this case, we proposed GloFF-Net, a convolutional neural network architecture. Our model exploits an encoder-decoder structure converging with custom attention mechanisms that fuse global features and local features. We have preformed excellent results and validated the improvements using several publicly attainable benchmark datasets. Furthermore, we compare our model with other state-of-the-art methods. Our approach indicates strong abilities of generalization, accomplishing great performance under limited training data.

Index Terms: Colonoscopy, Polyp Segmentation, Convolutional Neural Network, medical imaging.

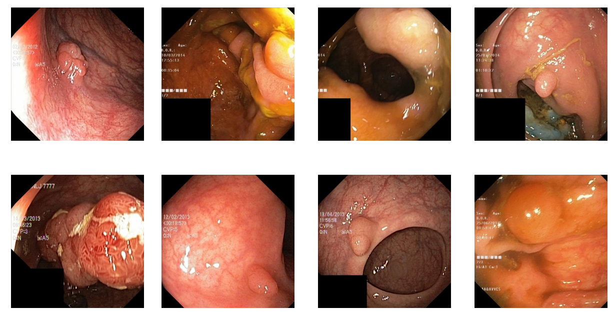

Figure 1: Sample images demonstrating the variations in polyps from the Kvasir-SEG [6].

1 Introduction

Cancer has become a significant health problem contemporarily, and within it, colorectal cancer-namely CRC-is one of the most prevailing cancers based on mortality globally and cancer incidence [1]: it is the second largest cause of cancer deaths that led to about 156,000 deaths in 27 EU nations overall in 2020 in Europe [2] and ranked the third for the cause of cancer deaths with 52,500 deaths predicted in the United States in 2022 [3]. Most CRCs are evolved from colorectal polyps, resulting in the importance of their early detection and removal for CRC diagnosis and treatment [4]. One colorectal polyp's attributes and structure varies over time at distinct stages of florishment. These polyps might have different configurations, volumes, colors and visuals, making them difficult to discover (see 1. In this case, colonoscopy became a primary method for detecting polyps before they further develop [5].

While successfully alleviating cancer burden with colonoscopy, around a quarter of all polyps, specifically 22 to 28 percent of polyps and 20 to 24 percent of adenomas are missed [7, 8]. One reason is that polyps that are too small are most likely to be missed [9]. Another element that might raise the risk of polyp-missing is the discovering of more than two polyps during colonoscopy [7]. The adenoma miss rate could therefore be decreased through six-minute or nine-minute withdrawal time [10]. More importantly, by evaluation enhancing polyp detection rates by 1% decreases the risk of CRC by about 3%, making it essential to improve polyp detectability [11].

In summary, our main contributions are as follows:

1) We propose a new encoder-decoder architecture that utilizes global features through attention mechasims to guide the process of decoding.

2) We combines global feature and local features so as to bridge the semantic gap between the encoder and the decoder.

3) We validate the proposed architecture on several

benchmark datasets e.g. Kvasir-SEG [6]. What's more, we have contrasted our work with other works, which is usually absent in similar studies.

2 Related Work

We seperate the ralated work into two parts, i.e., convolutional neural network and medical image segmentation.

2.1 Convolutional Neural Network

In recent years, machine learning algorithms, especially convolutional neural networks (CNNs) have developed quickly and demonstrated excellent results. They have driven remarkable advancements in tasks such as object detection and localization [12] and object classification [13]. Semantic segmentation tasks have also been transformed by CNNs. Currently, most modern CNN-based semantic segmentation architectures are based on either Fully Convolutional Network [14] or an encoder-decoder architecture such as U-Net [15]. FCN was upgraded by proposing a novel architectture named SegNet [16], which includes a 13-layer network that take obtain spatial features from the images with a corresponding 13-layer network that upsamples the feature maps to predict segmentation masks. Exploiting atrous convolutions that enable a dilated field of view and multi-scale feature extraction, enhancing the capability of capturing long-range contextual dependencies, DeepLab [17] and its extended version DeepLabV3+ [18] were proposed.

Despite making a breakthrough in computer vision tasks, CNN architectures shared the drawback of requiring vast amount of training data. Nonetheless, CNNs remained great performance in medical image segmentation in recent years, with much of the credit attributed to U-Net [15]. In U-Net++ [19], a nested ensemble of U-Nets with different depths and the incorporation of deep supervision are employed. MultiResUnet [20] is presented with the proposed MultiRes blocks to augment U-Net with the capacity of multi-resolutional analysis. ResUNet [21], which integrates residual blocks to improve location accuracy for polyps, can harness the strong expressive capabilities of these blocks but may demand more data and computational resources to accomplish optimal performance.

2.2 Medical Image Segmentation

Medical image segmentation plays a crucial role in pre-treatment diagnoses, treatment planning, and post-treatment assessments for various kinds of diseases, which is typically framed as a dense prediction problem, involving pixel-wise classification to generate segmentation maps of lesions or organs. CNNs have already been employed for medical image segmentation tasks widely [15, 19, 20, 22, 23, 24, 25]. To be specific, U-Net [15] has demonstrated outstanding performance in medical image segmentation by producing high-resolution segmentation maps that aggregate multi-stage features through the utilization of skip connections. Its variants such as UNet++ [19], MultiResUnet [20], UNet 3+ [22], DC-UNet [23] illustrated better performance in medical image segmentation as well. In spite of the impressive performance of such CNN-based architectures, they face limitations in taking in long-range dependencies among pixels under the spatial constraints of the convolution operation [26]. To approach this, some models such as Pranet [24] and Attention U-Net [25]-integrating attention modules into their architectures and therefore enhancing the feature maps for improved pixel-level classification in medical images-are proposed.

Simultaneously, transformer-based architectures have also been popular recently for medical image segmentation. As debut for sequence-to-sequence prediction in natural language processing (NLP), transformers [27] rely on an attention-based network structure, in which using self-attention to study correlations among all the input tokens, enabling the model to effectively capture dependencies over long distances. Later following the former success of transformers in NLP, the vision transformer (ViT) [28] segments an image into non-overlapping patches, which are then put into the transformer module along with positional embeddings. Nevertheless, their ability to capture local contextual relationships between pixels is restricted by the self-attention mechanism in transformers [29].

Overall, considering CNN's efficiency with smaller datasets and great performance for localized feature extraction compared to transformers, we exploited a U-Net-based architecture that incorporates attention modules.

3 Methodology

3.1 Overview

Based on the encoder-decoder architecture, this paper proposes a deep network framework for colorectal cancer segmentation that leverages global-local feature collaborative learning. The proposed method integrates features of different granularities from the encoding stage to generate global features, which guide

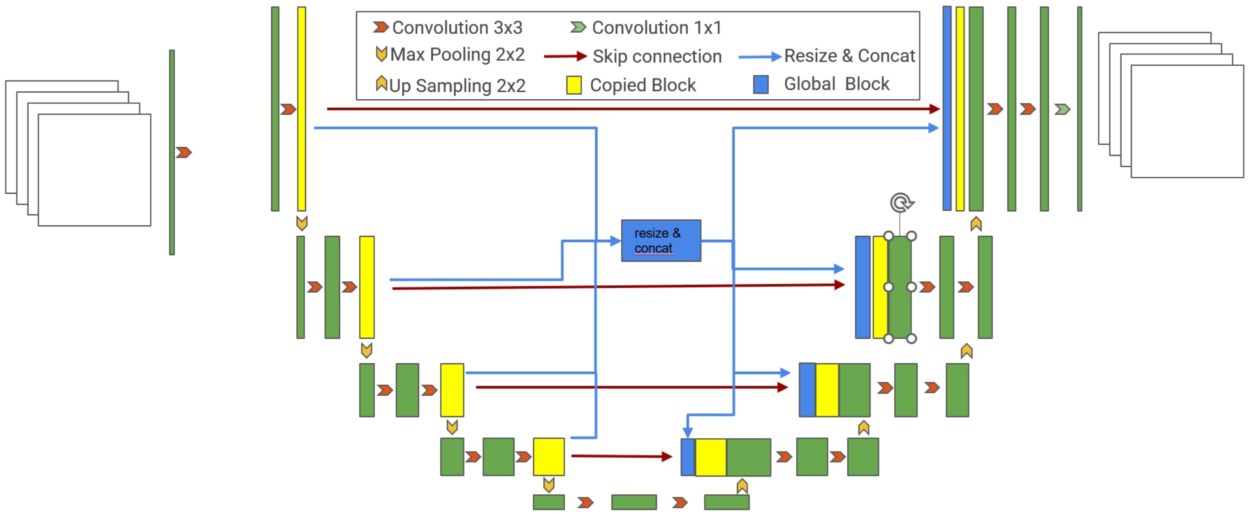

Figure 2: GLoFF-Net Architecture

the learning process in the decoding stage. Additionally, it incorporates an attention mechanism to effectively fuse global and local features, thereby complementing each other and enhancing the model's perception and recognition capabilities for various lesions. The overall architecture diagram is shown in Figure 2.

3.2 Initial Feature Extractor

As illustrated in Figure 2, our architecture utilizes an encoder-decoder architecture with a flexible, interchangeable backbone model. While U-Net is chosen as the CNN backbone, it is crucial to highlight that the implementation of our feature fusion design operates independently of the backbone network. Given $\{X, Y\}_{i=1}^{N}$ with $X \in \mathbb{R}^{H\times W\times C}$ and $Y \in \{0,1\}^{H\times W}$, where X is the input image and Y means the corresponding type. We employ multi-layer convolutional modules to extract initial encoding features:

$$F_i = \mathrm{Conv}(F_{i-1}, \theta_{i-1}) \quad i \in \{1,2,3,4,5\} \tag{1}$$

where $F_0 = X$ and θ is the parameters of the corresponding convolutional layer.

3.3 Global Feature Generation

Due to the different semantic and spatial detail information extracted at various stages, the initial stage features have higher resolution and richer spatial detail information. As convolution progresses, deeper features possess stronger semantic information. Therefore, this paper proposes to combine the initial encoding features from different stages to construct a global feature that guides the decoding stage:

$$F_i = \mathrm{SA}(\mathrm{Re}(F_i)) \quad i \in \{1,2,3,4\} \tag{2}$$

$$G = \mathrm{Conv}(\mathrm{Cat}(\mathrm{F}_1, \mathrm{F}_2, \mathrm{F}_3, \mathrm{F}_4), \mathrm{Q_G}) \tag{3}$$

in which SA corresponds to spatial attention, $\mathrm{Re}(\cdot)$ corresponds to resizing operation so as to adjust the size of each feature, and $\mathrm{Cat}(\cdot)$ corresponds to concatenation.

3.4 Global-Local Feature Fusion

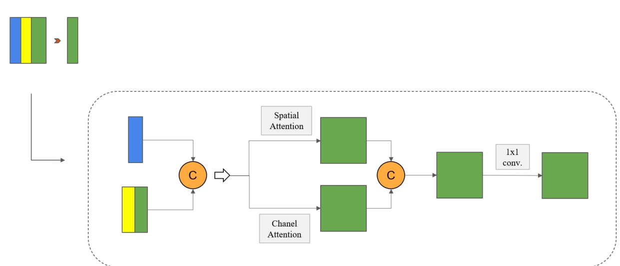

Figure 3: Global-Local Feature Fusion Block

Global features generated in the former stage can effectively support decoding process of local features in decoding phase. In order to better consolidate the global features generated, we apply attention blocks:

$$\begin{aligned} D_i = \mathrm{Conv}(&\mathrm{Cat}(\mathrm{SA}(\mathrm{Cat}(\mathrm{Re}(G), F_i))), \\ &\mathrm{CA}(\mathrm{Cat}(\mathrm{Re}(G), F_i)), \theta_i) \end{aligned} \tag{4}$$

with $i \in \{1,2,3,4\}$.

3.5 Optimization

We simultaneously use Cross-Entropy Loss [30] and Dice Loss [31] to do the optimization, which is defined as follows:

$$\mathcal{L}(P,Y) = \beta\mathcal{L}_{\text{CE}}(P,Y)\alpha\mathcal{L}_{\text{dice}}(P,Y) \tag{5}$$

Specifically, we have:

$$\mathcal{L}_{\text{CE}}(P,Y) = -\sum_{i=0}^{H-1}\sum_{j=0}^{W-1} y_{i,j}\cdot\log(p_{i,j}) \tag{6}$$

$$\mathcal{L}_{\text{dice}}(P,Y) = \frac{\sum_{i=0}^{H-1}\sum_{j=0}^{W-1} p_{i,j}y_{i,j}}{\sum_{i=0}^{H-1}\sum_{j=0}^{W-1} p_{i,j}^2 + \sum_{i=0}^{H-1}\sum_{j=0}^{W-1} y_{i,j}^2} \tag{7}$$

where $p_{i,j}$ means the predicted value of spatial position (i,j) and $y_{i,j}$ means the corresponding ground truth label.

4 Experiments

In this section, we initially elaborate the setting of experiments. Then, we compare the results of our proposed GLoFF-Net, with SOTA methods to indicate the superiority of our method.

4.1 Dataset and Evaluation Metrices

To prove the gratifying improvements that GLoFF-Net has gained, various experiments are conducted on Kvasir-SEG dataset [6] and CVC-ColonDB dataset [32] to evaluate its performance. The seperation of training set and testing set follows similar settings in PraNet [24].

We employ four commonly used evaluation metrics to evaluate GLoFF-Net and other methods quantitatively: the Jaccard Index (denoted as JA, also known as mIoU), Dice coefficient (denoted as DI), Sensitivity (denoted as SE) and Specificity (denoted as SP). The metrics for evaluating model performance can be computed using the following formulas:

$$\begin{aligned} JA &= \frac{TP}{TP+FN+FP} \\ DI &= \frac{2TP}{2TP+FN+FP} \\ SE &= \frac{TP}{TP+FN} \\ SP &= \frac{TN}{TN+FP} \end{aligned} \tag{8}$$

4.2 Implementation Details

All our experiments are performed in Pytorch 2.4.0. The models are trained on a single NVIDIA A100 GPU on Google Colab. The Adams optimizer [33] with weight decay and learning rate of 1e-3 and 2e-4 is chosen to optimize GLoFF-Net for Kvasir-SEG dataset and CVC-ColonDB dataset. Each model tested is optimized for 100 epochs after being initialized with random weights, and the batch size of training set is 4. The evaluation metrics are obtained by calculating the mean result from three successive tests on the entire validation set.

4.3 Results

Quantitative Evaluation

Figure 4 and Figure 5 records the data of the four metrics tested on models. Compared to other models, our network remains competitive and gain great results. An average improvement of 19.6% mIoU, 14.9% Dice, 18.2% sensitivity, 0.6% specificity over the baseline is accomplished.

Kvasir-SEG				
Method	JA	DI	SE	SP
U-Net	0.7010	0.7846	0.8044	0.9757
U-Net++	0.8210	0.7753	0.8283	0.9817
GLoFF-Net (Ours)	**0.8384**	**0.7769**	**0.8324**	**0.9823**

Figure 4: Segmentation evaluation on the Kvasir-SEG dataset

CVC-ColonDB				
Method	JA	DI	SE	SP
U-Net	0.5099	0.5479	0.5193	0.9397
U-Net++	0.4878	0.5570	5239	**0.9412**
GLoFF-Net (Ours)	**0.5832**	**0.6299**	**0.6128**	0.9406

Figure 5: Segmentation evaluation on the CVC-ColonDB dataset

Qualitative Evaluation

In Figure 6, four samples are selcted at random from the Kvasir-SEG dataset [6]. Note that in the second example, U-Net demonstrates obvious errors, cutting off healthy areas. Same for the fourth example, U-Net identifies two polyps, while our model correctly identify one and produce accurate boundary prediction.

Our model also performs well in other benchmark datasets e.g. ColonDB dataset [32]. For several randomly-chosen samples from it, U-Net neglects part of the area containing polyps in the second sample

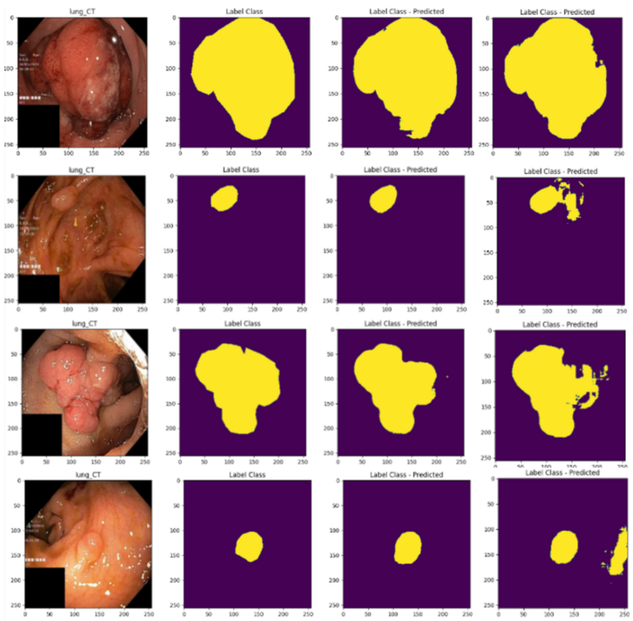

Figure 6: Segmentation evaluation on the Kvasir-SEG dataset [6] for different networks, with left-to-right original image, ground truth, our method, U-Net(baseline)

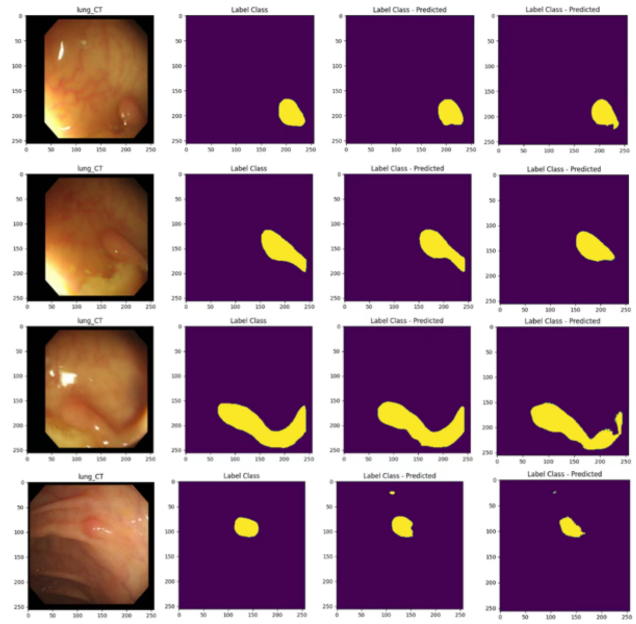

Figure 7: Segmentation example on the CVC-ColonDB dataset [32] for different networks, with left-to-right original image, ground truth, our method, U-Net(baseline)

and generates unsmooth boundaries in the third one; but GLoFF-Net overcomes these problems and generates similar image as the ground truth, making it a credible techinique for polyp segmentation.

5 Conclusion and Discussion

In summary, colorectal cancer (CRC) remains a significant health issue globally, with colonoscopy serving as the primary method for early detection and prevention. However, current methods using deep learning models, while showing progress, still suffer from high miss rates and a focus on local features at the expense of global contextual information. To address these limitations, we introduced GloFF-Net, a novel encoder-decoder architecture that integrates both global and local features using custom attention mechanisms. Our model demonstrated excellent results across multiple benchmark datasets, outperforming other state-of-the-art methods, and exhibited strong generalization abilities even with limited training data. By bridging the semantic gap between encoding and decoding, GloFF-Net significantly enhances polyp detection, which can contribute to reducing colorectal cancer mortality.

Nevertheless, although GLoFF-Net has indicated superior performance in the experiments conducted compared against existing state-of-the-art methods, there are still limitations for the model: limited combination of local and global features and relatively not ideal performance with limited training data. Inspired by Duck-Net [34], we are looking forward to achieving better performance through revising the current 3×3 convolutional block into new block that consolidates different blocks such as Midscope block [34] and Residual block [35] in future work.

References

[1] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.

[2] Tadeusz Dyba, Giorgia Randi, Freddie Bray, Carmen Martos, Francesco Giusti, Nicholas Nicholson, Anna Gavin, Manuela Flego, Luciana Neamtiu, Nadya Dimitrova, et al. The european cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. *European Journal of Cancer*, 157:308–347, 2021.

[3] Rebecca L Siegel, Kimberly D Miller, Hannah E Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: a cancer journal for clinicians*, 72(1), 2022.

[4] Finlay A Macrae, Johanna Bendell, KK Tanabe, D Savarese, and S Grover. Clinical presentation, diagnosis, and staging of colorectal cancer. *UpToDate Retrieved from (https://www. uptodate. com/contents/clinical-presentation-diagnosis-and-stagingof-colorectal-cancer). Accessed on*, 2:2016, 2016.

[5] Shaukat Aasma, Charles J Kahi, Carol A Burke, Linda Rabeneck, Bryan G Sauer, and Douglas K Rex. Acg clinical guidelines: Colorectal cancer screening 2021. *The American Journal of Gastroenterology*, 116(3):458–479, 2021.

[6] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas De Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26*, pages 451–462. Springer, 2020.

[7] Anke M Leufkens, Martijn G Van Oijen, Frank P Vleggaar, and Peter D Siersema. 1178 factors affecting miss rate of polyps during colonoscopy: Results from a prospective, multicenter back-to-back colonoscopy study. *Gastrointestinal Endoscopy*, 73(4):AB165, 2011.

[8] Sang Bong Ahn, Dong Soo Han, Joong Ho Bae, Tae Jun Byun, Jong Pyo Kim, and Chang Soo Eun. The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. *Gut and liver*, 6(1):64, 2012.

[9] Katharina Zimmermann-Fraedrich, Susanne Sehner, Douglas K Rex, Tonya Kaltenbach, Roy Soetikno, Michael Wallace, Wai K Leung, Chuanguo Guo, Ian M Gralnek, Eelco C Brand, et al. Right-sided location not associated with missed colorectal adenomas in an individual-level reanalysis of tandem colonoscopy studies. *Gastroenterology*, 157(3):660–671, 2019.

[10] Shengbing Zhao, Yihang Song, Shuling Wang, Rong Wang, Zhijie Feng, Aixia Gong, Xia Yang, Peng Pan, Dongmei Yao, Jingwen Zhang, et al. Reduced adenoma miss rate with 9-minute vs 6-minute withdrawal times for screening colonoscopy: a multicenter randomized tandem trial. *Official journal of the American College of Gastroenterology— ACG*, 118(5):802–811, 2023.

[11] Douglas A Corley, Christopher D Jensen, Amy R Marks, Wei K Zhao, Jeffrey K Lee, Chyke A Doubeni, Ann G Zauber, Jolanda De Boer, Bruce H Fireman, Joanne E Schottinger, et al. Adenoma detection rate and risk of colorectal cancer and death. *New england journal of medicine*, 370(14):1298–1306, 2014.

[12] P Sermanet. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[14] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[16] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[17] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, 2017.

[18] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611, 2018.

[19] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.

[20] Nabil Ibtehaz and M. Sohel Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, January 2020.

[21] Foivos I. Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, April 2020.

[22] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059, 2020.

[23] Ange Lou, Shuyue Guan, and Murray Loew. Dc-unet: Rethinking the u-net architecture with dual channel efficient cnn for medical images segmentation, 2020.

[24] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 263–273. Springer, 2020.

[25] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.

[26] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation, 2021.

[27] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[28] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[29] Xiangxiang Chu, Zhi Tian, Bo Zhang, Xinlong Wang, and Chunhua Shen. Conditional positional encodings for vision transformers. *arXiv preprint arXiv:2102.10882*, 2021.

[30] Anqi Mao, Mehryar Mohri, and Yutao Zhong. Cross-entropy loss functions: Theoretical analysis and applications, 2023.

[31] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso. *Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations*, page 240–248. Springer International Publishing, 2017.

[32] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE transactions on medical imaging*, 35(2):630–644, 2015.

[33] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.

[34] Razvan-Gabriel Dumitru, Darius Peteleaza, and Catalin Craciun. Using duck-net for polyp image segmentation. *Scientific Reports*, 13(1), June 2023.

[35] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Dag Johansen, Thomas de Lange, Pal Halvorsen, and Havard D. Johansen. Resunet++: An advanced architecture for medical image segmentation, 2019.

致谢信

“Hello World! ”自从打出这第一行代码，我对计算机科学便产生了浓厚的兴趣。从 C++到 JAVA，再到 Python，一步步走来，我的能力也随之提高。近两年 AI 的急速发展使我好奇，我便开始通过 Andrew Ng 的线上课程和 Aston Zhang Et al.的 Dive Into Deep Learning 进行自主学习，并随后在学校选修了胡剑老师开展的学术研究准备课程。

在胡剑老师的鼓励和推动下，我从刚开始的阅读博客和书籍逐渐转变到阅读各类机器学习相关的文献，并发现了对医学影像的图像处理方向的兴趣。关注到当前结直肠癌对于全球人民健康的危害，我逐步开始具向阅读文章，提出了一种方法以提升其前兆息肉分割的准确率。

在此项目中，我独立完成了模型设计、论文撰写和各个实验；胡剑老师作为指导老师，对我进行了一部分的理论指导，使得我对于论文方向的把握更加精确。他在我中期困惑与停滞时给予了我充分的信心和勇气，使我坚持下来，同时也使我对此方向的兴趣更加浓厚。在此，我也要感谢在我这一路上一直支持我的家人与朋友。

最后，我也要感谢丘成桐中学科学奖。其提供了我展现个人能力的平台，给予了我不断前进的动力。这次探索不仅在于结果，也在于过程；我从中受益良多，对各类编程语言都有了更深的认识，对深度学习也有了更多的感悟。也希望自己砥砺前行，创造出属于自己的未来。