Polyp Segmentation Model using UNet & PSPNet

Mahesh Balasaheb Raut
Guelph, Ontario

***Abstract*—Medical image analysis is crucial for disease understanding and treatment. Detecting early issues like polyps in the stomach and intestine is vital to prevent conditions like colorectal cancer. Limited annotated images challenge computer-aided diagnosis. We utilize the Kvasir-SEG dataset to enhance polyp detection. This project presents a pipeline for developing and evaluating polyp segmentation models using UNet and PSPNet architectures. Models leverage feature extraction and domain knowledge. Training involves hyperparameter tuning, loss configuration, and optimization. Evaluation metrics include IoU score and Dice Score. Results show strong segmentation by both models. UNet achieves IoU 0.8897, Dice Loss 0.0705, and Dice Score 0.9374. PSPNet obtains IoU 0.8600, Dice Loss 0.0828, and Dice Score 0.9210. Both models advance polyp detection and this approach enhances diagnostic precision in medical imaging, contributing to accurate polyp detection and diagnosis.**

***Index Terms*—polyp, UNet, PSPNet, optimizer, Dice, IOU**

I. Introduction

Medical image analysis, a foundation of modern healthcare, is important for accurate disease diagnosis, effective treatment planning, and enhanced patient care. Within this context, pixel-wise image segmentation plays a crucial role, enabling precise depiction of anatomical structures, anomalies, and abnormalities in medical images. This is particularly essential in the realm of gastrointestinal (GI) tract disorders, where early detection of growths like polyps is of paramount importance, given their potential to develop into life-threatening conditions, including colorectal cancer. Despite its significance, obtaining a comprehensive dataset with expertly annotated images for training segmentation models remains a challenge.

A. Polyp Segmentation and Diagnostic Significance

Polyps, abnormal tissue growths, can be precursors to colorectal cancer—a global health concern characterized by high prevalence and substantial public health impact. Detecting and diagnosing polyps at an early stage is critical for preventing the onset of GI diseases, especially colorectal cancer. Colonoscopy, a widely used diagnostic procedure, serves as the standard for detecting, assessing, and removing polyps. However, there are challenges associated with polyp detection, including the potential for oversight and variability in human observation, leading to polyp miss rates during colonoscopies. Improving the accuracy and efficiency of polyp detection is a pressing need to enhance patient outcomes and reduce the burden of colorectal cancer.

B. The Kvasir-SEG Dataset

Addressing the scarcity of annotated medical images specifically tailored for polyp segmentation, the Kvasir-SEG dataset emerges as a valuable asset. This dataset comprises gastrointestinal polyp images accompanied by meticulously annotated segmentation masks, expertly created and verified by experienced gastroenterologists. By introducing segmentation annotations to the existing Kvasir dataset, which previously focused on frame-wise annotations, this initiative opens doors for multimedia and computer vision researchers to contribute to the field of polyp segmentation and automated analysis of colonoscopy videos.

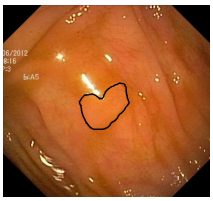

Fig. 1. Sample Frame from the Kvasir Dataset with Enhanced Polyp Tissue Highlighting Using Black Border

The Kvasir-SEG dataset plays a complex role. Firstly, it facilitates the replication of project results, enabling data enthusiasts to build upon and validate existing findings. Secondly, the dataset serves as a benchmark for method comparison, allowing us to assess the performance of their polyp segmentation approaches against a standardized evaluation set. Additionally, the Kvasir-SEG dataset helps in the field by providing a robust foundation for the development of polyp segmentation algorithms.

C. Automated Image Segmentation in Medical Context

Automated image segmentation techniques offer promise in revolutionizing medical image analysis. They have the potential to significantly enhance diagnostic precision, reduce manual intervention, and expedite the process of pathology

monitoring. However, the development of such techniques is not without challenges. Annotated medical data, particularly high-quality annotations from expert medical professionals, is essential for training and validating segmentation models. Manual annotation is time-consuming, labor-intensive, and subject to inter-observer variation. Moreover, concerns related to patient privacy and data security impose additional complexities on data collection.

D. Project Objectives and Structure

In light of these challenges and opportunities, this project seeks to contribute to the automated polyp segmentation in medical imaging. The following sections of this paper delve into the methodology employed for implementing deep learning-based polyp segmentation models utilizing UNet and PSPNet architectures. We detail the dataset preparation, training process, evaluation metrics, and comparative analysis of model performance. The outcomes of this project shed light on the segmentation capabilities of the UNet and PSPNet models, providing insights into their strengths and potential applications. Ultimately, this project aims to underscore the significance of accurate polyp segmentation in the context of medical image analysis and help to the broader goal of improving early disease detection and patient outcomes.

II. Related Work

In recent times, the realm of medical image analysis, particularly in the context of detecting and segmenting gastrointestinal polyps, has experienced a notable surge in research endeavors aimed at bolstering diagnostic precision and automated disease detection forward. Noteworthy contributions within this sphere encompass a comprehensive survey conducted by Ronchi et al. (2020), which categorizes methods for polyp detection and segmentation based on conventional image processing, machine learning, and deep learning paradigms. Another significant undertaking by Cai et al. (2018) entails the introduction of the CVC-ClinicDB dataset, an invaluable resource replete with annotated endoscopic images featuring various gastric ailments, including polyps.

Furthermore, Pogorelov et al. (2017) have presented the ETIS-Larib Polyp DB v2, an influential benchmark dataset encompassing a diverse array of polyp manifestations, thereby serving as a foundational platform for evaluating cutting-edge methodologies. The work of Bi et al. (2020) introduces an elevated deep learning approach catering to polyp segmentation, characterized by the infusion of transfer learning and attention mechanisms to achieve an elevated level of segmentation accuracy. Berthelot et al. (2021) contribute significantly through the creation of the AIChallenger Gastrointestinal Endoscopy Segmentation Dataset, a resource that provides valuable insights into the creation of datasets and the metrics used for evaluating endoscopy image segmentation tasks, including those concerning polyps.

While not exclusively centered on polyps, the endeavor by Wang et al. (2018) puts forth the DeepLesion dataset alongside a transferrable deep learning strategy targeting lesion detection, thereby demonstrating the broader applicability of their approach in the realm of lesion detection. In a parallel vein, Yu et al. (2017) showcase a method grounded in deep learning for polyp segmentation within colonoscopy images, thus spotlighting the inherent advantages of deep learning when going against conventional computer vision techniques.

III. Model Architecture

The UNet architecture, a reliable in medical image segmentation, bears a structural resemblance to the letter 'U'. This design inherently fosters both expansive and contractive pathways within the network, enabling effective feature extraction and accurate segmentation. The contracting path consists of convolutional and pooling layers, which progressively capture and distill pertinent features from the input image. These features are then propagated through a series of upsampling and convolutional layers to achieve precise segmentation.

A. UNet

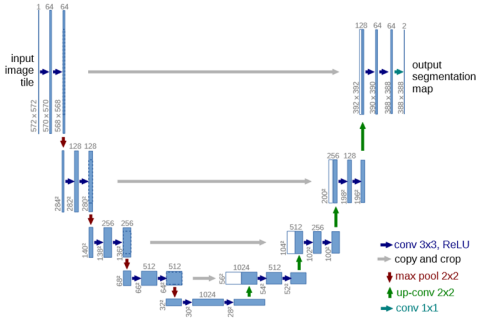

Fig. 2. UNet Model Architecture (Source: https://arxiv.org/abs/1505.04597)

Pretrained encoders, typically based on popular architectures like VGG or ResNet, are seamlessly integrated into UNet. These encoders bring a wealth of domain knowledge, facilitating the recognition of complex patterns and textures within medical images. Additionally, the UNet's innovative skip connections establish a direct link between the contracting and expanding pathways, allowing the model to retain fine details while simultaneously incorporating context. This combination of architectural ingenuity, pretrained encoders, and skip connections empowers the UNet to capture intricate details and contextual information vital for polyp segmentation in medical images. Its adaptability and performance make it a cornerstone in our quest for accurate and effective polyp detection.

B. PSPNet

The PSPNet (Pyramid Scene Parsing Network) architecture introduces an approach to semantic segmentation, particularly suited for tasks like polyp segmentation in medical imaging. At its core, PSPNet incorporates a pyramid pooling module, strategically designed to capture both local and global context from an input image. This module enables the model to comprehend fine-grained details as well as larger-scale patterns, crucial for accurate segmentation. The architecture also

integrates pretrained encoders, often derived from established networks such as ResNet or VGG, which furnish the model with foundational knowledge about features in various spatial hierarchies. Customized activation functions further enhance the model's ability to discriminate between different classes, optimizing its performance for the specific task of polyp detection.

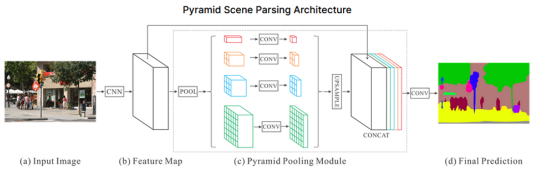

Fig. 3. PSPNet Model Architecture (Source: https://arxiv.org/abs/1612.01105)

One of PSPNet's defining features is its proficiency in capturing global context. The pyramid pooling module facilitates the extraction of features at multiple scales, enabling the model to comprehend contextual cues that aid in distinguishing polyps from surrounding tissues. This focus on holistic scene understanding, combined with the power of pretrained encoders, renders PSPNet an intriguing candidate for our polyp segmentation endeavor.

IV. Model Training and Evaluation

The model training process for the UNet and PSPNet architectures is outlined, both equipped with a 'resnet50' encoder. The dataset utilized for training and validation is sourced from the Kvasir dataset, known for its rich collection of gastrointestinal images. The model training pipeline includes data preprocessing, loss functions, optimization, scheduling, and progress monitoring.

A. Data Preparation

The Kvasir dataset is divided into a training set and a validation set using a 90/10 split. This partitioning ensures an adequate amount of data for training while maintaining a separate subset for model evaluation. Preprocessing steps involve data augmentation and normalization. Augmentation techniques such as random rotations, flips, and zooms are applied to enhance model generalization. Normalization ensures that pixel values are scaled to a standard range, facilitating model convergence.

B. Architecture and Setup

The UNet and PSPNet architectures serve as the backbone for our model design. The 'resnet50' encoder is integrated into both architectures to leverage its pretrained features, enhancing the model's ability to capture relevant information from medical images. The PSPNet is configured similarly to the UNet, creating a consistent experimental setup for comparison.

C. Metrics

Model's performance is evaluated using multiple metrics, including Dice Loss, Intersection over Union (IoU), and Dice Score. The Dice Loss quantifies the dissimilarity between predicted and ground truth segmentation masks.

IoU measures the overlap between predicted and true regions, defined as:

$$\text{IoU} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives} + \text{False Negatives}}$$

The Dice Score assesses the model's segmentation accuracy, defined as:

$$\text{Dice Score} = \frac{2 \cdot \text{Intersection}}{\text{Pixels in Prediction} + \text{Pixels in Ground Truth}}$$

These metrics collectively provide a comprehensive understanding of our model's segmentation accuracy.

D. Optimization and Scheduling

To optimize our models, we employ the Adam optimizer with a learning rate of 0.0001. Additionally, we implement the Cosine Annealing Warm Restarts scheduling strategy, which dynamically adjusts the learning rate throughout training. This technique enables our models to navigate the optimization landscape effectively and converge towards optimal solutions.

E. Training and Progress Monitoring

The training process spans five epochs, allowing the models to iteratively refine their segmentation capabilities. We monitor training and validation logs to assess the convergence trajectory and detect potential issues such as overfitting or underfitting. Regular checkpoints are saved after each epoch, ensuring that the best-performing model is preserved for future use and analysis.

V. Results

A. UNet

The performance of the UNet model is evaluated and presented in this section. The training process demonstrates progressive improvement across epochs, as evidenced by the decreasing Dice Loss values. The Dice Loss decreases from an initial value of 0.1533 to a final value of 0.0569, indicating an enhancement in segmentation accuracy over the training iterations. This reduction in Dice Loss is a strong indicator of the model's ability to generate more accurate predictions.

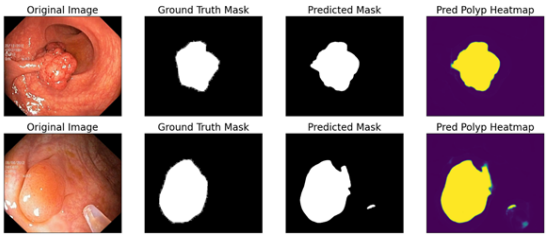

Fig. 4. UNET Example Predictions

Additionally, the Intersection over Union (IoU) Score, also known as the Jaccard Index, exhibits consistent growth throughout the training epochs. Starting at 0.8746 and concluding at 0.9119, the increasing IoU Score signifies a continuous improvement in the overlap between the predicted segmentation and the ground truth masks. This augmentation in IoU Score substantiates the model's progress in delineating accurate boundaries of polyp regions.

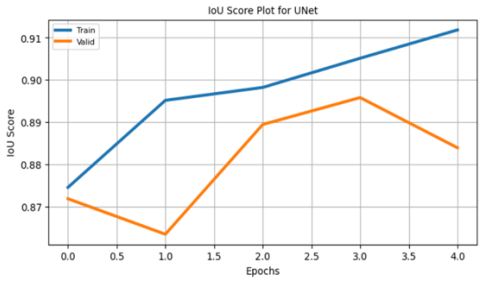

Fig. 5. IOU Score Plot for UNet

The Dice Score, follows a similar upward trajectory. The Dice Score starts at 0.9321 and steadily rises to 0.9529 as training advances. This consistent improvement in the Dice Score further emphasizes the UNet model's capability in precise image segmentation. The higher Dice Score values indicate a stronger correspondence between the predicted and ground truth polyp regions.

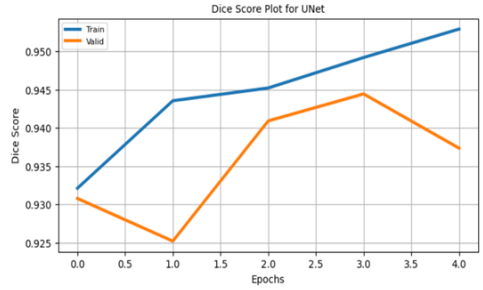

Fig. 6. Dice Score Plot for UNet

B. PSPNet

The performance of the PSPNet model is demonstrated in this section. Throughout the training epochs, the model exhibits gradual enhancement, as reflected in the declining Dice Loss values. The Dice Loss decreases from an initial value of 0.1570 to a final value of 0.0798, showcasing an improved precision in segmentation as the model learns from the data.

Moreover, the Intersection over Union (IoU) Score, or Jaccard Index, shows a steady increase from 0.7565 to 0.8607 across epochs. This upward trajectory of the IoU Score signifies an augmented overlap between the predicted segmentation and the ground truth masks. This consistent improvement in IoU Score indicates the PSPNet model's ability to accurately delineate polyp regions.

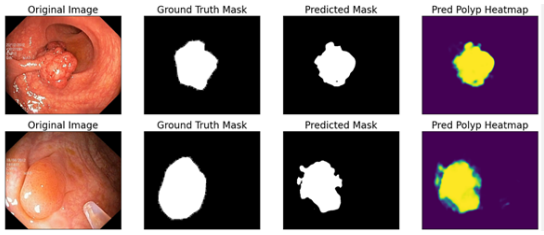

Fig. 7. PSPNet Example Predictions

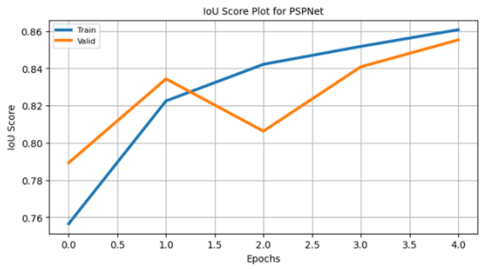

Fig. 8. IOU Score Plot for PSPNet

The Dice Score, also follows an upward trend. Starting at 0.8542 and gradually elevating to 0.9237, the Dice Score showcases the PSPNet model's proficiency in precise image segmentation. The increasing Dice Score values highlight the model's success in capturing the correspondence between the predicted and ground truth polyp regions.

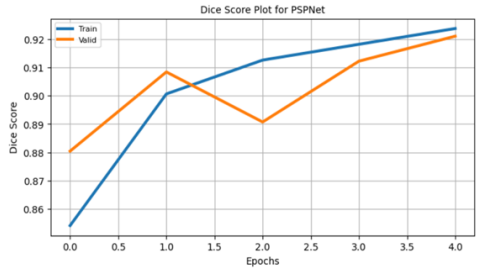

Fig. 9. Dice Score Plot for PSPNet

VI. Comparative Analysis

Comparing the performance metrics of both models, it becomes clear that UNet consistently outperforms PSPNet across all metrics and epochs. UNet's lower Dice Loss values indicate its ability to generate more accurate predictions, while the higher IoU Scores and Dice Scores underscore its proficiency in accurate polyp region delineation. These findings suggest that UNet excels in capturing more details and boundaries, resulting in high segmentation accuracy when compared to PSPNet.

Also, the comparative analysis underscores the distinct strengths of both UNet and PSPNet models in polyp segmentation. While both models show progressive improvement, UNet demonstrates little better performance across key metrics, indicating its potential for enhanced accuracy in automated polyp detection and diagnosis tasks.

TABLE I
MODEL PERFORMANCE METRICS FOR UNET AND PSPNET

Model	Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5
UNet					
Dice Loss	0.153295	0.098130	0.077641	0.065151	0.056876
IoU Score	0.874588	0.895190	0.898244	0.905154	0.911858
Dice Score	0.932122	0.943563	0.945232	0.949217	0.952936
PSPNet					
Dice Loss	0.156990	0.105502	0.092216	0.085852	0.079802
IoU Score	0.756549	0.822493	0.842161	0.851748	0.860717
Dice Score	0.854204	0.900609	0.912532	0.918100	0.923667

VII. CONCLUSION

The project focused on the development and evaluation of a "Polyp Segmentation Model using UNet and PSPNet." The assessment shed light on the segmentation capabilities of both architectures. UNet exhibited better performance, achieving a high Mean IoU Score of 0.8897, indicating substantial overlap, a low Mean Dice Loss of 0.0705, reflecting precise accuracy, and a Mean Dice Score of 0.9374, effectively capturing polyp regions.

PSPNet also demonstrated strong performance, boasting a Mean IoU Score of 0.8600 with significant overlap, a Mean Dice Loss of 0.0828 showcasing accurate segmentation, and a Mean Dice Score of 0.9210 indicating confident identification of polyp regions. While UNet slightly surpassed PSPNet in terms of IoU and Dice Scores, both models contribute significantly to medical imaging, elevating polyp segmentation and advancing diagnostic precision.

The project's structured approach, showing data preparation, model training, and evaluation, underscores the importance of architecture selection and training techniques. The pipeline shows to drive polyp segmentation methodologies, fostering detection and diagnosis within the realm of medical imaging.

REFERENCES

[1] Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., & Johansen, H. D. (2020). Kvasir-seg: A segmented polyp dataset. In International Conference on Multimedia Modeling (pp. 451-462). Springer.

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR, abs/1505.04597. Retrieved from http://arxiv.org/abs/1505.04597

[3] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2016). Pyramid Scene Parsing Network. CoRR, abs/1612.01105. Retrieved from http://arxiv.org/abs/1612.01105

[4] Qiu, Z., Wang, Z., Zhang, M., Xu, Z., Fan, J., & Xu, L. (2022, April). BDG-Net: Boundary distribution guided network for accurate polyp segmentation. In Ivana Išgum & Olivier Colliot (Eds.), Medical Imaging 2022: Image Processing. SPIE. doi:10.1117/12.2606785

[5] Xiangxiang Chu, Zhi Tian, Bo Zhang, Xinlong Wang, Xiaolin Wei, Huaxia Xia, and Chunhua Shen. Conditional positional encodings for vision transformers. arXiv preprint arXiv:2102.10882, 2021.

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

[7] Cai, W., Chen, H., Zhang, D., & Chen, X. (2018). CVC-ClinicDB: Gastrointestinal Image Database for Diagnosis of Gastric Diseases. Digestive Endoscopy, 30(3), 379-383.

[8] Berthelot, M., Drouin, S., & Gelly, S. (2021). The AIChallenger Gastrointestinal Endoscopy Segmentation Dataset. arXiv preprint arXiv:2101.00855.

[9] Bi, L., Kim, J., Ahn, E., & Feng, D. (2020). An Improved Deep Learning Approach for Polyp Segmentation. IEEE Access, 8, 7689-7699.

[10] Pogorelov, K., Khlif, M. S., & Bhattacharjee, S. (2017). ETIS-Larib Polyp DB v2. IEEE Transactions on Medical Imaging, 37(1), 190-199.

[11] Ronchi, S., Berta, D., & De Luca, M. (2020). Polyp Detection and Segmentation in Colonoscopy Frames: A Comprehensive Survey. Computers in Biology and Medicine, 126, 104004.

[12] Goldbloom, A., Hamner, B., et al.: Kaggle: Your home for data science. Competition, Kaggle, Inc, https://www.kaggle.com (2019), accessed: 2019-07-12

[13] Haggar, F.A., Boushey, R.P.: Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors. Clinics in colon and rectal surgery 22(04), 191–197 (2009)

[14] Kaminski, M.F., Wieszczy, P., Rupinski, M., Wojciechowska, U., Didkowska, J., Kraszewska, E., Kobiela, J., Franczyk, R., Rupinska, M., Kocot, B., Chaber-Ciopinska, A., Pachlewski, J., Polkowski, M., Regula, J.: Increased Rate of Adenoma Detection Associates With Reduced Risk of Colorectal Cancer and Death. Gastroenterology 153(1), 98–105 (2017)

[15] Kang, J., Gwak, J.: Ensemble of instance segmentation models for polyp segmentation in colonoscopy images. IEEE Access 7, 26440–26447 (2019)

[16] Otsu, N.: A threshold Selection Method from Gray-Level Histograms. IEEE transactions on systems, man, and cybernetics 9(1), 62–66 (1979)

[17] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2018). DeepLesion: Automated Mining of Large-Scale Lesion Annotations and Universal Lesion Detection with Deep Learning. Journal of Medical Imaging, 5(3), 036501.

[18] Yu, L., & Chen, H. (2017). Polyp Segmentation in Colonoscopy Using Deep Learning. In Proceedings of SPIE Medical Imaging (Vol. 10134, p. 101340H). International Society for Optics and Photonics.