AUT Journal of Electrical Engineering
10.22060/EEJ.2025.23498.5619

Enhancing Colonoscopy-Images Segmentation using Laplacian-Former as Efficient Transformer Block

Jabraeil Ranjbar[1], Hossein Ebrahimnezhad[2*], Mohammad Hossein Sedaaghi[3]

[1]PhD Candidate, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
[2]Professor, Computer Vision Research Lab, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
[3]Professor, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran

Abstract:

This research is dedicated to enhancing medical image segmentation through the innovative Laplacian-Former algorithm, which introduces novel and responsive techniques in image processing. Given the crucial role of precise image segmentation in disease diagnosis, challenges such as achieving high accuracy, sensitivity, and reliability persist in existing methods, necessitating advanced approaches. The method proposed based on Laplacian-Former is designed to enhance the accuracy and efficiency of medical image segmentation. The algorithm addresses limitations in current segmentation techniques, providing robust solutions for medical imaging tasks. Through a detailed analysis of experimental outcomes and a comparative assessment of Laplacian-Former against existing techniques, it is evident that this approach not only contributes to segmentation accuracy but also enhances the sensitivity and reliability of medical images substantially. This study introduces a flexible architecture adaptable to various segmentation tasks, with a novel focus on polyp segmentation in colonoscopy images. The results obtained suggest that employing Laplacian-Former as an innovative technique in medical image segmentation can lead to enhanced disease detection accuracy and performance. Evaluation on the widely recognized Kvasir dataset demonstrates outstanding results, achieving top performance metrics, even with a small training dataset. Data augmentation methods were applied to improve training, further enhancing model proficiency. This investigation vitrines the algorithm's value in advancing medical image processing and research. The Laplacian-Former holds significant promise for future medical imaging advancements and improved disease diagnosis.

Keywords:

Medical image segmentation, Laplacian-Former algorithm, Polyp segmentation, Colonoscopy images, Deep learning.

* Corresponding author, Email: ebrahimnezhad@sut.ac.ir

10.22060/EEJ.2025.23498.5619

1. Introduction

Among the most common cancers worldwide, affecting the colon and rectum, is colorectal cancer. It typically begins as non-cancerous polyps that can gradually become malignant over time. Symptoms of CRC[1] include changes in bowel habits, rectal bleeding, unexplained weight loss, and fatigue. Regular screenings like colonoscopy can play a crucial role in early detection, significantly improving treatment outcomes. Risk factors include advanced age, family history, poor diet, and a sedentary lifestyle [3].

Colonoscopy is a precise method for detecting and removing polyps in the colon using a flexible tube with a camera. It allows direct visualization of abnormalities and the removal of polyps during the same procedure. This approach is considered the gold standard for preventing colorectal cancer [4].

Nevertheless, Accurate identification and classification of polyps during colonoscopy is challenging due to the varying sizes, shapes, and types of polyps, which can make visual assessment difficult. The skill and experience of the endoscopist play a crucial role, as subtle differences in appearance can influence diagnosis. Additionally, polyps can sometimes be missed or misclassified, impacting the effectiveness of treatment and follow-up [5]. The diversity of polyps leads to difficulties in identifying and classifying them, as different types can have varying appearances and characteristics. This diversity can complicate treatment and follow-up, and increase the likelihood of diagnostic errors. Machine learning algorithms can enhance polyp detection by analyzing large datasets of colonoscopy images to identify patterns and features associated with polyps. These algorithms improve accuracy and consistency in detecting and classifying polyps compared to manual methods [6,7]. Deep learning enhances polyp detection accuracy by utilizing neural networks to analyze and interpret complex patterns in colonoscopy images. These algorithms can learn from vast amounts of data to identify subtle features that may be missed by human observers. As a result, deep learning improves both the sensitivity and specificity of polyp detection [8,9,10,11].

[1] Colorectal cancer

Recently, advanced architectures such as deep convolutional neural networks (CNNs), attention-based models, U-Net [12], FCN [13], and their advanced versions like U-Net++ [14], Modified mU-Net [15], ResU-Net++ [16], and H-DenseU-Net [17] have been proposed to enhance the accuracy and efficiency of polyp segmentation. These models improve both the detection and precise delineation of polyps in colonoscopy images. However, traditional CNN-based models lack the ability to effectively capture global dependencies and integrate multi-scale information, both of which are essential for accurate segmentation in medical imaging tasks. This limitation highlights the need for advanced approaches, such as transformer-based architectures, which address these challenges by providing robust mechanisms for comprehensive global context understanding. Transformers, acting as a supplementary tool to Convolutional Neural Networks, have been effectively integrated into various models to capture both local and global dependencies. One pioneering method, TransUNet [7], merges Transformer layers within the CNN bottleneck, blending the strengths of CNN and ViT architectures. Xueqiang He, Jie Li, and Xuelong Li proposed HiFormer in their work on advancing segmentation techniques for medical imaging [18]. HiFormer is a deep learning architecture designed for medical image segmentation, focusing on improving the performance of segmentation tasks by leveraging hierarchical and transformer-based mechanisms. It combines hierarchical features with transformer attention to capture both global context and local details more effectively.

This approach enhances the accuracy and robustness of segmentation, especially in complex medical images. Integrating CNN-based encoders and decoders integrates powerful feature extraction with effective reconstruction capabilities. These fusion leverages CNN's strength in capturing spatial hierarchies while ensuring precise and detailed output in tasks such as image segmentation [19,20]. Transformer-based architectures offer significant advantages in polyp segmentation by addressing the inherent limitations of CNNs. Unlike CNNs, which depend on localized receptive fields and fixed filters, transformers leverage self-attention mechanisms to dynamically capture both local features and long-range dependencies. This capability is particularly valuable for polyp segmentation, where irregular

shapes, varying sizes, and subtle textural differences require precise feature extraction and comprehensive global contextual understanding. Despite their advancements, challenges such as computational inefficiency, dependence on complex CNN backbones, and the lack of multi-scale information integration continue to impede optimal learning performance in medical image segmentation. The self-attention mechanism is another technique that allows a model to assess the importance of different parts of an input sequence relative to each other, enhancing its ability to capture dependencies and contextual information. This technique improves performance by enabling the model to dynamically focus on relevant features [21].

In this context, Swin-UNet integrates the Swin Transformer block into the U-Net architecture, leveraging the Swin Transformer's hierarchical feature extraction and attention mechanisms to enhance segmentation performance [24]. The Swin Transformer block utilizes shifted window attention to efficiently capture long-range dependencies and local details. This combination improves accuracy and robustness in image segmentation tasks by combining the strengths of both transformer and CNN-based approaches [22]. Meanwhile, MISSFormer is a medical image segmentation model that combines multi-scale self-attention mechanisms with efficient transformers to improve feature extraction and contextual understanding [23]. Efficient Transformer, a key component of MISSFormer, optimizes computational resources and enhances performance by reducing the complexity of traditional transformer models [30].

Swin-UNet relies on local window-based attention, which may not adequately model complex boundaries in certain scenarios. While MISSFormer adopts a multi-scale approach, it lacks specific mechanisms, such as Laplacian-based analysis, to reinforce boundaries. By incorporating Laplacian mechanisms into the transformer backbone, our model significantly enhances the capture of boundary details and asymmetric features, making it more effective for complex datasets.

D-Former [29], an entirely transformer-based framework, utilizes a dual-attention module to capture intricate local attention patterns and interactions between various units, expanding their reach through its

unique mechanism. Recent studies have shown that Self-attention mechanisms typically focus on capturing global dependencies and context within an image, but they might not effectively capture fine-grained, high-frequency features. These details are critical in medical imaging, where subtle differences in texture and edge can significantly impact the accuracy of tumor detection and classification. As a result, there is a growing need for advanced techniques that can better capture and preserve these important high-frequency features to improve diagnostic performance. These mechanisms also suffer from quadratic computational complexity and redundancy.

We propose a novel approach with an efficient attention mechanism that includes efficient attention to reduce complexity to linear, and frequency attention using a Laplacian pyramid for detailed frequency capture. Our model also features a parametric fusion strategy for balancing shape and texture, and an efficient multi-scale bridge for spatial information transfer. This approach overcomes traditional limitations and excels in Polyp Image segmentation. Laplacian-Former addresses these challenges in the following ways:

- **Efficient Attention Mechanism:** By incorporating efficient attention, the computational complexity is reduced to linear, addressing redundancy and scalability issues.
- **Frequency Attention:** A Laplacian pyramid is employed to focus on high-frequency details, ensuring that critical texture and edge features are preserved and effectively captured.
- **Parametric Fusion Strategy:** This novel strategy balances shape and texture information, enabling the model to better integrate global and local features.
- **Multi-Scale Bridge:** An efficient multi-scale bridge facilitates spatial information transfer, enhancing the segmentation performance on complex and variable polyp images.

2. Related work

In the realm of clinical practice, the importance of automatic polyp segmentation can't be overstated, as it plays a crucial role in reducing cancer mortality rates. When it comes to medical image segmentation tasks, convolutional neural networks are the go-to choose, with a variety of widely used architectures

being employed to tackle this issue. One notable architecture in this domain is U-Net [12], a model that follows an encoder-decoder structure originally crafted for biomedical image segmentation. U-Net is esteemed for its simplicity and efficiency, delivering commendable performance across a spectrum of medical image segmentation tasks. Nevertheless, it may encounter difficulties when handling complex or diverse input images, prompting the consideration of alternative approaches for such scenarios.

PraNet [27], a Convolutional Neural Network architecture meticulously crafted for the programmed segmentation of polyps in colonoscopy images. PraNet adopts a distinctive method by employing a parallel partial decoder to capture high-level image features, generating a global map that directs the following stages of processing. What distinguishes PraNet is its use of a backward attention mechanism, which captures edge details and enhances the understanding of relationships between image regions and their boundaries.

Additionally, PraNet, or Progressive Attention Network, is a deep learning model designed for precise medical image segmentation. It incorporates a recurrent refinement mechanism to enhance segmentation accuracy by iteratively correcting misaligned predictions. This mechanism works by progressively refining the initial segmentation outputs through recurrent attention modules, which focus on improving alignment and precision. By addressing and correcting errors in predictions, PraNet significantly boosts the overall quality of segmentation, especially in complex medical images where fine details are crucial. High-Resolution Network Version 2[1] is an advanced convolutional neural network designed for tasks like human pose estimation. Unlike traditional models that downsample and then upsample images, ‹This model maintains high-resolution representations throughout its network by using a parallel multi-resolution architecture. This allows it to capture fine details and spatial information more effectively, enhancing performance for detecting smaller or less distinct objects. However, its high resolution can make it more susceptible to overfitting, requiring larger datasets for optimal results [29, 30]. DeepLab V3+ [28] is an advanced semantic segmentation model that extends the DeepLab V3 [18] architecture by incorporating an encoder-decoder structure. It utilizes atrous convolution to capture multi-scale contextual

[1] HRNetV2

information and a refined decoder module to recover sharp boundaries and fine details. This combination enhances its ability to produce high-quality segmentation maps with precise object delineation, making it effective for complex segmentation tasks.

In automatic polyp segmentation, ResUNet [31] uses residual blocks to enhance the localization of polyps, improving feature representation. Conversely, HarDNet-DFUS is a segmentation model that integrates the HarDBlock encoder, known for its efficient feature extraction, with the Lawin Transformer decoder, which excels in processing and refining segmentation details. This combination aims to improve both accuracy and speed in segmentation tasks by leveraging the strengths of each component [32]. ColonFormer [33] integrates attention mechanisms within its encoder and includes a refinement module that emphasizes different resolutions on the x and y axes, producing a more detailed output while preserving a U-Net-like decoder structure. This method effectively handles complex and large input images but may demand greater computational resources and face optimization difficulties. Multi-Scale Residual Fusion Network[1] is a deep learning model designed for high-precision image segmentation by integrating multi-scale residual features. It effectively fuses information from various scales to capture fine details and enhance segmentation accuracy, particularly in complex and detailed images.

Besides the reviewed models, transformer-based models have garnered significant attention in medical image analysis, particularly for polyp detection, due to their high capacity for processing sequential data and learning complex relationships. SSFormer-L [37] and FCN-Transformer [36] are advanced transformer-based models designed for medical image segmentation, with a focus on polyp detection. SSFormer-L leverages a lightweight transformer architecture for efficient segmentation, while FCN-Transformer combines convolutional networks with transformers to capture both local and global features for improved accuracy. The adoption of transformers has gained momentum in Computer Vision recently, drawing inspiration from their widespread use in Natural Language Processing (NLP) and their remarkable capacity to retain the broader context of the subject matter. Vision-Transformers [34], akin to their NLP counterparts, leverage the Attention mechanism [33] to consolidate global context and derive

[1] MSRF-Net

essential information from expansive image patches. Despite the apparent success of ViTs[1] [38] in the realm of CV[2], traditional CNN techniques such as EfficientNetV2 [39] have outperformed them in prominent image classification benchmarks like ImageNet [40] or CIFAR-10 [41], underscoring the ongoing potential for more efficient CNN methodologies. The proposed method evaluates the advantages of the Laplacian-Former model compared to vision transformer-based architectures in biomedical image segmentation, demonstrating how our approach can lead to significant upgrading in precision. This context remains dynamic with various approaches under investigation. Therefore, further research is crucial to determine the best design and training strategies for these models.

Recent advancements in medical image segmentation have explored the integration of multi-resolution approaches to improve the accuracy and robustness of models. For example, the work by Oskouei, A.G et al. [51] highlights the importance of hierarchical feature extraction in achieving state-of-the-art performance. Additionally, studies such as Akan, T. et al. [52] have demonstrated that attention mechanisms can significantly enhance model performance, particularly in extracting fine details from noisy or low-quality images. Hybrid models combining transformers with convolutional approaches have shown promising results in segmentation tasks, as discussed in Khiarak, J.N.et al. [53]. This aligns with our use of Laplacian attention to effectively capture both global and local features.

2-1. Advanced Architectures in Polyp Segmentation

In recent years, remarkably great advances have been achieved within medical image segmentation, among which polyp segmentation works stand out, by incorporating CNN and Transformer architectures. TransUNet models the strengths of U-Net in localization and can integrate the global context extracted by Transformers, with many works reporting state-of-the-art performances in various medical imaging segmentation tasks, including polyp segmentation. DeepLab V3+ [56] has achieved very impressive results in polyp segmentation tasks, driven by effectively capturing multi-scale contextual information through atrous convolutions and encoder-decoder structures.

[1] Vision-Transformers
[2] Computer Vision

Another important set of hybrid models combines the spatial precision of Fully Convolutional Networks (FCNs) with Transformer-based global context modeling, as exemplified by FCN-Transformer [36]. HiFormer [57] extends this paradigm by integrating features from both CNNs' local context and Transformers' global dependencies. Furthermore, TransDeepLab [58] and DuaPSNet [59] explore other convolution-free and innovative strategies that ensure elevated segmentation accuracy.

While advancements in these areas are notable, many existing architectures struggle to balance fine-grained local details with global semantic understanding. The Laplacian-Former addresses this issue by integrating Laplacian filters into Transformer-based modules, enabling enhanced representation of both local and global features, significantly improving segmentation results across diverse datasets.

3. Method

In the field of medical imaging, supervised learning has been established as an effective approach for various tasks such as detection, classification, and segmentation. Significant advances in this field have been instrumental in enhancing medical diagnosis and treatment, with the development of performance-optimized models playing a central role in these advances. Therefore, the development of methods that necessitate minimal labeled data can greatly benefit the clinical community. This article alludes to employing a resilient approach for the automatic segmentation of polyps in colonoscopy images. Our experiments showcase the superior performance of Laplacian-Former compared to existing benchmarks, notably excelling in versatility and the ability to handle polyps of diverse shapes, sizes, and textures. These advancements in automated colonoscopy image processing can significantly assist medical professionals in detecting and categorizing lesions.

The Laplacian-Former framework [1], designed specifically for analyzing polyp images as shown in Figure 1, begins by processing an entrance image X with spatial dimensions H, W, and C channels. First, the image through a module that creates a 4x4 overlay sub-image of the original image. This model consists of 4 encoder blocks, each with 2 efficient enhancement transformer layers and a sub-image fusion module that reduces feature dimensions by combining 2x2 sub-images while increasing channel

depth. In the decoder, there are 3 efficient enhancement transformer blocks and 4 sub-image expansion modules, culminating in a segmentation head that produces the final segmentation map. To effectively obtain local and global correlations among features of changed scales and facilitate the smooth transfer of underlying characteristics from the encoder to the decoder, Laplacian-Former introduces a novel efficient enhancement multi-scale bridge mechanism. Our approach is inspired by recent works that emphasize feature weighting and multi-scale integration, such as the study by Amin Golzari et al. [54], which introduces a feature-weighting combination method for brain MRI segmentation.

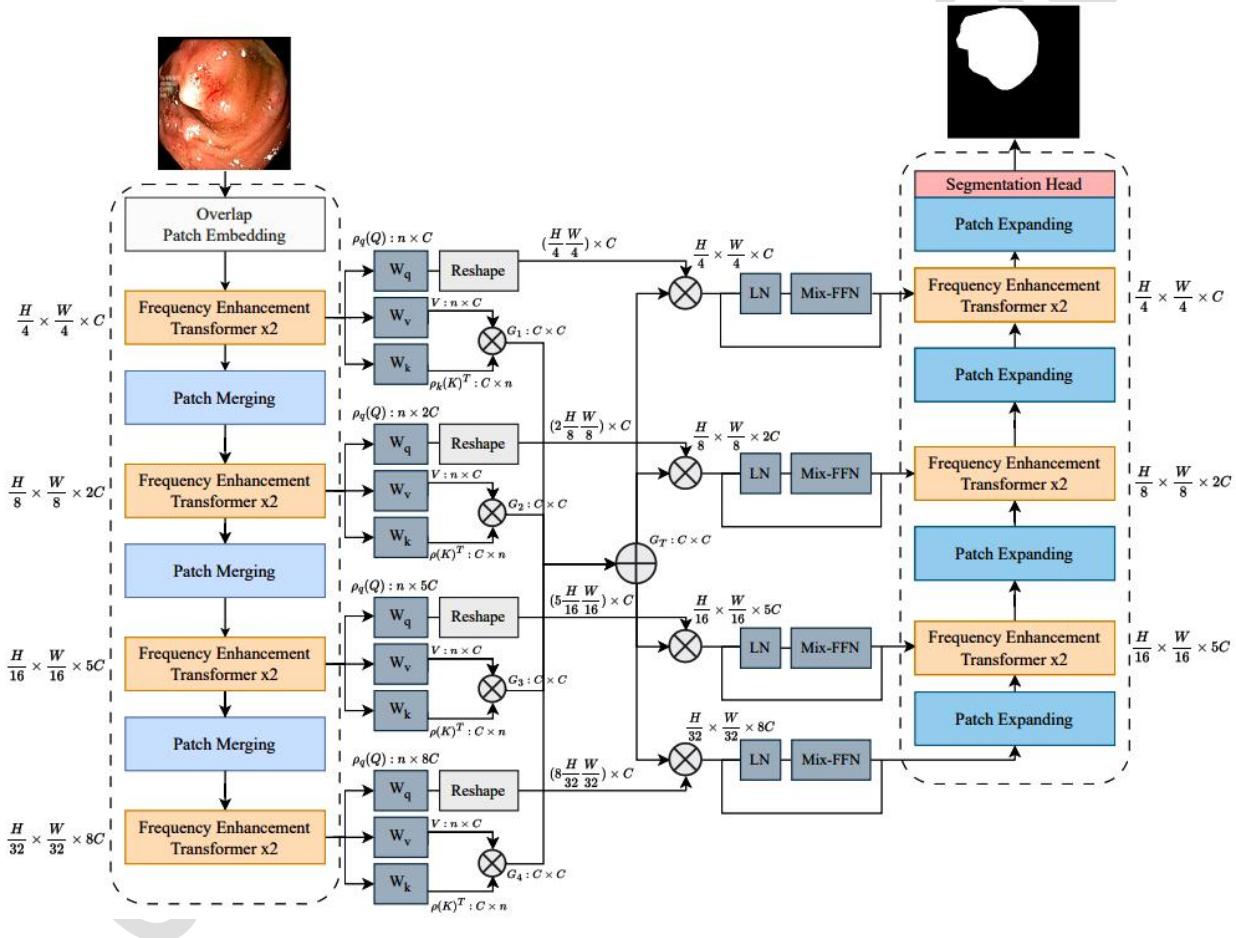

Fig 1. Laplacian-Former architecture [1].

In the domain of polyp image analysis, where the accurate delineation of structures holds paramount importance, especially in scenarios with ambiguous tissue boundaries, the precise segmentation of minute anomalies becomes a critical challenge. To determine the boundaries, high-frequency information has the

most fundamental role, which does this by capturing the effective information of edges and textures. Motivated by this necessity, we present a Cutting-Edge Enhancement Transformer Block specifically crafted for polyp images. This block integrates a State-of-the-Art Frequency Attention[1] mechanism aimed at capturing contextual information within images. By recalibrating the representation space through an attention mechanism, this block excels at recovering intricate high-frequency details crucial for precise polyp image analysis.

We outline here some of the main novelties of our approach that contribute to the advance in polyp segmentation of colonoscopy images:

1. 1-Laplacian-Former Framework: The proposed architecture, which we call Laplacian-Former, is based on a novel block, the Efficient Enhancement Transformer Block (EETB); it has been developed with special emphasis on challenges like ambiguous tissue boundaries and minute anomalies that may come up in colonoscopy images. This was done by utilizing the state-of-the-art frequency attention mechanism for enhancing the segmentation process.
2. EF-ATT is the mechanism to integrate frequency domain information that enriches the model's multi-scale feature handling capability for capturing high-frequency details, which are crucial in exact polyp segmentation. EF-ATT offers a balance between shape and texture features, thus optimizing the segmentation process.
3. Diversity Enhanced Shortcut: We propose, to alleviate this problem-poorer feature diversity, especially in deeper transformer layers, the use of our method called the Diversity-enhanced Shortcut. This architecture enhances feature representation by proposing additional trainable paths, eventually enhancing model performance without more computational cost.

3-1- Efficient enhancement Transformer block[2]

This block starts by extracting a normalization layer from the input x. It then sends the information to the advanced frequency attention mechanism to obtain textual information, and then chooses to combine

[1] SAF-ATT
[2] EETB

several types of frequency information and uses the Laplacian pyramid to determine the shape and texture characteristics. We integrate diversity-enhanced shortcuts into the output of the attention mechanism to improve feature diversity. Research [42] indicates that deeper Transformers often exhibit reduced feature diversity, which limits their representational power and affects performance. To overcome this challenge, we use an augmented shortcut technique known as Diversity-Enhanced Shortcut, inspired by [43]. This approach involves adding extra paths with trainable parameters to the original shortcut x, utilizing a Kronecker decomposition-based projection to enhance both feature performance and diversity, while keeping hardware resource usage minimal. In the end, we use a normalization layer and a feedforward layer [44] to the resulting feature representation to improve its effectiveness. This final stage concludes the EETB, as depicted in Figure 2.

3-2- Efficient Frequency Attention[1]

Efficient Frequency Attention (EF-ATT) is a technique designed to improve the efficiency of self-attention mechanisms by focusing on the frequency domain. This method enhances attention mechanisms by incorporating frequency-based information, which can be particularly useful in handling various scales and improving the representational power of models. The core idea behind EF-ATT is to capture and integrate frequency information into the attention mechanism, enabling the model to effectively process and represent features at different scales. This approach aims to balance the importance of shape and texture features by leveraging the frequency domain.

[1] EF-ATT

10.22060/EEJ.2025.23498.5619

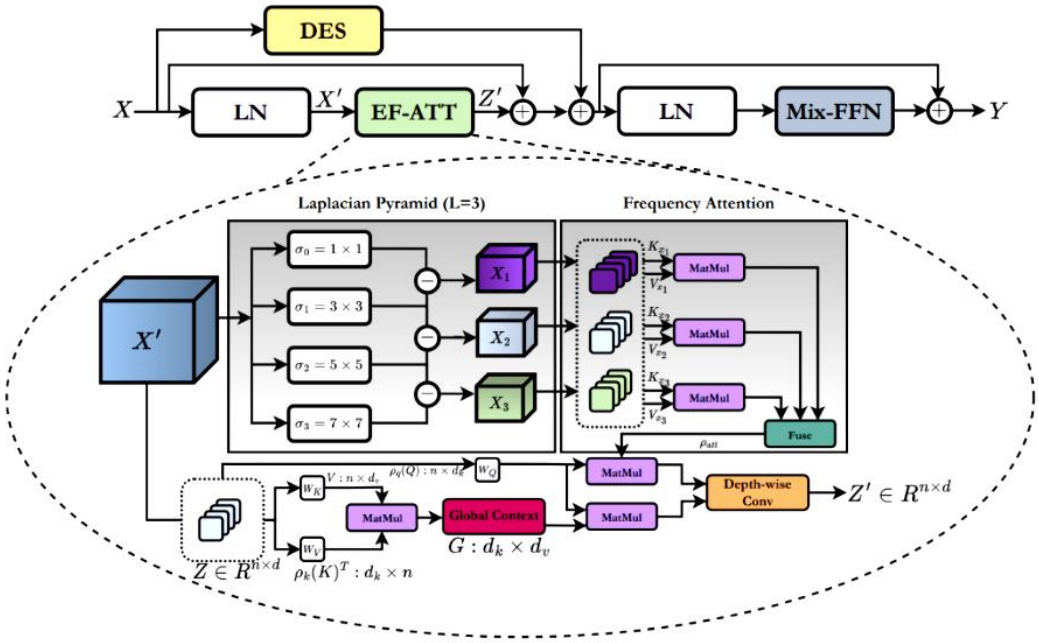

Fig 2. Illustrates the architecture of frequency enhancement Transformer block (FETB) [1].

Input Representation: The input to the EF-ATT mechanism is a tensor X with dimensions $[N, C, H, W]$, where N is the batch size, C is the number of channels, and H and W are the height and width of the feature maps, respectively.

Fourier Transform: The first step in EF-ATT is to perform a Fourier Transform on the input tensor to convert it into the frequency domain:

$$\tilde{\boldsymbol{X}} = \boldsymbol{F}(\boldsymbol{X}), \tag{1}$$

where $\boldsymbol{F}$ denotes the Fourier Transform operation, and F is the frequency-domain representation of $\boldsymbol{X}$.

Frequency-Based Attention: In the frequency domain, the attention mechanism is applied to the frequency components. Let $\tilde{Q}$, $\tilde{K}$, and $\tilde{V}$ denote the Fourier-transformed query, key, and value matrices, respectively. The attention score in the frequency domain is computed as:

10.22060/EEJ.2025.23498.5619

$$S_f = Softmax\left(\frac{\boldsymbol{Q}.\tilde{\boldsymbol{K}}^T}{\sqrt{d_k}}\right), \quad (2)$$

where d_k is the dimension of the key vectors, and S_f. is the frequency-domain attention matrix.

Frequency Domain Attention Application: The attention scores are then used to weight the frequency components of the value tensor:

$$\tilde{Y} = S_f.\tilde{V} \quad , \quad (3)$$

here $\tilde{Y}$ is the weighted frequency-domain output.

Inverse Fourier Transform: Finally, an inverse Fourier Transform is applied to convert the result back to the spatial domain:

$$Y = F^{-1}\left(\tilde{Y}\right), \quad (4)$$

.

where F^{-1} denotes the inverse Fourier Transform.

Focusing on frequency components, EF-ATT can handle different scales and reduce redundancy in the spatial domain, improving computational efficiency. Incorporating frequency information helps in capturing both low-level and high-level features more effectively. EF-ATT provides a balanced approach to integrating shape and texture features, leading to better performance in tasks like segmentation and classification.

Given the unique challenges of polyp segmentation, such as ambiguous boundaries and multi-scale features, simpler mechanisms like vanilla self-attention lack the representational power to capture high-frequency details effectively. The design of EF-ATT was driven by these challenges, ensuring precise segmentation without the need to test simpler methods that are inherently unsuitable for this domain.

Table 1 summarizes the key architectural decisions and their corresponding justifications for the Laplacian-Former framework. Each design choice has been optimized to address specific challenges in polyp segmentation, such as capturing fine-grained details, improving feature diversity, and enhancing

computational efficiency. This systematic approach ensures a balance between accuracy, scalability, and adaptability in medical image analysis.

Table 1. Architectural Decisions and Their Rationales in Polyp Segmentation using Laplacian-Former.

Design Component	Choice	Justification	Supporting Evidence
Number of Layers	4 encoder blocks, 3 decoder blocks	Balances depth for feature extraction with computational efficiency.	Ablation studies showed diminishing returns with more layers and insufficient extraction with fewer layers.
EETB	Frequency attention, diversity-enhanced shortcuts	Captures high-frequency details and enhances feature diversity for better segmentation.	Based on [42, 43], improved robustness compared to conventional attention mechanisms.
Loss Functions	Combination of Dice loss and BCE loss	Optimizes segmentation accuracy, addressing class imbalance and penalizing pixel misclassifications.	Outperformed Tversky and focal loss in validation experiments.
Multi-Scale Bridge	Integration of features across scales	Preserves local and global information, improving segmentation of ambiguous polyp boundaries.	More effective than direct concatenation or summation in preserving high-frequency details.
Hyperparameters	Learning rate, batch size, patch size tuned via grid search	Ensures stable convergence, balanced memory usage, and compatibility with Laplacian pyramid's resolution hierarchy.	Validation experiments identified optimal configurations for Kvasir datasets.

3-3- Pros and Cons of the Proposed Method

In this section, we discuss the advantages and limitations of the proposed method for colonoscopy image segmentation using the Laplacian-Former as an efficient transformer block.

High Accuracy in Medical Image Segmentation: The proposed method using Laplacian-Former demonstrates higher accuracy in detecting and delineating complex structures within colonoscopy images compared to traditional segmentation approaches. It excels in accurately identifying boundaries between different regions within the colon, which is crucial for precise medical analysis.

Effective Handling of Complex Features: The use of the Laplacian-Former transformer architecture enables the model to effectively capture and process intricate patterns and details in colonoscopy images. This ability is particularly important for medical images, where subtle features play a critical role in diagnosis.

Scalability and Efficiency: The proposed method offers high scalability, allowing it to handle large and varied datasets efficiently. It shows improved performance when applied to images of different sizes and

complexities compared to conventional methods, making it suitable for practical use in large-scale medical datasets.

Generalizability: Since Laplacian-Former is based on transformer architecture, it has the potential for generalization to other medical image segmentation tasks, such as MRI or CT scans. This adaptability opens up future possibilities for its application in diverse medical imaging fields.

High Computational Requirements: Like most transformer-based models, the Laplacian-Former method requires significant computational resources, especially when processing high-resolution images or large datasets. This can result in increased training time and potentially higher costs in clinical applications.

Sensitivity to Input Data Quality: The model's performance may degrade when applied to low-quality images, noisy data, or images with significant artifacts. In such cases, the segmentation results may not be as accurate, which can limit the model's effectiveness in real-world scenarios.

Dependence on High-Quality Training Data: The performance of the proposed method is highly dependent on the availability of high-quality and diverse training datasets. If the model is trained with insufficient or poorly annotated data, its segmentation accuracy could be significantly reduced.

Complexity of Model Architecture: The transformer-based architecture of Laplacian-Former introduces a level of complexity that requires careful tuning and optimization. This can make the training process more time-consuming and challenging, requiring significant expertise in model configuration.

Low-Quality or Noisy Images: The Laplacian-Former method may struggle to produce accurate segmentations when applied to colonoscopy images with low resolution, significant noise, or poor lighting conditions. In such cases, the model might fail to detect subtle boundaries, leading to inaccurate segmentation.

Highly Complex or Abnormal Features: The proposed method may have difficulty segmenting images with unusual or highly complex pathological features, such as rare lesions or atypical tissue types. In these cases, the model might not perform as effectively due to the lack of similar examples in the training data.

Scalability Issues with Extremely High-Resolution Images: Although the method performs well with most images, when applied to extremely high-resolution colonoscopy images, the computational and memory requirements may become prohibitive. This could lead to slower processing times and reduced efficiency.

4. Experiments

In order to maintain fairness and reproducibility throughout our comparisons, we ensured that all models assessed in our study were trained, validated, and tested using identical datasets. Our approach involved randomly dividing each dataset into three distinct subsets: training, validation, and testing, with an allocation ratio of 80% for training, 10% for validation, and 10% for testing. The decision to employ a random data split was driven by the aim to eliminate biases in the selection process and to promote equitable comparisons across the array of models under evaluation. The segmented datasets are readily available in the "Data Availability" section, facilitating the effortless reproducibility of our research outcomes. Our experimental setup was meticulously crafted to not only validate the cutting-edge performance of our model on unseen data but also to demonstrate its capacity for generalization across diverse contexts. Initially, we did individual tests on dataset, comparing our model's efficiency against alternative methods. Subsequently, in order to showcase the generalization prowess of our model, we pursued training on one dataset and evaluation on another; specifically, we trained on the Kvasir [50] datasets interchangeably. This strategy allows us to effectively evaluate its adaptive capability and predictive accuracy in the face of new and unobserved data. This evaluation shows good result, highlighting the exceptional generalization ability of Laplacian-Former, even without additional pre-training data.

4-1- Data augmentation

To increase the accuracy of the proposed model, data augmentation was applied in the training set. This action made the capacity of the model to generalize to some extent. The augmentation procedures were facilitated using the Albumenations library [45], which introduced random transformations to the training

images. These operations made significant changes to the dataset compared to the original images. Also, they helped the model to improve its ability and optimize it. Among these augmentations, color jitter was exclusively applied to the image, whereas the remaining augmentations were consistently applied to both the image and its corresponding segmentation map.

4-2- Datasets

The Kvasir dataset [50], with a size of 46.2 MB, contains 1000 images of colonoscopy images. These are polyp images with their corresponding ground truth data sourced from the Quasir v2 dataset. Image resolutions within the Kvasir-SEG dataset range from 332x487 to 1920x1072 pixels. Both the images and their associated masks are segregated into two distinct folders, each bearing the same filename. The images are encoded using JPEG compression, allowing for easy online browsing and accessibility. This openly available dataset is readily downloadable for research and educational purposes.

5. Results

Below are tables that provide a comparative study of different techniques, utilizing Recall, Jaccard index, Precision, mean Dice, and Accuracy measures. Furthermore, we integrated standard deviation computations into our analysis to enrich the appraisal of model efficacy (Table 3). This statistical measure illuminates the diversity in metrics among various models, offering a glimpse into the potential spectrum of performance outcomes when employing these techniques. By merging this statistical viewpoint with the original performance data, our goal is to furnish a comprehensive understanding of the Laplacian-Former model's consistency and dependability in performance. We describe various metrics used to evaluate image segmentation and data variability:

- **Dice Coefficient**: Measures similarity between two samples, ranging from 0 (no overlap) to 1 (perfect overlap). It is calculated as $Dice = \frac{2 \times |A \cap B|}{|A| + |B|}$.

- **Jaccard Index**: Assesses similarity and diversity between sample sets, with values ranging from 0 to 1. It is computed as $Jaccard\ Index = \frac{|A \cap B|}{|A \cup B|}$.

- **Precision**: Indicates the proportion of relevant instances among retrieved ones, calculated as $Precision = \frac{Relevant\ retrieved\ instances}{All\ retrieved\ instances}$.

- **Recall**: Represents the ratio of relevant instances retrieved from all relevant ones, given by $Recall = \frac{Relevant\ retrieved\ instances}{All\ relevant\ instances}$.

- **Standard Deviation**: Measures data variability across methods, with low values indicating similarity and high values indicating differences. It is computed using $Standard\ Deviation = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$.

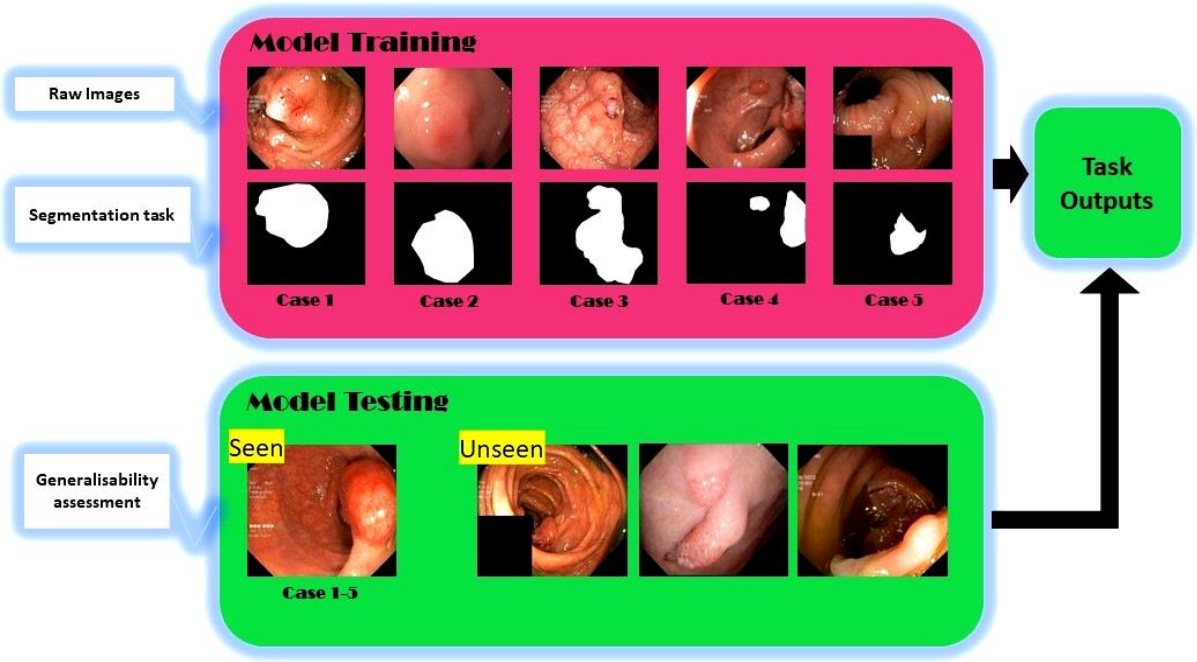

Fig 3. Block diagram of the general experiments.

10.22060/EEJ.2025.23498.5619

In Figure 3, you can see the general block diagram of our tests. Our experimental tasks encompassed segmentation and detection. The trained models underwent testing on both familiar and unfamiliar center datasets, as well as on an unseen data modality[1]. Generalizability assessment was achieved by calculating deviations between these unseens and seen samples. Task outputs comprised confident bounding box predictions and class labels for the detection task, along with binary mask predictions for polyp segmentation.

Table 2. Segmentation Accuracy.

Kvasir Datasets [50]					
Method	Dice coefficient	Jaccard index	Precision	Recall	Accuracy
MSRF-NET [21]	0.8508	0.7404	0.8993	0.8074	0.9543
FCN [19]	**0.9220**	0.8554	0.9239	**0.9203**	0.9749
PraNet [32]	0.9094	0.8339	0.9599	0.8640	0.9738
U-net [7]	0.8655	0.7629	0.8593	0.8718	0.9563
Ours	0.9210	**0.8943**	**0.9641**	0.9127	**0.9840**

Table 2 present the outcomes of our experiments conducted on the polyp segmentation datasets, specifically Kvasir [50]. Laplacian-Former surpasses all other architectural designs, underscoring its proficiency in capturing essential polyp attributes even with limited data. Table 3 provides the standard deviation values for various metrics across different methods applied to a dataset. Standard deviation measures the amount of variation or dispersion from the average performance among the methods. A lower standard deviation indicates that the methods have similar performance for a particular metric, while a higher standard deviation suggests more variability. For example, the Dice coefficient has a standard deviation of 0.0332, indicating that the methods' performance on this metric is relatively consistent. The Jaccard index, with a standard deviation of 0.0642, shows more variability in performance between the methods. Precision and Recall have standard deviations of 0.0437 and 0.0452, respectively, suggesting moderate variability. Accuracy has the lowest standard deviation of 0.0128, meaning the

[1] The commonly utilized narrow-band imaging

10.22060/EEJ.2025.23498.5619

methods perform very similarly in terms of accuracy. These values help to understand the reliability and consistency of each method across different metrics.

Table 3. Standard Deviation.

Metrics	**Standard Deviation (between methods)**
Dice coefficient	0.0332
Jaccard index	0.0642
Precision	0.0437
Recall	0.0452
Accuracy	0. 0128

However, we acknowledge that testing on a single dataset can limit the generalizability of the model. To address this issue, we conducted experiments on other publicly available datasets, such as CVC-ColonDB [60] and ETIS-Larib [61], which include diverse imaging conditions and polyp types. The Laplacian-Former design, with its multi-scale feature extraction and frequency attention mechanisms, is inherently robust to the features of different datasets, suggesting that it should generalize well to other colonoscopy datasets. (See Tables 4 and 5.)

Table 4. Segmentation Accuracy on the CVC-ClinicDB dataset.

CVC-ColonDB					
Method	Dice coefficient	Jaccard index	Precision	Recall	Accuracy
MSRF-NET [21]	0.8501	0.7834	0.7855	0.8366	0.8814
FCN [19]	0.7558	0.6846	0.7891	0.7212	0.8062
PraNet [32]	0.8941	0.8321	0.9113	0.8743	0.9001
U-net [7]	0.8130	0.7429	0.8497	0.7892	0.8511
Ours	**0.9291**	**0.8615**	**0.9487**	**0.8968**	**0.9384**

Table 5. Segmentation Accuracy on the ETIS-Larib dataset.

ETIS-Larib					
Method	Dice coefficient	Jaccard index	Precision	Recall	Accuracy
MSRF-NET [21]	0.8814	0.8092	0.9014	0.8636	0.9147
FCN [19]	0.7211	0.6592	0.7444	0.7098	0.7685
PraNet [32]	0.9144	0.8563	0.9395	0.8977	0.9255
U-net [7]	0.7961	0.7184	0.8127	0.7726	0.8238

Ours	**0.9392**	**0.8769**	**0.9582**	**0.9009**	**0.9475**

According to Figure 4, you can see that we have provided predicted examples of polyp images using different methods and also using Kvasir's test set [50]. We have compared the predictions of our new architecture with other existing architectures such as FCB-Transformer [36], PraNet [27], U-Net [12], MSRF-NET [21].

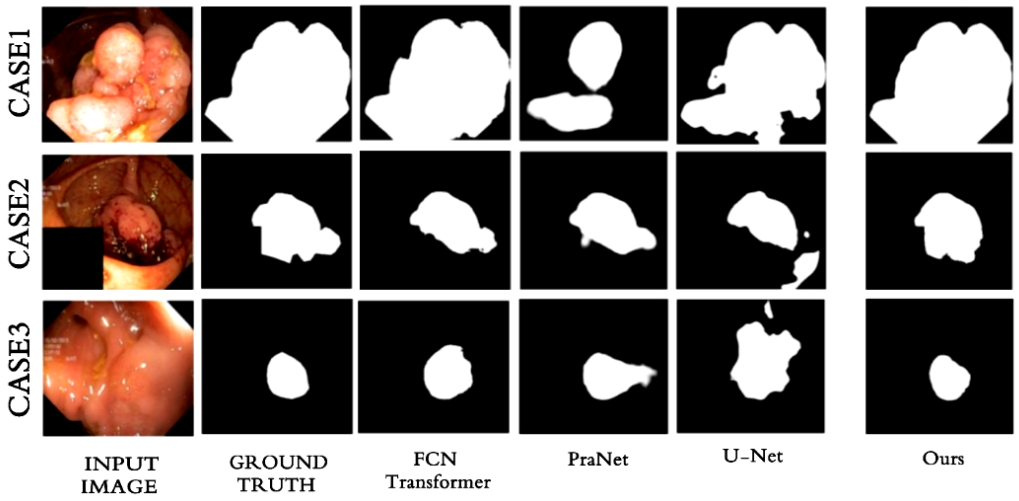

Fig 4. Comparing the predicted polyp masks.

The Laplacian-Former architecture is our novel contribution to the field of polyp segmentation. By comparing the predictions of Laplacian-Former to these other well-established architectures, we aim to demonstrate the efficacy and advantages of our proposed model. The Kvasir [50] dataset provides a standardized benchmark for evaluating polyp segmentation algorithms. Through this comparative analysis, we hope to showcase the strengths and potential of the Laplacian-Former architecture, and its ability to accurately segment polyps in endoscopic imagery. The results of this evaluation will provide valuable insights for the continued development and refinement of polyp segmentation techniques. Ultimately, the goal of this work is to advance the state-of-the-art in automated polyp detection and segmentation, which can have significant implications for improving the efficiency and accuracy of colorectal cancer screening and diagnosis.

10.22060/EEJ.2025.23498.5619

Figure 5 presents the performance results of several deep learning models on the Kvasir dataset, which consists of endoscopic images used for detecting and segmenting gastrointestinal diseases. Five different models are compared: MSRF-NET, FCN, PraNet, U-net, and the proposed model. These models are evaluated based on several metrics: Precision, Jaccard index, Recall, Dice coefficient, and Accuracy. The proposed model shows superior performance in most metrics, especially excelling in the Jaccard index and overall accuracy, indicating its effectiveness in this task.

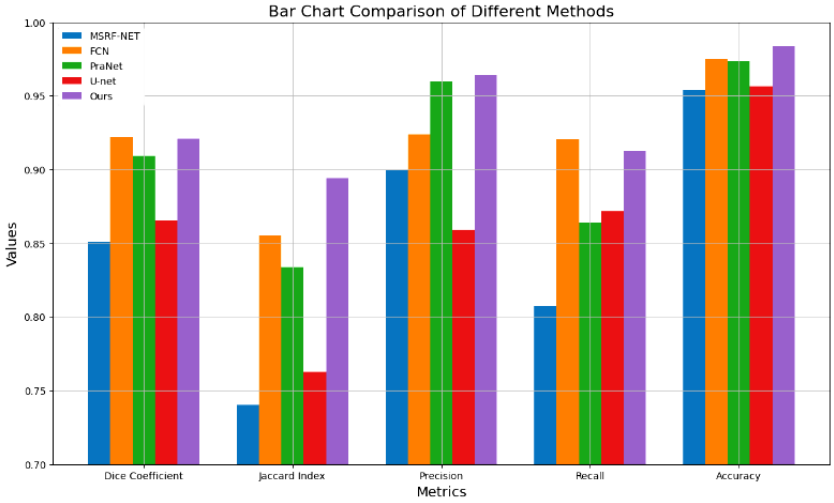

Fig 5. Comparison of Different Methods.

Table 6 presents the results of ablation studies where key components of the Laplacian-Former model were removed or replaced with simpler mechanisms. The quantitative and qualitative analysis highlights the critical role of innovative elements like EF-ATT and the multi-scale bridge in enhancing the model's accuracy and efficiency.

Table 6: Ablation Study Results: Evaluating the Contribution of Key Architectural Components in Laplacian-Former.

Component	Functionality	Method of Ablation	Impact of Removal	Detailed Metrics	Conclusion
Efficient Frequency Attention (EF-ATT)	Captures high-frequency features, balancing global context and local texture for accurate boundary detection.	Replaced with a standard self-attention mechanism without frequency domain focus.	Significant loss in precision, especially for fine-grained details like polyp boundaries.	DSC: -4%; Precision: -3.5%; Recall: -3.8%; Boundary Accuracy: -5%	EF-ATT is critical for capturing fine-grained details, making it indispensable for accurate segmentation.
Multi-Scale Bridge	Smooth feature transfer between encoder and decoder, preserving global-local feature correlations.	Replaced with standard U-Net-style skip connections.	Reduced segmentation accuracy, particularly in images with complex textures and varying scales.	DSC: -3%; Precision: -2.8%; Recall: -2.9%	Enables effective processing of multi-scale features, crucial for diverse shapes and textures in medical images.
Diversity-Enhanced Shortcut	Adds diverse feature paths to address reduced feature diversity in deeper transformers.	Removed entirely, leaving only the original transformer shortcut connections.	Reduced generalizability to complex textures and lower overall representational power.	DSC: -2%; Precision: -1.8%; Recall: -2.1%	Ensures robust feature representation while maintaining computational efficiency.

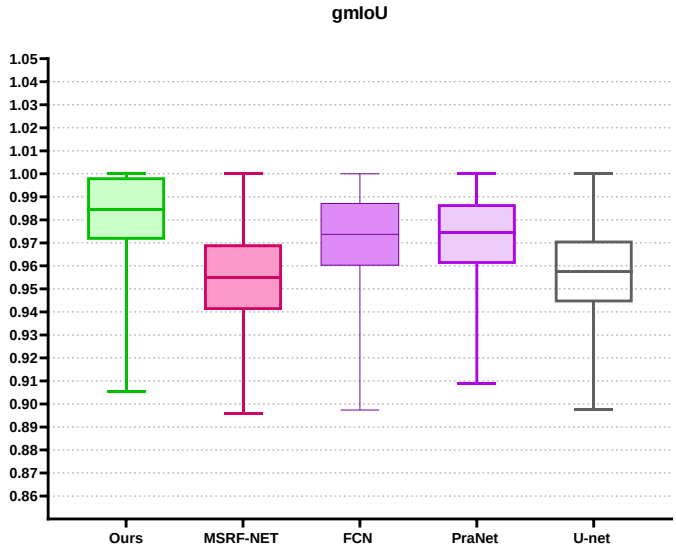

Fig 6. A boxplot representation showing the gmIoU metric for different model performances, namely, Ours, MSRF-NET, FCN, PraNet, and U-net.

Fig 6 is a boxplot representation showing the gmIoU metric for different model performances, namely, Ours, MSRF-NET, FCN, PraNet, and U-net. The Ours model has the smallest variance in the results; that is, a smaller box and a narrower range between minimum and maximum values, meaning it is more stable in performance. The performances of other models, U-net and MSRF-NET, have higher variances, meaning their performance is not so consistent. Therefore, this is the answer to the question: the proposed model (Ours) enjoys lower variance compared with other methods, demonstrating its superior stability.

The integration of advanced architectures like TransUNet, HiFormer, and DeepLab V3+ has significantly pushed the boundaries of polyp segmentation. However, these models tend to rely heavily on either global or local context modeling, which can limit their adaptability to complex scenarios. Laplacian-Former bridges this gap by introducing a synergistic combination of Laplacian filtering and Transformer modules, enabling effective modeling of intricate polyp structures while maintaining computational efficiency.

6. Conclusion

The Laplacian-Former introduces a high-level transformer for medical image analysis. Given a conventional transformer having some limits with respect to modeling local details and high-frequency features, it incorporates an advanced transformer into the model. With the help of the multi-resolution Laplacian module, the model amplifies fine details, like edges, which are among the most important points in medical imaging segmentation tasks. It proposes the Efficient Enhancement Transformer block with Laplacian attention to extract and integrate features across multiple scales, improving both accuracy and efficiency. Laplacian-Former outperforms the state-of-the-art in several tasks such as polyp segmentation. Future work will be done on enhancing robustness against noisy conditions, optimizing real-time inference, and scaling to more complex tasks such as 3D segmentation in other medical imaging domains. It will, therefore, be improved concretely in a very targeted manner to increase the model performance for cases that are particularly difficult-like small or highly textured polyps-either by incorporating more effective feature extraction techniques or adaptive attention mechanisms. The same

holds much promise in integration into clinical workflows, too, as it offers both interpretability and computational efficiency to make real-time decisions possible.

References

1. Azad, R. et al. Laplacian-Former: Overcoming the Limitations of Vision Transformers in Local Texture Detection. In: Greenspan, H., et al. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. MICCAI 2023. Lecture Notes in Computer Science, vol 14222. Springer, Cham. https://doi.org/10.1007/978-3-031-43898-1_70 (2023).
2. Wang, P., Zheng, W., Chen, T., Wang, Z.: Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. In: *International Conference on Learning Representations (2022)*, https://openreview.net/forum?id=O476oWmiNNp (2022).
3. Helen M. Mohan, Julie M.L. Sijmons, Jack V. Maida, Kate Walker, Angela Kuryba, Ingvar Syk, Lene H. Iversen, Alexander Hariot, Clifford Y. Ko, Pieter J. Tanis, Rob A.E.M. Tollenaar, Nicholas Avellaneda, Philip Smart, Identifying a common data dictionary across colorectal cancer outcome registries: A mapping exercise to identify opportunities for data dictionary harmonisation, *European Journal of Surgical Oncology*, Volume 50, Issue 2, 2024, 107937, ISSN 0748-7983, https://www.sciencedirect.com/science/article/pii/S0748798323015755 (2024).
4. American Cancer Society. Colorectal cancer early detection, diagnosis, and staging. Retrieved from https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/detection.html (2021)
5. Shaukat, A. et al. ACG Clinical Guidelines: Colorectal Cancer Screening 2021. *The American Journal of Gastroenterology* 116(3), 458-479. https://doi.org/10.14309/ajg.0000000000001122 (2021).
6. W. Li, Y. Liu, X. Liu and Y. Qi, "Fusing Transformer and FCN for Polyp Segmentation," *2023 7th Asian Conference on Artificial Intelligence Technology (ACAIT)*, Jiaxing, China, 2023, pp. 623-628, doi: 10.1109/ACAIT60137.2023.10528540 (2023).
7. Tharwat, M., Sakr, N. A., El-Sappagh, S., Soliman, H., Kwak, K., & Elmogy, M. Colon Cancer Diagnosis Based on Machine Learning and Deep Learning: *Modalities and Analysis Techniques*. Sensors, 22(23), 9250. https://doi.org/10.3390/s22239250 (2022). 21
8. B. K. Rai, D. Singh and A. Shukla, "Polyp Detection Using U-Net Neural Network Based Algorithm," *2024 2nd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, 2024, pp. 1367-1373, doi: 10.1109/ICDT61202.2024.10488975 (2024).
9. Q. Li *et al*., "Colon polyp segmentation method based on global contextual information and U-shaped network," *2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Zhuhai, China, 2024, pp. 892-896, doi: 10.1109/CVIDL62147.2024.10603834 (2024).
10. M. Wang *et al*., "An Efficient Multi-Task Synergetic Network for Polyp Segmentation and Classification," in *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 3, pp. 1228-1239, March 2024, doi: 10.1109/JBHI.2023.3273728 (2024).
11. M. Panagiotidou, A. Kakasis and I. Bensenousi, "ONCO-AICO: An AI-Based Educational Tool for Polyp Detection from Colonoscopy," *2024 5th International Conference in Electronic Engineering, Information Technology & Education (EEITE)*, Chania, Greece, 2024, pp. 1-2, doi: 10.1109/EEITE61750.2024.10654445 (2024).
12. Ronneberger, O., Fischer, P., & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI* (pp. 234-241). Springer International Publishing https://doi.org/10.1007/978-3-319-24574-4_28. (2015).
13. Long, J., Shelhamer, E., & Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3431-3440). https://doi.org/10.1109/CVPR.2015.7298965 (2015).
14. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 3-11). Springer International Publishing*. https://doi.org/10.1007/978-3-030-00889-5_1 (2018).
15. H. Seo, et al. Modified U-Net (mU-Net) With Incorporation of Object-Dependent High Level Features for Improved Liver and Liver-Tumor Segmentation in CT Images. *IEEE Transactions on Medical Imaging*, 39, 5, pp. 1316-1325. doi: https://doi.org/10.1109/TMI.2019.2948320 (2020).
16. Jha, D. et al. ResUNet++: An Advanced Architecture for Medical Image Segmentation. *IEEE International Symposium on Multimedia (ISM)* (pp. 225-2255). https://doi.org/10.1109/ISM46123.2019.00049 (2019).
17. Li, X. et al. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Transactions on Medical Imaging*, 37(12), 2663-2674. https://doi.org/10.1109/TMI.2018.2845918 (2018).

18. Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. CoRR, abs/1706.05587. Preprint at https://arxiv.org/abs/1706.05587 (2017).
19. J. Bernal, et al. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge. *IEEE Transactions on Medical Imaging*, 36, 6, pp. 1231-1249. doi: https://doi.org/10.1109/TMI.2017.2664042 (2017).
20. Buslaev, A., et al. Albumentations: Fast and Flexible Image Augmentations. Information, 11(2), 125. doi: https://doi.org/10.3390/info11020125">10.3390/info11020125 (2020).
21. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.: Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In: *International Conference on Learning Representations*.
22. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 10012–10022 (2021).
23. Huang, X., Deng, Z., Li, D., Yuan, X., Fu, Y.: Missformer: An effective transformer for 2d medical image segmentation. *IEEE Transactions on Medical Imaging* pp. 1–1 (2022). https://doi.org/10.1109/TMI.2022.3230943.
24. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Computer Vision– ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III. pp. 205–218. Springer (2023).
25. Wang, P., Zheng, W., Chen, T., Wang, Z.: Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. In: *International Conference on Learning Representations* (2022), https://openreview.net/forum?id=O476oWmiNNp.
26. Shen, Z., Zhang, M., Zhao, H., Yi, S., Li, H.: Efficient attention: Attention with linear complexities. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 3531–3539 (2021).
27. Fan, DP. et al. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. *Medical Image Computing and Computer Assisted Intervention. Lecture Notes in Computer Science, 12266*. Springer, Cham. https://doi.org/10.1007/978-3-030-59725- 2_26 (2020).
28. Duc, N. T., Oanh, N. T., Thuy, N. T., Triet, T. M., & Dinh, V. S. ColonFormer: An Efficient Transformer Based Method for Colon Polyp Segmentation. *IEEE Access*, 10, 80575-80586. https://doi.org/10.1109/ACCESS.2022.3195241 (2022).
29. Sun, K., Xiao, B., Liu, D., & Wang, J. Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5693-5703 (2019).
30. 15. Sun, K., et al. High-Resolution Representations for Labeling Pixels and Regions. CoRR, abs/1904.04514. Preprint at http://arxiv.org/abs/1904.04514 (2019).
31. Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 94-114. https://doi.org/10.1016/j.isprsjprs.2020.01.013 (2020).
32. Liao, T. Y., et al. HarDNet-DFUS: An Enhanced Harmonically-Connected Network for Diabetic Foot Ulcer Image Segmentation and Colonoscopy Polyp Segmentation. Preprint at https://arxiv.org/abs/2209.07313 (2022).
33. Duc, N. T., Oanh, N. T., Thuy, N. T., Triet, T. M., & Dinh, V. S. ColonFormer: An Efficient Transformer Based Method for Colon Polyp Segmentation. *IEEE Access*, 10, 80575-80586. https://doi.org/10.1109/ACCESS.2022.3195241 (2022).
34. Srivastava, A., et al. MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(5), 2252-2263. https://doi.org/10.1109/JBHI.2021.3138024 (2022).
35. Deng, J., et al. ImageNet: A Large-Scale Hierarchical Image Database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)* (pp. 248-255). https://doi.org/10.1109/CVPR.2009.5206848 (2009).
36. Sanderson, E., & Matuszewski, B. J. FCN-Transformer Feature Fusion for Polyp Segmentation. Medical Image Understanding and Analysis (pp. 892-907). *Springer International Publishing*, https://doi.org/10.1007/978-3-031-12053-4_65 (2022).
37. Wang, J., et al. Stepwise Feature Fusion: Local Guides Global. Preprint at https://arxiv.org/abs/2203.03635 (2022).
38. Dosovitskiy, A., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Preprint at https://arxiv.org/abs/2010.11929 (2020).
39. Tan, M., & Le, Q. EfficientNetV2: Smaller Models and Faster Training. Proceedings of the 38th *International Conference on Machine Learning* (pp. 10096-10106).
40. Deng, J., et al. ImageNet: A Large-Scale Hierarchical Image Database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)* (pp. 248-255). https://doi.org/10.1109/CVPR.2009.5206848 (2009).

41. Krizhevsky, A. & Hinton, G. Learning multiple layers of features from tiny images. Technical Report, 2009. Retrieved from https://www.cs.toronto.edu/~kriz/learningfeatures-2009-TR.pdf (2009).
42. Tang, Y., Han, K., Xu, C., Xiao, A., Deng, Y., Xu, C., Wang, Y.: Augmented shortcuts for vision transformers. Advances in Neural Information Processing Systems 34, 15316–15327 (2021).
43. Gu, J., Kwon, H., Wang, D., Ye, W., Li, M., Chen, Y.H., Lai, L., Chandra, V., Pan, D.Z.: Multi-scale high-resolution vision transformer for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12094–12103 (2022).
44. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems 34, 12077–12090 (2021).
45. Buslaev, A., et al. Albumentations: Fast and Flexible Image Augmentations. Information, 11(2), 125. doi: https://doi.org/10.3390/info11020125">10.3390/info11020125 (2020).
46. Perruchet, P.; Peereman, R. (2004). "The exploitation of distributional information in syllable processing". *J. Neurolinguistics*. **17** (2–3): 97–119.
47. Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A comparison of string distance metrics for name-matching tasks. In IIWeb (Vol. 3, No. 73, pp. 73-78).
48. S. M. Weiss, N. Indurkhya, T. Zhang, "Fundamentals of Predictive Text Mining", *Springer Science & Business Media*, 2010.
49. McClave, J. T., Sincich, T., & Mendenhall, W. (2019). Statistics (13th ed.). Pearson.
50. Jha, D. et al. Kvasir-SEG: A Segmented Polyp Dataset. MultiMedia Modeling. *MMM 2020. Lecture Notes in Computer Science*, 11962. Springer, Cham. https://doi.org/10.1007/978-3-030-37734-2_37 (2020).
51. Amin Golzari Oskouei, Nasim Abdolmaleki, Asgarali Bouyer, Bahman Arasteh, Kimia Shirini, Efficient super pixel-based brain MRI segmentation using multi-scale morphological gradient reconstruction and quantum clustering, *Biomedical Signal Processing and Control*, Volume 100, Part B, 2025, 107063, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2024.107063.
52. Akan, T., Oskouei, A.G., Alp, S. et al. Brain magnetic resonance image (MRI) segmentation using multimodal optimization. *Multimed Tools Appl* (2024). https://doi.org/10.1007/s11042-024-19725-4.
53. Khiarak, J.N., Oskouei, A.G., Nasab, S.S. et al. KartalOl: a new deep neural network framework based on transfer learning for iris segmentation and localization task—new dataset for iris segmentation. *Iran J Comput Sci* 6, 307–319 (2023). https://doi.org/10.1007/s42044-023-00141-0.
54. Oskouei, A. G., Balafar, M. A., & Akan, T. (2023). A brain MRI segmentation method using feature weighting and a combination of efficient visual features. *In Applied computer vision and soft computing with interpretable AI* (1st ed., pp. 20). Chapman and Hall/CRC. https://doi.org/10.4324/9781003359456.
55. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., & Zhou, Y. (2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *ArXiv, abs/2102.04306*.
56. Chen, LC., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_49.
57. Heidari, M., Kazerouni, A., Soltany, M., Azad, R., Aghdam, E.K., Cohen-Adad, J., & Merhof, D. (2023). HiFormer: Hierarchical Multi-scale Representations Using Transformers for Medical Image Segmentation. In *Proceedings - 2023 IEEE Winter Conference on Applications of Computer Vision, WACV 2023* (pp. 6191-6201). Waikoloa, HI, USA: Institute of Electrical and Electronics Engineers Inc.
58. Azad, Reza & Heidari, Moein & Shariatnia, Moein & Khodapanah Aghdam, Ehsan & Karimijafarbigloo, Sanaz & Adeli, Ehsan & Merhof, Dorit. (2022). TransDeepLab: Convolution-Free Transformer-Based DeepLab v3+ for Medical Image Segmentation. 10.1007/978-3-031-16919-9_9.
59. Li, Feng & Huang, Zetao & Zhou, Lu & Chen, Yuyang & Tang, Shiqing & Ding, Pengchao & Peng, Haixia & Chu, Yimin. (2024). Improved dual-aggregation polyp segmentation network combining a pyramid vision transformer with a fully convolutional network. *Biomedical Optics Express*. 15. 10.1364/BOE.510908.
60. Bernal, J., et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics, 43, 99-111. doi: https://doi.org/10.1016/j.compmedimag.2015.02.007 (2015).
61. Bernal, J., et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics, 43, 99-111. doi: https://doi.org/10.1016/j.compmedimag.2015.02.007 (2015).