Kolmogorov-Arnold Networks for Medical Image Segmentation

Deep Bhattacharyya

A Thesis
in
The Concordia Institute
for
Information Systems Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science (Quality Systems Engineering) at
Concordia University
Montreal, QC, Canada

December 2025

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Deep Bhattacharyya

Entitled: Kolmogorov-Arnold Networks for Medical Image Segmentation

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality Systems Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Yong Zeng	Chair
Dr. Yong Zeng	Examiner
Dr. Anjali Awasthi	Examiner
Dr. Abdessamad Ben Hamza	Thesis Supervisor(s)
Dr. Ali Ayub	Thesis Supervisor(s)

Approved by
Dr. Chun Wang Chair of Department or Graduate Program Director

Dr. Mourad Debbabi Dean of Gina Cody School

Abstract

Kolmogorov-Arnold Networks for Medical Image Segmentation

Deep Bhattacharyya

Medical image segmentation plays a vital role in diagnosis and treatment planning, but remains challenging due to the inherent complexity, variability, and diversity of medical imaging modalities, particularly in capturing non-linear relationships and long-range dependencies within the data. To address these issues, this thesis first proposes U-KABS, a novel hybrid framework that integrates the expressive power of Kolmogorov-Arnold Networks (KANs) with a U-shaped encoder-decoder architecture to enhance segmentation performance. The U-KABS model combines a convolutional and squeeze-and-excitation stage that enhances channel-wise feature representations, and a KAN Bernstein Spline (KABS) stage that employs learnable activation functions based on Bernstein polynomials for global smoothness and B-splines for local adaptability. This hybrid design effectively captures both broad contextual trends and fine-grained patterns critical for delineating complex structures, supported by skip connections for multi-scale feature fusion and spatial detail preservation. Building upon these foundations, the thesis also introduces AdaKAN, an adaptive U-shaped model architecture that integrates an efficient attention mechanism with an adaptive KAN (AdaptKAN) block. The AdaKAN model consists of a convolutional stage and efficient KAN (EffiKAN) stage. The latter is comprised of an efficient attention for global feature extraction across resolutions and a dual-branch AdaptKAN block, with the right branch using a KAN layer with Bernstein polynomials for stable, smooth approximations, and the left branch incorporating up- and down-projections with adaptive scaling for channel-wise refinement. This encoder-decoder design captures higher-order dependencies essential for complex segmentations, with skip connections ensuring detail restoration during decoding. Evaluated across diverse benchmark datasets, including multi-class and multi-modal scenarios, both U-KABS and AdaKAN consistently outperform strong baselines, demonstrating improved accuracy, robustness to blurry boundaries, and efficiency in segmenting complex anatomical structures.

Acknowledgments

I am deeply grateful to my supervisors, Professors Abdessamad Ben Hamza and Ali Ayub, for their invaluable guidance, insightful feedback, and unwavering support throughout this research journey. Their expertise and encouragement have been instrumental in shaping this work. I extend my sincere thanks to my lab colleague Dalia Alzu'bi for her collaboration, thoughtful discussions, and constant assistance in the laboratory. Above all, I am forever indebted to my family for their boundless love and encouragement. To my father Tapan Bhattacharyya and mother Debjani Bhattacharyya, thank you for your lifelong sacrifices and belief in me. To my elder sister Madhura Bhattacharyya, thank you for being a constant source of strength, motivation, and inspiration. This thesis would not have been possible without each of you.

Table of Contents

List of Figures

List of Tables

List of Acronyms

FCN	Fully Convolutional Network
BN	Batch Normalization
ReLU	Rectified Linear Unit
SiLU	Sigmoid Linear Unit
MLP	Multi-Layer Perceptron
U-KABS	U-shaped KAN Bernstein Spline
KAB	KAN Bernstein Layer
KAS	KAN Splines Layer
DwConv	Depth-wise Convolution
SE	Squeeze-Excitation
LN	Layer normalization
AdaKAN	Adaptive Kolmogorov-Arnold Network
EA	Efficient Attention
EffiKAN	Efficient Kolmogorov-Arnold Network

CHAPTER 1

Introduction

This chapter outlines the motivations for our research, focusing on the challenges in medical image segmentation that inspired our approach. We start with the background and key reasons behind the research, followed by an overview of critical components such as encoder, decoder, and skip connections that form the backbone of medical image segmentation frameworks. A detailed literature review covers existing methods and their gaps, paving the way for our proposed solutions. The chapter ends by summarizing the research objectives and the thesis contributions to the field of medical image segmentation.

1.1 Challenges and Motivation

Medical image segmentation remains central to quantitative medical imaging, offering structured information crucial for clinical interpretation and computational diagnosis. Its challenges are rooted in the segmentation of non-linear nature of biological data, where lesion boundaries are often ambiguous and texture variations are subtle.

Fully Convolutional Networks (FCNs) such as the U-Net [8], UNet++ [9], Attention U-Net [1], have achieved remarkable progress by mapping local pixel dependencies through hierarchical convolutional filters. Yet, their inherently localized operations limit their capacity to represent long-range relationships across the image plane, often resulting in fragmented or incomplete segmentations in anatomically complex regions.

The advent of Transformer-based models such as TransUNet [10], Swin-Unet [11], UNETR [12], with self-attention mechanisms has addressed this limitation by introducing global contextual modeling, enabling the capture of distant dependencies between image tokens. However, the accompanying quadratic computational overhead and reliance on large-scale annotated datasets restrict their scalability and clinical applicability. MLP-based models, although computationally efficient, often fail to approximate the intricate high-dimensional functional relationships characteristic of medical imagery, leading to reduced representational expressiveness.

Kolmogorov-Arnold Networks (KANs) [13], derived from the Kolmogorov-Arnold representation theorem, present a mathematically elegant solution by approximating arbitrary multivariate functions through compositions of adaptive one-dimensional mappings. This functional decomposition allows KANs to model highly non-linear and discontinuous structures with improved generalization and interpretability, qualities essential for capturing the diverse morphological and textural patterns encountered in medical image segmentation.

1.2 Core Architectural Components

The key components that frame the architectures of our proposed models are

Encoder. In a medical image segmentation network, the encoder functions as the core analytical component responsible for interpreting and transforming the raw image into a meaningful, high-level representation. It serves as the model's perceptive mechanism, capable of distinguishing between different visual patterns present in complex medical images, such as Breast ultrasound images. The encoder gradually abstracts the image information through a hierarchy of feature extraction and dimensionality reduction stages.

At the initial levels of the encoding process, the network captures low-level visual characteristics such as edges, boundaries, and textural variations, features that form the foundation for higher-level understanding. As the data progresses through deeper layers, the encoder extracts increasingly abstract and semantically rich features that describe the regions of interest.

Pooling layers, commonly implemented as max-pooling operations, play a crucial role in this transformation. They act as spatial downsampling mechanisms, reducing the dimensionality of feature maps. For example, from 256×256 to 128×128 pixels, while preserving the most significant features. The encoder, therefore, trades spatial precision for semantic depth, producing a condensed but information-rich feature map that encapsulates both global and local contexts.

The final stage of the encoder yields what is known as the bottleneck representation, a compact representation of the original image. This bottleneck captures the most discriminative features necessary for accurate segmentation, serving as the foundation for the decoder to reconstruct the spatial details lost during downsampling. In essence, the encoder functions as the network's visual understanding unit, distilling complex medical imagery into a structured form that balances computational efficiency with semantic richness.

Decoder. The decoder serves as the reconstructive counterpart to the encoder, tasked with transforming the compact, abstract representation back into a high-resolution segmentation map. It functions as the generative phase of the network, gradually recovering the spatial information that was compressed during encoding. Instead of relying on a single operation type, the decoder can employ a variety of upsampling and refinement mechanisms.

Through successive decoding stages, the feature representations are progressively refined to regain fine-grained spatial precision. At each stage, contextual information from deeper layers is combined with localized details from earlier encoding levels through skip connections, allowing the model to recover subtle boundaries and spatial textures that may have been lost during abstraction. This integration ensures a balance between global semantic understanding and local spatial accuracy. The final output of the decoder is a dense, pixel-level segmentation map in which each spatial position corresponds to a distinct anatomical or pathological region, providing an interpretable and clinically meaningful reconstruction of the input image.

Skip Connections. The skip connections act as the bridges between the encoder and decoder, helping the network retain important information that might otherwise be lost during compression. They pass detailed features from earlier layers directly to later decoding stages, ensuring both fine and broad details are preserved. In architectures such as U-Net, these connections link encoder feature maps with corresponding decoder layers, allowing the model to combine low-level textures with high-level context. For example, while the encoder may reduce a 256×256 MRI scan to a 32×32 feature map, skip connections transfer fine boundary details to the decoder, leading to more accurate and detailed segmentation masks.

1.3 Problem Statement

In this section, we present the core problem statement of our research. The primary objective is to achieve accurate semantic segmentation of medical images by leveraging available ground truth annotations to generate precise segmentation masks. Semantic segmentation involves assigning a specific class label to each pixel in an image [14–16], effectively partitioning it into meaningful regions, such as organs, lesions, or abnormalities in medical imaging. Specifically, given an input medical image be represented as $I \in \mathbb{R}^{H \times W \times 3}$, where H and W denote the image height and width, and the three channels correspond to the RGB components, the goal of semantic image segmentation is to predict a segmentation mask $M \in [C]^{H \times W}$, where $[C] = \{0, 1, \ldots, C-1\}$ and C represents the total number of semantic classes. Each pixel in M is assigned an integer label from 0 to $C-1$, indicating the class to which that pixel belongs.

1.4 Objectives

The main goal of this thesis is to develop Kolmogorov-Arnold Network (KAN)-based models for medical image segmentation. Specifically, the key objectives are to:

- Develop a U-shaped Kolmogorov-Arnold Bernstein Splines (U-KABS) model designed to effectively capture both nonlinear global and local features in medical image segmentation. The model should maintain moderate computational complexity to ensure efficiency while achieving robust performance in extracting intricate feature representations.
- Develop an Adaptive Kolmogorov-Arnold Network (AdaKAN) model, that integrates efficient attention mechanisms and adaptive Kolmogorov-Arnold Networks. The model should effectively capture both global and local dependencies in data while maintaining moderate computational complexity, ensuring computational efficiency and robust performance in feature representation.

1.5 Literature Review

Fully Convolutional Network (FCN)-based Methods. The U-Net model is a widely used fully convolutional network (FCN) for medical image segmentation, with a symmetric encoder-decoder design and skip connections to preserve spatial details [8]. UNet++ improves U-Net by adding nested skip connections for better feature integration [9]. Attention U-Net uses attention to highlight important features [1], while MALUNet captures features at multiple scales [17]. These FCN methods are good for local details but struggle with long-range connections, limiting their ability to understand the full context of complex medical images.

Transformer-based Methods. Inspired by Vision Transformers [18], Transformer-based models are adapted for medical image segmentation to capture long-range connections. TransUNet merges transformer blocks with U-Net for better performance through self-attention [10]. MedT uses a dual-branch encoder for fine spatial details [2], and Swin-Unet applies shifted window attention for long-range connections [11]. UNETR treats 3D image segmentation as a sequence task [12].

These methods excel at global context but require significant computational resources and large datasets.

Multi-Layer Perceptron (MLP)-based Methods. To reduce the computing needs of Transformers, MLP-based methods are gaining popularity for their efficiency. UNeXt uses tokenized MLP blocks to capture local and global connections with a shifted design [3]. Rolling-Unet adds a Rolling-MLP block to handle long-range connections while keeping local details in a U-Net structure [4]. These methods offer a good balance of performance and efficiency but may not handle complex medical imaging tasks well.

Recent KAN-based Methods. Kolmogorov-Arnold Networks (KANs) are an emerging alternative to MLPs, using adjustable activation functions to model complex patterns [13]. U-KAN enhances U-Net with KAN blocks and B-spline functions for better segmentation [5]. ResU-KAN adds residual attention and atrous pooling to capture long-range connections and expand the receptive field [6]. However, their use of B-splines limits their ability to balance global and local features.

With regards to the existing methods, we propose the U-KABS framework, which integrates a U-shaped encoder-decoder architecture with a KAN Bernstein Spline (KABS) block. This hybrid design combines the global smoothness of Bernstein polynomials with the local adaptability of B-splines, enhancing the model's ability to capture both global contextual information and fine-grained local details. By incorporating squeeze-and-excitation mechanisms and skip connections, U-KABS achieves superior segmentation performance while maintaining computational efficiency, making it well-suited to the complexities of medical image segmentation.

Furthermore, we introduce EffiKAN, an encoder-decoder architecture that combines convolutional layers, Efficient Attention, and an AdaKAN block with bidirectional projections and Bernstein polynomials. This design balances local and global context, reduces computational complexity, and outperforms state-of-the-art baselines in binary and multi-class segmentation tasks across diverse datasets. The AdaKAN block enhances boundary accuracy, while Efficient Attention improves global feature extraction, addressing the limitations of prior methods in handling complex anatomical structures.

1.6 Contributions

The rest of the thesis is structured as follows:

- In chapter 2, we propose U-KABS, a hybrid Kolmogorov-Arnold Network that integrates a U-shaped encoder-decoder architecture with a KAN Bernstein Spline (KABS) block. This framework leverages the global smoothness of Bernstein polynomials and the local adaptability of B-splines to effectively cpture the on-linear features. The integration of convolutional with squeeze-and-excitation stages enhances channel-wise feature representations, while skip connections facilitate preserving spatial details.

- In chapter 3, we introduce AdaKAN, an Adaptive Kolmogorov-Arnold Network that integrates Efficient Attention, and an Adaptive KAN block within an encoder-decoder framework. The proposed model architecture balances local and global dependencies by employing up/down projections and Bernstein polynomials, while reducing computational complexity through Efficient Attention.

- Chapter 4 presents the key contributions of the thesis, discusses its limitations, and highlights possible directions for future research.

CHAPTER 2

Hybrid Kolmogorov-Arnold Networks for Medical Image Segmentation

Medical image segmentation plays a vital role in diagnosis and treatment planning, but remains challenging due to the inherent complexity and variability of medical images, especially in capturing non-linear relationships within the data. In this chapter, we propose U-KABS, a novel hybrid framework that integrates the expressive power of Kolmogorov-Arnold Networks (KANs) with a U-shaped encoder-decoder architecture to enhance segmentation performance. The U-KABS model combines the convolutional and squeeze-and-excitation stage, which enhances channel-wise feature representations, and the KAN Bernstein Spline (KABS) stage, which employs learnable activation functions based on Bernstein polynomials and B-splines. This hybrid design leverages the global smoothness of Bernstein polynomials and the local adaptability of B-splines, enabling the model to effectively capture both broad contextual trends and fine-grained patterns critical for delineating complex structures in medical images. Skip connections between encoder and decoder layers support effective multi-scale feature fusion and preserve spatial details. Evaluated across diverse medical imaging benchmark datasets, U-KABS demonstrates superior performance compared to strong baselines, particularly in segmenting complex anatomical structures.

2.1 Introduction

The aim of medical image segmentation is to partition medical images, such as ultrasound, MRI, or dermoscopic images, into meaningful regions corresponding to anatomical structures or pathological areas, assigning each pixel a class label to delineate organs, tissues, or lesions. This task is fundamental to computer-aided diagnosis, enabling precise identification of regions of interest for clinical analysis, and disease monitoring. Recent advances in deep learning have driven significant progress in medical image segmentation [19], with several approaches emerging to tackle its challenges. Fully Convolutional Network (FCN)-based methods such as U-Net [8] and its variants, UNet++ [9], Attention U-Net [1], UNet3+ [20], 3D UNet [21], V-Net [22], and Ki-UNet [23, 24], leverage convolutional layers to capture local spatial patterns, achieving robust performance. However, they often struggle with global context modeling, leading to over-segmentation or diffused boundaries in noisy images.

To address these challenges, Transformer-based U-shaped architectures, such as MedT [2] and Swin-Unet [11], have been developed to capture global spatial patterns through self-attention mechanisms. Yet, their quadratic computational complexity with respect to input length and reliance on large datasets limit their applicability in resource-constrained clinical settings or data-scarce scenarios. Multi-Layer Perceptron (MLP)-based methods, such as UNeXt [3] and MALUNet [17], prioritize efficiency, but their linear transformations often lack the expressiveness needed for complex lesion boundaries, resulting in under-segmentation. More recently, Kolmogorov-Arnold Networks (KANs) [13] have emerged as a promising alternative to MLPs, leveraging learnable activation functions to enhance interpretability and expressiveness. Building on this, U-KAN [5] incorporates B-spline activation functions within a U-shaped architecture to effectively model local spatial patterns, achieving competitive results across various medical image segmentation datasets. However, it struggles to capture global context.

The aforementioned limitations highlight the need for a segmentation framework that combines computational efficiency, robust global-local feature modeling, and adaptability to diverse medical imaging modalities. To this end, we propose a novel hybrid framework, U-KABS, which integrates a U-shaped encoder-decoder architecture with a KAN Bernstein spline (KABS) block as a core component, combining the global smoothness of Bernstein polynomials and the local adaptability of B-splines. This hybrid design enables our model to effectively capture both global contextual information and fine-grained local details, which are critical for the accurate segmentation of complex anatomical structures in medical images. Our main contributions can be summarized as follows:

- We propose a novel hybrid Kolmogorov-Arnold network, integrating a U-shaped encoder-decoder architecture, and leveraging skip connections and squeeze-and-excitation mechanisms to capture hierarchical features and prioritize salient regions.
- We introduce the KAN Bernstein spline block, combining Bernstein polynomials and B-

splines to balance global smoothness and local adaptability.

- We conduct extensive evaluations on four benchmark datasets with varying characteristics, demonstrating superior performance over state-of-the-art baselines and robustness to blurry boundaries, and multi-class tasks.

2.2 Related Work

FCN-based Methods: The U-Net architecture [8] is a foundational FCN-based model for medical image segmentation, featuring a symmetric encoder-decoder structure with skip connections to preserve spatial details across scales. Building on U-Net's success, several variants have been developed to further improve segmentation performance. For instance, UNet++ [9] enhances U-Net by incorporating nested and dense skip connections, which improve feature aggregation. To address the need for focusing on salient regions, attention-based FCN methods have emerged. Attention U-Net [1] integrates attention mechanisms to emphasize relevant features. Similarly, MALUNet [17] employs multi-scale attention to capture features at varying resolutions. While FCN-based methods are effective for local feature learning, they often struggle to model long-range dependencies, limiting their ability to capture global contextual information in complex medical images.

Transformer- and MLP-based Methods: Drawing inspiration from Vision Transformers [18], Transformer-based models have been recently adapted for medical image segmentation to overcome the limitations of FCN-based approaches in capturing global dependencies. TransUNet [10] integrates Transformer blocks with U-Net architecture, leveraging self-attention and cross-attention mechanisms to enhance performance. MedT [2] employs a dual-branch encoder to learn fine-grained spatial details. Swin-Unet [11] utilizes shifted window-based self-attention to model long-range dependencies. UNETR [12] introduces a Transformer-based architecture that reformulates volumetric medical image segmentation as a 1D sequence-to-sequence task. While Transformer-based methods excel in capturing global context, they often require substantial computational resources and large annotated datasets. To address this computational overhead, lightweight MLP-based methods have gained attention for their efficiency. For instance, UNeXt [3] introduces tokenized MLP blocks to capture local and global dependencies using a shifted MLP block to handle spatial relationships across image dimensions. Rolling-Unet [4] introduces the Rolling-MLP block to capture long-distance dependencies in multiple directions while preserving local features through a U-Net architecture. These MLP-based approaches offer a balance between performance and computational efficiency, but lack the expressive power needed for complex medical imaging tasks.

KAN-based Methods. KANs [13] have emerged as a compelling alternative to MLPs, leveraging their ability to approximate multivariate functions through compositions of learnable univariate functions. More recently, U-KAN [5], which builds on the U-Net backbone by incorporating tokenized KAN blocks with B-spline activation functions, has demonstrated improved performance in medical image segmentation. Similarly, ResU-KAN [6] integrates a residual convolutional attention and an atrous spatial pyramid pooling module into U-KAN in an effort to capture long-distance dependencies and expand expand the model's receptive field. However, their reliance on B-splines constrains the model's ability to effectively balance global context with local feature representations. In contrast, our proposed U-KABS framework integrates both Bernstein polynomials and B-splines as learnable activation functions, combining global smoothness with local adaptability, and offering a novel design that achieves superior segmentation performance while maintaining computational efficiency. By integrating this hybrid approach within a U-Net-inspired architecture and leveraging squeeze-and-excitation mechanisms, U-KABS provides a flexible and expressive model well-suited to the complexities of medical image segmentation.

2.3 Method

In this section, we begin by outlining the segmentation task, followed by a brief background on standard KANs using splines as basis functions. We then present an overview of our model architecture and introduce its core components.

Problem Description. Let $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ be an input image with height H and width W. Our goal is to predict a pixel-wise segmentation mask $\mathbf{M} \in [C]^{H \times W}$, where $[C] = \{0, \ldots, C-1\}$ and C is the number of semantic classes. Each pixel in $\mathbf{M}$ is assigned an integer label from 0 to $C-1$, indicating the class to which that pixel belongs.

Kolmogorov-Arnold Networks. KANs are inspired by the Kolmogorov-Arnold representation theorem [25, 26], which states that any continuous multivariate function on a bounded domain can be represented as a finite composition of continuous univariate functions of the input variables and the binary operation of addition. A KAN layer is a fundamental building block of KANs [13], and is given by a matrix $\mathbf{\Phi} = (\phi_{q,p})$ of 1D functions, where each trainable activation function ϕ is defined as a weighted combination of a sigmoid linear unit (SiLU) function and a spline function:

$$\phi(x) = w_b \mathrm{SiLU}(x) + w_s \mathrm{spline}(x), \tag{2.1}$$

where $\text{spline}(x) = \sum_i c_i B_i(x)$ is a weighted sum of B-splines basis functions with trainable coefficients c_i. During training, the weights w_b and w_s are learned to optimize performance. Given a input feature vector $\mathbf{x}^{(l)} \in \mathbb{R}^{F_l}$, the output of the l-th KAN layer is an F_{l+1}-dimensional feature vector given by

$$\mathbf{x}^{(l+1)} = \text{KAN}^{(l)}(\mathbf{x}^{(l)}). \tag{2.2}$$

Architecture Overview. The overall architecture of the proposed U-KABS framework is illustrated in Figure 2.1. It follows a symmetric encoder-decoder design, along with skip connections that bridge corresponding blocks, enabling effective feature fusion. U-KABS integrate the expressive capabilities of KANs with convolutional and attention-based squeeze-and-excitation mechanisms, and incorporates two distinct processing stages in both the encoder and decoder: the Convolutional and Squeeze-and-Excitation (ConvSE) stage and the KAN Bernstein Spline (KABS) stage. The ConvSE stage, comprised of three blocks, combines convolutional layers with SE blocks to enhance channel-wise feature representations. This stage is particularly effective for capturing local patterns, such as edges, and recalibrating feature channels to emphasize the most relevant features for segmentation. On the other hand, the KABS stage leverages the principles of KANs, employing learnable activation functions based on Bernstein polynomials and B-splines. These activation functions provide a unique balance between global smoothness and local adaptability, making the KABS stage particularly well-suited for modeling complex patterns in medical images, such as subtle tissue boundaries, tumors, or pathological structures. The decoder mirrors the encoder's structure to generate the segmentation map from the encoded features.

2.3.1 Encoder

The encoder consists of the following two stages:

Convolutional and squeeze-and-excitation (ConvSE) stage. This stage consists of L ConvSE blocks, each of which is comprised of a convolutional layer, a batch normalization (BN) layer, a Rectified Linear Unit (ReLU) activation function, and a squeeze-and-excitation (SE) block to enhance channel-wise feature representations. Formally, the process of the ConvSE stage can be expressed as

$$\mathbf{X}^{(\ell)} = \text{SE}(\text{ReLU}(\text{BN}(\text{Conv}(\mathbf{X}^{(\ell-1)})))), \tag{2.3}$$

where $\mathbf{X}^{(\ell)}$ is the output feature map of the ℓ-th ConvSE block for $\ell \in \{1, \ldots, L\}$, and $\mathbf{X}^{(0)} = \mathbf{I}$ is the input image. A 2×2 max-pooling operation is applied after each block for downsampling. The output feature map $\mathbf{X}^{(L)}$ of the last ConvSE block serves as input for the first block of the subsequent KAN Bernstein Spline (KABS) stage.

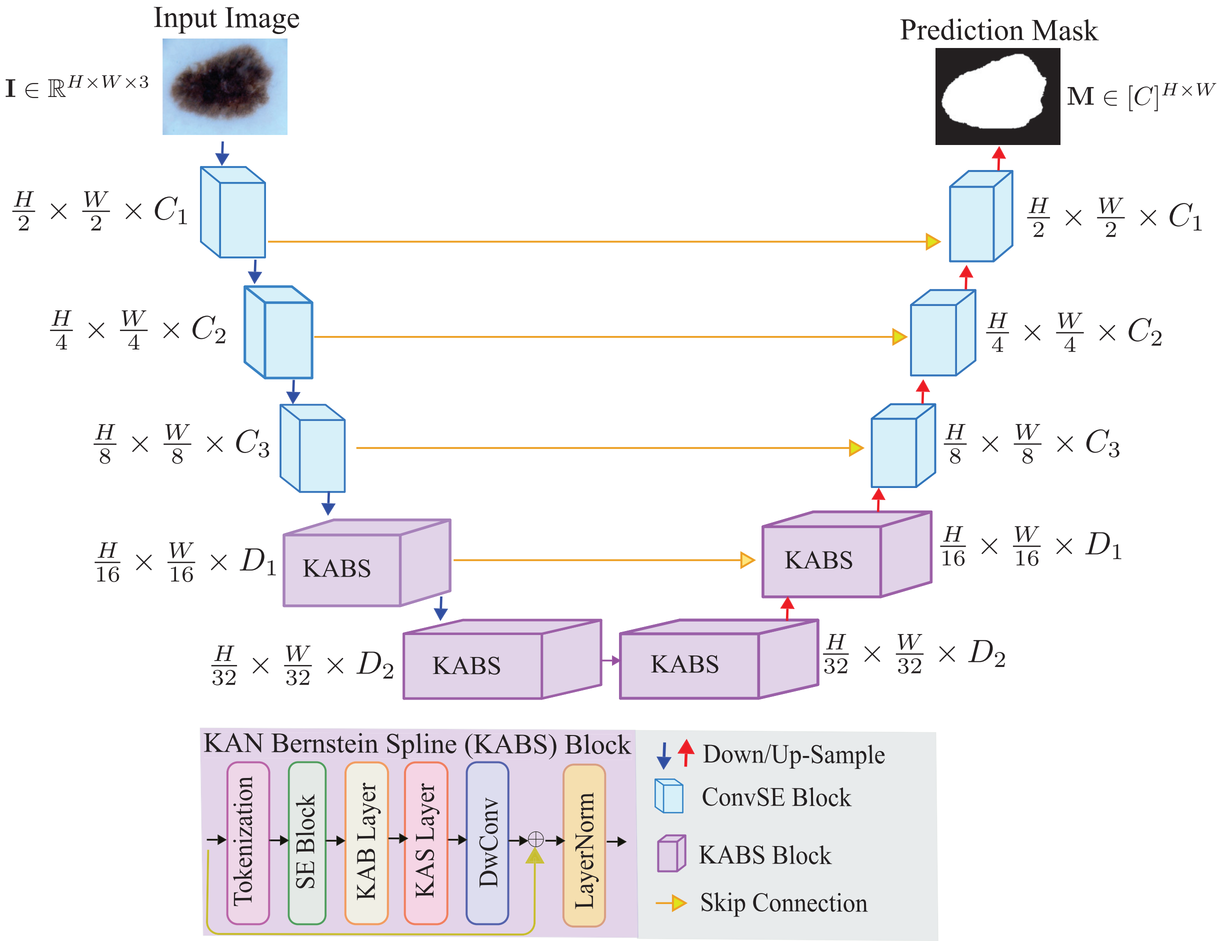

Figure 2.1: Overall architecture of the proposed U-KABS framework. The model follows a symmetrically U-shaped encoder-decoder structure, integrating Convolutional and Squeeze-and-Excitation (ConvSE) and KAN Bernstein Spline (KABS) stages. The encoder, comprising three ConvSE blocks followed by two KABS blocks, progressively downsamples the input image to learn hierarchical features, increasing the number of feature channels while halving spatial resolution at each block. The decoder mirrors the design of the encoder, upsampling features via bilinear interpolation to restore the feature map spatial resolution and generate a pixel-wise segmentation mask.

KAN Bernstein Spline (KABS) stage. This stage is comprised of K blocks, each of which consists of a tokenization layer, a squeeze-and-excitation block, two KAN layers with distinct basis functions as learnable activation functions, a depth-wise convolution layer, a residual connection, and a layer normalization, as illustrated in Figure 2.1.

Tokenization and squeeze-and-excitation. The feature map $\mathbf{X}^{(L)}$ of the L-th block of the ConvSE stage is converted into a sequence of tokens, yielding an intermediate feature map $\mathbf{Z}^{(0)}$. This tokenization process follows the Vision Transformer methodology [18], where the input feature map is partitioned into a sequence of flattened patches, and projected into a latent embedding space using a trainable linear projection. The features are then processed through a squeeze-and-excitation block before being fed into two subsequent KAN layers with different basis functions for further feature enhancement.

KAN Bernstein (KAB) Layer. We approximate each trainable activation function of the KAB layer over the interval $[0, 1]$ using a weighted combination of a Sigmoid Linear Unit (SiLU) function and a Bernstein polynomial function:

$$\tilde{\phi}(z) = \tilde{w}_b \text{SiLU}(z) + \tilde{w}_s \text{Bernstein}(z), \tag{2.4}$$

where

$$\begin{aligned} \text{Bernstein}(z) &= \sum_{r=0}^{R} \theta_r P_{r,R}(z) \\ &= \sum_{r=0}^{R} \theta_r \binom{R}{r} z^r (1-z)^{R-r} \end{aligned} \tag{2.5}$$

is a weighted sum of Bernstein basis polynomials $P_{r,R}$ of degree R, and θ_r are the learnable Bernstein coefficients. Bernstein-based KAN offers global, smooth function approximation, while being memory efficient due to the recurrence avoiding explicit storage of all polynomial terms.

KAN Spline (KAS) Layer. This layer uses learnable, locally-supported B-spline basis functions to form activation functions, as defined in Eq. (3.1). Each activation function is expressed as a linear combination of these splines, enabling the layer to adaptively model fine-grained patterns in the data. This localized, data-driven transformation improves the model's expressiveness, making it especially effective in medical image segmentation tasks where capturing subtle local variations is important.

The choice of Bernstein polynomials and splines as basis functions in our proposed model is motivated by their complementary approximation properties. Bernstein polynomials are globally supported over the interval $[0, 1]$, meaning that each basis function influences the entire input domain. This global support results in smooth and stable approximations, where any change to a single coefficient affects the output across the full range. Such behavior is well-suited for modeling broad trends and ensuring continuity. In contrast, spline basis functions are locally supported, with each function active only within a limited region defined by a set of knots. This local support enables fine-grained adjustments: modifying a coefficient only alters the function's shape in a specific subregion, without impacting the rest of the domain. As a result, splines provide greater flexibility for capturing localized structures or sharp transitions. By integrating both types of basis functions, our model balances global smoothness with local adaptability, allowing for expressive and controllable function representations.

After the tokenized feature maps are processed by the KAB and KAS layers, they go through the depth-wise convolution (DwConv), which applies a single filter per embedding dimension of the tokenized feature map, capturing local spatial patterns. The residual connection in the KABS block adds the input to the transformed output, preserving critical features. Layer normalization (LN) standardizes the embedding dimensions. The process of the KABS stage can be expressed as follows:

$$\mathbf{Z}^{(k)} = \mathrm{LN}(\mathrm{DwConv}(\mathrm{KAS}(\mathrm{KAB}(\mathrm{SE}(\mathbf{Z}^{(k-1)})))) + \mathbf{Z}^{(k-1)}), \tag{2.6}$$

where $\mathbf{Z}^{(k)}$ is the feature map of the kth KABS block for $k \in \{1, \ldots, K\}$, and $\mathbf{Z}^{(0)}$ is the tokenized feature map of the last ConvSE block.

2.3.2 Decoder

The U-KABS decoder is designed as a symmetric counterpart to the encoder, processing features from the encoder's final block to produce a segmentation mask $\mathrm{M} \in [C]^{H \times W}$, where C is the number of semantic classes in the input image. The mask generation process is carefully structured to gradually restore spatial resolution and refine features.

The decoder is composed of two primary stages, namely KABS and ConvSE, mirroring the encoder's design for architectural consistency. In the KABS stage, each of the two blocks begins by applying bilinear interpolation to upsample the input feature maps by a factor of two, ensuring a smooth and continuous increase in spatial resolution. The upsampled features are then concatenated with skip-connected features from the corresponding encoder layers to preserve fine-grained spatial information that may have been lost during downsampling. This skip connection strategy helps mitigate the semantic gap between encoder and decoder representations, enhancing localization accuracy. The subsequent ConvSE stage comprises three sequential blocks, each performing additional upsampling through bilinear interpolation and integrating encoder features via skip connections. These ConvSE blocks further refine the representations by emphasizing salient spatial and channel-wise patterns. To generate the segmentation mask, the output of the last ConvSE block is passed through a 1×1 convolutional layer with C output channels, mapping the features to a segmentation map. A softmax activation is applied along the channel dimension to compute class-wise probabilities for each pixel. The final segmentation mask M is obtained by taking the argmax over these probabilities, assigning each pixel the class with the highest probability.

2.4 Experiments

2.4.1 Experimental Setup

Datasets. We conduct comprehensive experiments on four benchmark datasets: Breast Ultrasound Images (BUSI) [27], Gland Segmentation (GlaS) [28], International Skin Imaging Collaboration (ISIC 2018) [29], and Automated Cardiac Diagnosis Challenge (ACDC) [30].

- **BUSI** includes ultrasound images of breast cancer, categorized as normal, benign, or malignant, with ground truth segmentation masks. We use 647 benign and malignant images, resized to 256×256.
- **GlaS** contains 165 histological colorectal tissue images, benign or malignant, with segmentation masks. Variability in staining and structural complexity challenges gland segmentation. Images are resized to 512×512.
- **ISIC 2018** comprises 2594 dermoscopic skin lesion images with binary segmentation masks. Blurry boundaries and diverse lesion types (benign, nevi, seborrheic keratoses, melanomas) pose challenges. Images are resized to 256×256.
- **ACDC** includes cine MRI heart scans with 5–8 mm slice thickness and 0.83–1.75 mm²/pixel resolution. Scans are annotated for left ventricle, right ventricle, myocardium, and background (four classes). It has 80 training and 20 validation cases.

Evaluation Metrics. To compare the performance of U-KABS against the state-of-the-art (SOTA) methods, we adopt three standard evaluation metrics: Intersection Over Union (IoU), Dice Similarity Coefficient (DSC), and Hausdorff Distance (HD95). IoU measures the overlap between the predicted (P) and ground truth (G) segmentation masks:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|}, \tag{2.7}$$

and DSC quantifies how well the predicted mask matches the ground truth mask:

$$DSC = \frac{2 \times |P \cap G|}{|P| + |G|}, \tag{2.8}$$

while HD95 measures the maximum deviation between the two masks, focusing on the 95th percentile of the nearest point distances:

$$\text{HD95}(P, G) = \max\{d_{95}(P, G), d_{95}(G, P)\}. \tag{2.9}$$

Higher (resp. smaller) values of IoU and DSC (resp. HD95) indicate better segmentation performance.

Baselines. We benchmark the performance of U-KABS against several SOTA methods, including U-Net [8], U-Net++ [9], Attention-UNet [1], MedT [2], UNeXt [3], Rolling-Unet [4], U-KAN [5], TPFIANet [31], ResU-KAN [6], MALUNet [17] , ViT-CUP [18], Swin-Unet [11], MC-Former [32], FCT [33], UNETR [12], TransUNet [10], SCM-UNET [34] , WTCM-UNet [35], and CTO [36] .

Implementation Details. All experiments are conducted using PyTorch on a Linux workstation equipped with a single NVIDIA GeForce RTX 3070 GPU (8GB memory). For binary segmentation datasets, we use a batch size of 8 and a learning rate of 10^{-4}, while for the multi-class ACDC dataset, the batch size is set to 2 with the same learning rate. The number of channels are set to $C_1 = 32$, $C_2 = 64$, $C_3 = 256$, $D_1 = 320$, and $D_2 = 512$. We set the Bernstein polynomial order to 4, spline order to 3, and grid size to 5. The model is trained using the Adam optimizer with a cosine annealing learning rate scheduler, decaying to a minimum of 10^{-5}. Each dataset is randomly split into 80% training and 20% validation. For binary segmentation tasks, we employ the average of Binary Cross-Entropy (BCE) and Dice loss, while on the multi-class ACDC dataset, we use the average of Cross-Entropy and Dice loss. For all datasets, the model is trained for 400 epochs, and to ensure robustness, we report the mean and standard deviation over 5 independent runs with different random seeds.

2.4.2 Results and Analysis

Quantitative Results. Table 2.1 compares our U-KABS model with SOTA methods on the BUSI and GlaS datasets, assessing model size (parameters), computational complexity (FLOPS, in gigaflops), and segmentation performance. The reported results for the baselines are taken from their respective papers. U-KABS has a moderate parameter count, and achieves notably lower FLOPS, operating at 6.92 FLOPS(G), which is more efficient compared to the strongest baseline ResU-KAN. Our model demonstrates the most effective segmentation performance across both datasets. On the BUSI dataset, U-KABS achieves better performance than ResU-KAN in terms of both IoU and DSC. Similarly, on the GlaS dataset, U-KABS outperforms ResU-KAN in terms of both IoU and DSC . These improvements stem from U-KABS's hybrid architecture, enabling superior global and local feature approximation compared to KAN's B-spline-only approach. In addition, the integration of squeeze-and-excitation mechanisms into the convolutional blocks helps enhance feature prioritization in ultrasound (BUSI) and histological (GlaS) images.

Table 2.2 shows that our U-KABS model achieves the best performance compared to baselines in terms of both IoU and DSC on the ISIC 2018 dataset. U-KABS maintains a consistent performance advantage, with relative improvements of approximately 3.22% in IoU and 1.9% in DSC over

Table 2.1: Quantitative comparison of our method against benchmark baselines on the BUSI and GlaS datasets in terms of IoU, DSC, and HD95. Best results are in bold, and the second best results are underlined.

Method	Params(M) ↓	FLOPS(G)↓	BUSI			GlaS		
			IoU↑	DSC↑	HD95↓	IoU↑	DSC↑	HD95↓
U-Net [8]	7.76	13.78	57.22±4.74	71.91±3.54	7.57±2.44	86.66±0.91	92.79±0.56	0.83±0.18
U-Net++ [9]	9.16	34.90	57.41±4.77	72.11±3.90	7.72±2.16	87.07±0.76	92.96±0.65	**0.81±0.16**
Att-UNet [1]	50.27	60.06	55.18±3.61	70.22±2.88	8.36±2.11	86.84±1.19	92.89±0.65	0.82±0.29
MedT [2]	1.60	21.24	52.15±3.47	67.68±3.18	10.23±1.17	75.47±3.46	85.92±2.93	–
UNeXt [3]	**1.47**	**4.58**	61.78±1.46	75.52±0.91	8.33±0.42	83.95±1.09	91.22±0.67	1.04±0.10
Rolling-Unet [4]	25.32	28.32	61.00±0.64	74.67±1.24	6.19±0.62	86.42±0.96	92.63±0.62	1.00±0.08
U-KAN [5]	9.38	6.89	63.38±2.83	76.42±2.90	–	87.64±0.32	93.37±0.16	–
TPFIANet [31]	–	–	66.20±0.28	75.84±0.26	–	84.88±0.37	91.32±0.25	–
ResU-KAN [6]	20.06	8.57	67.74±1.35	79.92±0.73	–	87.99±0.44	93.61±0.24	–
U-KABS (Ours)	9.54	6.92	**67.98±0.73**	**80.37±0.56**	**5.67±1.04**	**88.59±0.36**	**93.94±0.18**	0.82±0.09

the lightweight UNeXt. Similarly, U-KABS outperforms the strongest baseline, Rolling-Unet, achieving a relative improvement 0.55% in terms of DSC. These improvements, though modest, are significant given the challenging nature of dermoscopic images in the ISIC 2018 dataset, which feature blurry boundaries and diverse lesion types. U-KABS's superior performance is largely attributed to its hybrid approach combining Bernstein polynomials and B-splines, enabling it to capture both global contextual trends and local fine-grained details more effectively than Rolling-Unet's blocks, which focus primarily on lightweight MLP-based feature learning. Moreover, the incorporation of squeeze-and-excitation mechanisms in U-KABS enhances its ability to prioritize salient features in complex skin lesion images.

Table 2.2: Quantitative comparison of our method with benchmark baselines on the ISIC 2018 dataset.

Method	IoU	DSC
U-Net [8]	77.86	87.55
U-Net++ [9]	78.31	87.83
Att-UNet [1]	78.43	87.91
MedT [2]	81.48	89.49
UNeXt [3]	81.70	89.70
MALUNet [17]	80.25	89.04
Rolling-Unet [4]	83.74	90.90
SCM-UNET [34]	81.88	90.14
WTCM-UNet [35]	80.26	89.05
CTO [36]	84.00	90.60
Ours	**84.33**	**91.40**

Table 2.3 presents the performance comparison of U-KABS against baselines on the ACDC dataset, evaluated using the DSC metric for three cardiac structures (Right Ventricle (RV), Myocardium (Myo), and Left Ventricle (LV)), along with their average. U-KABS achieves a 0.53% higher average DSC than Swin-Unet, which is the best-performing baseline. When compared to TransUNet, U-KABS improves by 0.82% in average DSC, and the gain over UNETR is 1.92%. The difference becomes even more significant when compared to ViT-CUP, with U-KABS outperforming it by a large margin of 9.08% in average DSC. On a per-organ level, U-KABS shows a 2.18% improvement over Swin-Unet for myocardium segmentation and a 0.42% improvement for the right ventricle. Although Swin-Unet slightly outperforms U-KABS on the left ventricle, U-KABS shows a more balanced performance across all the three structures.

Table 2.3: Quantitative comparison of our model with benchmark baselines on the ACDC dataset in terms of DSC, highlighting per-organ performance for the right ventricle (RV), myocardium (Myo), and left ventricle (LV), along with the average DSC scores.

	ACDC			
Method	RV	Myo	LV	Avg (%)
ViT-CUP [18]	81.46	70.71	92.18	81.45
Swin-Unet [11]	88.55	85.62	**95.83**	90.00
UNETR [12]	85.29	86.52	94.02	88.61
MCFormer [32]	87.60	79.20	93.40	87.70
FCT [33]	86.10	81.90	94.80	87.60
TransUNet [10]	88.86	84.53	95.73	89.71
Ours	**88.97**	**87.80**	94.76	**90.53**

Qualitative Results. To assess the effectiveness of U-KABS qualitatively, we conduct visual comparisons of results on the BUSI, ISIC 2018, and ACDC datasets. Figure 2.2 presents a visual comparison of segmentation outputs generated by U-KABS and baselines on the BUSI dataset using heatmaps. The figure includes three representative cases, each displaying the input ultrasound image, the ground truth mask, and segmentation heatmaps. The ground truth mask serves as the reference for evaluating segmentation accuracy. U-KABS demonstrates superior alignment with the ground truth, accurately capturing lesion boundaries with minimal false positives or under-segmented regions, even in challenging cases with low-contrast or noisy ultrasound images. For instance, in the top row, U-KABS precisely delineates irregular lesion contours, whereas U-Net and UNet++ produce diffused heatmaps with over-segmentation around lesion edges, leading to higher false positives. Similarly, in the middle row, U-KABS captures fine-grained boundary details. In the bottom row, U-KABS outperforms U-KAN by reducing under-segmentation in heterogeneous lesion regions, attributed to the hybrid KABS block combining Bernstein polynomials for global smoothness and B-splines for local adaptability. Attention U-Net and Rolling-Unet show improved boundary focus but still exhibit artifacts, such as jagged edges, compared to U-KABS's smooth delineations. These qualitative results highlight U-KABS's robustness in handling ultrasound-specific challenges, such as low contrast, supporting its quantitative superiority (e.g., 80.37% DSC on BUSI, Table 2.1). The integration of squeeze-and-excitation mechanisms enhances feature prioritization, enabling U-KABS to emphasize salient lesion regions over background noise, as evidenced by the focused heatmaps.

Figure 2.3 illustrates the qualitative segmentation performance of U-KABS on representative dermoscopic images from the ISIC 2018 skin lesion dataset, highlighting its ability to delineate lesion boundaries accurately. The figure includes the input dermoscopic image, the corresponding ground truth mask, and segmentation outputs for U-KABS and baselines. U-KABS most accurately matches the ground truth, outlining lesions with more precise boundaries. These qualitative results highlight U-KABS's robustness in addressing ISIC 2018's challenges, such as blurry boundaries, diverse lesion appearances, and imaging artifacts, reinforcing its quantitative superiority.

Figure 2.4 shows the visualization performance of our model compared to strong baselines on the ACDC dataset. The visualizations reveal that U-KABS achieves more precise delineation of the LV, Myo, and RV boundaries, with fewer segmentation artifacts and improved boundary adherence compared to TransUNet, Swin-Unet, and UNETR. Specifically, U-KABS accurately captures the thin, complex Myo boundaries and the irregular RV contours, even in regions with low contrast or motion blur, resulting in balanced performance. While TransUNet and Swin-Unet exhibit accurate LV segmentation, they over-segment the Myo, producing thickened boundaries that encroach into adjacent RV or LV regions, as evidenced by excess Myo pixels in the segmentation maps. This over-segmentation is attributed, in part, to their Transformer-based architectures prioritizing global context over local boundary precision. UNETR struggles with significant RV over-segmentation, with erroneous pixels extending into the surrounding tissue.

2.4.3 Model Efficiency Analysis

Figure 2.5 presents a circle chart comparing the segmentation performance and model efficiency of U-KABS against state-of-the-art baselines on the BUSI dataset, evaluated using the DSC metric against floating-point operations per second (FLOPS, in gigaflops). The circle size represents the number of learnable parameters, reflecting model size. U-KABS achieves a superior DSC of 80.37% with a moderate 9.54M parameters, striking a good balance between performance and efficiency, as visualized in the chart. Compared to lightweight models, UNeXt and MedT offer

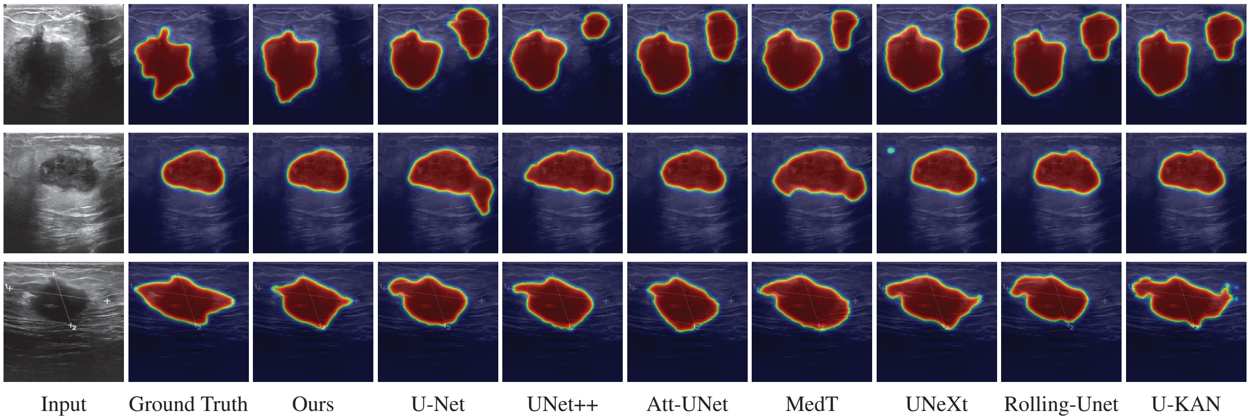

Figure 2.2: Qualitative evaluation of our method using heatmaps compared to benchmark baselines on the BUSI dataset.

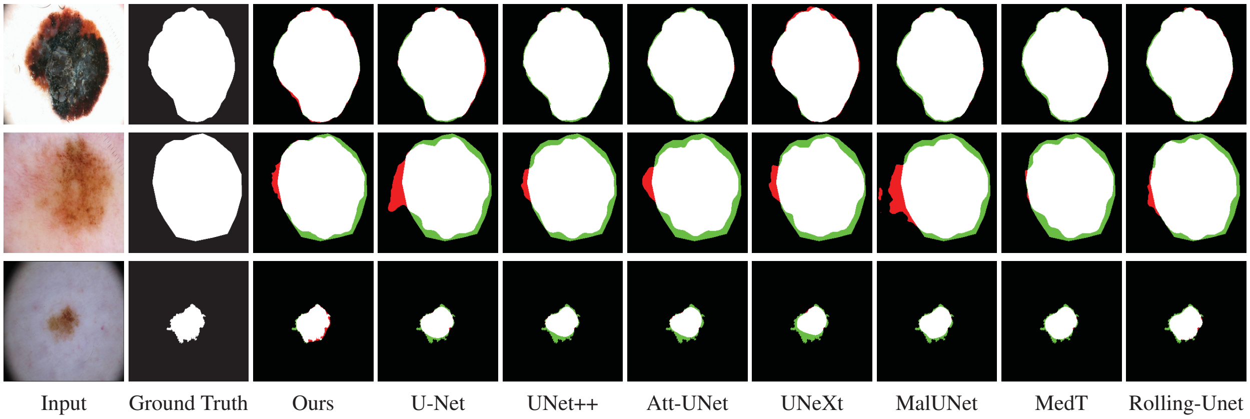

Figure 2.3: Qualitative comparison of our method with benchmark baselines on the ISIC 2018 dataset, highlighting segmentation errors. White, red, and green regions represent predicted segmentation, over-segmentation, and under-segmentation, respectively

efficiency but exhibit under-segmentation due to limited expressiveness of their MLP-based architectures. U-KABS's moderate parameter count is justified by its performance gains, driven by the hybrid design of the KABS block, which integrates Bernstein polynomials for global smoothness and B-splines for local adaptability.

2.4.4 Ablation Study

Table 2.4 assesses the impact of the core components of our network architecture on model performance using the BUSI dataset. The complete model, incorporating all components, yields the best performance. Excluding the KAB layer leads to the most significant performance decline, with IoU and DSC scores dropping to 61.74% and 75.98%, respectively, underscoring its pivotal role as a global approximator. In contrast, excluding the KAS layer while retaining its KAB counterpart

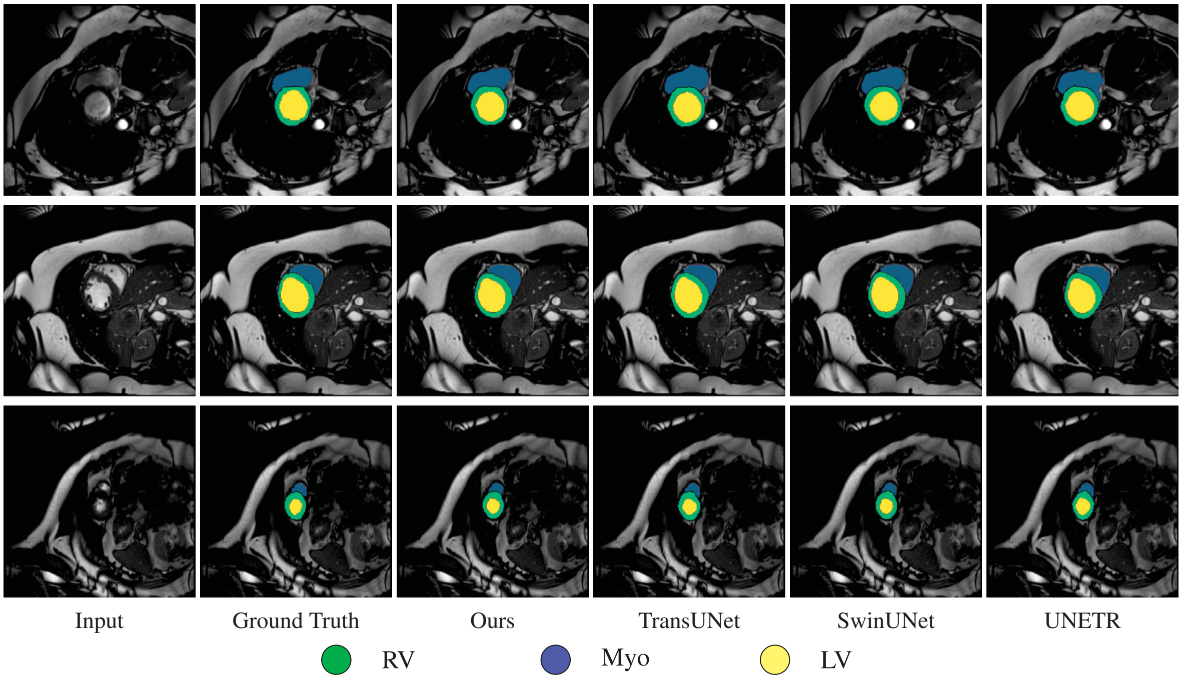

Figure 2.4: Qualitative comparison of our model with baselines on the ACDC dataset, where RV, Myo, LV represent Right Ventricle, Myocardium and Left Ventricle, respectively.

results in superior performance. Omitting other components reduces IoU by 0.91-6.24% and DSC by 0.48-4.39%, indicating their substantial contributions. Notably, the global approximation capability of the KAB layer, when combined with the local adaptability of other components, leads to a synergistic improvement in segmentation accuracy.

Table 2.4: Effect of individual core components on model performance using the BUSI dataset.

Component					Metric (%)	
KAB	KAS	Tokenize	SE	DWConv	IoU	DSC
✗	✓	✓	✓	✓	61.74	75.98
✓	✗	✓	✓	✓	64.28	77.92
✓	✓	✗	✓	✓	62.48	76.25
✓	✓	✓	✗	✓	65.61	78.34
✓	✓	✓	✓	✗	67.07	79.89
✓	✓	✓	✓	✓	**67.98**	**80.37**

Effect of Squeeze-and-Excitation Mechanism. Table 2.5 presents the impact of incorporating squeeze-and-excitation (SE) mechanisms within the convolutional stage of the U-KABS framework, evaluated on the BUSI dataset using the IoU and DSC metrics. The table compares U-KABS with and without SE in its convolutional stages, demonstrating that the inclusion of SE improves

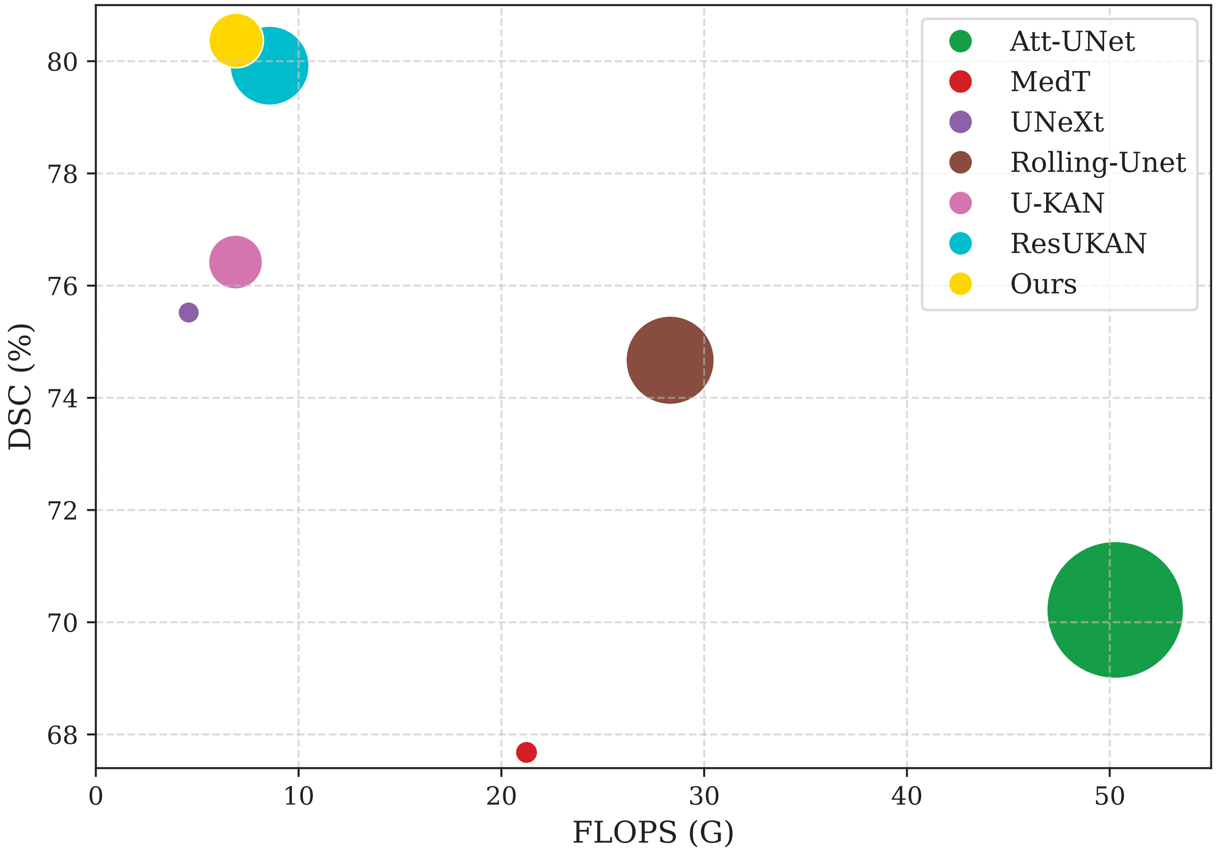

Figure 2.5: Comparison of performance and model efficiency on the BUSI dataset. Our U-KABS model is benchmarked against state-of-the-art methods, including Attention-UNet [1], MedT [2], UNeXt [3], Rolling-Unet [4], U-KAN [5], and ResU-KAN [6]. Performance is assessed using the DSC metric, where higher values indicate superior segmentation performance. The size of the circle indicate the number of learnable parameters..

both metrics. These gains highlight the SE mechanism's role in enhancing U-KABS's segmentation performance, particularly for challenging ultrasound images characterized by low contrast, and heterogeneous lesion textures.

Table 2.5: Effect of squeeze-and-excitation (SE) on model performance using the BUSI dataset.

Convolutional Stage	IoU	DSC
Without SE	66.91	79.73
With SE	**67.98**	**80.37**

Effect of Bernstein Polynomial Order. Table 2.6 evaluates the impact of varying the Bernstein polynomial order in the KAB layer of U-KABS on segmentation performance, assessed on the BUSI dataset using the IoU and DSC metrics. The Bernstein polynomial order in the KAB layer defines the degree of the basis polynomials, which govern the global smoothness of feature approximations in U-KABS. Higher orders increase the polynomial's capacity to model complex functions but risk overfitting or numerical instability due to increased parameters. The table shows that order 4 achieves the best performance, indicating a good balance between expressiveness and stability for ultrasound image segmentation thanks, in large part, to its ability to capture smooth, continuous lesion shapes in BUSI images. This polynomial order effectively models global anatomical trends, such as heterogeneous lesion textures, without over-smoothing or underfitting, unlike order 3, which underfits due to limited flexibility, resulting in under-segmentation. Order 5's slight performance drop suggests overfitting or excessive smoothness, reducing precision for fine-grained boundaries, critical for ultrasound images with low contrast.

Table 2.6: Effect of Bernstein polynomial order on model performance using the BUSI dataset.

Bernstein Polynomial Order	IoU	DSC
2	65.92	79.08
3	65.80	78.99
4	**67.98**	**80.37**
5	65.34	78.67

Effect of Spline Order and Grid Size. Table 2.7 investigates the impact of varying the B-spline order and grid size in the KAS layer of the U-KABS framework on segmentation performance, evaluated on the BUSI dataset using the IoU and DSC metrics. The spline order controls the polynomial degree, affecting smoothness and flexibility, while grid size governs the granularity of local approximations, with larger grid sizes increasing the number of basis functions for finer control. The table shows that the combination of spline order 3 and grid size 5 achieves the best performance, demonstrating its effectiveness in capturing fine-grained lesion boundaries in BUSI images. The KAS layer's B-splines complement the KAB layer's Bernstein polynomials, which provide global smoothness, by refining local details critical for ultrasound challenges low contrast.

2.5 Discussion

As demonstrated in the experiments, U-KABS achieves strong performance across diverse datasets while maintaining computational efficiency. The results highlight its strengths in robust global-local feature modeling, enabled by the hybrid KAN Bernstein spline block, which combines Bernstein polynomials for smooth global approximations and B-splines for precise local refinements. This design addresses key challenges in medical imaging, such as noise in ultrasound and blurry boundaries in dermoscopy, as evidenced by the qualitative results, where U-KABS exhibits min-

Table 2.7: Effect of spline order and grid size on model performance using the BUSI dataset.

Spline Order	Grid Size	IoU	DSC
2	3	66.00	79.17
	4	66.41	79.40
	5	65.15	78.74
3	3	65.98	79.07
	4	67.19	80.10
	5	**67.98**	**80.37**
4	3	64.62	78.13
	4	67.42	79.93
	5	66.97	79.61

imal over- or under-segmentation compared to baselines like U-Net and TransUNet. The ablation studies further underscore these strengths: the KAB layer's global approximation is pivotal, while the KAS layer's local adaptability and squeeze-and-excitation mechanisms enhance feature prioritization, reducing artifacts in noisy regions. Unlike Transformer-based models, U-KABS achieves higher accuracy with substantially lower computational cost. Despite these strengths, U-KABS has two main limitations that warrant consideration. First, its parameter count is higher than lightweight baselines such as UNeXt, potentially limiting deployment on edge devices with memory constraints, although its FLOPS remains competitive. Second, while ablation studies confirm component contributions, broader generalization across unseen modalities (e.g., PET) requires additional validation. These limitations suggest avenues for future enhancements, such as lightweight variants or federated learning integration.

CHAPTER 3

Adaptive Kolmogorov-Arnold Network for Medical Image Segmentation

Medical image segmentation is essential for accurate diagnosis and treatment planning, yet it remains a challenging task due to the diversity of medical imaging modalities and the difficulty in capturing non-linear relationships within these images. In this chapter, we introduce an Adaptive Kolmogorov-Arnold Network (AdaKAN), a U-shaped model architecture comprised of a convolutional stage and an efficient KAN (EffiKAN) stage. The two key components of the EffiKAN stage is the efficient attention which enhances global feature extraction across multiple resolutions and the AdaptKAN block which further consists of two sub-branches where the right one consisting of a KAN layer with Bernstein polynomials as its basis activation function to provide a globally stable and smooth approximation and the left branch constitutes of the up and down projections to handle channel-wise representations. This adaptive design effectively captures higher-order dependencies which are crucial in segmenting complex structures. Skip-connections ensure to preserve details during the downsampling process from the encoder and pass it up to effectively restore spatial details during the decoding process. Through extensive experiments, we evaluate the proposed framework on five diverse medical imaging benchmark datasets, demonstrating strong performance over competing baseline models, particularly in segmenting complex anatomical structures.

3.1 Introduction

The aim of medical image segmentation in delineating anatomical structures and pathological regions from diverse imaging modalities such as ultrasound, MRI, and dermoscopy. Accurate pixel-wise labeling of organs, lesions, and tissues is essential for diagnosis, treatment planning, and disease progression monitoring. The evolution of deep learning has significantly advanced this field, with Fully Convolutional Networks (FCNs) forming the cornerstone of modern approaches. Among these, U-Net [8] and its variants including UNet++ [9], Attention U-Net [1], UNet3+ [20], V-Net [22], and Ki-UNet [23] leverage hierarchical convolutional feature extraction within a symmetric encoder-decoder framework to model local spatial patterns effectively. Despite their success, these models are inherently limited in capturing long-range dependencies due to the localized receptive field of convolutions, often resulting in blurred boundaries or incomplete segmentation of complex structures.

To overcome these shortcomings, Transformer-based architectures such as MedT [2], Swin-Unet [11], TransUNet [10], and UNETR [12] have introduced self-attention mechanisms to integrate global context. While effective, their quadratic computational complexity restricts deployment in resource-constrained clinical environments. Lightweight MLP-based architectures such as UNeXt [3] and MALUNet [17] emphasize computational efficiency but often lack the expressive capacity required to accurately segment irregular lesion boundaries. Recently, Kolmogorov-Arnold Networks (KANs) [13] have been introduced as a compelling alternative to traditional MLPs, employing learnable activation functions along edges to enhance interpretability and capture complex non-linear relationships. In continuation, U-KAN [5] adopts B-spline-based KAN layers within a U-shaped network, successfully modeling local features but still limited in representing global contextual information. The above mentioned limitations underscore the need for a balanced framework that unifies local precision, global awareness, and computational efficiency across diverse medical imaging scenarios.

To this end, inspired by the Adpatformer [37] we introduce an Adaptive Kolmogorov-Arnold Network (AdaKAN) featuring an adaptive KAN block (AdaptKAN) in conjunction with Efficient Attention to effectively capture broad contextual dependencies. Within the EffiKAN module, a dual-branch design is employed, where the right branch consists of a KAN layer utilizing Bernstein polynomials as basis functions for global adaptability, and the left branch incorporates projection operations combined with ReLU activation and an adaptive scaling coefficient for fine-grained feature refinement. The main contributions of this work can be summarized as follows:

- We propose a novel AdaptKAN block that seamlessly integrates Efficient Attention with an

Adaptive Kolmogorov-Arnold Network (AdaptKAN) block. This design enables the model to effectively capture global contextual dependencies while maintaining computational efficiency.

- We introduce a dual-branch AdaptKAN module, where the right branch uses a KAN layer with Bernstein polynomials for global adaptability, and the left branch applies projection operations with ReLU and an adaptive scaling coefficient for local refinement, enhancing both global smoothness and local precision.
- We validate the effectiveness of our model by conducting comprehensive experiments on five diverse benchmark datasets, outperforming state-of-the-art methods and exhibiting robustness to blurry boundaries and multi-class tasks.

3.2 Related Work

FCN-based Methods: Fully Convolutional Networks (FCNs) have long served as the cornerstone of medical image segmentation due to their ability to preserve spatial hierarchies through end-to-end pixel-wise prediction. Among attention-augmented FCNs, Attention U-Net [1] introduces gated signal propagation to suppress irrelevant regions and emphasize salient features during upsampling. MALUNet [17] further advances this paradigm by integrating multi-scale attention mechanisms to capture cross-resolution contextual dependencies. nnUNet [38] establishes a robust empirical benchmark via automated configuration and ensemble strategies tailored to diverse datasets. More recently, SelfRegularized-UNet [39] incorporates self-supervised consistency constraints to enhance generalization under limited annotations. Despite their strength in local feature extraction and hierarchical modeling, FCN-based approaches remain inherently limited in capturing long-range spatial relationships due to the locality of convolutional operations.

Transformer- and MLP-based Methods: Drawing inspiration from Vision Transformers [18], Transformer-based models have been recently adapted for medical image segmentation to overcome the limitations of FCN-based approaches in capturing global dependencies. TransUNet [10] integrates Transformer blocks with U-Net architecture, leveraging self-attention and cross-attention mechanisms to enhance performance. Swin-Unet [11] utilizes shifted window-based self-attention to model long-range dependencies efficiently. UNETR [12] reformulates volumetric segmentation as a sequence-to-sequence task using pure Transformer encoders. MedSAM [40] extends segment-anything principles to medical domains via prompt-driven adaptation. Lightweight alternatives include MobileUNETR [41], which optimizes Transformer efficiency for edge deployment, and QMaxViT-UNet+ [42], which combines quantized attention with ViT backbones. MLP-based methods offer further efficiency gains: UNeXt [3] introduces tokenized MLP blocks with shifted operations to balance local and global modeling, while Rolling-Unet [4] employs directional rolling MLPs to capture multi-axis long-range interactions. Despite their computational advantages, both Transformer- and MLP-based methods often trade off expressive capacity or require extensive pre-training to match the nuanced feature representation needed in medical imaging.

KAN-based Methods. KANs [13] have recently emerged as a powerful alternative to MLPs, leveraging compositions of learnable univariate functions to approximate complex multivariate relationships with enhanced interpretability and flexibility. Building upon this foundation, U-KAN [5] adapts the U-Net architecture by replacing convolutional layers with tokenized KAN blocks using B-spline activations, achieving improved segmentation accuracy through expressive non-linear modeling. ResU-KAN [6] further extends U-KAN by integrating residual convolutional attention and atrous spatial pyramid pooling to expand receptive fields and capture multi-scale contextual dependencies. PDS-UKAN [43] introduces a progressive down-sampling strategy within a KAN-based encoder-decoder framework to preserve fine-grained spatial details across scales. TP-FIANet [31] combines KANs with dual-path feature interaction to balance local and global feature learning. Despite these advances, existing KAN-based methods often rely heavily on B-spline parameterizations, which, while smooth, limit adaptive expressiveness and struggle to efficiently integrate long-range contextual modeling with lightweight computation. However, their dependence on B-splines limits the model's capacity to maintain an effective balance between global contextual understanding and local feature representation. In contrast, the proposed AdaKAN framework introduces the Efficient Kolmogorov-Arnold Network (EffiKAN) block, which seamlessly integrates Efficient Attention with the Adaptive Kolmogorov–Arnold Network (AdaptKAN). Within this block, the KAN layer employs Bernstein polynomials as its basis activation, enabling enhanced global adaptability and smooth functional approximation. This design allows the network to effectively balance global context modeling with local feature refinement, fostering robust information propagation across encoder–decoder stages. By integrating these components within a symmetric U-Net-inspired architecture, AdaKAN achieves better segmentation performance while maintaining computational efficiency, making it highly suitable for the nuanced demands of medical image segmentation.

3.3 Proposed Method

In this section, we outline our approach for semantic segmentation task, introducing the key components of the proposed method, along with a concise overview of the KANs.

Problem Description. Given an input image $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ with height H and W width, our goal is to generate pixel-wise segmentation mask $\mathbf{M} \in C^{H \times W}$, where $[C] = \{0, \ldots, C-1\}$ and C is the number of semantic classes. Each pixel in $\mathbf{M}$ is assigned an integer label from 0 to $C-1$, denoting the class of each pixel.

Kolmogorov-Arnold Networks. KANs are based on the Kolmogorov-Arnold representation theorem [25, 26], which asserts that any continuous multivariate function defined on a bounded domain can be expressed as a finite combination of continuous univariate functions of the input variables, combined using the binary operation of addition. The KAN layer is a fundamental element of KANs [13], represented by a matrix $\mathbf{\Phi} = (\phi_{q,p})$ of one-dimensional functions, where each trainable activation function ϕ is expressed as a weighted combination of a sigmoid linear unit (SiLU) function and a spline function with adjustable weights.

$$\phi(x) = w_b \text{SiLU}(x) + w_s \text{spline}(x), \tag{3.1}$$

where $\text{spline}(x) = \sum_i c_i B_i(x)$ represents a weighted sum of B-spline basis functions with trainable coefficients c_i. During the training process, the weights w_b and w_s are optimized to enhance performance. For an input feature vector $\mathbf{x}^{(l)} \in \mathbb{R}^{F_l}$, the output of the l-th KAN layer is a feature vector of dimension F_{l+1}, computed as:

$$\mathbf{x}^{(l+1)} = \text{KAN}^{(l)}(\mathbf{x}^{(l)}). \tag{3.2}$$

Architecture Overview. Figure 3.1 outlines the comprehensive structure of the AdaKAN framework, which features a symmetric U-shaped encoder-decoder configuration incorporating skip connections to preserve spatial information despite downsampling. The process comprises two main stages: the Convolution Stage and the EffiKAN Stage. The Convolution Stage, with three consecutive blocks, extracts local features such as textures and edges. The EffiKAN Stage follows, employing layer normalization, Efficient Attention, residual connections with an AdaptKAN Block, which utilizes Bernstein Polynomials as its basis activation function inside the KAN layer to capture global non-linear relationships.

3.3.1 Encoder

The encoder is composed of the following two stages:

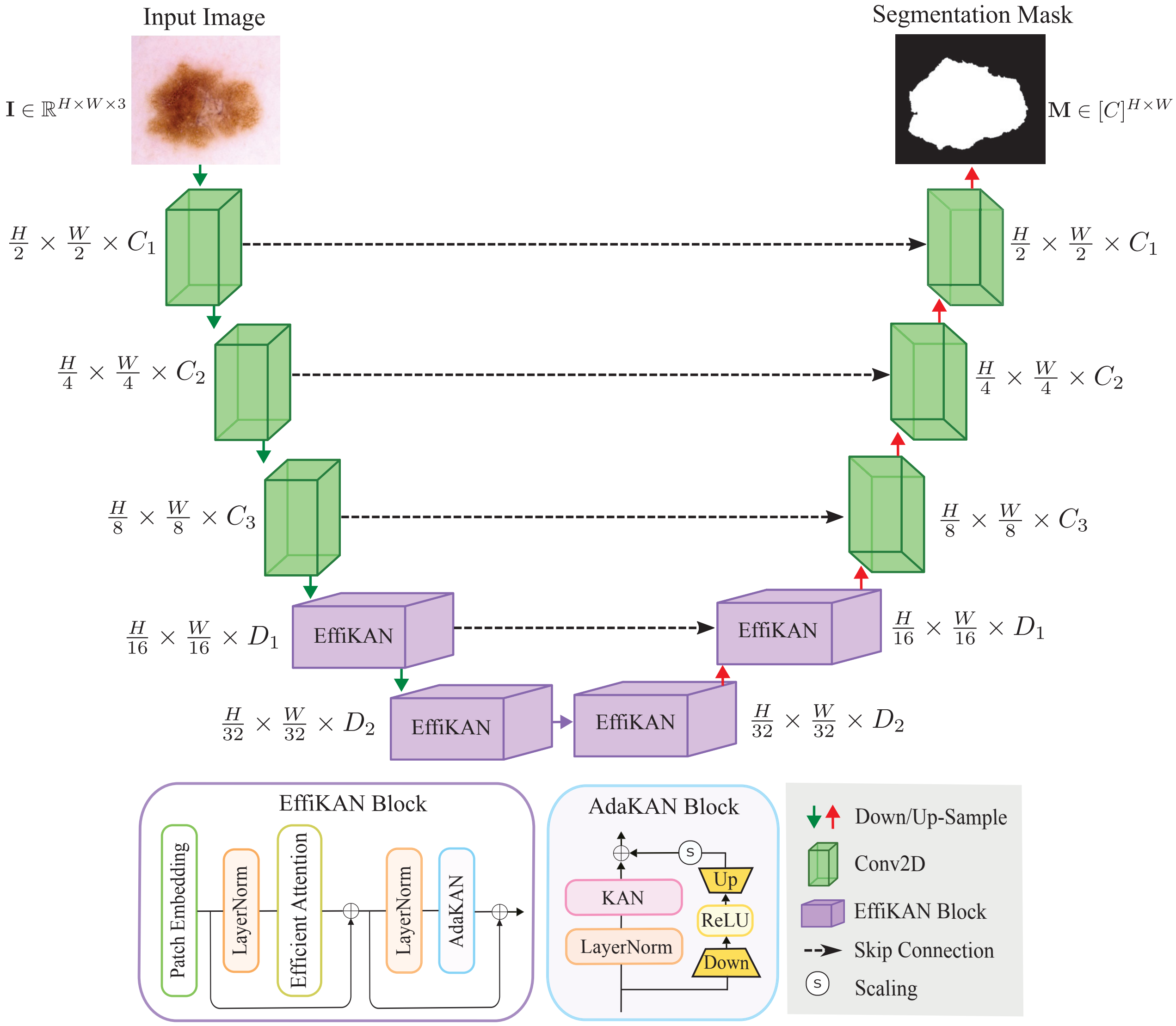

Figure 3.1: Overall architecture of the proposed AdaKAN framework. The model follows a symmetrically U-shaped encoder-decoder structure, integrating Convolutional 2D (Conv2D) and EffiKAN stages. The encoder downsamples the input image through three Conv2D stages followed by two EffiKAN Blocks, progressively reducing spatial resolution while expanding feature channels to capture hierarchical representations. The decoder mirrors this design, upsampling features via bilinear interpolation and integrating high-resolution details from the encoder through skip connections to produce a precise segmentation mask.

Convolutional stage. In this stage we have L Conv2D blocks each of which consists of 3×3 convolutional layer, batch normalization and a ReLU activation function. The process of this block can be expressed as

$$\mathbf{X}^{(\ell)} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}(\mathbf{X}^{(\ell-1)}))) \tag{3.3}$$

where $\mathbf{X}^{(\ell)}$ represents the output feature map of the ℓ-th Conv2D block for $\ell \in 1, \ldots, L$, where $\mathbf{X}^{(0)} = \mathbf{I}$ is the original input image. After each block, a 2×2 max-pooling layer is applied to downsample the feature map. The output feature map $\mathbf{X}^{(L)}$ from the final Conv2D block is passed on to the first block of the subsequent EffiKAN stage.

EffiKAN Stage. This stage comprises K blocks, each featuring a patch embedding layer, a layer normalization, efficient attention with a residual connection, a second layer normalization, and an AdaKAN Block.

Patch Embedding. The output from the Lth convolutional stage, $\mathbf{X}^{(L)} \in \mathbb{R}^{\frac{H}{2^L} \times \frac{W}{2^L} \times C_L}$, serves as the input to the EffiKAN block. It is converted into a tokenized input feature matrix $\mathbf{Y} \in \mathbb{R}^{n \times d}$, where n represents the number of patches and d denotes the embedding dimension. This layer follows the method used in ViT [18], where input feature map is divided into non-overlapping flattened patches and then projected into a latent embedding space using a learnable linear projection.

Layer Normalization and Efficient Attention. The tokenized input feature matrix $\mathbf{Y}$ first passes through the Layer Normalization layer which ensures stable distribution of features keeping the same dimension. The output from the layer normalization $LN(\mathbf{Y}) \in \mathbb{R}^{n \times d}$ acts as the input feature matrix to the Efficient attention layer. The efficient attention layer enables global feature interaction with linear computational complexity, in contrast to the quadratic computational complexity standard attention mechanisms. Given an input feature matrix $\mathbf{G} \in \mathbb{R}^{n \times d}$, where $\mathbf{G} = \mathrm{LN}(\mathbf{Y})$. The efficient attention mechanism comprises of three linear projections to generate the query $\mathbf{Q} \in \mathbb{R}^{n \times d_k}$, key $\mathbf{K} \in \mathbb{R}^{n \times d_k}$, and value $\mathbf{V} \in \mathbb{R}^{n \times d_k}$ matrices, where d_k is the dimensionality of each projected feature. The efficient attention (EA) operation is then formulated as

$$EA(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \sigma_{\text{row}}(\mathbf{Q})(\sigma_{\text{col}}(\mathbf{K})^T \mathbf{V}) \tag{3.4}$$

where σ_{row} and σ_{col} represent the row-wise and column-wise softmax functions respectively. The output of the efficient attention layer, $\mathbf{EA}(\mathbf{LN}(\mathbf{Y})) \in \mathbb{R}^{n \times d}$, passes through a residual connection that adds the initial feature $\mathbf{Y}$ which can be expressed as

$$\mathbf{Z} = \mathbf{Y} + \mathrm{EA}(\mathrm{LN}(\mathbf{Y})) \tag{3.5}$$

where $\mathbf{Z} \in \mathbb{R}^{n \times d}$ acts as the input to the AdaptKAN block.

AdaptKAN Block. The Adaptive Kolmogorov-Arnold Network (AdaptKAN) Block is a core component of the AdaKAN architecture, designed to enhance representational power while preserving computational efficiency. The AdaptKAN block comprises of two sub-branches the left branch integrates layer normalization and a KAN Layer with bernstein polynomials as its basis activation function and the right branch uses up and down projections with ReLU sequentially and an adaptive scaling coefficient for task-specific fine tuning.

We use Bernstein polynomials as the basis activation function inside the KAN layer which is

$$\begin{aligned} \text{Bernstein}(z) &= \sum_{r=0}^{R} \theta_r P_{r,R}(z) \\ &= \sum_{r=0}^{R} \theta_r \binom{R}{r} z^r (1-z)^{R-r}, \end{aligned} \tag{3.6}$$

where $P_{r,R}(z)$ is the r-th Bernstein basis polynomial of order R and θ_r are learnable coefficients. Mathematically, the KAN layer is defined over an interval $[0, 1]$ using a a weighted combination of a SiLU function and a Bernstein polynomial function:

$$\tilde{\phi}(z) = \tilde{w}_b \, \text{SiLU}(z) + \tilde{w}_s \, \text{Bernstein}(z), \tag{3.7}$$

the choice of Bernstein polynomials as its basis activation function which is driven by their globally stable and smooth approximation properties. These polynomials, globally defined, can approximate any continuous function on an interval of $[0, 1]$. The equation for the AdaptKAN block can be expressed as

$$\tilde{\mathbf{Z}} = \text{ReLU}\big(\mathbf{Z} \cdot W_{\text{down}}\big) \cdot W_{\text{up}} \tag{3.8}$$

$$\text{AdaptKAN}(\mathbf{Z}) = \text{KAN}\big(\text{LN}(\mathbf{Z}\big) + s \cdot \tilde{\mathbf{Z}} \tag{3.9}$$

where $\mathbf{Z} \in \mathbb{R}^{n \times d}$ denotes the input feature matrix obtained from the residual connection applied after the efficient attention layer. The matrix $\tilde{\mathbf{Z}}$ represents the intermediate feature matrix obtained from the right branch. The parameters W_{down} and W_{up} correspond to learnable down-projection and up-projection weight matrices, respectively, which are applied sequentially with a ReLU activation to capture nonlinear channel dependencies. A learnable adaptive scaling coefficient s modulates the contribution of the right branch before both branches are combined to using the residual connection.

By combining these components, the EffiKAN block is able to capture both local and global dependencies in a computationally efficient manner. Patch Embedding and LayerNorm provide a stable input representation, Efficient Attention ensures long-range contextual reasoning at linear-time complexity, and AdaptKAN enriches the representation with global function approximation and multi-scale feature fusion. This makes the EffiKAN block efficiently capture non-linear features with , making it well-suited for high-resolution segmentation tasks. The final equation for this block can be written as:

$$\text{EffiKAN}(\mathbf{Y}) = \text{AdaptKAN}(\mathbf{Z}) + \mathbf{Z} \tag{3.10}$$

The bottleneck acts as the bridge between the encoder and decoder, enabling deep feature extraction before the upsampling process begins.

3.3.2 Decoder

The AdaKAN decoder functions as a symmetric counterpart to the encoder, reconstructing the encoded feature representations into a segmentation mask $\mathbf{M} \in [C]^{H\times W}$, where C denotes the total number of semantic classes in the input image. The segmentation mask emerges from a structured approach of restoring the spatial details and refine the features.

The decoder is organized into two main stages, EffiKAN and Conv2D, following a structure that mirrors the encoder for architectural coherence. In the EffiKAN stage, each of the two blocks begins by upsampling the input feature maps using bilinear interpolation by a factor of two, enabling a smooth and gradual increase in spatial resolution. The upsampled maps are then concatenated with the corresponding encoder features through skip connections, which help preserve fine-grained spatial details that may have been lost during downsampling. This skip connection mechanism also reduces the semantic gap between encoder and decoder features, thereby improving boundary precision and localization accuracy.

The subsequent Conv2D stage comprises three additional upsampling blocks that further refine and reconstruct the feature representations. Each block employs bilinear interpolation and concatenation with encoder features, progressively enhancing the spatial and semantic consistency of the decoded maps. Finally, a 1×1 convolutional layer with C output channels is applied to project the refined features into the class space, followed by a softmax activation to produce pixel-wise class probabilities. The segmentation mask $\mathbf{M}$ is obtained by taking the argmax across the channel dimension, assigning each pixel to the class with the highest probability.

3.4 Experiments

3.4.1 Experimental setup

Datasets. We evaluate and assess the effectiveness of the proposed image segmentation model across five distinct datasets: Breast Ultrasound Images (BUSI) [27], Gland Segmentation GlaS [28], CVC [44], International Skin Imaging Collaboration (ISIC 2018) [29], and and Automated Cardiac Diagnosis Challenge (ACDC) [30].

- **BUSI** contains ultrasound images of normal, benign, and malignant breast cancer cases, along with corresponding ground truth segmentation masks. Following [6], we use 647 benign and malignant images, resized to 256×256.
- **GlaS** includes 165 histological images of colorectal tissue, labeled as benign or malignant, with segmentation masks. Variations in structural complexity pose significant challenges for gland segmentation. Following [4, 5], all images are resized from 384×288 to 512×512.
- **CVC-ClinicDB** comprises 612 images, each with a resolution of 384×288, sourced from 31 colonoscopy procedures, for polyp identification in colonoscopy footage. Following [5], all images are resized to a uniform resolution of 256×256.
- **ISIC** comprises 2594 dermoscopic images of skin lesions, each accompanied by binary segmentation masks. Challenges arise from blurred lesion boundaries and the diversity of lesion types (including benign, nevi, seborrheic keratoses, and melanomas). Following [4, 5], all images are resized to 256×256.
- **ACDC** includes heart MRI scans with 5–8 mm slice thickness and 0.83–1.75 mm²/pixel resolution. Illustrating cardiac structures across short-axis slices from the base to the apex of the Left Ventricle (LV), Right Ventricle (RV), Myocardium (Myo), and background.

Evaluation Metrics. To evaluate the performance of AdaKAN against state-of-the-art (SOTA) methods, we employ two standard evaluation metrics commonly used for training and testing [45]: Intersection over Union (IoU) and Dice Similarity Coefficient (DSC). IoU measures the overlap between the predicted (P) and ground truth (G) segmentation masks as follows:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|}, \tag{3.11}$$

and DSC quantifies how well the predicted mask matches the ground truth mask:

$$DSC = \frac{2 \times |P \cap G|}{|P| + |G|}, \tag{3.12}$$

Baselines. We evaluate the performance of AdaKAN by comparing it against SOTA methods, including Att-UNet [1], MedT [2], UNeXt [3], Rolling-Unet [4], U-KAN [5], TPFIANet [31], PDS-UKAN [43], ResU-KAN [6], MALUNet [17], MobileUNETR [41], TransUNet [10], Swin-Unet [11], UNETR [12], UTNet [46], MISSFormer [47], CycleMix [48], MixFormer [49], QMaxViT-UNet+ [42], ScribFormer [50], nnUNet [38], nnFormer [51], EFEH-Net [52], T-ACD [53], DFEDC [54], TransCASCADE [55], SelfRegularized-UNet [39], PVT-EMCAD-B2 [56], SET [57], SCM-UNET [34], WTCM-UNet [35], U-NET V2 [58], CTO [36], and Med-SAM [40].

Implementation Details. All experiments are conducted using PyTorch on PyTorch on two Linux systems, each equipped with a single GPU: an NVIDIA GeForce RTX 3080 (10 GB memory) and an NVIDIA GeForce RTX A4500 (20 GB memory). For the binary segmentation datasets (BUSI, CVC-ClinicDB, and ISIC 2018), we set a batch size of 8. For the GlaS dataset, we used a batch size of 6, while for the multi-class ACDC dataset, a batch size of 16. The input size set to 256×256 for the BUSI, CVC-ClinicDB, ISIC 2018, and ACDC datasets, while it set to 512×512 for the GlaS dataset. The learning rate set to 10^{-4} across all datasets. Channel configurations are defined as $C_1 = 32$, $C_2 = 64$, $C_3 = 256$, $D_1 = 320$, and $D_2 = 512$. We set the Bernstein polynomial order to 3. The model is trained using the Adam optimizer with a cosine annealing learning rate scheduler, decaying to a minimum of 10^{-5}. Each dataset is randomly split into an 80% training and 20% validation. For binary-class datasets (BUSI, GlaS, CVC-ClinicDB, and ISIC 2018), we employed the average of Binary Cross-Entropy (BCE) and Dice loss. Whereas for the multi-class ACDC dataset, we use the average of Cross-Entropy (CE) and Dice loss. For all datasets, the model is trained for 400 epochs.

3.4.2 Results and Analysis

Table 3.1: Quantitative comparison of our method against benchmark baselines on three heterogeneous medical scenarios using IoU and DSC as evaluation metrics. Best results are in bold, and the second-best results are underlined.

			BUSI		GlaS		CVC	
Method	Params(M) ↓	FLOPS(G)↓	IoU↑	DSC↑	IoU↑	DSC↑	IoU↑	DSC↑
Att-UNet [1]	50.27	60.06	55.18±3.61	70.22±2.88	86.84±1.19	92.89±0.65	84.52±0.51	91.46±0.25
MedT [2]	1.60	21.24	52.15±3.47	67.68±3.18	75.47±3.46	85.92±2.93	60.08±0.92	79.18±1.05
UNeXt [3]	**1.47**	**4.58**	61.78±1.46	75.52±0.91	83.95±1.09	91.22±0.67	74.83±0.24	85.36±0.17
Rolling-Unet [4]	25.06	28.32	61.00±0.64	74.67±1.24	86.42±0.96	92.63±0.62	82.87±1.47	90.48±0.83
U-KAN [5]	9.38	6.889	63.38±2.83	76.42±2.90	87.64±0.32	93.37±0.16	85.05±0.53	91.88±0.29
TPFIANet [31]	–	–	66.20±0.28	75.84±0.26	–	–	84.88±0.37	91.32±0.25
PDS-UKAN [43]	–	–	62.74± -	76.99± -	88.55± -	93.93± -	–	–
ResU-KAN [6]	20.06	8.57	67.74±1.35	79.92±0.73	87.99±0.44	93.61±0.24	85.38±0.87	92.12±0.56
RWKV-UNet [7]	17.40	14.64	60.51±1.46	75.49±0.45	84.51±3.97	91.04±2.12	75.79±0.85	82.96±0.55
Ours	14.21	8.12	**68.84±1.04**	**80.95±0.75**	**88.58±0.21**	**93.94±0.13**	**85.73±0.49**	**92.26±0.29**

Quantitative Results on BUSI, GlaS, and CVC. Table 3.1 compares our AdaKAN model with SOTA methods on the BUSI, GlaS, and CVC datasets, assessing model size (parameters), computational complexity (FLOPS, in gigaflops), and segmentation performance. AdaKAN has a moderate parameter count and FLOPS(G), which is more efficient compared to the strongest baseline ResU-KAN and PDS-UKAN. Our model demonstrates the most effective segmentation performance across three datasets. AdaKAN achieves relative improvements of 1.1% and 1.03% in IoU and DSC, respectively, on the BUSI dataset compared to the second-best method, ResU-KAN. It also exhibits an improvement of 0.03% and 0.01% in IoU and DSC, respectively, on the GlaS dataset compared to PDS-UKAN. Moreover, AdaKAN achieves improvements of 0.35% and 0.14% in IoU and DSC, respectively, on the CVC dataset compared to ResU-KAN.

Table 3.2: Quantitative comparison of our method with benchmark baselines on the ISIC 2018 dataset using IoU and DSC.

Method	IoU	DSC
Att-UNet [1]	78.43	87.91
MedT [2]	81.48	89.49
UNeXt [3]	81.70	89.70
MALUNet [17]	80.25	89.04
Rolling-Unet [4]	83.74	90.90
MobileUNETR [41]	84.56	90.74
SET [57]	84.64	90.98
SCM-UNET [34]	81.88	90.14
WTCM-UNet [35]	80.26	89.05
U-NET V2 [58]	84.15	91.52
CTO [36]	84.00	90.60
MedSAM [40]	77.40	83.60
Ours	**84.88**	**91.71**

Quantitative Results on ISIC. We report the performance comparison of AdaKAN compared to baselines in terms of both IoU and DSC on the ISIC 2018 dataset, as shown in Table 3.2. The results demonstrate that AdaKAN outperforms all baselines and achieves relative improvements of 0.24% in IoU compared to the second-best method, SET, and 0.19% in DSC compared to U-NET V2. Given the difficult dermoscopic images in the ISIC dataset, which have hazy borders and a variety of lesion types, the improvements are modest.

Quantitative Results on ACDC. Table 3.3 presents the performance comparison of AdaKAN against baselines on the ACDC dataset, evaluated using the DSC metric for three cardiac structures (Right Ventricle (RV), Myocardium (Myo), and Left Ventricle (LV)). AdaKAN achieves a 0.88% higher average DSC than PVT-EMCAD-B2, which is the second-best performing baseline. When compared to TransUNet, AdaKAN improves by 3.29% in average DSC, and the gain over UNETR is 4.39%. Notably, AdaKAN achieves improvements of 2.53% in RV compared to the second-best method, nnFormer, and 0.41% in LV compared to the second-best method, PVT-EMCAD-B2. However, for the Myocardium (Myo), AdaKAN shows a relative decrease of 1.16% compared to the best methods DFEDC and TransCASCADE, due to the small size and inherent difficulty in segmenting the Myocardium region in cardiac MRI, which often suffers from low contrast and motion artifacts. Nevertheless, it achieves the highest average DSC with an improvement of 0.88% over the second-best method, PVT-EMCAD-B2, and shows a more balanced performance across all three structures.

Table 3.3: Performance comparison of our model and baseline methods on the ACDC dataset using DSC as the evaluation metric for the Right Ventricle (RV), Myocardium (Myo), and Left Ventricle (LV).

	ACDC			
Method	RV	Myo	LV	Avg (%)
Rolling-UNet [4]	88.00	89.00	93.36	90.12
TransUNet [10]	88.86	84.53	95.73	89.71
Swin-Unet [11]	88.55	85.62	95.83	90.00
UNETR [12]	85.29	86.52	94.02	88.61
UTNet [46]	83.78	87.34	94.90	88.67
MISSFormer [47]	86.36	85.75	91.59	87.90
LeViT-UNet [59]	89.55	87.64	93.76	90.32
CycleMix [48]	88.20	85.80	91.90	88.63
MixFormer [49]	89.02	88.46	95.55	91.01
QMaxViT-UNet+ [42]	84.50	86.80	91.90	87.60
ScribFormer [50]	84.50	84.80	90.30	86.50
nnUNet [38]	90.24	89.24	95.36	91.61
nnFormer [51]	90.94	89.58	9.65	92.06
EFEH-Net [52]	90.70	88.09	94.70	91.17
T-ACD [53]	90.68	89.56	95.81	92.02
DFEDC [54]	89.79	**90.25**	95.97	91.89
TransCASCADE [55]	89.14	**90.25**	95.50	91.63
SelfRegularized-UNet [39]	88.92	89.49	95.88	91.43
PVT-EMCAD-B2 [56]	90.65	89.68	96.02	92.12
Ours	**93.47**	89.09	**96.43**	**93.00**

Qualitative Results. To assess the effectiveness of AdaKAN qualitatively, we conduct visual comparisons of results on the BUSI, GlaS, CVC, ISIC 2018, and ACDC datasets. Figure 3.2 presents a qualitative comparison of segmentation outputs generated by AdaKAN and baselines on the BUSI dataset using heatmaps. The figure includes four representative cases, each displaying the input ultrasound image, the ground truth mask, and segmentation heatmaps. AdaKAN shows improved shape preservation compared to other methods, even with diverse and irregular lesion shapes, achieving high accuracy despite these variations. However, the Intersection over Union (IoU) remains low across all methods due to the dataset's challenging characteristics, such as varied lesion sizes and textures. U-KAN offers reasonable results but shows minor over-segmentation and under-segmentation along the edges. Att-UNet, MedT, and UNeXt capture the general lesion structure yet suffer from over-segmentation and under-segmentation and noticeable boundary errors. Rolling-UNet and U-KAN struggle to delineate lesion contours accurately. ResU-KAN approximates lesion shapes but fails to preserve fine boundary details, leading to considerable over-segmentation.

Figure 3.3 illustrates the qualitative segmentation performance of AdaKAN on the Gland Segmentation in Colon Histology Images Challenge (GlaS) dataset, highlighting its ability to delineate gland boundaries accurately. The predicted segmentations are shown in white, under-segmentation errors in green, and over-segmentation errors in red. AdaKAN produces the most precise segmentation, with minimal false positives and false negatives. Its outputs closely align with the ground truth boundaries, particularly for complex and irregular gland structures in colon histology images. Att-UNet introduces minor over-segmentation and under-segmentation along the edges. MedT and UNeXt, while capturing the general gland morphology, produce more evident boundary discrepancies, marked by enlarged red and green regions. Rolling-U-Net and U-KAN face similar challenges, often failing to precisely outline gland contours, leading to notable excesses or omissions. ResU-KAN, though approximating overall gland shapes, struggles to maintain fine boundary fidelity, frequently resulting in over-segmentation. Baseline approaches have difficulty effectively fusing global and local gland edge features, which could undermine segmentation reliability and affect pathological assessments. Our AdaKAN accurately delineates the glands with minimal segmentation errors.

In Figure 3.4, we present visualization comparisons of segmentation results generated by AdaKAN and baseline methods, on the CVC-ClinicDB dataset for colonoscopy polyp segmentation. The segmentation probability maps are shown in red, highlighting predicted polyp regions against the input endoscopy images. AdaKAN better handles the varying shapes, edges, and sizes of colon polyps compared to the baseline methods. AdaKAN segmentation maps exhibit the most precise contours, closely aligning with the ground truth and minimizing gaps in challenging curved polyp structures. Att-U-Net and MedT produce reasonable probability maps but often show under-segmentation along polyp boundaries. UNeXt and Rolling-U-Net display more pronounced irregularities, with disconnected segments or excessive blurring that fails to capture fine edge details.

U-KAN and ResU-KAN approximate polyp shapes but suffer from notable under-segmentation, leading to less reliable edge delineation. Baseline methods generally struggle to integrate global context with local polyp edge features due to architectural limitations, making them vulnerable to the variability in colonoscopy images, such as lighting variations and motion artifacts, which can compromise diagnostic accuracy.

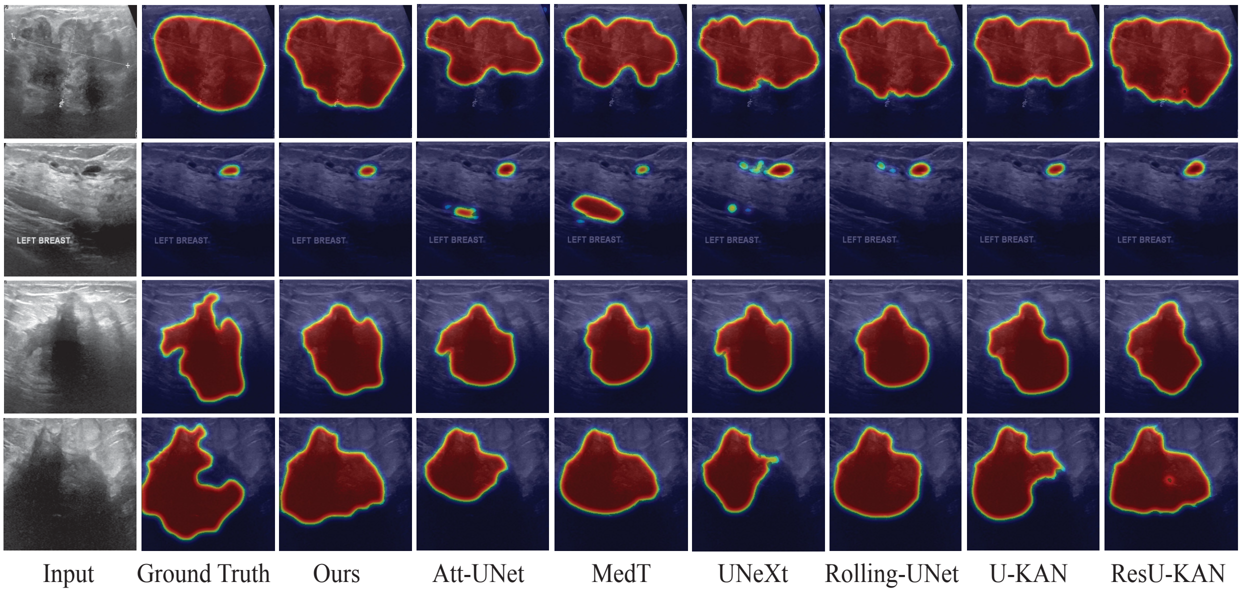

Figure 3.2: **Qualitative Comparisons of Segmentation Heatmap Results on BUSI dataset.** Ours (AdaKAN) demonstrates the closest alignment with the ground truth and the fewest segmentation errors across varying tumor shapes and sizes.

Figure 3.5 presents a qualitative comparison of the segmentation performance of AdaKAN against baseline methods on dermoscopic images from the ISIC 2018 dataset. The predicted segmentations are depicted in white, under-segmentation errors in green, and over-segmentation errors in red. AdaKAN produces the most precise segmentation, with minimal false positives and false negatives. Its outputs closely match the ground truth contours, particularly for intricate and irregular lesion shapes. Att-UNet introduces over-segmentation at the boundaries. MedT and MALUNet, while capturing the overall lesion shape, produce more prominent edge inaccuracies, as evidenced by larger red and green regions. UNeXt and Rolling-U-Net face similar challenges, often failing to accurately trace lesion edges, resulting in omissions. SET delivers reasonably strong results, although it outlines approximate lesion shapes and struggles to preserve detailed boundary integrity. Baseline approaches struggle to effectively integrate global and local lesion edge features. As a result, they often overlook boundary pixels, which can compromise segmentation precision and affect clinical decisions.

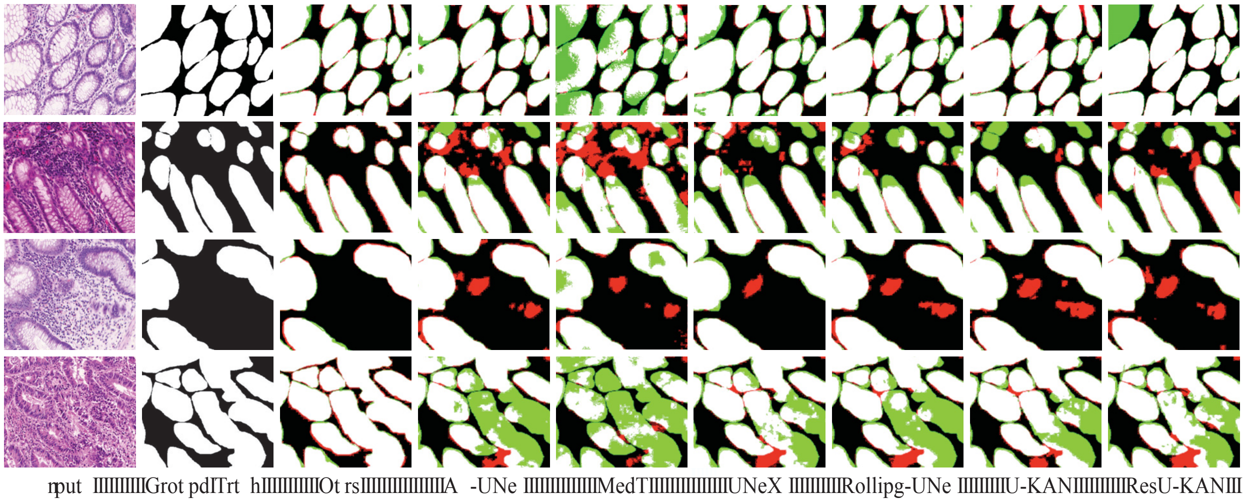

Figure 3.3: **Qualitative Comparisons of Segmentation Results on the GlaS dataset.** Visualizations highlight segmentation errors. White, green, and red regions indicate predicted segmentation, under-segmentation, and over-segmentation, respectively. Ours (AdaKAN) is capable of handling complex irregular gland shapes and edges compared to the baseline methods.

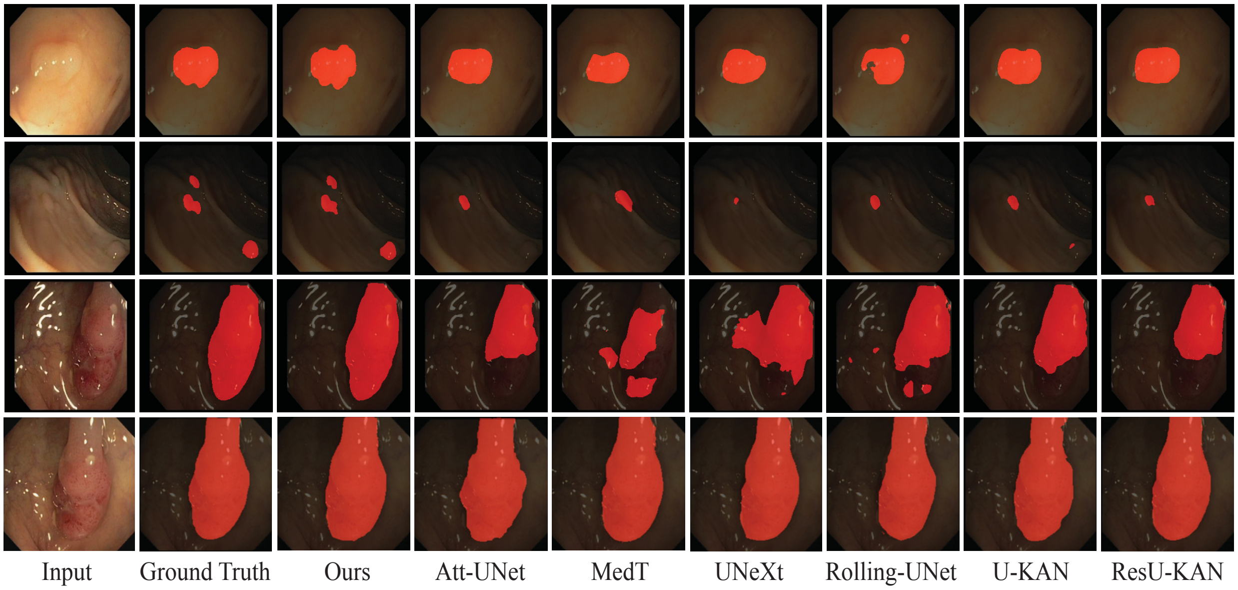

Figure 3.4: **Qualitative Comparisons of Segmentation Results on the CVC-ClinicDB dataset.** Segmentation probability maps are shown in red. Ours (AdaKAN) better handles different shapes and sizes of colon polyps compared to the baseline methods.

Figure 3.6 shows the visualization performance of AdaKAN compared to strong baselines on the cardiac ACDC dataset. The visualizations reveal that AdaKAN achieves more precise delineation boundaries of the Right Ventricle (RV) in red, the Myocardium (Myo) in blue, and the Left Ventricle (LV) in yellow. AdaKAN delineates closer to the ground truth and maintains sharper contours, with minimal over- or under-segmentation artifacts. It shows fewer segmentation artifacts and im-

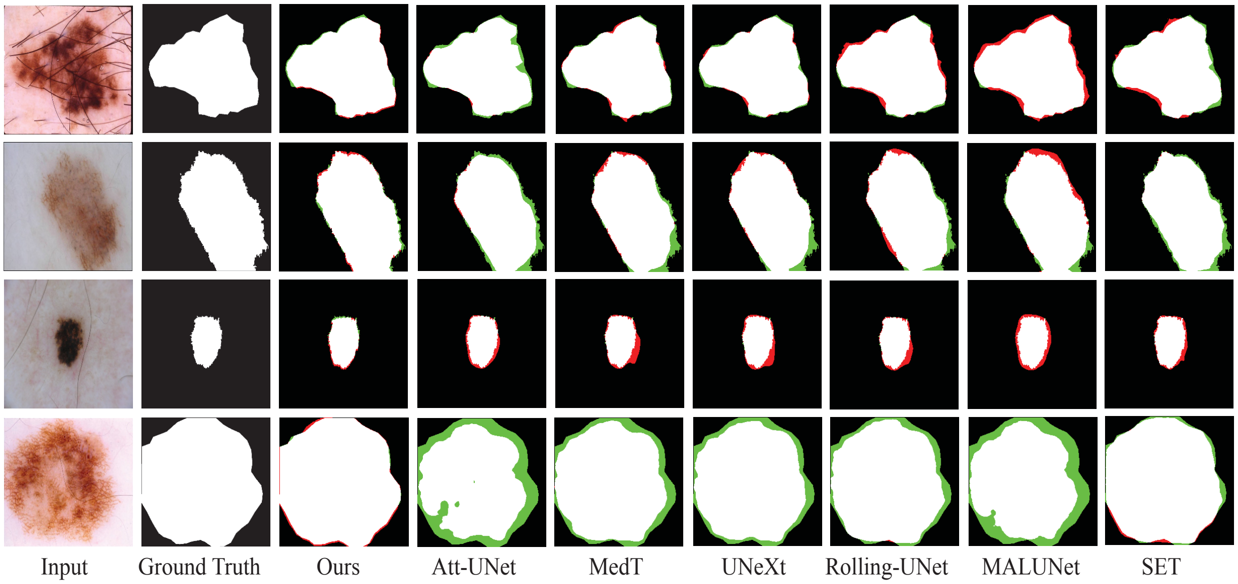

Figure 3.5: **Qualitative Comparison of Segmentation Results on the ISIC 2018 Dataset.** The qualitative results highlight segmentation errors, where white, red, and green regions represent predicted segmentation, over-segmentation, and under-segmentation, respectively. Ours (AdaKAN) exhibits the least segmentation error compared to the baseline methods.

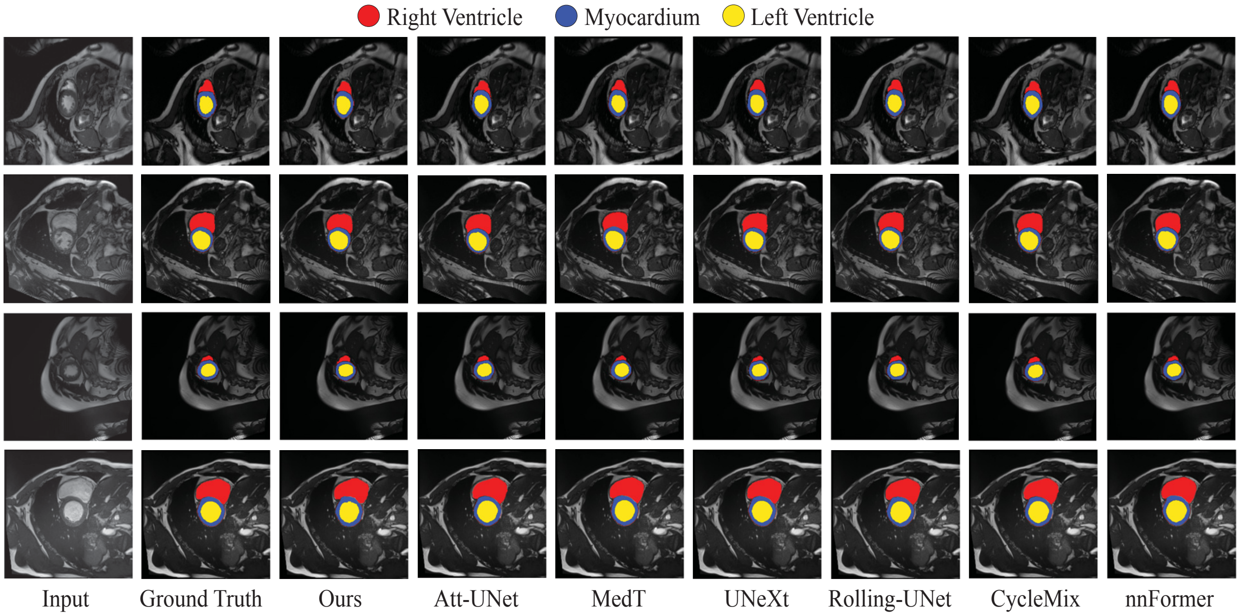

Figure 3.6: **Qualitative Comparisons of Segmentation Results on ACDC dataset.** Ours (AdaKAN) better handles complex cardiac structures compared to baseline methods.

proved boundary adherence compared to UNeXt, Att-UNet, and MedT. While Rolling-UNet and CycleMix exhibit reasonable segmentation performance, they occasionally show boundary inconsistencies. Furthermore, nnFormer struggles to capture finer structural details accurately. Notably, AdaKAN preserves fine boundary, edge, and shape details, effectively capturing cardiac structure shapes while outperforming state-of-the-art methods.

3.4.3 Model Efficiency Analysis

Figure 3.7 presents a circle chart comparing the segmentation performance and model efficiency of AdaKAN against SOTA baselines on the BUSI dataset, evaluated using the DSC metric against floating-point operations per second (FLOPS) in gigaflops. The circle represents a method, and the circle size represents the number of learnable parameters, reflecting model size. Notably, our method (yellow point) achieves a superior DSC of 80.95% with a moderate 8.12 gigaflops, maintaining a good balance between performance and efficiency.

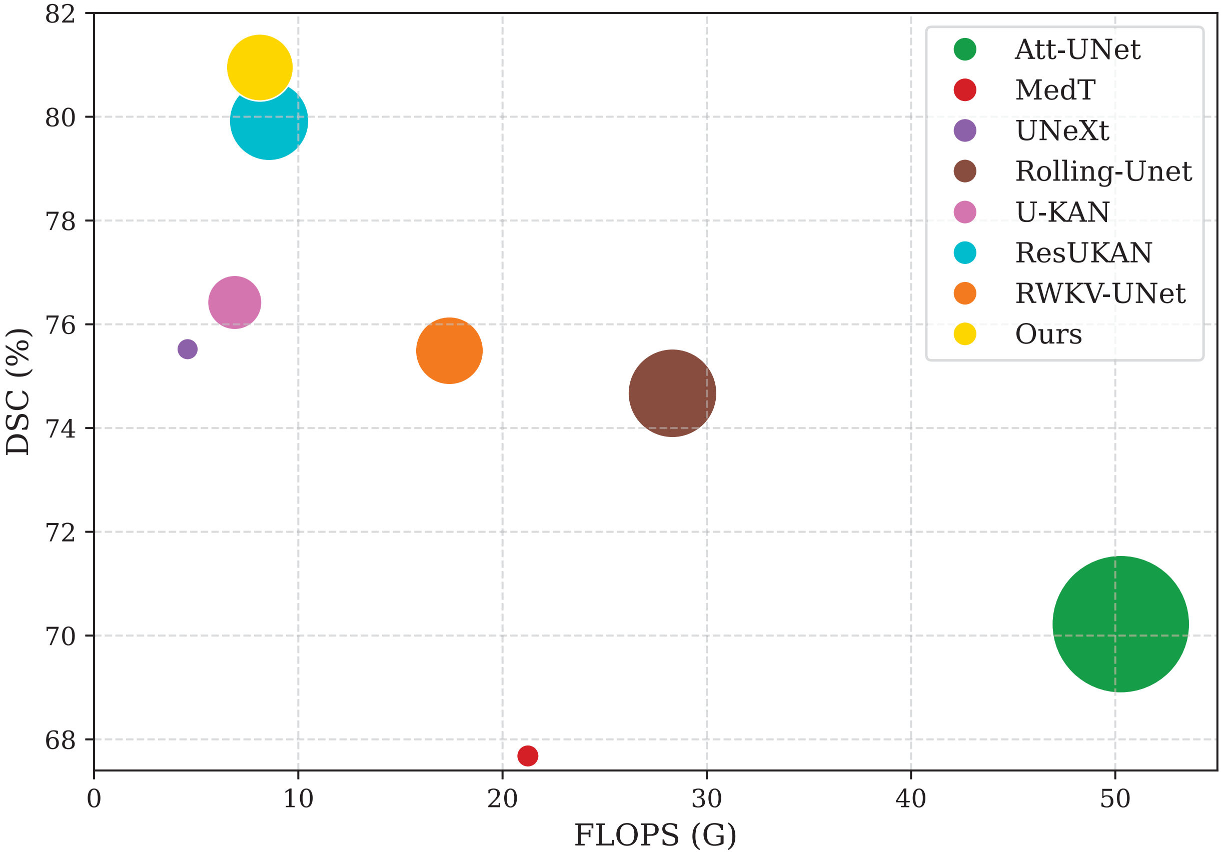

Figure 3.7: Comparison of performance and model efficiency on the BUSI dataset. Our AdaKAN model is benchmarked against state-of-the-art methods, including Attention-UNet [1], MedT [2], UNeXt [3], Rolling-Unet [4], U-KAN [5], ResU-KAN [6], RWKV-UNet [7]. Performance is assessed using the DSC metric, where higher values indicate superior segmentation performance. The size of the circle indicate the number of learnable parameters.

3.4.4 Ablation Studies

Table 3.4 assesses the impact of the core components of our network architecture on model performance using the BUSI dataset. The complete model, incorporating all components, yields the best performance. Excluding the Efficient Attention block leads to a performance decline, with IoU and DSC scores dropping to 2.13% and 1.09%, respectively, underscoring its pivotal role in capturing global dependencies. In contrast, removing the AdaptKAN block while retaining the Efficient Attention block led to a performance decline, with IoU and DSC scores dropping by 4.75% and 3.44%, respectively. Omitting the Efficient Attention and AdaptKAN reduces performance, underscoring their significant contributions.

Table 3.4: Effect of individual core components on model performance using the BUSI dataset.

Efficient Attention	AdaptKAN			
	Up-projection	Down-projection	IoU	DSC
✓	✓	✗	65.28	78.54
✓	✗	✓	66.49	79.33
✗	✓	✓	66.71	79.86
✓	✗	✗	64.09	77.51
✓	✓	✓	**68.84**	**80.95**

Effect of the Bernstein Polynomials. Table 3.5 presents the impact of varying the Bernstein polynomial order in the KAN layer of the AdaptKAN block on segmentation performance, evaluated on the BUSI dataset using IoU and DSC metrics. The global smoothness of feature approximations in AdaptKAN is controlled by the degree of the basis polynomials, which are defined by the Bernstein polynomial order in the KAN layer. Higher orders improve the polynomial's ability to represent complex functions, but because of the additional parameters, there is a chance of overfitting or numerical instability. The results show that a Bernstein polynomial order of three achieves the best performance, striking an optimal trade-off between expressiveness and stability in ultrasound image segmentation due to its ability to capture smooth and continuous lesion boundaries in the BUSI dataset. Reducing the order to two decreases accuracy by 3.26% in IoU and 2.19% in DSC, while increasing the order to four results in declines of 2.04% in IoU and 1.44% in DSC. Order five further reduces performance by 2.74% in IoU and 1.98% in DSC. These findings indicate that a moderate order, specifically three, provides the best balance between expressive power and over-parameterization.

Effect of Multi-head Attention and Efficient Attention. Table 3.6 presents the impact of Multi-head Attention and Efficient Attention within the AdaptKAN block. The results indicate that

Table 3.5: Effect of Bernstein polynomial order in the KAN layer of the AdaptKAN block on model performance using the BUSI dataset.

Bernstein Order	IoU	DSC
2	65.58	78.76
3	**68.84**	**80.95**
4	66.80	79.51
5	66.10	78.97

Efficient Attention outperforms Multi-head Attention, with a relative improvement of 1.35% in IoU and 0.92% in DSC. This yields enhanced global contextual modeling with reduced complexity, demonstrating that Efficient Attention boosts segmentation accuracy effectively.

Table 3.6: Effect of using Multi-head Attention and Efficient Attention on model performance using the BUSI dataset.

Method	IoU	DSC
Multi-head Attention	67.49	80.03
Efficient Attention	**68.84**	**80.95**

Effect of KAN vs MLP in AdaptKAN. Table 3.7 investigates the impact of replacing the standard MLP with the KAN module within the AdaptKAN block. The results demonstrate that replacing the MLP with KAN significantly improves performance, yielding relative improvements of 4.35% in IoU and 2.99% in DSC. This confirms that the KAN's superior ability to model complex global relationships enables substantially enhanced feature representation.

Table 3.7: Effect of KAN layers versus MLP layers within the AdaptKAN block on model performance using the BUSI dataset.

Method	IoU	DSC
MLP	64.49	77.96
KAN	**68.84**	**80.95**

CHAPTER

4

Conclusions and Future Work

This thesis investigated the integration of Kolmogorov–Arnold Networks (KANs) into deep learning architectures for medical image segmentation. The aim is to design effective models capable of capturing complex non-linear relationships while balancing global context, local detail preservation, and computational efficiency. Through this research, two major frameworks were developed the Hybrid Kolmogorov–Arnold Network (U-KABS) presented in Chapter 2, and the Adaptive Kolmogorov–Arnold Network (AdaKAN) presented in Chapter 3. Together, these frameworks establish a learnable function-based deep segmentation paradigm that merges theoretical interpretability with strong empirical performance across multiple medical imaging modalities. Further, Section 4.1 summarizes the primary findings and contributions of the research detailed in the earlier chapters. Section 4.2 discusses the key limitations of the proposed method, and Section 4.3 suggests avenues for future research extending this study.

4.1 Contributions of the Thesis

4.1.1 Hybrid Kolmogorov-Arnold Networks for Medical Image Segmentation

In Chapter 2, we introduced U-KABS, a novel hybrid framework that integrates KANs with a U-shaped encoder-decoder architecture, leveraging the global smoothness of Bernstein polynomials and the local adaptability of B-splines to effectively capture both broad contextual trends and fine-grained patterns in medical images. Extensive experiments on various benchmark datasets demonstrate that U-KABS consistently outperforms state-of-the-art methods. Notably, U-KABS excels in segmenting complex anatomical structures and performs robustly in both binary and multi-class segmentation tasks, all while maintaining a lightweight architecture. Qualitative results further demonstrate U-KABS's ability to produce accurate segmentation masks, with improved boundary delineation. Moreover, ablation results confirm the importance of the core architectural components, especially the KAB layer for global approximation. For future work, we aim to further optimize U-KABS for efficiency, and incorporate self-supervised learning to improve performance on datasets with limited annotated data, a common challenge in medical imaging.

4.1.2 Adaptive Kolmogorov-Arnold Network for Medical Image Segmentation

In Chapter 3, we introduce AdaKAN, an adaptive segmentation framework that integrates the Adaptive Kolmogorov–Arnold Network (AdaptKAN) and Efficient Attention within the EffiKAN block. The framework aims to generate precise segmentation masks by employing a dual-branch structure in the AdaptKAN module, where the KAN layer uses Bernstein polynomials as global approximators, and the projection branch captures fine-grained details. Meanwhile, the Efficient Attention mechanism models global dependencies with linear computational complexity, addressing the limitations of standard quadratic attention. We evaluate AdaKAN on five diverse, publicly available medical imaging datasets containing both binary and multi-class segmentation tasks. The results demonstrate that AdaKAN achieves accurate segmentation with improved boundary delineation while maintaining a lightweight architecture. Our ablation studies further validate the contribution of each core component, particularly the AdaptKAN and Efficient Attention modules, in capturing broader contextual information. For future work, we plan to enhance AdaKAN's efficiency and explore self-supervised learning strategies to improve performance under limited annotation settings.

4.2 Limitations

Both U-KABS and AdaKAN have a strong performance on the benchmark datasets against SOTA baselines but there are several areas of imporvement with respect tot both the frameworks which are as follows:

- **U-KABS .** U-KABS performs well on multi-class datasets like ACDC, but the hybrid use of Bernstein polynomials and B-splines in KABS blocks, combined with squeeze-and-excitation, results in higher parameter counts and inference times than pure CNN or MLP models. Reducing parameters for greater efficiency is key, especially for deployment on edge devices, the supervised nature of U-KABS limits its applicability to datasets with limited labels. Incorporating self-supervised learning could mitigate this, as the model may overfit or generalize poorly in data-scarce medical scenarios.

- **AdaKAN .** Although, AdaKAN achieves strong performance on diverse datasets like ACDC (a multi-class cardiac MRI dataset), its integration of Efficient Attention and AdaptKAN blocks with Bernstein polynomials introduces additional computational overhead compared to simpler CNN-based models. The model could benefit from further optimization to reduce the number of parameters, making it more lightweight and suitable for resource-constrained clinical environments or real-time applications. Although, robust across five benchmarks, AdaKAN may struggle with extreme noise, artifacts, or modalities not represented in the training data.

4.3 Future Work

Building upon the contributions of this thesis, several promising research directions can be pursued:

- **3D and Multimodal Extensions.** Current implementations of AdaKAN and U-KABS operate on 2D slices, limiting their ability to capture full volumetric context critical for clinical in vivo clinical analysis. Extending both architectures to native 3D processing—by replacing 2D convolutions with 3D counterparts and adapting KAN layers to volumetric inputs would enable more accurate delineation of complex anatomical structures. Furthermore, integrating multimodal fusion within the EffiKAN and KABS blocks could leverage complementary imaging contrasts. For instance, early fusion via shared Conv2D stages followed by modality-specific KAN pathways could enhance robustness in cross-modal registration and joint segmentation tasks, paving the way for comprehensive diagnostic pipelines.

- **Self Supervised Learning.** Both models currently rely on fully supervised training, which is constrained by the scarcity and cost of pixel-level annotations in medical imaging. Incorporating self-supervised pretraining such as masked image modeling, contrastive learning, or reconstruction-based objective directly into the KAN-based feature space could significantly improve generalization. The learnable univariate activations in Bernstein and B-spline bases offer a natural mechanism for functional pretext tasks, such as predicting activation coefficients or reconstructing local polynomial approximations. This would not only reduce dependency on labeled data but also enhance domain adaptation across imaging modalities and institutions, making AdaKAN and U-KABS more practical in real-world, low-annotation clinical environments.

- **Theoretical Analysis of Functional Adaptivity.** While empirical results demonstrate the efficacy of Bernstein polynomials (AdaKAN) and hybrid Bernstein–B-spline designs (U-KABS), a rigorous mathematical characterization of how basis function adaptivity influences segmentation performance remains unexplored. Future work should focus on establishing formal bounds on approximation error, stability, and expressive power of KAN layers under different polynomial degrees and grid resolutions. Analyzing the spectral properties of learned activation functions and their impact on gradient flow and feature propagation could yield deeper insights into why KANs outperform MLPs in boundary-sensitive tasks. Additionally, studying the trade-off between global smoothness and local flexibility in terms of generalization and overfitting possibly via information-theoretic measures or curvature analysis would strengthen the theoretical foundation of function based deep learning and guide optimal architectural design.

References

[1] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention U-Net: Learning where to look for the pancreas,” in *Proc. Medical Imaging with Deep Learning*, 2022.

[2] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, “Medical Transformer: Gated axial-attention for medical image segmentation,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 12901, pp. 109–119, 2021.

[3] J. M. J. Valanarasu and V. M. Patel, “UNeXt: MLP-based rapid medical image segmentation network,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 23–33, 2022.

[4] Y. Liu, H. Zhu, M. Liu, H. Yu, Z. Chen, and J. Gao, “Rolling-Unet: Revitalizing MLP’s ability to efficiently extract long-distance dependencies for medical image segmentation,” in *Proc. AAAI Conference on Artificial Intelligence*, 2024.

[5] C. Li, X. Liu, W. Li, C. Wang, H. Liu, and Y. Yuan, “U-KAN makes strong backbone for medical image segmentation and generation,” in *Proc. AAAI Conference on Artificial Intelligence*, 2025.

[6] H. Wang, Z. Zhao, Q. Liu, and S. Wang, “ResU-KAN: A medical image segmentation model integrating residual convolutional attention and atrous spatial pyramid pooling,” *Applied Intelligence*, vol. 55, no. 7, 2025.

[7] J. Jiang, J. Zhang, W. Liu, M. Gao, X. Hu, X. Yan, F. Huang, and Y. Liu, “RWKV-UNet: Improving UNet with long-range cooperation for effective medical image segmentation,” *arXiv preprint arXiv:2501.08458*, 2025.

[8] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015.

[9] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. International Workshop on Deep Learning in Medical Image Analysis*, pp. 3–11, 2018.

[10] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, *et al.*, "TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, vol. 97, 2024.

[11] 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 *Proc. European Conference on Computer Vision*, pp. 205–218, 2022.

[12] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," in *Proc. Conference on Applications of Computer Vision*, pp. 574–584, 2022.

[13] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov-Arnold networks," in *International Conference on Learning Representations*, 2025.

[14] H. Krim and A. Ben Hamza, *Geometric methods in signal and image analysis*. Cambridge University Press, 2015.

[15] H. Zunair and A. Ben Hamza, "Sharp U-Net: Depthwise convolutional network for biomedical image segmentation," *Computers in Biology and Medicine*, vol. 136, 2021.

[16] H. Zunair and A. Ben Hamza, "Masked supervised learning for semantic segmentation," in *Proc. British Machine Vision Conference*, pp. 327–334, 2022.

[17] J. Ruan, S. Xiang, M. Xie, T. Liu, and Y. Fu, "MALUNet: A multi-attention and light-weight UNet for skin lesion segmentation," in *Proc. IEEE International Conference on Bioinformatics and Biomedicine*, 2022.

[18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. International Conference on Learning Representations*, 2021.

[19] Q. Xu, H. Zhu, Y. Wang, Z. Shi, and W. Liu, "Cross-teaching with dual uncertainty awareness for semi-supervised medical image segmentation," *Multimedia Systems*, vol. 31, 2025.

[20] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, “UNet 3+: A full-scale connected UNet for medical image segmentation,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020.

[21] O. C. A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432, 2016.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Proc. Fourth International Conference on 3D Vision*, pp. 565–571, 2016.

[23] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, “Towards accurate segmentation of biomedical images using over-complete representations,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 363–373, 2020.

[24] J. M. J. Valanarasu, R. Yasarla, P. Wang, I. Hacihaliloglu, and V. M. Patel, “Learning to segment brain anatomy from 2D ultrasound with less data,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1221–1234, 2020.

[25] J. Braun and M. Griebel, “On a constructive proof of kolmogorov’s superposition theorem,” *Constructive Approximation*, vol. 30, pp. 653–675, 2009.

[26] J. Schmidt-Hieber, “The kolmogorov-arnold representation theorem revisited,” *Neural Networks*, vol. 137, pp. 119–126, 2021.

[27] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data in Brief*, vol. 28, 2020.

[28] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, *et al.*, “Gland segmentation in colon histology images: The glas challenge contest,” *Medical Image Analysis*, vol. 35, pp. 489–502, 2017.

[29] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, and H. Kittler, “Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging,” in *Proc. IEEE International Symposium Conference on Biomedical Imaging*, 2018.

[30] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, and M. A. G. Ballester, “Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[31] Y. Zhao, S. Wang, Y. Ren, J. Wang, S. Wang, S. Qiao, T. Liu, and S. Pang, “TPFIANet: Three path feature progressive interactive attention learning network for medical image segmentation,” *Knowledge-Based Systems*, 2025.

[32] J. Li, N. Chen, H. Zhou, R. Chen, C. Feng, H. Dong, T. Lai, C. Yang, L. Wei, and F. Cai, “MCRformer: Morphological constraint reticular transformer for 3D medical image segmentation,” *Expert Systems with Applications*, vol. 232, 2023.

[33] A. Tragakis, C. Kaul, R. Murray-Smith, and D. Husmeier, “The fully convolutional transformer for medical image segmentation,” in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3660–3669, 2023.

[34] H. Yan, Q. Hong, S. Wei, X. Zhang, and J. Yin, “SCM-UNet: Spatial-Channel Mamba UNet for medical image segmentation,” *Digital Signal Processing*, 2025.

[35] Z. Gan, Z. Xie, Y. Zhang, W. Han, B. Zhang, and X. Chai, “WTCM-UNet: A hybrid CNN-SSM framework combining wavelet transform for medical image segmentation,” *Biomedical Signal Processing and Control*, vol. 112, 2026.

[36] Y. Lin, D. Zhang, X. Fang, Y. Chen, K.-T. Cheng, and H. Chen, “Rethinking boundary detection in deep learning-based medical image segmentation,” *Medical Image Analysis*, 2025.

[37] S. Chen, C. Ge, Z. Tong, J. Wang, Y. Song, J. Wang, and P. Luo, “Adaptformer: Adapting vision transformers for scalable visual recognition,” in *Advances in Neural Information Processing Systems*, vol. 35, pp. 16664–16678, 2022.

[38] 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*, vol. 18, no. 2, pp. 203–211, 2021.

[39] W. Zhu, X. Chen, P. Qiu, M. Farazi, A. Sotiras, A. Razi, and Y. Wang, “SelfReg-UNet: Self-Regularized UNet for medical image segmentation,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 601–611, 2024.

[40] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, “Segment anything in medical images,” *Nature Communications*, vol. 15, no. 1, 2024.

[41] S. Perera, Y. Erzurumlu, D. Gulati, and A. Yilmaz, “MobileUNETR: A lightweight end-to-end hybrid vision transformer for efficient medical image segmentation,” in *Proc. European Conference on Computer Vision*, pp. 281–299, 2024.

[42] T. B. Nguyen-Tat, H.-A. Vo, and P.-S. Dang, “QMaxViT-Unet+: A query-based MaxViT-Unet with edge enhancement for scribble-supervised segmentation of medical images,” *Computers in Biology and Medicine*, vol. 187, p. 109762, 2025.

[43] L. Deng, W. Wang, S. Chen, X. Yang, S. Huang, and J. Wang, “PDS-UKAN: Subdivision hopping connected to the U-KAN network for medical image segmentation,” *Computerized Medical Imaging and Graphics*, 2025.

[44] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilarino, “WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians,” *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.

[45] D. Müller, I. Soto-Rey, and F. Kramer, “Towards a guideline for evaluation metrics in medical image segmentation,” *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022.

[46] Y. Gao, M. Zhou, and D. N. Metaxas, “UTNet: A hybrid transformer architecture for medical image segmentation,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 61–71, 2021.

[47] 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*, vol. 42, no. 5, pp. 1484–1494, 2022.

[48] K. Zhang and X. Zhuang, “CycleMix: A holistic strategy for medical image segmentation from scribble supervision,” in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 11656–11665, 2022.

[49] J. Liu, K. Li, C. Huang, H. Dong, Y. Song, and R. Li, “MixFormer: A mixed CNN-Transformer backbone for medical image segmentation,” *IEEE Transactions on Instrumentation and Measurement*, 2024.

[50] Z. Li, Y. Zheng, D. Shan, S. Yang, Q. Li, B. Wang, Y. Zhang, Q. Hong, and D. Shen, "Scribformer: Transformer makes cnn work better for scribble-based medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 43, no. 6, pp. 2254–2265, 2024.

[51] H.-Y. Zhou, J. Guo, Y. Zhang, X. Han, L. Yu, L. Wang, and Y. Yu, "nnFormer: Volumetric medical image segmentation via a 3D transformer," *IEEE Transactions on Image Processing*, vol. 32, pp. 4036–4045, 2023.

[52] Y. Liu, C. Chen, and Z. Liu, "EFEH-Net: Edge feature enhancement-based cardiac MRI image segmentation," in *Proc. International Conference on Computer Vision, Image and Deep Learning*, pp. 385–391, 2025.

[53] L. Wang, Q. Xu, C. Chen, H. Yang, and G. Deng, "Adaptive cascade decoders for segmenting challenging regions in medical images," *Computers in Biology and Medicine*, vol. 185, p. 109572, 2025.

[54] X. Fang, Y. Pan, and Q. Chen, "DFEDC: Dual fusion with enhanced deformable convolution for medical image segmentation," *Image and Vision Computing*, vol. 151, p. 105277, 2024.

[55] M. M. Rahman and R. Marculescu, "Medical image segmentation via cascaded attention decoding," in *Proc. Conference on Applications of Computer Vision*, pp. 6222–6231, 2023.

[56] M. M. Rahman, M. Munir, and R. Marculescu, "EMCAD: Efficient multi-scale convolutional attention decoding for medical Image Segmentation," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 11769–11779, 2024.

[57] Z. Wang, J. Lyu, and X. Tang, "SET: Superpixel embedded transformer for skin lesion segmentation," *Medical Image Analysis*, p. 103738, 2025.

[58] Y. Peng, D. Z. Chen, and M. Sonka, "U-Net V2: Rethinking the skip connections of U-Net for medical image segmentation," in *Proc. IEEE International Symposium on Biomedical Imaging*, pp. 1–5, 2025.

[59] G. Xu, X. Zhang, X. He, and X. Wu, "Levit-UNet: Make faster encoders with transformer for medical image segmentation," in *Proc. Chinese Conference on Pattern Recognition and Computer Vision*, pp. 42–53, 2023.