A Multiscale Edge-Guided Polynomial Approximation Network for Medical Image Segmentation

Fuxian Sui
Shcool of Computer Science and Technology, Shandong Technology and Business University
Shandong, Yantai, China
s15006588612@163.com

Hua Wang
School of Information and Electrical Engineering, Ludong University
Shandong, Yantai, China
hwa229@163.com

Fan Zhang
Shcool of Computer Science and Technology, Shandong Technology and Business University
Shandong Future Intelligent Financial Engineering Laboratory
Shandong, Yantai, China
zhangfan51@sina.com

Abstract

As the core cornerstone of building an efficient medical care system, especially promoting accurate disease diagnosis and treatment, medical image segmentation is of great importance. However, medical segmentation faces many challenges, including complex background, shape and size changes, resulting in inaccurate or fuzzy segmentation boundaries. To meet these challenges, this paper proposes a multiscale edge-guided polynomial approximation network (AMEPANet). The well-designed edge guided bridge module in this paper uses the Laplacian operator to accurately capture and strengthen the edge information in the image, and realizes the robust preservation of edge information across multiple scales. At the same time, by building an information mixed attention mechanism, the network can further mine and use the subtle features of the boundary area to further improve the segmentation accuracy. In order to maximize the use of rich feature information at different scales and stages, this paper combines Kolmogorov–Arnold theorem to build an efficient decoder architecture, which can seamlessly integrate multi-source features to achieve comprehensive fusion and optimization of feature information. In addition, this paper also proposes an innovative C^1 continuous activation function, which shows significant advantages in reducing the fluctuation of model calculation and promoting the stable convergence of the model, and further enhances the comprehensive processing ability of the model for complex medical image features. Through extensive and in-depth experiments on multiple authoritative data sets such as Synapse, the excellent performance of AMEPANet has been verified.

Keywords: ***Segmentation Boundary Detection Piecewise Polynomial Curve Medical Image Polynomial Approximation.***

1. Introduction

In medical image segmentation, while traditional methods have established a strong foundation, their performance is often hindered by factors such as inconsistent image quality, complex anatomical structures, and the high variability of lesion areas, making accurate and robust segmentation challenging in dynamic clinical environments. In recent years, with the vigorous rise of deep learning technology, especially the extensive application of convolutional neural networks (CNN)[8, 56], it has brought revolutionary progress for medical image segmentation. Among them, UNet, with its unique jumping connection mechanism, has shown outstanding performance in fusing multi-scale features to generate high-resolution segmentation images, which has inspired the emergence of many variants (such as CE-Net[17]), which have made remarkable achievements in further improving segmentation accuracy.

However, the inherent limitations of CNN are also grad-

ually emerging, especially its convolution operation is limited by the local receptive field, and it is difficult to effectively capture the long-distance dependency between pixels in the image, which is particularly obvious when dealing with pathological areas with complex backgrounds, significant changes in shape and size. In addition, CNN is also insufficient in dealing with fuzzy boundaries, which limits its further application in complex medical image segmentation tasks. In order to overcome the above problems, researchers tried to expand the receptive field by introducing dilated convolutions (such as CPFNet[16]), and to strengthen key feature mapping by integrating attention mechanisms (such as BCDU-Net[2], Ms-red[13]), but these methods failed to fundamentally solve the problems of remote dependency and local detail capture.

significant breakthroughs have been made not only in image field [41, 51, 52], but also in other fields [24, 53, 61, 27, 34, 58]. With the great success of Transformer in the field of natural language processing, its powerful global modeling capability has attracted widespread attention[49, 60]. The proposal of Vision Transformer (ViT)[14] marks a new chapter of Transformer in image recognition and analysis. Researchers began to explore its potential applications in image processing and medical image segmentation. In order to improve computing efficiency and meet the processing requirements of high resolution medical images, hierarchical transformer architectures such as Swin Transformer[29] based on window attention came into being. Parallel MERIT[37] and other models are designed with dual transformer encoders, which increases the accuracy and model complexity at the same time. In addition, Pyramid Vision Transformer (PVT)[47] based on space reduction attention and TransDeeplab[3] integrating Transformer and DeepLab further expand the application boundary of Transformer in medical image segmentation. MISSFormer[19], DAEformer[1], Swin-Unet[7] and TransUNet[8] use different transformer blocks to replace the convolution part in UNet, and enhance the remote capture capability of the model. These methods alleviate the remote dependency problem to a certain extent by combining the global view of Transformer and the local feature extraction ability of CNN. However, the pure Transformer architecture still has shortcomings in capturing local context information of images, and it is difficult to accurately process the complex details in medical images.

In order to remedy this defect, PVTv2[48] and other models embed convolutional layers in the Transformer encoder, aiming to enhance the ability of local feature learning. At the same time, CASCADE and its variants (e.g., EMCAD [38], G-CASCADE [36], PVT-CASCADE [35]) enhance the model's ability to detect multi-scale targets in complex medical images by integrating attention-driven decoders and multi-scale processing techniques. However, although these improvements have enhanced the local learning and global modeling capabilities of the model to a certain extent, they still fail to meet the challenges of complex background, shape and size changes, and fuzzy boundaries in medical images. At the same time, in recent years, the trend of revealing the black box behavior of neural networks has attracted extensive attention. The interpretability of neural networks is crucial[30].

In view of this, this paper aims to propose an innovative medical image segmentation method, which deeply mines the complementary advantages of Transformer and CNN, and introduces a new mechanism to better deal with the diversity and uncertainty in complex medical images. Specifically, we propose an edge guided bridge, which uses the Laplacian operator to retain the edge information, so that the model can emphasize the complete edge information on various scales. Mixed attention is used in the encoder and decoder, combining the advantages of channel attention and spatial attention to improve attention scores and obtain more boundary information. Based on Kolmogorov Arnold theorem, an efficient decoder architecture is designed to integrate multi-source features and achieve comprehensive fusion and optimization of feature information. At the same time, a new C^1 continuous activation function is designed. In general, the contributions of this paper are as follows:

1.We designed Edge Guided Bridge (EGB) module. The EGB module robustly maintains edge information on multiple scales, and retains initial high-frequency information, especially boundary information, through the Laplacian operator to better fill the semantic gap between low-level features extracted by the encoder and high-level features generated by the decoder.

2.This paper proposes an information mixed attention module (IMAM), which skillfully combines the internal Q, K and V of the two attention mechanisms for hybrid operations, so that it can focus on multiple levels and different parts of the input data at the same time. Furthermore, combining Kolmogorov Arnold representation theorem, a high-performance decoder architecture is carefully constructed. The multi-source feature information from the encoder is integrated to realize the deep fusion and optimization of feature information.

3.In this paper, we address issues such as the existence of non-differentiable points in commonly used activation functions like LeakyReLU. We propose a novel C^1 continuous activation function, which reduces fluctuations in the model's computational process, leading to more stable convergence and better integration of different features.

4.Our proposed AMEPANet has demonstrated superior performance on multiple different types of datasets.

2. Related works

The UNet architecture is renowned for its effectiveness in medical segmentation and has been widely applied in fields such as organ segmentation and polyp segmentation. In this section, we will focus on introducing the relevant background of our proposed model, including boundary detection algorithms, Kolmogorov-Arnold and Vision Transformer.

2.1. Boundary detection algorithm

Boundary detection algorithms aim to precisely locate object boundaries or contours in the field of images, with widespread applications including image segmentation, edge enhancement, and various other computer vision tasks[57]. To improve boundary perception accuracy, research has focused on optimizing network training with novel loss functions, including advanced techniques like Boundary Loss [25] and HD Loss [23], which directly guide the network to enhance boundary discrimination. Additionally, some classical techniques continue to play important roles in capturing boundary information. The Sobel operator, as a classical gradient-based algorithm, relies on computing image pixel gradients to identify edges, estimating horizontal and vertical gradients through simple convolution and synthesizing an edge intensity map. However, its high sensitivity to image noise may lead to edge misjudgment. In contrast, the Canny edge detection algorithm follows a more complex multi-step process, including Gaussian smoothing, gradient computation, non-maximum suppression, and hysteresis thresholding, to achieve precise edge detection and connection, although this process is computationally intensive and highly parameter-dependent. The Prewitt operator, as a homologous technique to the Sobel operator, employs different convolution kernel configurations to estimate gradients and also faces the issue of false-positive edges. The Laplacian operator, by performing second-order derivative operations on the image, exhibits higher sensitivity to details and is adept at revealing subtle edges, but this may also increase sensitivity to image noise. As multi-task learning paradigms are increasingly applied in medical image segmentation, new approaches have been explored to integrate boundary detection as an auxiliary task, boosting the model's boundary perception abilities [46, 32]. Furthermore, advanced network structures dynamically highlight key boundary features through the integration of spatial attention mechanisms[50], further enhancing the expressiveness of boundary regions and demonstrating fine-grained processing of boundary information. In this context, our research takes a novel approach, aiming to integrate the essence of traditional boundary detection with insights from modern deep learning, proposing an Edge-Guided Bridge mechanism that robustly maintains and enhances image boundary information across multiple scales.

2.2. Kolmogorov-Arnold

Kolmogorov-Arnold theory[30] is based on multivariable continuous functions, and any multivariable continuous function f can be expressed as a combination of finite single variable continuous functions.

$$f(x) = f(x_1, x_2, ...x_n) = \sum_{q=1}^{2n+1} \psi_q (\sum_{p=1}^{n} \phi_{q,p}(x_p)) \quad (1)$$

Where $\phi_{q,p} : [0,1] \rightarrow \mathbb{R}$, $\psi_q : \mathbb{R} \rightarrow \mathbb{R}$. This theorem has laid a solid theoretical foundation for the construction of Kolmogorov-Arnold Network (KAN). Unlike traditional neural networks, which use fixed activation functions, KANs introduce learnable activation functions at the edge of the network. This design allows each weight parameter in the network to be replaced by a highly flexible single variable function, which is usually parameterized in the form of a spline function. This feature not only significantly improves the flexibility of the model, but also effectively simulates complex functional relationships by reducing the number of parameters, thus enhancing the interpretability and generalization ability of the model. In the development process of KANs, Azam et al.[4] conducted in-depth research on their effectiveness in visual tasks such as image recognition, which promoted the application of KANs in this field. Vaca et al.[43] further expanded the application scope of KANs, introduced them into the time series prediction and control problem, and verified the strong learning ability and wide applicability of KANs. Although KANs have made significant progress in theoretical research and preliminary application, their deep integration and application in the general vision network architecture are still insufficient, especially in the complex real vision tasks, which lack extensive practice and verification. In view of this, this paper explores and designs a general vision network architecture integrating KANs, aiming to give full play to KANs' advantages in flexibility and interpretability, and promote their innovative applications in visual recognition, analysis and broader computer vision tasks.

2.3. Transformer

The Transformer architecture and its various variants have emerged prominently in the domain of medical imaging analysis, particularly in segmentation, demonstrating extraordinary potential applications. In particular, ViT represents a groundbreaking advancement in visual data processing, challenging the dominant paradigm of traditional Convolutional Neural Networks. ViT revolutionizes the image processing pipeline by fully adopting self-attention mechanisms: it first divides the image into small patches and then processes these sequences through the

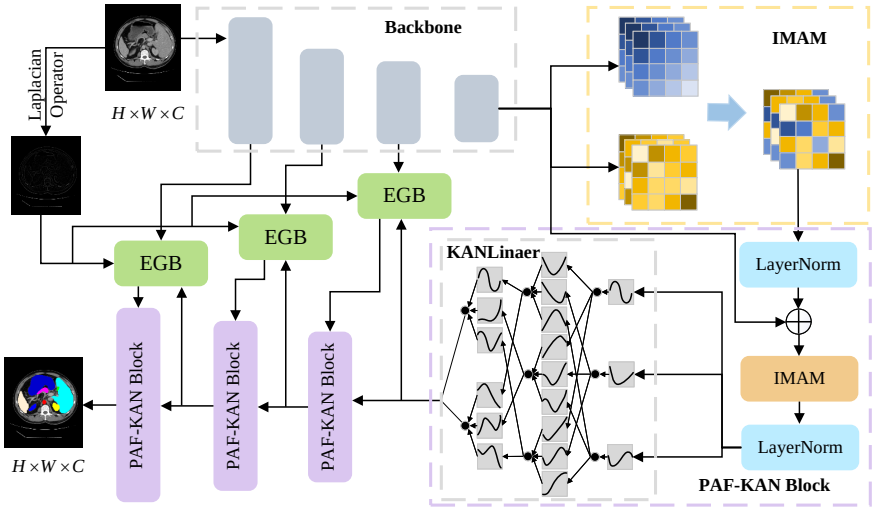


Figure 1. The main architecture of AMEPANet is shown in the figure.

Transformer architecture for in-depth analysis. This innovation not only competes with CNNs on core visual tasks such as object detection but in some scenarios surpasses them, redefining the paradigm of visual information processing. Building upon this foundation, TransUnet ingeniously integrates Transformer and Unet structures to broaden the applicability and efficacy of image segmentation tasks, showcasing the advantages of cross-architecture integration. However, the inherent high computational complexity of the original Transformer, especially the resource consumption of quadratic computations, hinders its broader applications. Addressing this challenge, Swin Transformer emerges with a core focus on introducing a sliding window attention mechanism, significantly reducing computational overhead by confining self-attention calculations within non-overlapping windows. This simultaneously ensures effective capturing of long-range dependencies, achieving dual optimization of efficiency and performance. Furthermore, MaxViT takes an important step in exploring a new dimension of fusion between Transformer and convolution, innovatively integrating self-attention mechanisms with convolutional elements to create a novel architectural unit. By integrating a streamlined multi-scale architecture, this framework empowers MaxViT to dynamically adjust to diverse vision challenges, demonstrating the efficacy of combined computational strategies in advancing visual perception systems.

3. Proposed Method

3.1. Network Architecture

The Vision Transformer (ViT) architecture has emerged as a prominent research focus within computer vision studies, with various ViT-derived methodologies demonstrating exceptional performance across multiple visual recognition tasks. MaxViT[42] effectively combines attention with convolution on the basis of ViT to generate a new architectural element, enhancing segmentation performance. Building upon MaxViT, we propose a multiscale edge-guided polynomial approximation network —AMEPANet to accurately segment medical images. The overall architecture of our model is illustrated in Figure 1.

3.2. Edge Guided Bridge

Currently, most methods[1, 8, 33] employ pooling or convolution in the encoder to downsample feature maps, reducing the amount of information to be processed. While this operation offers significant advantages in constructing deep architectures, it often leads to information loss as downsampling layers accumulate at deeper levels. LOD-Net[11] refines boundary prediction by learning adjustable directional derivatives for each pixel, selecting large-derivative pixels as boundary candidates, and combining high-level semantic features. BATFormer[28] generates boundary-aware windows through entropy-based adaptive windowing, applying local self-attention within these windows to preserve boundary details. CSAP-UNet[15] strengthens edge features by introducing a boundary enhancement module (BEM) in the shallow layers, integrating local and global features via an attention fusion module. These methods still have some limitations in preserving edge information and integrating boundary and background features. To better address this issue and robustly preserve edge information at multiple scales, we propose an Edge-Guided Bridge (EGB), as shown in the Figure 2. At the ith layer, this module takes three inputs: the features X_i^e from

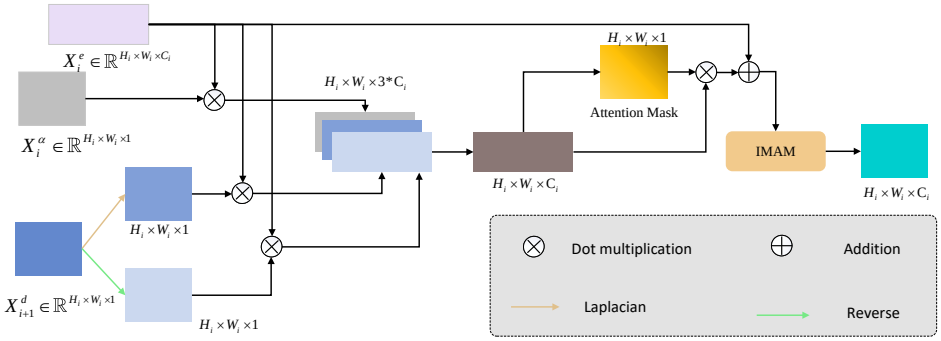


Figure 2. Edge Guided Bridge takes feature X_i^e from the encoder, edge information X_i^α extracted by Laplacian operator and more advanced prediction feature X_{i+1}^d as input.

the encoder, the high-frequency edge features X_i^α obtained through the Laplacian operator, and the more advanced features $X_{i+1}^d \in \mathbb{R}^{H_i \times W_i \times 1}$ from the decoder. To simplify calculations and reduce channel numbers, we apply a convolution to the feature maps at each layer, obtaining the final input $X_i^e \in \mathbb{R}^{H_i \times W_i \times C_i}$ for the EGB module.

We utilize Laplacian pyramid, an effective technique for preserving image edge information. The Laplacian operator is a gradient-based second-order derivative operator that can detect finer edges and is more sensitive to details. In practical applications, we initially smooth the original image using a Gaussian filter, i.e., employing the Laplacian of Gaussian function. The Laplacian pyramid encompasses crucial low-level details at different scales:

$$\begin{aligned} P_k &= P, \ if \ k = 0 \\ P_k &= d\left(g\left(P_{k-1}\right)\right), \ if \ k \geq 1 \\ G_k &= P_k - \ \mu(P_{k+1}) \end{aligned} \tag{2}$$

Where, P represents the input image, g represents the Gaussian filtering convolution operator, d represents downsampling by a factor of 2, G_k represents the k-th level of the Laplacian pyramid, and μ represents the corresponding upsampling operation. The Laplacian operator detects second-order variations in the image, such as edges, contours, and other high-frequency details, which are crucial for medical segmentation. Therefore, the feature $X_i^\alpha \in \mathbb{R}^{H_i \times W_i \times 1}$ from the i-th layer of the Laplacian pyramid is provided to the i-th layer of the EGB module. The calculation formula for X_i^α is as follows:

$$\begin{aligned} X_0^\alpha &= G_1 \\ X_i^\alpha &= \left(d(X_0^\alpha)\right)^i, \ if \ i \geq 1 \end{aligned} \tag{3}$$

Where $(d(X))^i$represents downsampling X by a factor of 2 for i times. Inspired by[10], we decompose the advanced features generated by the decoder into two different attention maps, X_{i+1}^{d1} and X_{i+1}^{d2}. We then perform element-wise multiplication and concatenation of X_i^e and X_i^α with these two features as shown in Figure 2. Finally, after passing through the attention module, we extract the feature relationships between the background and boundary regions.

3.3. Piecewise Activation Function

Activation functions play a crucial role in improving the accuracy of the model. In our model, LeakyReLU demonstrates a significant advantage over activation functions such as Sigmoid, Tanh, and ReLU in enhancing model accuracy. However, LeakyReLU is C^0 continuous, and its lack of C^1 continuity causes discontinuity during backpropagation, leading to fluctuations in computations and affecting the stability of model convergence. To address this issue, we propose an activation function, PAF (x), constructed from piecewise polynomial curves, as shown in Figure 3. The design principles of PAF (x) are as follows: it is similar in shape to LeakyReLU, it is C^1 continuous, and it consists of four segments of curves. Specifically, in the intervals $(-\infty, -1)$ and $(1, +\infty)$, it consists of two segments of linear polynomial functions, while in the intervals $[-1, 0)$ and $(0, 1]$, it consists of two segments of cubic polynomial functions. First, let's discuss the construction in the intervals $[-1, 0)$ and $(0, 1]$. From this, we can construct Hermite interpolation functions $P_1(x)$ and $P_2(x)$ over the intervals $[x_i, x_{i+1}]$, where i=1,2:

$$\begin{aligned} P_i(x) &= m_0(x)F_i + d_0(x)\frac{dF_i}{dx} + m_1(x)F_{i+1} + d_1(x)\frac{dF_{i+1}}{dx} \\ m_0(x) &= \left(x_{i+1} - x\right)^2 \left(2\left(x - x_i\right) + h\right) / h^3 \\ m_1(x) &= \left(x - x_i\right)^2 \left(2(x_{i+1} - x) + h\right) / h^3 \\ d_0(x) &= \left(x_{i+1} - x\right)^2 (x - x_i) / h^2 \\ d_1(x) &= -(x - x_i)^2 \left(x_{i+1} - x\right) / h^2 \\ h &= x_{i+1} - x_i \end{aligned} \tag{4}$$

Based on the similarity with LeakyReLU, the function values of P(x) at points $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$ are

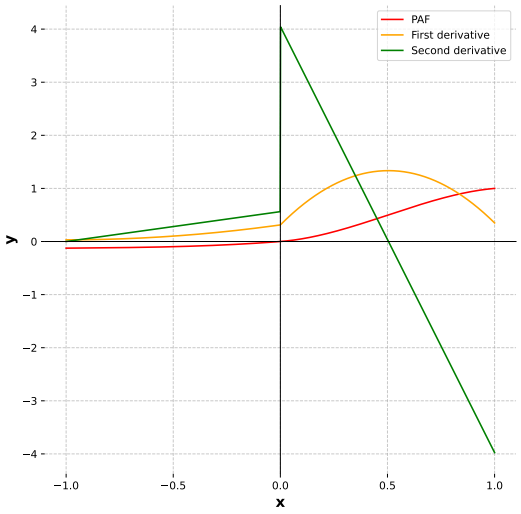


Figure 3. We use curves of different colors to represent PAF and its first and second derivatives.

$F_1 = -0.125$, $F_2 = 0$, and $F_3 = 1$, respectively. Then, based on the design by interaction, the first derivatives of PAF(x) at x_1, x_2 and x_3 are set to $\frac{F_1}{dx} = 0.0313$, $\frac{F_2}{dx} = 0.3125$, and $\frac{F_3}{dx} = 0.35$. From equations (4), $P_1(x)$ and $P_2(x)$ are uniquely determined. To reduce computational complexity, we express $P_1(x)$ and $P_2(x)$ in terms of power series:

$$\begin{aligned} P_1(x) &= ((0.0938x + 0.2813)x + 0.3125)x \\ P_2(x) &= ((-1.3375x + 2.025)x + 0.3125)x \end{aligned} \tag{5}$$

Let's discuss the construction of the linear function $P_0(x)$ on the interval $(-\infty, -1)$. The function value and the first derivative of $P_1(x)$ at $x = -1$ are -0.125 and 0.0313, respectively. Since $P_0(x)$ and $P_1(x)$ are C^1 continuous at $x = -1$, we have $P_0(x) = 0.0313(x + 1) - 0.125$. Simplifying this, we get $P_0(x) = 0.0313x - 0.0937$. Similarly, by ensuring the C^1 continuity of $P_2(x)$ and $P_3(x)$ at $x = 1$, we can obtain:

$$P_3(x) = 0.35x + 0.65 \tag{6}$$

Therefore, PAF(x) can be defined as:

$$PAF(x) = \begin{cases} P_0(x), & x < -1, \\ P_1(x), & -1 \le x < 0, \\ P_2(x), & 0 \le x \le 1, \\ P_3(x), & x > 1. \end{cases} \tag{7}$$

3.4. PAF-KAN Block

A significant amount of research has found that both spatial attention and channel attention play important roles in the field of segmentation. However, we have observed that the current usage of these two attention mechanisms is limited to simple serial or parallel implementations[13, 33], which may not fully exploit the advantages of both. Therefore, we propose an Information Mixed Attention Module, as shown in Figure 4. This module is based on the cross mixing of spatial attention and channel attention, which makes spatial attention and channel attention cross complement each other, and further excavates and utilizes the subtle features of boundary areas to further improve the accuracy of attention.

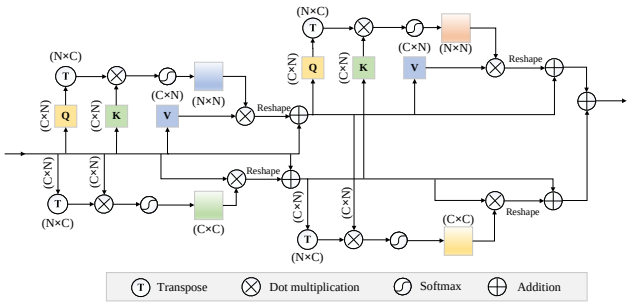


Figure 4. The architecture of Information Mixed Attention Module.

This module takes an input $Y \in \mathbb{R}^{H\times W\times C}$. First, spatial attention and channel attention are applied separately to the input data Y, resulting in $Y^s \in \mathbb{R}^{H\times W\times C}$ and $Y^c \in \mathbb{R}^{H\times W\times C}$. Y^s is chosen as the anchor data, and Y^c as the complementary data. The following formulas were used to calculate Q, K, and V:

$$\begin{aligned} Q^s &= Y^s \cdot P_1^Q \\ K^c &= Y^c \cdot P_1^K \\ V^s &= Y^s \cdot P_1^V \end{aligned} \tag{8}$$

where P_1^Q, P_1^K, $P_1^V \in \mathbb{R}^{C\times D}$ represent the learned weight matrices, where D is the number of channels after the linear transformation. The mixed attention was calculated using the following formula:

$$SA(Q^s, K^c, V^s) = SoftMax(Q^s \cdot (K^c)^T/\sqrt{D})\cdot V^s \tag{9}$$

where SoftMax represents the row-wise SoftMax operation. Similarly, X^c as the anchor data and X^s were chosen as the complementary data, and another set of Q, K, and V was calculated:

$$\begin{aligned} Q^c &= X^c \cdot P_2^Q \\ K^s &= X^s \cdot P_2^K \\ V^c &= X^c \cdot P_2^V \end{aligned} \tag{10}$$

The mixed attention for this instance can be calculated using the following formula:

$$CA(Q^c, K^s, V^c) = SoftMax(Q^c \cdot (K^s)^T/\sqrt{D})\cdot V^c \tag{11}$$

Therefore, the global-local cross-mix attention module can be represented as follows:

$$IMAM = SA(Q^s, K^c, V^s) + CA(Q^c, K^s, V^c) \quad (12)$$

In the segmentation task, effective use of multi-source information is crucial to improve the recognition and segmentation performance of various size objects. Multi-scale information fusion strategy has been proved to be the key to improve performance[54, 41, 45, 55]. For this reason, this paper uses the information mixed attention module to design an efficient decoder architecture, integrate multi-source features, achieve comprehensive fusion and optimization of feature information, and further enhance the model's ability to understand and represent complex scenes. Specifically, the decoding architecture accepts data $t1$ and $t2$ from different processing stages as inputs. Except for the first layer encoder, $t1$ in other layers is the edge information output from the edge guidance bridge, and $t2$ is the high-level semantic information from the upper layer decoder. First, the high-level semantic information $t2$ is normalized,

$$t2 = LayerNorm(t2) \quad (13)$$

Then it is spliced with the edge information t1, and IMAM is used to mine and utilize features in a deeper level, and normalization processing is carried out.

$$t = LayerNorm(IMAM(t1 + t2)) \quad (14)$$

When extracting key information, we innovatively use KAN to calculate, using the following formula definition.

$$\phi(t) = \omega(b(t) + spline(t)) \quad (15)$$

$$b(t) = PAF(t) \quad (16)$$

$$spline(t) = \sum_n c_n B_n(t) \quad (17)$$

Where c_n is a trainable parameter, ω is a constant, $B_n(x)$ is a B-spline, and PAF is the activation function above. The flexibility of spline enables it to adaptively model complex relationships in data by adjusting the shape, make full use of spline to optimize the extracted features, and enhance the learning ability of subtle features.

3.5. Loss function

We use a combination of binary cross-entropy (BCE) loss and Dice loss as the loss function for network training. The definitions of BCE loss and Dice loss are:

$$\begin{aligned} \mathcal{L}_{BCE} &= -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log\,(p_i) + (1 - y_i) \cdot log\,(1 - p_i) \\ Dice &= \frac{2\sum_{i=1}^{N} y_i \cdot p_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \epsilon} \\ \mathcal{L}_{Dice} &= 1 - Dice \end{aligned} \quad (18)$$

Where, $p_i \in 0, 1$ represents the predicted value of the model, $y_i \in 0, 1$ represents the true label, and $N = W * H$ represents the number of pixels.

4. Experiments

4.1. Datasets

Synapse Multi-Organ Segmentation: First, we use the Synapse Multi-organ Dataset [26] to evaluate the performance of our method. This dataset comprises 30 abdominal CT scans consisting of a total of 3779 axial contrast-enhanced abdominal CT images. Each CT scan contains 85 to 198 slices with a size of 512×512 pixels. The voxel spacing is ([0:54-0:54]×[0:98-0:98]×[2:55:0]) mm^3. Our evaluation follows the settings outlined in [8], where we segment eight anatomical structures including the gallbladder (Gal), aorta (Aor), left kidney (LK), liver (Liv), right kidney (RK), stomach (Sto), pancreas (Pan) and spleen (Spl).

Automated Cardiac Diagnosis Challenge (ACDC): The ACDC dataset [6] is designed for automatic cardiac diagnosis and consists of 100 cardiac MRI scans, each containing three cardiac structures: the myocardium (Myo), right ventricle (RV) and left ventricle (LV). In line with the methodology presented in [38], we use 20 for testing, 20 samples for validation, and 70 for training.

Polyp Segmentation: To better validate the generalization ability of our proposed model, we further evaluated it on a polyp segmentation dataset. For fair comparison, we followed the same experimental settings as [59],selecting 1450 images from the Kvasir[22] and CVC-ClinicDB[5] datasets as the training set, and testing on the EITS and CVC-300 datasets[40].

Skin Lesion Image Segmentation: The ISIC2018 dataset [12] is a public collection for skin lesion segmentation, consisting of 2594 skin lesion images with varying resolutions. The PH2 dataset [31] contains 200 color images of skin lesions. For fair model comparison, in line with previous work [13], both datasets are resampled to 224×320 pixels. PH2 is split into 80 training images, 100 test images, and 20 validation images, while The ISIC2018 dataset is divided into training, test sets, and validation in a 7:2:1 ratio.

4.2. Implementation Details

The network uses AdamW, with an initial learning rate of 0.001, momentum of 0.001, and weight attenuation of 0.0001. The batch size is set to 12. For CT datasets such as Synapse and ACDC, we train 400 epochs. For polyp segmentation datasets and skin lesion image segmentation-datasets, we train 200 epochs to prevent network overfitting. All experiments were conducted on a single NVIDIA A100.

Table 1. We compare AMEPANet with several competing models on the Synapse dataset, emphasizing the best-performing results in bold.

Methods	Spl	RK	LK	Gal	Liv	Sto	Aor	Pan	Average	
									DSC	HD95
UNet [39]	81.48	62.64	72.41	56.70	86.98	67.96	84.00	48.73	70.11	44.69
TransUNet [8]	85.08	77.02	81.87	63.16	94.08	75.62	87.23	55.86	77.49	31.69
Swin-UNet [7]	90.66	79.61	83.28	66.53	94.29	76.60	85.47	56.58	79.13	21.55
TransDeeplab [3]	89.00	79.88	84.08	69.16	93.53	78.4	86.04	61.19	80.16	21.25
MISSFormer [20]	91.92	82.00	85.21	68.65	94.41	80.81	86.99	65.67	81.96	18.20
HiFormer [18]	90.99	79.77	85.23	65.23	94.61	81.08	86.21	59.52	80.39	14.70
DAEFormer [1]	91.82	82.39	87.66	71.65	95.08	80.77	87.84	63.93	82.63	16.39
PVT-CASCADE [35]	90.10	80.37	82.23	70.59	94.08	83.69	83.01	64.43	81.06	20.23
TransUnet [9]	88.14	69.73	81.11	76.95	93.64	77.84	88.01	61.22	79.58	28.73
Parallel MERIT [37]	91.21	84.31	87.21	73.48	95.06	84.15	88.38	69.97	84.22	16.51
EMCAD [38]	**92.17**	84.10	88.08	68.87	95.26	83.92	88.14	68.51	83.63	15.68
G-CASCADE [36]	90.52	82.38	85.64	74.86	**95.33**	83.65	88.27	**71.99**	84.08	18.89
Ours	91.19	**85.61**	**89.29**	**77.13**	95.16	**85.99**	**88.52**	70.79	**85.55**	**14.39**

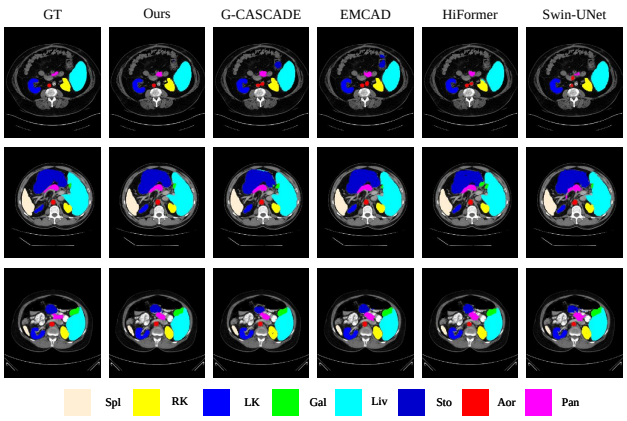


Figure 5. A qualitative comparison of various visualization methods on the Synapse dataset is shown. Our approach demonstrates fewer false predictions and preserves more detailed information.

4.3. Performance Comparisons

Synapse Dataset: Synapse Dataset: Table 1 presents a comprehensive comparison of the proposed method's performance with other SOTA approaches. The results show that our method significantly surpasses the previous SOTA method. Specifically, compared with the EMCAD method, our method achieves 1.92% performance improvement; Compared with G-CASCADE method, it shows an advantage of 1.47%. In the task of segmentation of specific regions such as kidney, aorta and stomach, our method has also achieved remarkable results. It is particularly worth mentioning that our method has achieved a performance leap of up to 2.27% compared with the second ranking method for small and difficult to accurately segment pancreatic regions, which marks a critical step in achieving more accurate segmentation results. The qualitative comparison results between different methods are shown in Figure 5.

Polyp Segmentation: The comparison results of this method and other SOTA methods in polyp segmentation dataset are shown in Table 3. It is worth noting that our method is superior to the competitive method in all indicators of ETIS and CVC-300 data sets. This advantage observed on different data sets emphasizes the powerful generalization ability of this method. The qualitative comparison of results is shown in Figure 6. It can be seen that our method has excellent performance in maintaining a low false positive rate, which accurately avoids mistakenly classifying healthy areas as tumors.

ACDC Dataset: Table 2 shows the DICE scores of our method and other SOTA methods for heart organ segmentation on MRI images of ACDC dataset. Our method achieves the highest average DICE score of 92.68%, which is about 0.45% higher than that of G-CASCADE. At the same time, the segmentation of the left ventricle, the right ventricle and the myocardium achieves about the same result.

Skin Lesion Segmentation: The data presented in Table 3 highlights the excellent performance of our method in PH2 and ISIC2018 skin lesion segmentation data sets, especially in PH2 data sets. Our method outperforms other competitive methods in all evaluation indicators. Specifically, we achieved 90.37% of the DICE score on the ISIC2018 dataset, 0.8% higher than the nearest competitive method G-CASCADE; In another data set, the DICE score reached 94.82%, 0.42% ahead of G-CASCADE. In

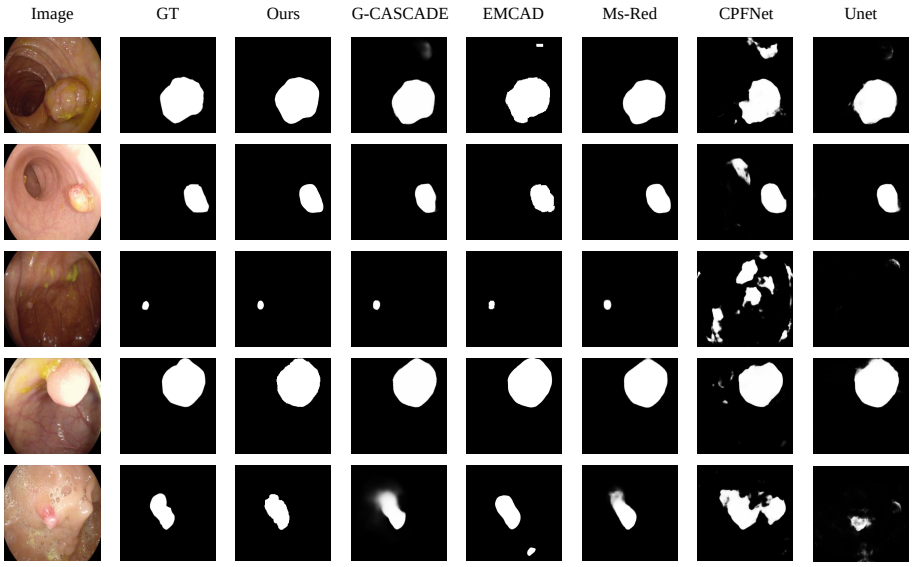


Figure 6. Qualitative comparison of different visualization methods on polyp segmentation dataset. It can be seen that our method can accurately locate the lesion area and achieve more accurate segmentation.

Table 2. Our Method compared with various competing models on the ACDC dataset.

Methods	Avg Dice	RV	Myo	LV
UNet [39]	87.55	87.10	80.63	94.92
TransUNet [8]	89.71	86.67	87.27	95.18
nnU-Net [21]	92.34	90.67	90.30	96.04
MISSFormer [20]	90.86	89.55	88.04	94.99
Swin-UNet [7]	88.07	85.77	84.42	94.03
MT-UNet [44]	90.43	86.64	89.04	95.62
BATFormer [28]	91.14	87.99	89.48	95.97
PVT-CASCADE [35]	91.46	89.97	88.90	95.50
TransCASCADE [35]	91.63	90.25	89.14	95.50
nnFormer [62]	91.78	90.22	89.53	95.59
Parallel MERIT [37]	92.32	90.87	90.00	96.08
EMCAD [38]	92.14	90.65	89.68	96.02
G-CASCADE [36]	92.23	90.64	89.96	96.08
Ours	**92.68**	**91.64**	**90.33**	**96.08**

addition, in terms of accuracy, our method also showed significant advantages. Compared with G-CASCADE, it improved 0.31% on ISIC2018 dataset and achieved 1.22% growth on PH2 dataset. Figure 7 visualizes the results of our method alongside other competing approaches on both datasets. The figure clearly shows that our method exhibits superior generalization performance compared to the others on both datasets.

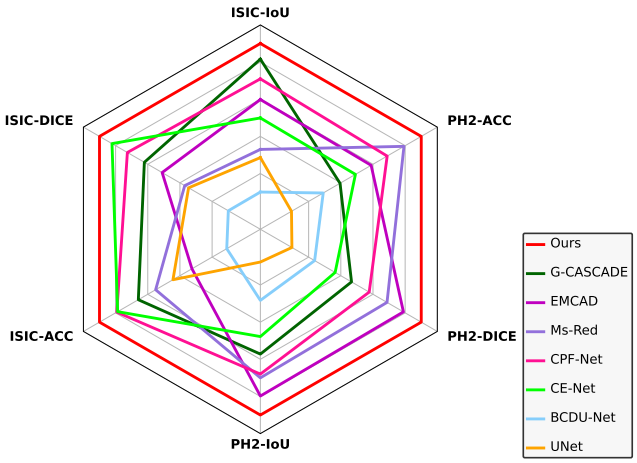


Figure 7. Visual comparison of indicators between our method and other competitive methods on ISIC2018 and PH2.

4.4. Ablation Study

To evaluate the effectiveness of each module in our proposed network, we conducted a large number of ablation experiments. We referred to the network without the EGB, IMAM, and PAF modules as the baseline model.

Qualitative Study: Figure 8 illustrates the advantages of the baseline model and the other three models mentioned by us on the Synapse dataset. It can be clearly seen from the figure that the organic combination of various modules effectively improves the segmentation performance, can more accurately locate the organ position, and make the segmentation boundary more accurate and clear.

Quantitative Study: Table 4 shows the experimental re-

Table 3. The comparison results between our method and other competitive methods on polyp segmentation dataset and skin lesion segmentation dataset.

Methods	ETIS			CVC-300			ISIC2018				PH2			
	DSC	IoU	MAE	DSC	IoU	MAE	IOU	DICE	ACC	Recall	IOU	DICE	ACC	Recall
UNet [39]	39.80	33.50	3.60	71.00	62.70	2.20	81.69	88.81	95.68	88.58	87.07	92.62	95.57	92.86
BCDU-Net [2]	64.71	26.72	3.10	74.01	42.86	2.05	80.84	88.33	95.48	89.13	87.41	93.06	95.61	91.11
CE-Net [17]	65.65	32.62	13.16	85.10	61.57	2.79	82.82	89.59	95.97	90.54	89.62	94.36	96.68	94.51
CPF-Net [16]	77.75	50.82	4.32	88.08	69.41	1.55	82.92	89.63	96.02	90.62	89.91	94.52	96.72	93.74
Ms-Red [13]	73.95	44.17	4.72	90.45	73.83	1.77	82.13	89.05	95.71	90.82	90.14	94.65	96.80	94.73
PVT-CASCADE [35]	79.28	68.43	2.03	88.34	82.30	0.84	82.83	90.21	95.62	92.38	89.87	94.51	96.30	95.91
Parallel MERIT [37]	69.73	62.82	1.74	88.21	81.23	1.12	83.44	90.05	95.66	91.92	89.70	94.30	96.81	95.43
EMCAD [38]	80.68	68.53	1.72	88.71	81.33	0.83	82.81	89.46	95.55	**93.81**	90.25	94.69	96.69	96.11
G-CASCADE [36]	78.13	67.34	2.24	88.15	80.69	1.42	82.98	89.57	95.74	93.27	89.69	94.40	95.69	94.93
Ours	**81.51**	**68.82**	**1.70**	**90.92**	**83.34**	**0.61**	**83.82**	**90.37**	**96.05**	93.25	**90.41**	**94.82**	**96.91**	**96.24**

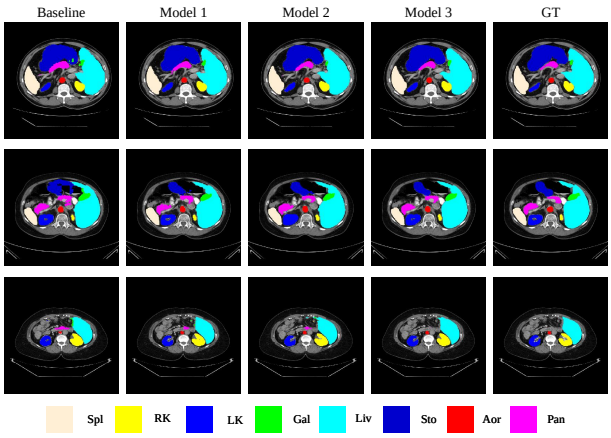


Figure 8. Visual comparison of segmentation by different models on the Synapse dataset, where Model 1 represents Baseline + IMAM; Model 2 represents Baseline + IMAM + EGB; Model 3 represents Baseline + IMAM + EGB + PAF.

Table 4. Research on ablation of each module on Synapse dataset.

Network			Dice	HD95
IMAM	EGB	PAF		
×	×	×	82.22	18.82
✓	×	×	82.59	17.98
×	✓	×	83.74	16.99
×	×	✓	83.22	17.87
✓	✓	×	84.27	15.71
✓	×	✓	84.17	16.83
×	✓	✓	84.36	16.91
✓	✓	✓	85.55	14.39

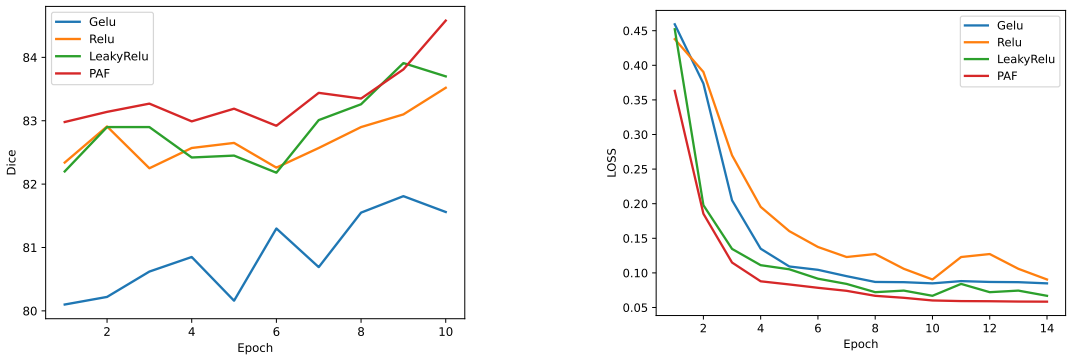
sults of the detailed quantitative study of the baseline model and the three modules proposed by us. In order to ensure that the effectiveness of our proposed method is fully verified, we have implemented a thorough experimental strategy, including the independent application of these three modules and the pairwise combination between them. From the experimental data in the table, it can be seen that the proposed method can significantly improve the segmentation accuracy of the model, whether used alone or in combination. Finally, compared with Baseline, Dice increased by 3.33% and HD95 decreased by 4.43%, significantly improving the performance of the entire model. Table 5 shows the study on parameters and time efficiency.

Table 5. Research on parameters and time efficiency.

Network		Params(M)	Flops(G)	Infer-Times(ms)
IMAM	EGB			
×	×	86.92	17.88	49.69
✓	×	87.95	18.17	51.64
✓	✓	93.21	20.95	57.30

Significance Analysis: We conducted five experiments with two models: one using the method proposed in this paper and the other without it. The five sets of DSC data obtained were [85.53, 85.50, 85.48, 85.55, 85.52] for the model using the proposed method, and [82.10, 81.95, 82.00, 82.22, 82.15] for the model without it. We performed an analysis of variance (ANOVA) on these two groups of data, calculating an F-statistic of 4615.45 and a p-value of 2.45×10^{-12}. The p-value is far smaller than the significance level (0.05), which means we can reject the null hypothesis. The null hypothesis suggests there is no difference between the groups, but the results indicate that the performance of the model with the proposed method is significantly different from that of the model without it. The extremely small p-value further supports that the effectiveness of the proposed model is highly reliable and stable.

Activation Function: Our proposed PAF is a novel C^1 continuous activation function. To fully validate its per-



(a) (b)

Figure 9. The figure (a) shows the impact of our proposed PAF and several other commonly used activation functions on model training. Figure (b) shows the impact of several methods on the loss during the training process. It can be clearly seen that our PAF can better reduce the fluctuation in the model training process.

Table 6. A comparative study of our proposed PAF and several commonly used activation functions on the Synapse dataset.

Activation Function				Dice	HD95
Relu	LeakyRelu	Gelu	PAF		
✓	×	×	×	83.57	16.33
×	✓	×	×	83.91	15.55
×	×	✓	×	81.81	31.23
×	×	×	✓	85.55	14.39

formance, we compared it with activation functions like LeakyRelu, Relu, Gelu, etc. As shown in Table 6, it is evident that our PAF achieves better segmentation performance. As can be seen from Figure 9, compared with these commonly used activation functions, PAF can better reduce the fluctuation of the model, make the training process more stable, and improve the segmentation performance better.

Research on ablation within the IMAM: Research on ablation within the IMAM module: In the exploration of the internal mechanism of IMAM, we innovatively combined two different types of attention mechanisms for hybrid computing. In order to comprehensively evaluate the effectiveness of this design decision, we carefully designed a group of experiments to compare the performance of different attention combinations. The experimental results are shown in Table 7, from which we can intuitively see that the IMAM design proposed in this paper shows the most outstanding performance compared with other architecture configurations, fully verifying the correctness and superiority of our design ideas.

Comparison with Different Backbone Networks: To validate our choice, we expanded ResNet101 and ResNet50 networks to have a comparable number of parameters to MaxVit and applied the same pre-training regimen. We then replaced the backbone with these networks. The results in

Table 7. Research on ablation within the IMAM on the Synapse dataset.

Methods	Attention Parallel	Attention Serial	Ours
Dice	83.64	83.94	85.55
HD95	19.34	17.46	14.39

Table 8. An analysis of the effect of different backbones on the network performance.

Backbone	ResNet50	ResNet101	Maxvit
Dice	77.39	82.89	85.55
HD95	20.87	20.06	14.39

Table 8 demonstrate the soundness of our decision to use MaxVit.

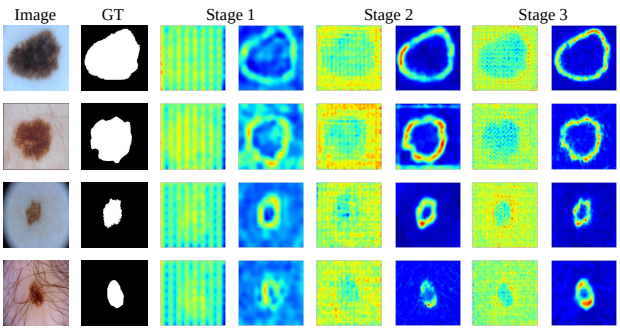


Figure 10. Visualization of the EGB module's specific functions is presented. In every stage, the left side depicts the model's output without EGB, while the right side shows the result after applying EGB. Stages 1 to 3 correspond to the three EGBs in the deeper to shallower layers of the network.

Research on the EGB Module: This paper proposes an edge guided bridge module to capture edge information robustly on multiple scales. We show the specific function of this module in the form of thermodynamic diagram, as shown in Figure 10. It can be clearly seen from the figure that EGB can make the model focus more on the boundary of the lesion area in each stage, and pay more attention to the edge information, so that the model can more accurately segment the boundary.

Table 9. Study on sensitivity analysis of super parameters.

lr \ batch size	4	8	12	16
0.00001	94.08	92.24	92.07	91.76
0.0001	94.54	94.76	94.82	94.79
0.001	94.73	94.78	94.55	94.60
0.01	82.34	90.78	83.62	89.62

Sensitivity Analysis: In this study, we conducted a sensitivity analysis on the learning rate and batch size to evaluate the impact of different hyperparameter combinations on model performance. The experimental results are shown in Table 9. Overall, this sensitivity analysis confirms the appropriateness of the learning rate and batch size settings used in this study.

5. Conclusion

In this paper, we propose a multiscale edge guided polynomial approximation network (AMEPANet). The well-designed EGB module in the network uses the Laplacian operator to accurately capture and enhance the edge information in the image, and realizes the robust preservation of edge information across multiple scales. By building an information mixed attention mechanism, the network can further mine and use the subtle features of the boundary area, and combine it with Kolmogorov-Arnold theorem to build an efficient decoder architecture, seamlessly integrate multi-source features, and achieve comprehensive fusion and optimization of feature information. In addition, this paper also proposes a C^1 continuous activation function using polynomial approximation, which shows significant advantages in reducing model calculation fluctuations and promoting model stability and convergence. Through extensive and in-depth experiments on several authoritative data sets such as Synapse, the advanced performance of our proposed model is proved. However, our model also has some limitations. Our main concern is how to improve the segmentation performance, while ignoring whether it can be deployed in real life scenarios. Therefore, our future work will improve this aspect.

Acknowledgement

This work was supported in part by the following: the Joint Fund of the National Natural Science Foundation of China under Grant No. U24A20219, the National Natural Science Foundation of China under Grant No. 62272281, the Special Funds for Taishan Scholars Project(tsqn202306274), and the Youth Innovation Technology Project of Higher School in Shandong Province under Grant No. 2023KJ212.

References

[1] R. Azad, R. Arimond, E. K. Aghdam, A. Kazerouni, and D. Merhof. Dae-former: Dual attention-guided efficient transformer for medical image segmentation. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 83–95. Springer, 2023. 2, 4, 8

[2] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera. Bi-directional convlstm u-net with densley connected convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0, 2019. 2, 10

[3] R. Azad, M. Heidari, M. Shariatnia, E. K. Aghdam, S. Karimijafarbigloo, E. Adeli, and D. Merhof. Transdeeplab: Convolution-free transformer-based deeplab v3+ for medical image segmentation. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 91–102. Springer, 2022. 2, 8

[4] B. Azam and N. Akhtar. Suitability of kans for computer vision: A preliminary investigation. *arXiv preprint arXiv:2406.09087*, 2024. 3

[5] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015. 7

[6] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018. 7

[7] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022. 2, 8, 9

[8] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 1, 2, 4, 7, 8, 9

[9] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, M. P. Lungren, S. Zhang, L. Xing, L. Lu, A. Yuille, and Y. Zhou. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, 97:103280, 2024. 8

[10] S. Chen, X. Tan, B. Wang, and X. Hu. Reverse attention for salient object detection. In *Proceedings of the European conference on computer vision (ECCV)*, pages 234–250, 2018. 5

[11] M. Cheng, Z. Kong, G. Song, Y. Tian, Y. Liang, and J. Chen. Learnable oriented-derivative network for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 720–730. Springer, 2021. 4

[12] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 7

[13] D. Dai, C. Dong, S. Xu, Q. Yan, Z. Li, C. Zhang, and N. Luo. Ms red: A novel multi-scale residual encoding and decoding network for skin lesion segmentation. *Medical image analysis*, 75:102293, 2022. 2, 6, 7, 10

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[15] X. Fan, J. Zhou, X. Jiang, M. Xin, and L. Hou. Csap-unet: Convolution and self-attention paralleling network for medical image segmentation with edge enhancement. *Computers in Biology and Medicine*, 172:108265, 2024. 4

[16] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen. Cpfnet: Context pyramid fusion network for medical image segmentation. *IEEE transactions on medical imaging*, 39(10):3008–3018, 2020. 2, 10

[17] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, 38(10):2281–2292, 2019. 1, 10

[18] M. Heidari, A. Kazerouni, M. Soltany, R. Azad, E. K. Aghdam, J. Cohen-Adad, and D. Merhof. Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6202–6212, 2023. 8

[19] X. Huang, Z. Deng, D. Li, and X. Yuan. Missformer: An effective medical image segmentation transformer. *arXiv preprint arXiv:2109.07162*, 2021. 2

[20] X. Huang, Z. Deng, D. Li, X. Yuan, and Y. Fu. Missformer: An effective transformer for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, 2022. 8, 9

[21] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021. 9

[22] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. 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

[23] D. Karimi and S. E. Salcudean. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on medical imaging*, 39(2):499–513, 2019. 3

[24] Z. Ke and Y. Yin. Tail risk alert based on conditional autoregressive var by regression quantiles and machine learning algorithms. In *2024 5th International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, pages 527–532. IEEE, 2024. 2

[25] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed. Boundary loss for highly unbalanced segmentation. In *International conference on medical imaging with deep learning*, pages 285–296. PMLR, 2019. 3

[26] B. Landman, Z. Xu, J. Igelsias, M. Styner, T. Langerak, and A. Klein. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge*, volume 5, page 12, 2015. 7

[27] Z. Li, J. Cui, H. Chen, H. Lu, F. Zhou, P. R. Rocha, and C. Yang. Research progress of all-fiber optic current transformers in novel power systems: A review. *Microwave and Optical Technology Letters*, 67(1):e70061, 2025. 2

[28] X. Lin, L. Yu, K.-T. Cheng, and Z. Yan. Batformer: Towards boundary-aware lightweight transformer for efficient medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 27(7):3501–3512, 2023. 4, 9

[29] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 2

[30] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*, 2024. 2, 3

[31] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira. Ph 2-a dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 5437–5440. IEEE, 2013. 7

[32] Y. Meng, H. Zhang, Y. Zhao, X. Yang, Y. Qiao, I. J. MacCormick, X. Huang, and Y. Zheng. Graph-based region and boundary aggregation for biomedical image segmentation. *IEEE transactions on medical imaging*, 41(3):690–701, 2021. 3

[33] L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi, et al. Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical image analysis*, 67:101874, 2021. 4, 6

[34] J. Peng, Q. Chang, H. Yin, X. Bu, J. Sun, L. Xie, X. Zhang, Q. Tian, and Z. Zhang. Gaia-universe: Everything is supernetify. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):11856–11868, 2023. 2

[35] M. M. Rahman and R. Marculescu. Medical image segmentation via cascaded attention decoding. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6222–6231, 2023. 2, 8, 9, 10

[36] M. M. Rahman and R. Marculescu. G-cascade: Efficient cascaded graph convolutional decoding for 2d medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7728–7737, 2024. 2, 8, 9, 10

[37] M. M. Rahman and R. Marculescu. Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation. In *Medical Imaging with Deep Learning*, pages 1526–1544. PMLR, 2024. 2, 8, 9, 10

[38] M. M. Rahman, M. Munir, and R. Marculescu. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11769–11779, 2024. 2, 7, 8, 9, 10

[39] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 8, 9, 10

[40] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International journal of computer assisted radiology and surgery*, 9:283–293, 2014. 7

[41] H. Tao, J. Li, Z. Hua, and F. Zhang. Dudb: Deep unfolding based dual-branch feature fusion network for pan-sharpening remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 2, 7

[42] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li. Maxvit: Multi-axis vision transformer. In *European conference on computer vision*, pages 459–479. Springer, 2022. 4

[43] C. J. Vaca-Rubio, L. Blanco, R. Pereira, and M. Caus. Kolmogorov-arnold networks (kans) for time series analysis. *arXiv preprint arXiv:2405.08790*, 2024. 3

[44] H. Wang, S. Xie, L. Lin, Y. Iwamoto, X.-H. Han, Y.-W. Chen, and R. Tong. Mixed transformer u-net for medical image segmentation. In *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 2390–2394. IEEE, 2022. 9

[45] M. Wang, H. Wang, and F. Zhang. Famc-net: Frequency domain parity correction attention and multi-scale dilated convolution for time series forecasting. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 2554–2563, 2023. 7

[46] S. Wang, K. He, D. Nie, S. Zhou, Y. Gao, and D. Shen. Ct male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. *Medical image analysis*, 54:168–178, 2019. 3

[47] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 568–578, 2021. 2

[48] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao. Pvt v2: Improved baselines with pyramid vision transformer. *Computational Visual Media*, 8(3):415–424, 2022. 2

[49] X. Wang, H. Wang, M. Zhang, and F. Zhang. Combining optical flow and swin transformer for space-time video super-resolution. *Engineering Applications of Artificial Intelligence*, 137:109227, 2024. 2

[50] H. Wu, J. Pan, Z. Li, Z. Wen, and J. Qin. Automated skin lesion segmentation via an adaptive dual attention module. *IEEE transactions on medical imaging*, 40(1):357–370, 2020. 3

[51] Y. Xin, J. Du, Q. Wang, Z. Lin, and K. Yan. Vmt-adapter: Parameter-efficient transfer learning for multi-task dense scene understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16085–16093, 2024. 2

[52] Y. Xin, S. Luo, X. Liu, H. Zhou, X. Cheng, C. E. Lee, J. Du, H. Wang, M. Chen, T. Liu, et al. V-petl bench: A unified visual parameter-efficient transfer learning benchmark. *Advances in Neural Information Processing Systems*, 37:80522–80535, 2025. 2

[53] H. Yu, L. Zhang, W. Wang, S. Li, S. Chen, S. Yang, J. Li, and X. Liu. State of charge estimation method by using a simplified electrochemical model in deep learning framework for lithium-ion batteries. *Energy*, 278:127846, 2023. 2

[54] F. Zhang, G. Chen, H. Wang, J. Li, and C. Zhang. Multi-scale video super-resolution transformer with polynomial approximation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 7

[55] F. Zhang, G. Chen, H. Wang, and C. Zhang. Cf-dan: Facial-expression recognition based on cross-fusion dual-attention network. *Computational Visual Media*, pages 1–16, 2024. 7

[56] F. Zhang, T. Guo, and H. Wang. Dfnet: Decomposition fusion model for long sequence time-series forecasting. *Knowledge-Based Systems*, 277:110794, 2023. 1

[57] F. Zhang, H. Wang, H. Fan, and C. Zhang. Rational polynomial image magnification based on edge and distance constraints. *Sci. Sin. Inform.*, 51:1270–1286, 2021. 3

[58] F. Zhang, M. Wang, W. Zhang, and H. Wang. Thatsn: Temporal hierarchical aggregation tree structure network for long-term time-series forecasting. *Information Sciences*, 692:121659, 2025. 2

[59] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, and Y. Yu. Adaptive context selection for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23*, pages 253–262. Springer, 2020. 7

[60] W. Zhang, H. Wang, and F. Zhang. Skip-timeformer: Skip-time interaction transformer for long sequence time-series forecasting. 2

[61] W. Zhang, H. Wang, and F. Zhang. Skip-timeformer: Skip-time interaction transformer for long sequence time-series forecasting. In *International joint conference on artificial intelligence*, pages 5499–5507, 2024. 2

[62] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu. nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*, 2021. 9