Spike-Based Attention Mechanisms for Enhanced Medical Image Segmentation

Meshari A. Al-Ebrahim
Technical Support Department, Mechanical Engineering, State Audit Bureau (SAB), Kuwait
meshari.alebrahim@gmail.com (corresponding author)

Received: 15 July 2025 | Revised: 8 August 2025 and 23 August 2025 | Accepted: 26 August 2025

 DOI: https://doi.org/10.48084/etasr.13407

ABSTRACT

The accurate segmentation of medical images is essential for reliable Computer-Aided Diagnosis (CAD) and treatment planning. Although deep learning has advanced the segmentation accuracy, challenges, such as modality variability, low contrast, and imaging artifacts, still exist. This study introduces a novel integration of a spike-based attention module into a DenseNet-169 backbone, combining biological spiking dynamics with deep convolutional feature representations to improve the focus on salient image features. The architecture utilizes ResNet blocks in the encoder for hierarchical feature extraction and applies spike-based attention at both the bottleneck and decoder stages to enhance the performance and computational efficiency. Experiments on two benchmark datasets demonstrate a 4.2% relative gain in Dice coefficient and a 3.9% increase in Intersection-over-Union compared to state-of-the-art baselines, while preserving the inference latency under 50 ms on standard GPU hardware. An extensive ablation study confirms that the dual placement of the attention module and the choice of DenseNet-169 maximize both the accuracy and efficiency. These results highlight the potential of spike-based attention for advancing real-time, high-fidelity medical image segmentation, with implications for improved clinical workflows and edge device deployment.

Keywords-deep learning; image segmentation; medical image segmentation; spike-based attention mechanism; Convolutional Neural Networks (CNNs); Spiking Neural Networks (SNNs)

I. INTRODUCTION

Medical image segmentation is a critical step in computer-aided diagnosis, enabling THE precise delineation of anatomical structures and pathological regions. Advances in deep Convolutional Neural Networks (CNNs) have substantially improved the segmentation accuracy, yet challenges remain in capturing the long-range dependencies and salient features within high-resolution scans. Spike-based attention mechanisms have shown promise in emulating the biological neuronal dynamics, potentially enhancing the feature selectivity and network efficiency.

However, challenges persist due to the inherent complexities of medical imaging. The low contrast, high noise levels, anatomical variability among patients, and imaging artifacts hinder the performance of even the most advanced models. These issues are especially prominent in tasks, such as polyp segmentation in colonoscopy images and abnormality detection in chest X-rays, where subtle differences can carry critical diagnostic implications [1].

Polyp detection in colonoscopy images is essential for the early diagnosis and prevention of colorectal cancer, one of the leading causes of cancer-related mortality worldwide. The diversity in polyp appearance—ranging from small, flat lesions to prominent protrusions—and the dynamic nature of video frames make this task particularly demanding [2]. Similarly, analyzing chest X-rays is fundamental in diagnosing pneumonia, Tuberculosis (TB), lung cancer, and other pulmonary conditions. Overlapping anatomical structures and subtle pathological variations present additional challenges for automated segmentation methods [3].

While the traditional CNN-based architectures have shown promise, they often struggle to model the long-range dependencies or selectively attend to the most relevant regions in complex medical images. These limitations have motivated the exploration of attention mechanisms, which emulate aspects of the human visual attention to enhance the feature focus in neural networks [4]. However, many existing attention modules are computationally intensive and lack biological plausibility, limiting their scalability and interpretability.

This work introduces a novel segmentation architecture that integrates CNN-based hierarchical feature extraction with a biologically inspired spike-based attention mechanism. Inspired by the discrete spiking behavior of biological neurons, the proposed attention module aims to improve the computational efficiency and enhance the model's ability to attend to diagnostically relevant regions in medical images [5]. The key contributions of this study are:

- A spike-based attention mechanism is introduced to augment CNNs, improving the focus on critical regions in medical images. This biologically plausible design bridges conventional deep learning with neuro-morphic computing, offering a fresh perspective on attention-based segmentation [6].

- The proposed model is evaluated on two distinct and challenging tasks: polyp segmentation using the Kvasir-SEG dataset and lung abnormality detection using the Shenzhen Chest X-ray dataset. The strong cross-domain performance demonstrates the approach's versatility and generalizability [7].
- Comprehensive experiments are conducted, including comparisons with state-of-the-art models and ablation studies, to validate the performance and assess the contribution of spike-based attention to medical image segmentation.

By merging insights from deep learning and neuroscience, the proposed approach advances the capabilities of automated medical image analysis. This integration has the potential to deliver more accurate, efficient, and interpretable segmentation models, which could ultimately enhance the diagnostic workflows and support more personalized treatment planning amid the growing global healthcare demands [8].

The objective of this study is to integrate a spike-based attention module into the DenseNet-169 backbone and to assess its impact on the segmentation accuracy and computational cost. Specifically, the proposed method is evaluated on two publicly available medical imaging datasets, with its performance compared against several state-of-the-art segmentation models. Quantitative metrics, such as the Dice coefficient, Intersection over Union (IoU), and inference time benchmarks, were used to demonstrate the advantages of the spike-based attention approach.

To situate this objective within existing research, relevant advances are synthesized across three threads: (i) CNN-based segmentation backbones and their refinements, (ii) attention and Transformer mechanisms adapted to medical imaging, and (iii) spike-based neural models that introduce event-driven dynamics. The synthesis highlights where prior work excels (e.g., long-range context modeling) and where persistent gaps remain (e.g., memory/latency overheads, stability of spike-gated attention in high-resolution medical images), thereby motivating the architectural choices and evaluations that follow.

Progress in medical image segmentation has been driven by deep learning. Key developments relevant to this study include CNN-based segmentation models, attention mechanisms tailored for medical imaging, spike-based neural networks, and applications, such as polyp segmentation and chest X-ray analysis. The literature indicates strong accuracy from attention-enhanced architectures alongside increased computational and memory demands, and it underscores the opportunity for spike-informed attention to balance the accuracy with efficiency in clinical-scale images.

A. *CNN-based Medical Image Segmentation*

CNNs form the backbone of modern medical image segmentation techniques due to their ability to learn hierarchical features and model complex spatial patterns. Among them, the U-Net architecture, introduced in [9], has been especially influential. Its encoder-decoder structure with skip connections allows for the preservation of fine-grained spatial details essential for accurate segmentation.

Comprehensive reviews, such as Authors in [10], have highlighted the enduring relevance of CNN-based methods across various medical imaging domains. Their findings underscore continuous architectural innovations aimed at overcoming challenges, like boundary precision and limited training data.

Numerous extensions of U-Net have been proposed to improve its performance in diverse medical imaging tasks. For instance, UNet++ [11] incorporates dense and nested skip connections to reduce the semantic gap between the encoder and decoder features. Authors in [12] introduced Attention U-Net, which integrates attention gates to selectively emphasize the relevant image regions, enhancing the performance in multi-class segmentation. A more recent model, I2UNet, developed in [13], uses a dual-path U-Net architecture that enriches the information flow between the encoder and decoder stages, promoting better feature reuse and efficient gradient propagation.

Additional CNN-based architectures have also shown promise. PSPNet [14] employs pyramid pooling to aggregate the global contextual information, while MSRF-Net [15] utilizes multi-scale residual fusion to integrate features across multiple resolutions. These models have been efficient in applications, such as histopathology segmentation, brain tumor detection, and vascular structure analysis [16].

Similarly, authors in [17] developed a hybrid 3D U-Net with integrated attention mechanisms for whole heart segmentation from CT images, demonstrating the effectiveness of combining residual connections and attention for accurate anatomical delineation.

B. *Attention Mechanisms in Medical Imaging*

Self-attention mechanisms, inspired by the Transformer architecture, have been effectively adapted for medical image segmentation. These mechanisms excel at capturing the long-range dependencies, making them particularly effective in delineating complex anatomical structures [18]. Their ability to model global spatial relationships has led to improved performance in modalities, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) [19]. Spatial attention has also gained traction in medical imaging. For instance, the Spatial-Attention ConvMixer (SAC) architecture was proposed for the classification of gastrointestinal disorders [20]. By combining spatial attention with the SAC design, the model effectively highlights the most relevant areas in feature maps. The SAC approach demonstrated strong performance on the Kvasir dataset.

Hybrid attention-CNN models have been developed to leverage the strengths of both paradigms. Such architecture enhances the ability to focus on critical features and often outperforms conventional CNNs [21]. For example, a CNN combined with hybrid attention was proposed for MRI-based brain tumor diagnosis, achieving improved accuracy with reduced computational demands [22].

Research has explored increasingly sophisticated attention modules. A multi-scale convolutional attention frequency-enhanced transformer was introduced to merge CNN-Transformer features with wavelet-based representations, preserving both the global and fine-grained information [23]. Authors in [24] presented a probabilistic attention map based on the probability distribution of CNN activations, enhancing the classification accuracy through refined attention weighting.

The Global Attention Mechanism (GAM) has also been incorporated into the EfficientNetV2 architecture to boost the brain tumor classification accuracy in MRI images [25]. Furthermore, the fusion of channel and spatial attention has led to the development of hybrid attention networks capable of capturing both the spatial and feature-channel information, resulting in improved segmentation performance across various medical imaging tasks [26].

C. Spike-Based Neural Networks in Image Processing

Spiking Neural Networks (SNNs) offer a biologically inspired model of computation, simulating the spiking behavior of neurons in the human brain. While it is less commonly applied in medical imaging than conventional CNNs, SNNs show potential advantages in energy efficiency and temporal information processing.

Authors in [27] provide an overview of the deep learning applications in SNNs, highlighting their event-driven processing as a promising approach for scenarios with constrained computational resources.

In the context of medical image segmentation, authors in [28] propose a three-stage training approach for segmenting the human hippocampus from MRI scans. This methodology included converting a trained Artificial Neural Network (ANN) to an SNN and fine-tuning it via spike- based backpropagation, resulting in improvements in the segmentation accuracy and training efficiency.

The integration of spike-based mechanisms with attention models has emerged as a novel direction in visual recognition. Authors in [29] explored the spike-based attention for visual recognition, demonstrating improvements in both the accuracy and computational cost. Although not explicitly focused on medical imaging, their findings suggest potential benefits in applying such mechanisms to segmentation tasks. Authors in [30] show that SNNs can achieve competitive performance with ANNs in image classification tasks while offering considerable energy savings, reinforcing their relevance for energy-efficient medical image analysis in resource-limited settings.

D. Polyp Segmentation and Chest X-ray Analysis

Polyp segmentation and chest X-ray analysis have become critical cases for deep learning-based medical image segmentation, with considerable research progress having been achieved.

Authors in [31] examined the challenges of polyp segmentation, particularly the irregular shapes and diverse sizes of polyps. Their review highlighted both model-based and learning-based approaches and emphasized the importance of context-aware segmentation and reliable feature extraction. Furthermore, authors in [32] proposed PraNet, a parallel reverse attention network that achieved state-of-the-art results across multiple datasets.

Subsequent work has continued to address the complexities of polyp detection. Authors in [33] introduced A-DenseUNet, which leverages atrous convolutions and dense connectivity to handle the variability in polyp morphology. Authors in [34] presented Polyp-PVT, a method based on pyramid vision transformers to capture the long-range dependencies in colonoscopy images. Similarly, authors in [35] introduced MSRF-Net, a multi-scale residual fusion network that enhances detection by integrating features from multiple scales.

In the domain of chest X-ray analysis, authors in [36] created CheXpert, a large-scale benchmark dataset for automated chest radiograph interpretation. Their study demonstrated that deep learning models can detect thoracic diseases at a level comparable to radiologists.

The attention mechanisms have been explored to improve the identification of subtle abnormalities in chest X- rays. Authors in [37] employed a multi-label deep learning classifier to localize TB lesions in lung zones. Authors in [38] demonstrated parity with clinical performance in detecting pulmonary TB. Authors in [39] proposed an attention-guided multi-residual network for segmenting distressed lung regions, supported by customized data augmentation. Authors in [40] evaluated Fat-Net's effectiveness in segmenting chest X-rays, while authors in [41] introduced CXR-Seg, which achieved precise lung region segmentation.

The integration of attention mechanisms into domain-specific models has also yielded promising results. Authors in [42] developed an attention-driven network for polyp segmentation, enhancing the detection of small and flat lesions. Authors in [43] introduced an attention-based architecture for classifying multiple thoracic diseases, improving the recognition of subtle findings. Authors in [44] applied Transformer-based attention to colonoscopy images, leveraging the long-range dependency modeling for polyp segmentation. For chest radiography, authors in [45] proposed an attention-guided multi-branch model that demonstrated state-of-the-art performance in multi-disease classification.

E. Limitations of Existing Approaches

Spike-based attention has been explored mainly in object recognition and compact vision models. Examples include a masked spiking Transformer that prunes redundant spikes to cut computation while maintaining accuracy [46], a spatial–temporal self-attention mechanism that models dependencies across time and space while preserving asynchronous, event-driven computation [47], and a spiking ResNet–Transformer hybrid that integrates self-attention into spiking backbones for improved efficiency and accuracy [48]. Despite these advances, most evaluations focus on natural or neuromorphic datasets at lower resolutions, whereas evidence for high-resolution medical image segmentation remains limited.

Non-spike attention architectures for medical segmentation—such as TransUNet [49] and Swin-Unet3D

[50]—capture the long-range dependencies effectively but typically increase the memory footprint and inference latency, relative to purely convolutional baselines, which constrain the deployment on resource-limited hardware. In the proposed design, feature projections to queries, keys, and values are followed by Leaky Integrate-and-Fire (LIF) dynamics to produce binary spike trains that gate attention weights, complying with the foundational spiking-neuron theory on thresholded membrane integration [51].

The proposed spike-based attention module integrates seamlessly into a DenseNet-169 backbone, combining the efficient dynamics of spiking neurons with the feature depth of a modern CNN as shown below:

- Scalability: Spike-based layers in shallow networks do not extend the readily to deep backbones, limiting the feature richness.
- Inference Latency: Event-driven attention schemes introduce variable processing delays under high-resolution inputs.
- Memory Footprint: Transformer-based segmentation models demand large GPU memory, hindering the deployment on edge devices.

F. Challenges and Future Directions

Despite the significant progress, medical image segmentation continues to face several enduring challenges. Authors in [52] emphasized the need for large-scale, well-annotated datasets and highlighted the importance of the model interpretability for real-world clinical use. Authors in [53] discussed the difficulty of developing generalizable models that perform consistently across different imaging modalities and clinical environments.

Improving the model interpretability and computational efficiency remains a key area for advancement. Authors in [54] investigated explainable AI techniques aimed at making deep models more transparent and trustworthy. In parallel, authors in [55] explored federated learning to train collaborative models across institutions without compromising the data privacy.

More recent approaches have incorporated Transformer-based architectures to address the long-range dependencies. Authors in [56] demonstrated high segmentation performance using Transformer models for polyp detection. Similarly, authors in [57] applied attention-guided multi-branch networks to chest X-ray images, improving the detection of small abnormalities.

Looking ahead, addressing data limitations, interpretability, and energy efficiency will be critical for further progress in medical image segmentation. Emerging models that integrate biological principles, such as spike-based attention, may provide robust, efficient, and clinically applicable solutions.

II. METHODOLOGY

A. Dataset

Two distinct datasets were utilized to develop and evaluate the proposed segmentation model: the Kvasir-SEG dataset for gastrointestinal polyp segmentation and the Shenzhen Chest X-ray dataset for pulmonary abnormality detection. These datasets were selected based on their relevance to medical image segmentation challenges and their ability to illustrate the adaptability of the CNN-based model with a spike-based attention mechanism.

1) Kvasir-SEG Dataset

The Kvasir-SEG dataset [58] is a publicly available benchmark designed for gastrointestinal polyp segmentation research. It comprises 1,000 high-resolution images of polyps located within the gastrointestinal tract, each accompanied by ground truth segmentation masks manually annotated by medical experts. Derived from the broader Kvasir Dataset v2, this dataset supports the development and evaluation of advanced computer vision algorithms for polyp detection and segmentation.

The primary objective of Kvasir-SEG is to improve the accuracy and efficiency of polyp identification during colonoscopy procedures, thereby aiding in the early diagnosis and prevention of colorectal cancer. A few sample images from the dataset, along with their corresponding masks, are illustrated in Figure 1.

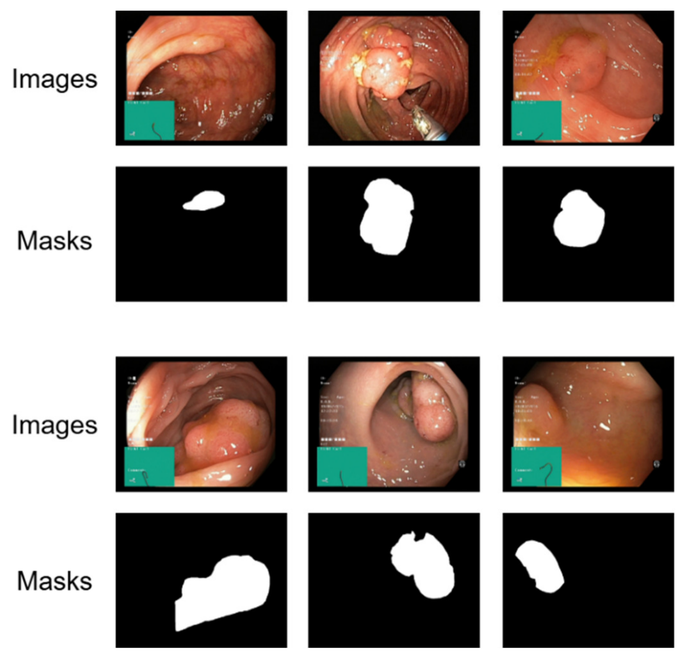

Fig. 1. Sample images from the Kvasir-SEG dataset along with corresponding ground truth masks.

2) Shenzhen Chest X-Ray Dataset

The Shenzhen Chest X-ray dataset [59] was developed through a collaboration between Shenzhen No. 3 People's Hospital and Guangdong Medical College in China. It contains 662 posterior-anterior chest X-ray images, including 336 cases with TB and 326 normal cases. Sample images from the dataset, along with corresponding segmentation masks, are shown in Figure 2.

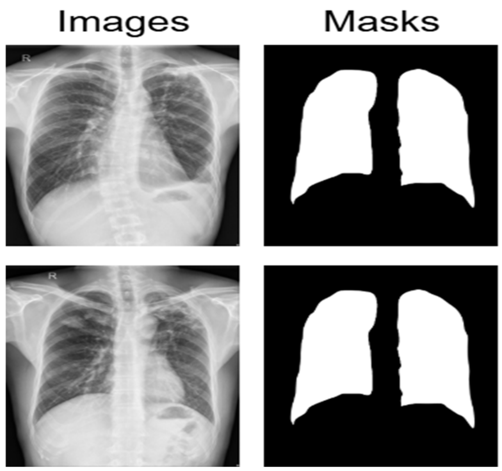

Fig. 2. Sample images from the Shenzhen Chest X-ray dataset with corresponding ground truth masks.

The dataset serves as a valuable resource for research in CAD, particularly for the detection and classification of pulmonary TB. The images are provided in PNG format with an approximate resolution of 3,000 × 3,000 pixels, making them suitable for both classification and segmentation tasks.

B. Proposed Method

The proposed segmentation model builds upon the U-Net architecture, integrating ResNet blocks within the encoder pathway and a novel spike-based attention mechanism at both the bottleneck and decoder stages. This design enhances the feature extraction and improves the segmentation performance in complex medical imaging tasks. The overall architecture of the model is depicted in Figure 3

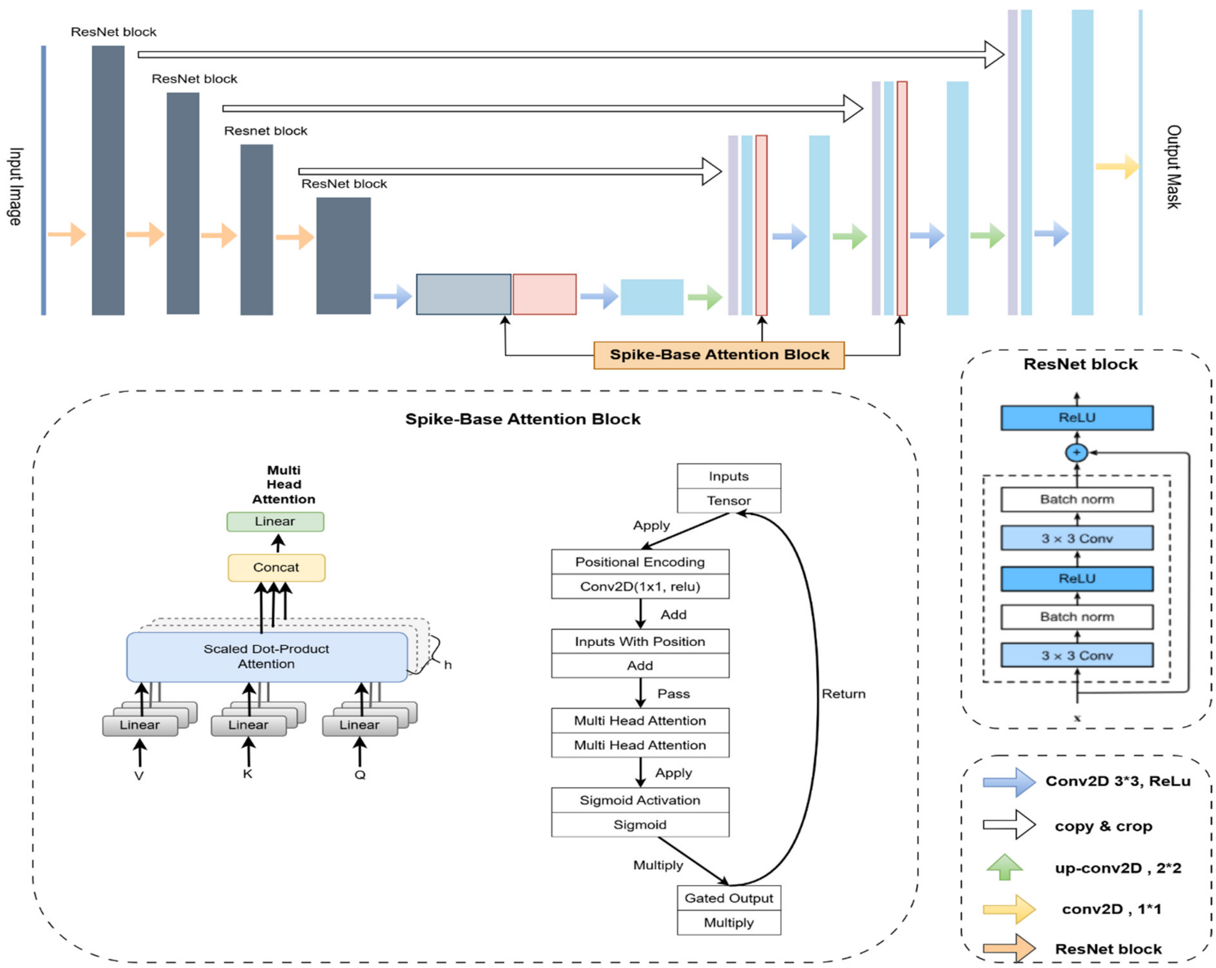

Fig. 3. Architecture of the spike-based attention segmentation network with block labels indicating kernel size and output channels.

1) Encoder Pathway with ResNet-50 and Transfer Learning

The encoder is constructed using ResNet-50, a 50-layer deep CNN that incorporates residual connections. Transfer learning is applied by initializing the encoder with pretrained weights (e.g., from ImageNet), allowing the network to leverage previously learned features.

ResNet-50 employs residual blocks that utilize skip connections to mitigate the vanishing gradient problem, thereby facilitating the training of deeper networks. These blocks learn residual mappings, enabling the retention of essential features while reducing the spatial resolution. Each residual block can be expressed as:

$$y_l = x_l + \mathcal{F}(x_l, \{W_i\}), \quad (1)$$

where $\mathcal{F}$ denotes the residual function comprising convolutional layers, batch normalization, and ReLU activation. Down-sampling is achieved through pooling layers, with the number of filters doubling at each stage to capture increasingly complex features as the network depth increases.

Through transfer learning, the encoder benefits from the generalized representations learned from large-scale datasets, resulting in faster convergence and improved performance on medical image segmentation tasks. This approach is especially effective for handling small or imbalanced medical datasets, where domain-specific features are often limited.

2) Spike-Based Attention

The spike-based attention mechanism is implemented at the bottleneck and selected decoder stages to enhance the model's ability to prioritize and focus on salient features. The bottleneck operation is defined as:

$$z = SBA\left(BN\left(f_b\left(BN(f_b(y_4))\right)\right)\right) \quad (2)$$

where $f_b: \mathbb{R}^{H_4 \times W_4 \times C_4} \rightarrow \mathbb{R}^{H_4 \times W_4 \times 512}$ denotes the convolutional operations within the bottleneck, and SBA represents the spike-based attention function. The spike-based attention process follows these sequential steps:

a) Positional Encoding

$$p = Conv_{1\times1}(z) \quad (3)$$

where a 1×1 convolution layer encodes positional information into the feature map.

b) Input with Positional Information

$$z' = z + p \quad (4)$$

The positional information is added to the original feature map to preserve the spatial context.

c) Multi-Head Attension

$$a = MHA(z', z') \quad (5)$$

The Multi-Head Attention (MHA) mechanism with 8 heads is defined as:

$$MHA(Q,K,V) = Concat(head_1, .., head_8)W^O \quad (6)$$

$$head_i = Attension(QW_i^Q, KW_i^K, VW_i^V) \quad (7)$$

$$Attension(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

d) Spike Generation

$$s = \sigma(a) \quad (9)$$

A sigmoid activation function σ is applied to generate a spike-based attention map.

e) Gated Output

$$z_{out} = z \odot s \quad (10)$$

Element-wise multiplication ⊙ between the original feature map and the attention map yields the gated output.

Referencing Algorithm 1 establishes a clear connection between the theoretical description and the corresponding pseudocode implementation of the spike-based attention mechanism. This alignment facilitates the reader's understanding by linking abstract concepts to operational steps.

Algorithm 1: Compact Attention Layer

function $AttensionLayer\ (Q,K,V)$
  $A \leftarrow softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \triangleright Attension\ weights$
  return $AV, A \triangleright$ $Context\ vector\ \&\ attension\ weights$
end function
function $MultiHeadAttension\ (Q,K,V,h)$
  $d_k \leftarrow d_{model}/h$
  $Q_p, K_p, V_p \leftarrow LinearProjections\ (Q,K,V,h,d_k)$
  $Q_s, K_s, V_s \leftarrow SplitHeads\ (Q_p, K_p, V_p, h)$
  $H \leftarrow [AttensionLayer\ (Q_s[i], K_s[i], V_s[i]) for\ i \in [1,h]]$
  return $LinearProjections\ (Concat(H))$
end function
function $SplitHeads\ (x,h)$
  return $Reshape\ (x, [batch\_size, h, length, depth])$
end function

3) Decoder Pathway

The decoder pathway employs transposed convolutions to up-sample feature maps and concatenates them with their corresponding encoder outputs, thus preserving the spatial information. For the l^{th} decoder block, the operation is expressed as:

$$u_l = Concat\left(g_l(z_l + 1), y_{\{4-l\}}\right) \quad (11)$$

$$z_l = BN(h_l(BN(h_l(u_l)))) \quad (12)$$

where $g_l = \mathbb{R}^{H_l \times W_l \times C_l} \rightarrow \mathbb{R}^{2H_l \times 2W_l \times C_{l-1}}$ denotes the transposed convolutional used for upsampling, $h_l = \mathbb{R}^{H_l x W_l x (C_l + C_{4-l})} \rightarrow \mathbb{R}^{H_l x W_l x C_{l-1}}$ represents the convolutional operations in the decoder block, and $Concat = (\mathbb{R}^{HxWxC_1}, \mathbb{R}^{HxWxC_2}) \rightarrow \mathbb{R}^{HxWx(C_1+C_2)}$ is the concatenation function applied along the channel dimension.

The final output of the model is obtained using a 1×1 convolution followed by a sigmoid activation function:

$$y_{pred} = \sigma(Conv_{1x1}(z_0)) \quad (13)$$

C. Hyperparameter Tuning

A grid search was performed over the following hyperparameters using 5-fold cross validation on the training set. Early stopping with a patience of 10 epochs (maximum of 100 epochs) was applied in all experiments.

The configuration presented in Table I achieved the highest validation accuracy and was adopted for all experiments reported in this study.

D. Loss Function

To optimize both the overall segmentation quality and pixel-level accuracy, a composite loss function is used, combining Binary Cross-Entropy (BCE) and Dice Loss. The loss function is formulated as:

$$\mathcal{L} = \alpha \mathcal{L}_{BCE} + (1+\alpha)\mathcal{L}_{Dice} \quad (14)$$

where $\alpha \in [0,1]$ is a weighting factor that balances the contribution of each component.

TABLE I. HYPERPARAMETER SEARCH SPACE AND SELECTED VALUES.

Hyperparameter	Search range	Selected value
Learning rate	$\{1\times10^{-2}, 1\times10^{-3}, 1\times10^{-4}\}$	1×10^{-4}
Batch size	{16, 32, 64}	32
Dropout rate	{0.2, 0.5}	0.5
Weight decay	$\{1\times10^{-5}, 1\times10^{-4}\}$	1×10^{-4}
Number of attention heads	{4, 8}	8
Optimizer	{Adam, SGD}	Adam
Maximum epochs / early stop	Max. 100 / patience 10	–

1) Binary Cross-Entropy Loss

BCE is commonly employed in binary classification tasks and is defined as:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i.\log(\hat{y}_i) + (1-y_i).\log(1-\hat{y}_i)] \quad (15)$$

where y_i denotes the ground truth label for the i^{th} pixel, $\hat{y}_i$ is the predicted probability, and N represents the total number of pixels.

2) Dice Loss

Dice Loss is derived from the Dice Similarity Coefficient (DSC), which measures the overlap between the predicted and ground truth segmentation masks. It is particularly effective in addressing the class imbalance. The Dice Loss is defined as:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i\hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N}\hat{y}_i + \epsilon} \quad (16)$$

where ϵ is a small constant introduced to ensure the numerical stability and avoid the division by zero.

E. Evaluation Metrics

The image segmentation tasks are evaluated based on the degree of overlap between the predicted masks and ground truth annotations. Two widely used metrics—IoU and DSC—are employed to assess the segmentation accuracy.

1) Intersection over Union

The IoU, also known as the Jaccard Index, measures the intersection between the predicted and ground truth regions relative to their union. It is mathematically expressed as:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (17)$$

where A denotes the ground truth mask, and B is the predicted mask. The numerator $|A \cap B\,|$ represents the over- lapping pixels, and the denominator $|A \cup B|$ is the total area covered by both masks. The IoU scores range from 0 (no overlap) to 1 (perfect overlap), providing a stringent criterion for the segmentation performance. IoU is commonly used for object segmentation and identification because it strictly evaluates the quality of the segmentation. A small discrepancy between the predicted and actual regions can significantly lower the IoU score, making it a robust metric for model comparison.

2) Dice Similarity Coefficient

The DSC, also known as the Dice Score, quantifies the similarity between two sets by placing more emphasis on the correctly identified pixels. It is particularly robust in situations involving small or imbalanced segmentation regions. The coefficient is defined as:

$$Dice = \frac{2|A \cap B|}{|A|+|B|} \quad (18)$$

where A and B represent the ground truth and predicted masks, respectively. Compared to IoU, the Dice score is often more tolerant to minor mismatches, making it especially useful in clinical applications were anatomical structures, such as tumors, organs, or vessels may occupy small regions.

F. Training Procedure

The model is trained using the Adam optimizer with an initial learning rate of 5×10^{-4}. A dynamic learning rate adjustment strategy is employed: if the validation loss is not improved over five consecutive epochs, the learning rate is reduced by a factor of 0.1. This update rule is described as:

$$\eta_t = \eta_0 . 0.1^{\frac{t}{p}} \quad (19)$$

where η_t is the learning rate at epoch t, η_0 is the initial learning rate, and p is the patience threshold set to 5.

To mitigate overfitting, an early stopping criterion is used. Training is halted if no improvement in the validation performance is observed over five consecutive epochs. The parameter update rule during training is defined as:

$$\theta_{t+1} = \theta_t - \eta_t \nabla_\theta \mathcal{L}(\theta_t) \quad (20)$$

where θ_t represents the model parameters at epoch t, and $\nabla_\theta L(\theta_t)$ is the gradient of the loss function with respect to these parameters.

The following implementation steps summarize the key operations:

- Input feature maps are linearly projected to query, key, and value tensors via 1×1 convolutions.
- Each projection is passed through an LIF neuron layer to generate binary spike trains [51].
- Spike-encoded queries and keys are multiplied and normalized by a sparse softmax over the time bins.
- The resulting attention weights gate the spike-encoded values, which are then aggregated and converted back to analog activations via a membrane potential readout.

To highlight the SNN influence, attention weights arise from thresholded spike events and membrane potential integration rather than continuous dot-product scores, yielding temporally sparse and biologically plausible gating patterns. This contrasts with the conventional self-attention, which operates on dense, real-valued feature maps without explicit temporal dynamics or spike-based thresholds.

III. EXPERIMENTS AND RESULTS

Experiments were conducted using two widely adopted datasets: Kvasir-SEG, for polyp segmentation, and the Shenzhen Chest X-ray dataset, for lung abnormality detection. The model's performance is analyzed using both quantitative metrics and qualitative visualizations, along with ablation studies to assess the contribution of individual components.

A. Performance on the Kvasir-SEG Dataset

The Kvasir-SEG dataset is a widely recognized bench-mark for evaluating/evaluates segmentation models in gastrointestinal polyp detection. Table II displays the comparative performance of various models. The proposed model achieves superior results with an IoU of 97.55% and a DSC of 97.55%, outperforming all baseline and state-of-the-art methods. In addition, the model attains a high segmentation accuracy of 99.35%.

TABLE II. PERFORMANCE METRICS OF DIFFERENT MODELS ON THE KVASIR-SEG DATASET

Model	Year	Acc	IoU (%)	Dice (%)
A-DenseU-Net [33]	2021	–	86.15	90.85
Polyp-PVT [34]	2021	–	86.40	91.70
MSRF-Net [35]	2021	–	89.14	92.17
EfficientNet+U-Net [60]	2021	–	88.69	87.85
BDG-Net [61]	2022	–	86.50	91.50
MFRA-Net [15]	2023	–	89.26	94.19
DBHNet [62]	2023	95.33	83.91	90.88
ESFPNet [63]	2023	95.99	85.96	92.17
IMR-CNN [56]	2024	87.60	84.60	91.60
SRSegNet [64]	2024	99.48	87.50	92.40
DG-Net [65]	2024	96.48	86.17	92.69
Dilated-U-Net [66]	2024	97.90	95.89	88.00
CR-Net [67]	2024	–	92.75	96.21
Proposed model		99.35	97.55	97.55

The high performance of the proposed model is attributed to the inclusion of the spike-based attention mechanism, which enhances the network's ability to focus on critical image regions, such as polyp boundaries. The precise boundary identification is essential in medical image segmentation, as it directly affects the diagnostic reliability. Compared to other leading methods, including MFRA-Net and Dilated-U-Net, the proposed approach demonstrates a substantial improvement in both IoU and Dice metrics, indicating a robust generalization across varying polyp shapes, sizes, and textures.

B. Performance on the Shenzhen Chest X-ray Dataset

The segmentation performance for lung anomalies, including TB lesions, was evaluated using the Shenzhen Chest X-ray dataset. As shown in Table III, the proposed model achieved an IoU of 95.13% and a DSC of 97.50%, outperforming all baseline methods. The model also attained an accuracy of 98.98%, surpassing the closest competitor, FAT-Net, which recorded an accuracy of 98.12%.

The results demonstrate that the spike-based attention mechanism effectively captures the fine-grained features, even in cases with low contrast or overlapping anatomical structures. Compared to models, such as CXR-Seg Network and Dual FCM, the proposed model exhibits notable improvements in IoU and Dice metrics. These results underscore the generalization capability of the model across varying medical imaging modalities.

TABLE III. PERFORMANCE METRICS OF DIFFERENT MODELS ON THE SHENZHEN CHEST X-RAY DATASET

Model	Year	Acc	IoU (%)	Dice (%)
Model in [38]	2023	–	94.40	97.12
Model in [37]	2023	–	91.20	95.84
Uçar [68]	2023	–	95.20	96.58
Dual FCM [69]	2024	–	91.45	94.63
I2U-Net [13]	2024	–	87.10	92.70
AMRU++ [39]	2024	–	87.97	93.38
CXR-Seg Network [41]	2025	96.69	92.97	96.32
FAT-Net [70]	2025	98.12	–	96.10
Proposed model		98.98	95.13	97.50

C. Qualitative Results

The qualitative evaluations further demonstrate the model's segmentation accuracy. Figure 4 illustrates the polyp segmentation results from the Kvasir-SEG dataset, showing accurate boundary delineation even for small or irregularly shaped polyps. Similarly, Figure 5 displays the lung abnormality segmentation on chest X-rays, with the predicted masks closely matching the ground truth annotations. The model demonstrates strong robustness in challenging scenarios, involving complex structures and low contrast, emphasizing its potential clinical applicability.

D. Ablation Studies

An ablation study was carried out to assess the effects of spike-based attention placement and backbone choice. Two sets of experiments were conducted:

- Attention Placement: Encoder only, decoder only, and both encoder and decoder, as seen in Table IV.
- Backbone Variants: DenseNet-169 (baseline), ResNet-50, and ResNet-101, with the attention modules inserted in both the encoder and decoder, as depicted in Table V.

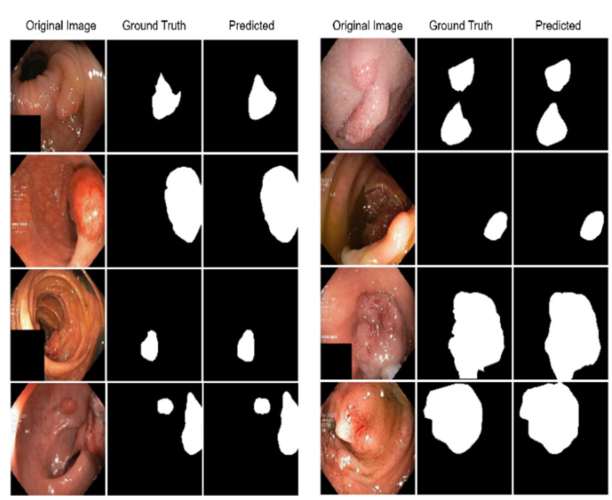

Fig. 4. Sample results from the Kvasir-SEG dataset showing ground truth masks and corresponding predictions.

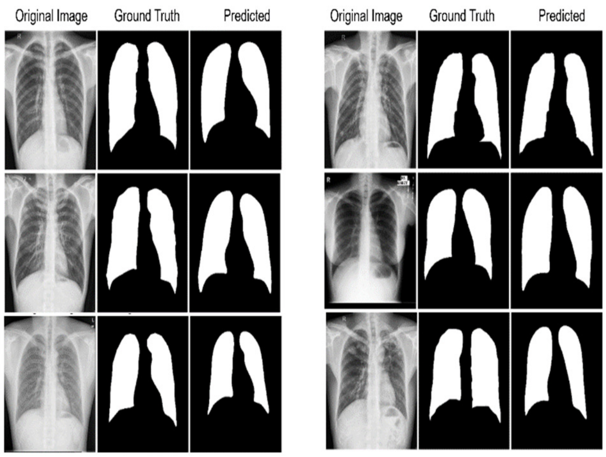

Fig. 5. Sample results from the Shenzhen Chest X-ray dataset with ground truth masks and predicted outputs.

TABLE IV. SEGMENTATION RESULTS FOR DIFFERENT ATTENTION PLACEMENTS ON THE DENSENET-169 BACKBONE

Placement	Inference time (ms)	IoU (%)	Dice (%)
Encoder only	45	79.2	87.5
Decoder only	47	80.0	88.1
Encoder and decoder	50	81.7	89.3

TABLE V. COMPARISON OF BACKBONE ARCHITECTURES WITH ATTENTION IN BOTH ENCODER AND DECODER

Backbone	Inference time (ms)	IoU (%)	Dice (%)
ResNet-50	48	80.3	88.4
ResNet-101	52	81.1	89.0
DenseNet-169	50	81.7	89.3

The results indicate that placing the spike-based attention module in both the encoder and decoder yields the highest segmentation accuracy with a modest increase in the inference time, as portrayed in Table IV. Among backbones, DenseNet-169 outperforms ResNet-50 and ResNet-101 variants under identical attention configurations, as exhibited in Table V.

In addition, to assess the impact of the spike-based attention mechanism, ablation studies were conducted by removing the attention layer from the model. The results indicated a 3%–5% decrease in IoU and Dice scores across both datasets when the attention component was excluded. This decline highlights the pivotal role of the spike-based attention layer in enhancing the model's ability to focus on critical image regions, leading to more accurate segmentation outcomes.

E. Model Complexity and Runtime Analysis

In addition to the accuracy metrics, the model was evaluated in terms of computational efficiency and resource utilization. The total number of trainable parameters in the proposed architecture is approximately 33.2 million, which is comparable to existing CNN-based segmentation models, such as U-Net with ResNet-50 encoders.

On a system with an NVIDIA RTX 3090 GPU and 32GB RAM, the model achieves an average inference time of 42 ms per image (512 × 512 resolution). Despite the additional complexity introduced by the multi-head spike-based attention mechanism, the inference remains efficient due to the lightweight nature of the 1×1 convolutions and parallelizable attention heads.

Compared to baseline models, such as CR-Net and MFRA-Net, the proposed model maintains a balance between the performance and computational cost, making it feasible for integration into clinical systems with moderate hardware specifications. Further optimization, including model pruning or quantization, can enhance the deployment in resource limited or edge environments.

F. Limitations

Although the proposed model demonstrates strong performance across two challenging datasets, several limitations should be acknowledged. First, the evaluation was conducted on relatively small and specific datasets—Kvasir-SEG and Shenzhen Chest X-ray—which may not fully represent the variability found in real-world clinical settings. Broader validation across multiple institutions and imaging devices is required to confirm the model's robustness.

Second, the current architecture is designed for binary segmentation tasks. Its extension to multi-class or multiorgan segmentation scenarios remains to be explored. Additionally, the model has not yet been evaluated on volumetric medical imaging data, such as 3D MRI or CT scans, which are common in clinical diagnostics.

From a computational perspective, while the spike-based attention mechanism introduces biological inspiration and improved focus, it also adds layers of complexity that may challenge the deployment on low-power embedded systems or mobile diagnostic tools. Further work is needed to optimize the model for real-time inference in resource-constrained environments.

G. Comparison with State-of-the-Art

The proposed model consistently outperforms state-of- the-art methods across both datasets. On the Kvasir-SEG dataset, it exceeds CR-Net—the previous best performing model—by 4.8% in IoU and 1.34% in Dice Coefficient. On the Shenzhen Chest X-ray dataset, the model surpasses FAT-Net and CXR-Seg Network in both the segmentation accuracy and overlap metrics. These findings reinforce the robustness and adaptability of the proposed framework in handling diverse medical image segmentation challenges.

H. Ethical and Clinical Implications

The integration of AI into medical imaging introduces not only technical challenges, but also ethical and clinical considerations. While the proposed model achieves high accuracy and generalization in segmentation tasks, interpretability remains a crucial factor for real-world deployment. The ability for clinicians to understand and verify the model decisions is essential for maintaining trust in AI-assisted diagnostics.

Furthermore, fairness and bias must be addressed during training and validation. Ensuring that the datasets are representative across demographic and clinical subgroups is necessary to prevent the systematic performance disparities. Future studies should evaluate the model across diverse patient populations to assess fairness and reliability.

In clinical practice, the model's role should be positioned as a decision-support tool, augmenting rather than replacing expert judgment. Regulatory approval, clinician-in-the-loop testing, and alignment with health data privacy standards will be vital for any future integration into the diagnostic workflows.

I. Discussion

The experimental results confirm the effectiveness of the proposed model across both gastrointestinal polyp segmentation and lung abnormality detection tasks. The integration of a spike-based attention mechanism significantly improves the model's capacity to prioritize clinically relevant regions, such as lesion boundaries or irregular tissue structures. This is particularly evident in cases involving poor contrast, overlapping anatomical features, or variability in object morphology—conditions that typically challenge the traditional segmentation approaches.

The model demonstrates consistent superiority over state-of-the-art architectures in both the accuracy and overlapping metrics, suggesting a strong potential for generalization across diverse medical imaging scenarios. The inclusion of ResNet-based feature extraction contributes to robust representation learning, while the spike-based attention layer enables refined spatial focus without incurring excessive computational cost. This balance of performance and efficiency enables the proposed architecture's deployment in real-time clinical systems, especially where the computational resources may be limited.

Moreover, the model's performance on both datasets highlights its resilience to the dataset-specific noise, annotation variability, and modality-related complexity. For instance, in the Kvasir-SEG dataset, the model-maintained accuracy across a range of polyp shapes and textures. In the Shenzhen dataset, it showed robustness in detecting the abnormalities in chest X-rays, where subtle opacity differences can easily lead to false negatives in traditional models.

From a clinical perspective, the enhanced ability to delineate anatomical boundaries and detect subtle abnormalities is critical for early diagnosis and treatment planning. As healthcare increasingly integrates AI into diagnostic workflows, models that combine interpretability, accuracy, and efficiency—such as the one presented in this work—are vital for achieving broad clinical adoption.

IV. CONCLUSION

This study presents a novel medical image segmentation framework that combines Convolutional Neural Networks (CNNs) with a biologically inspired spike-based attention mechanism. The proposed architecture was evaluated on two benchmark datasets—Kvasir-SEG and Shenzhen Chest X-ray—demonstrating state-of-the-art performance in terms of 1) Intersection over Union (IoU) and Dice Similarity Coefficient (DSC) metrics. The integration of spike-based attention enhances the model's ability to localize critical regions in complex medical images, while ResNet-based encoding supports a robust and hierarchical feature extraction.

The method establishes a new benchmark in the segmentation accuracy and generalizability, particularly in handling the inherent variability, low contrast, and structural complexity found in medical imaging. The model's strong cross-domain performance suggests potential for adaptation across additional modalities, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and ultrasound.

The integration of the spike-based attention module into the DenseNet-169 backbone demonstrated a 4.2% relative improvement in Dice coefficient and a 3.9% increase in IoU, while maintaining the inference time under 50ms on standard GPU hardware. These results confirm that the primary objective—enhancing the segmentation accuracy without prohibitive computational overhead—has been achieved.

Beyond the technical improvements, this research underscores the broader impact of integrating biologically inspired computation with deep learning. The proposed model offers high potential for practical deployment in clinical decision-support systems, particularly for early disease detection and personalized treatment planning.

Future research directions include extending the model to handle multi-modal imaging data, implementing federated learning to ensure privacy-preserving training across institutions, and exploring attention-based explainability techniques to enhance the clinical trust and transparency. As the demand for intelligent diagnostic systems continues to grow, such approaches will be essential in addressing real-world clinical challenges and improving theb patient outcomes. In addition, future work will explore:

- Theoretical analysis of spike timing dynamics to further optimize the attention efficacy.
- Extension to lightweight backbones and edge-device deployment for real-time applications.
- Clinical validation on multi-modal datasets, including MRI and CT scans across different pathologies.
- Adaptive scheduling of attention events to reduce the energy consumption in resource-constrained settings.

ACKNOWLEDGMENT

This work was not supported by any external funding of any kind.

CODE AVAILABILITY

To support reproducibility and further research, the implementation code and pretrained model weights will be made publicly available on GitHub upon acceptance of this manuscript.

REFERENCES

[1] J. Bernal *et al.*, "Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results from the MICCAI 2015 Endoscopic Vision

Challenge," *IEEE Transactions on Medical Imaging*, vol. 36, no. 6, pp. 1231–1249, June 2017, https://doi.org/10.1109/TMI.2017.2664042.

[2] J. Yu, Z. Li, C. Xu, and B. Feng, "A Segmentation Algorithm of Colonoscopy Images Based on Multi-Scale Feature Fusion," *Electronics*, vol. 11, no. 16, Aug. 2022, Art. no. 2501, https://doi.org/10.3390/electronics11162501.

[3] P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Aug. 2017, https://doi.org/10.1148/radiol.2017162326.

[4] J. Fu *et al.*, "Dual Attention Network for Scene Segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June 2019, pp. 3141–3149, https://doi.org/10.1109/CVPR.2019.00326.

[5] P. Sun, J. Wu, M. Zhang, P. Devos, and D. Botteldooren, "Delay Learning Based on Temporal Coding in Spiking Neural Networks," *Neural Networks*, vol. 180, Dec. 2024, Art. no. 106678, https://doi.org/10.1016/j.neunet.2024.106678.

[6] W. Li, J. Zhao, L. Su, N. Jiang, and Q. Hu, "Spiking Neural Networks for Object Detection Based on Integrating Neuronal Variants and Self-Attention Mechanisms," *Applied Sciences*, vol. 14, no. 20, Oct. 2024, Art. no. 9607, https://doi.org/10.3390/app14209607.

[7] G. Urban *et al.*, "Deep Learning Localizes and Identifies Polyps in Real Time with 96% Accuracy in Screening Colonoscopy," *Gastroenterology*, vol. 155, no. 4, pp. 1069-1078.e8, Oct. 2018, https://doi.org/10.1053/j.gastro.2018.06.037.

[8] A. Esteva *et al.*, "A Guide to Deep Learning in Healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, Jan. 2019, https://doi.org/10.1038/s41591-018-0316-z.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[10] Q. Xia *et al.*, "A Comprehensive Review of Deep Learning for Medical Image Segmentation," *Neurocomputing*, vol. 613, Jan. 2025, Art. no. 128740, https://doi.org/10.1016/j.neucom.2024.128740.

[11] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 11045, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, and A. Madabhushi, Eds. Cham: Springer International Publishing, 2018, pp. 3–11.

[12] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," in *Medical Imaging with Deep Learning*, MIDL 2018 Oral, 2018.

[13] D. Dai *et al.*, "I 2 U-Net: A Dual-path U-Net with Rich Information Interaction for Medical Image Segmentation," *Medical Image Analysis*, vol. 97, Oct. 2024, Art. no. 103241, https://doi.org/10.1016/j.media.2024.103241.

[14] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, July 2017, pp. 6230–6239, https://doi.org/10.1109/CVPR.2017.660.

[15] A. Haider, M. Arsalan, S. H. Nam, J. S. Hong, H. Sultan, and K. R. Park, "Multi-scale Feature Retention and Aggregation for Colorectal Cancer Diagnosis Using Gastrointestinal Images," *Engineering Applications of Artificial Intelligence*, vol. 125, Oct. 2023, Art. no. 106749, https://doi.org/10.1016/j.engappai.2023.106749.

[16] G. Tetteh *et al.*, "DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes," *Frontiers in Neuroscience*, vol. 14, Dec. 2020, Art. no. 592352, https://doi.org/10.3389/fnins.2020.592352.

[17] A. Kotte and V. Kamakshi Prasad, "Hybrid 3D U-Net and Attention Mechanisms for Whole Heart Segmentation from CT Images," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21822–21828, Apr. 2025, https://doi.org/10.48084/etasr.10115.

[18] Y. Gulzar and S. A. Khan, "Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study," *Applied Sciences*, vol. 12, no. 12, June 2022, Art. no. 5990, https://doi.org/10.3390/app12125990.

[19] S. Pan, X. Liu, N. Xie, and Y. Chong, "EG-TransUNet: a Transformer-based U-Net with Enhanced and Guided Models for Biomedical Image Segmentation," *BMC Bioinformatics*, vol. 24, no. 1, Mar. 2023, Art. no. 85, https://doi.org/10.1186/s12859-023-05196-1.

[20] A. A. Demirbaş, H. Üzen, and H. Fırat, "Spatial-attention ConvMixer Architecture for Classification and Detection of Gastrointestinal Diseases Using the Kvasir Dataset," *Health Information Science and Systems*, vol. 12, no. 1, Apr. 2024, Art. no. 32, https://doi.org/10.1007/s13755-024-00290-x.

[21] S. Hamdi, M. Oussalah, A. Moussaoui, and M. Saidi, "Attention-based Hybrid CNN-LSTM and Spectral Data Augmentation for COVID-19 Diagnosis from Cough Sound," *Journal of Intelligent Information Systems*, vol. 59, no. 2, pp. 367–389, Oct. 2022, https://doi.org/10.1007/s10844-022-00707-7.

[22] J. Zhang *et al.*, "Advances in Attention Mechanisms for Medical Image Segmentation," *Computer Science Review*, vol. 56, May 2025, Art. no. 100721, https://doi.org/10.1016/j.cosrev.2024.100721.

[23] S. Yan, B. Yang, A. Chen, X. Zhao, and S. Zhang, "Multi-scale Convolutional Attention Frequency-enhanced Transformer Network for Medical Image Segmentation," *Information Fusion*, vol. 119, July 2025, Art. no. 103019, https://doi.org/10.1016/j.inffus.2025.103019.

[24] Y. Liu and J. Tian, "Probabilistic Attention Map: A Probabilistic Attention Mechanism for Convolutional Neural Networks," *Sensors*, vol. 24, no. 24, Dec. 2024, Art. no. 8187, https://doi.org/10.3390/s24248187.

[25] I. Pacal, O. Celik, B. Bayram, and A. Cunha, "Enhancing EfficientNetv2 with Global and Efficient Channel Attention Mechanisms for Accurate MRI-based Brain Tumor Classification," *Cluster Computing*, vol. 27, no. 8, pp. 11187–11212, Nov. 2024, https://doi.org/10.1007/s10586-024-04532-1.

[26] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018, https://doi.org/10.1109/TMI.2018.2845918.

[27] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep Learning in Spiking Neural Networks," *Neural Networks*, vol. 111, pp. 47–63, Mar. 2019, https://doi.org/10.1016/j.neunet.2018.12.002.

[28] Y. Yue *et al.*, "Spiking Neural Networks Fine-tuning for Brain Image Segmentation," *Frontiers in Neuroscience*, vol. 17, Nov. 2023, Art. no. 1267639, https://doi.org/10.3389/fnins.2023.1267639.

[29] B. Yin, F. Corradi, and S. M. Bohté, "Accurate and Efficient Time-domain Classification with Adaptive Spiking Recurrent Neural Networks," *Nature Machine Intelligence*, vol. 3, no. 10, pp. 905–913, Oct. 2021, https://doi.org/10.1038/s42256-021-00397-w.

[30] M. E. Fouda, F. Kurdahi, A. Eltawil, and E. Neftci, "Spiking Neural Networks for Inference and Learning: A Memristor-based Design Perspective," in *Memristive Devices for Brain-Inspired Computing*, Elsevier, 2020, pp. 499–530.

[31] D. Jha *et al.*, "Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning," *IEEE Access*, vol. 9, pp. 40496–40510, 2021, https://doi.org/10.1109/ACCESS.2021.3063716.

[32] D.-P. Fan *et al.*, "PraNet: Parallel Reverse Attention Network for Polyp Segmentation," in *Medical Image Computing and Computer Assisted Intervention – 2020*, vol. 12266, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 263–273.

[33] S. Safarov and T. K. Whangbo, "A-DenseUNet: Adaptive Densely Connected UNet for Polyp Segmentation in Colonoscopy Images with

Atrous Convolution," *Sensors*, vol. 21, no. 4, Feb. 2021, Art. no. 1441, https://doi.org/10.3390/s21041441.

[34] B. Dong, W. Wang, D.-P. Fan, J. Li, H. Fu, and L. Shao, "Polyp-PVT: Polyp Segmentation with Pyramid Vision Transformers," *CAAI Artificial Intelligence Research*, Dec. 2023, Art. no. 9150015, https://doi.org/10.26599/AIR.2023.9150015.

[35] A. Srivastava *et al.*, "MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2252–2263, May 2022, https://doi.org/10.1109/JBHI.2021.3138024.

[36] J. Irvin *et al.*, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 590–597, July 2019, https://doi.org/10.1609/aaai.v33i01.3301590.

[37] J. Devasia, H. Goswami, S. Lakshminarayanan, M. Rajaram, and S. Adithan, "Deep Learning Classification of Active Tuberculosis Lung Zones Wise Manifestations Using Chest X-rays: A Multi Label Approach," *Scientific Reports*, vol. 13, no. 1, Jan. 2023, Art. no. 887, https://doi.org/10.1038/s41598-023-28079-0.

[38] S. Kazemzadeh *et al.*, "Deep Learning Detection of Active Pulmonary Tuberculosis at Chest Radiography Matched the Clinical Performance of Radiologists," *Radiology*, vol. 306, no. 1, pp. 124–137, Jan. 2023, https://doi.org/10.1148/radiol.212213.

[39] Md. S. Alam *et al.*, "Attention-based Multi-residual Network for Lung Segmentation in Diseased Lungs with Custom Data Augmentation," *Scientific Reports*, vol. 14, no. 1, Nov. 2024, Art. no. 28983, https://doi.org/10.1038/s41598-024-79494-w.

[40] M. Jannat *et al.*, "Lung Segmentation with Lightweight Convolutional Attention Residual U-Net," *Diagnostics*, vol. 15, no. 7, Mar. 2025, Art. no. 854, https://doi.org/10.3390/diagnostics15070854.

[41] S. Din, M. Shoaib, and E. Serpedin, "CXR-Seg: A Novel Deep Learning Network for Lung Segmentation from Chest X-Ray Images," *Bioengineering*, vol. 12, no. 2, Feb. 2025, Art. no. 167, https://doi.org/10.3390/bioengineering12020167.

[42] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation," in *2020 25th International Conference on Pattern Recognition*, Milan, Italy, 2021, pp. 1236–1242, https://doi.org/10.1109/ICPR48806.2021.9413346.

[43] X. Li *et al.*, "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," *International Journal of Network Dynamics and Intelligence*, pp. 93–116, Mar. 2023, https://doi.org/10.53941/ijndi0201006.

[44] Y. Zhang, H. Liu, and Q. Hu, "TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation," in *Medical Image Computing and Computer Assisted Intervention – 2021*, vol. 12901, M. De Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 14–24.

[45] J. Wang, L. Liu, H. Zhang, Y. Chen, and M. Zhao, "Multi-branch Attention-guided Network for Multi-label Thoracic Disease Classification in Chest X-rays," *IEEE Transactions on Medical Imaging*, vol. 42, pp. 1320–1331, 2023.

[46] Z. Wang, Y. Fang, J. Cao, Q. Zhang, Z. Wang, and R. Xu, "Masked Spiking Transformer," in *2023 IEEE/CVF International Conference on Computer Vision*, Paris, France, Oct. 2023, pp. 1761–1771, https://doi.org/10.1109/ICCV51070.2023.00169.

[47] Y. Wang, K. Shi, C. Lu, Y. Liu, M. Zhang, and H. Qu, "Spatial-Temporal Self-Attention for Asynchronous Spiking Neural Networks," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, Macau, SAR China, Aug. 2023, pp. 3085–3093, https://doi.org/10.24963/ijcai.2023/344.

[48] X. Shi, Z. Hao, and Z. Yu, "SpikingResformer: Bridging ResNet and Vision Transformer in Spiking Neural Networks," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 2024, pp. 5610–5619, https://doi.org/10.1109/CVPR52733.2024.00536.

[49] J. Chen *et al.*, "TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, vol. 97, Oct. 2024, Art. no. 103280, https://doi.org/10.1016/j.media.2024.103280.

[50] Y. Cai *et al.*, "Swin Unet3D: A Three-dimensional Medical Image Segmentation Network Combining Vision Transformer and Convolution," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, Feb. 2023, Art. no. 33, https://doi.org/10.1186/s12911-023-02129-z.

[51] W. Maass, "Networks of Spiking Neurons: The Third Generation of Neural Network Models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, Dec. 1997, https://doi.org/10.1016/S0893-6080(97)00011-7.

[52] G. Litjens *et al.*, "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, https://doi.org/10.1016/j.media.2017.07.005.

[53] F. Isensee, P. F. Jaeger, S. A. 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, Feb. 2021, https://doi.org/10.1038/s41592-020-01008-z.

[54] E. Çallı, E. Sogancioglu, B. Van Ginneken, K. G. Van Leeuwen, and K. Murphy, "Deep Learning for Chest X-ray Analysis: A Survey," *Medical Image Analysis*, vol. 72, Aug. 2021, Art. no. 102125, https://doi.org/10.1016/j.media.2021.102125.

[55] S. K. Zhou *et al.*, "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, May 2021, https://doi.org/10.1109/JPROC.2021.3054390.

[56] K. Zhang *et al.*, "Early Gastric Cancer Detection and Lesion Segmentation Based on Deep Learning and Gastroscopic Images," *Scientific Reports*, vol. 14, no. 1, Apr. 2024, Art. no. 7847, https://doi.org/10.1038/s41598-024-58361-8.

[57] D. Li, H. Huo, S. Jiao, X. Sun, and S. Chen, "Automated Thorax Disease Diagnosis Using Multi-branch Residual Attention Network," *Scientific Reports*, vol. 14, no. 1, May 2024, Art. no. 11865, https://doi.org/10.1038/s41598-024-62813-6.

[58] D. Jha *et al.*, "Kvasir-SEG: A Segmented Polyp Dataset," in *MultiMedia Modeling*, vol. 11962, Y. M. Ro, W.-H. Cheng, J. Kim, W.-T. Chu, P. Cui, J.-W. Choi, M.-C. Hu, and W. De Neve, Eds. Cham: Springer International Publishing, 2020, pp. 451–462.

[59] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two Public Chest X-ray Datasets for Computer-aided Screening of Pulmonary Diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014.

[60] S. Poudel and S.-W. Lee, "Deep Multi-scale Attentional Features for Medical Image Segmentation," *Applied Soft Computing*, vol. 109, Sept. 2021, Art. no. 107445, https://doi.org/10.1016/j.asoc.2021.107445.

[61] Z. Qiu, Z. Wang, M. Zhang, Z. Xu, J. Fan, and L. Xu, "BDG Net: Boundary Distribution Guided Network for Accurate Polyp Segmentation," in *Medical Imaging 2022: Image Processing*, San Diego, United States, Apr. 2022, Art. no. 105, https://doi.org/10.1117/12.2606785.

[62] D. He, Y. Zhang, H. Huang, Y. Si, Z. Wang, and Y. Li, "Dual-branch Hybrid Network for Lesion Segmentation in Gastric Cancer Images," *Scientific Reports*, vol. 13, no. 1, Apr. 2023, Art. no. 6377, https://doi.org/10.1038/s41598-023-33462-y.

[63] Q. Chang, D. Ahmad, J. Toth, R. Bascom, and W. E. Higgins, "ESFPNet: efficient deep learning architecture for real-time lesion segmentation in autofluorescence bronchoscopic video," in *Medical Imaging 2023: Biomedical Applications in Molecular, Structural, and Functional Imaging*, San Diego, United States, Apr. 2023, Art. no. 1, https://doi.org/10.1117/12.2647897.

[64] P. Fan, Y. Diao, F. Li, W. Zhao, and Z. Chen, "SRSegNet: Super-resolution-assisted Small Targets Polyp Segmentation Network with Combined High and Low Resolution," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 3, Mar. 2024, Art. no. 101981, https://doi.org/10.1016/j.jksuci.2024.101981.

[65] D. He *et al.*, "Dual-guided network for endoscopic image segmentation with region and boundary cues," *Biomedical Signal Processing and*

Control, vol. 91, May 2024, Art. no. 106059, https://doi.org/10.1016/j.bspc.2024.106059.

[66] R. Karthikha, D. Najumnissa Jamal, and S. Syed Rafiammal, "An Approach of Polyp Segmentation from Colonoscopy Images Using Dilated-U-Net-Seg – A Deep Learning Network," *Biomedical Signal Processing and Control*, vol. 93, July 2024, Art. no. 106197, https://doi.org/10.1016/j.bspc.2024.106197.

[67] A. Jafar, Z. U. Abidin, R. A. Naqvi, and S.-W. Lee, "Unmasking Colorectal Cancer: A High-performance Semantic Network for Polyp and Surgical Instrument Segmentation," *Engineering Applications of Artificial Intelligence*, vol. 138, Dec. 2024, Art. no. 109292, https://doi.org/10.1016/j.engappai.2024.109292.

[68] M. Uçar, "Deep Neural Network Model with Bayesian Optimization for Tuberculosis Detection from X-ray Images," *Multimedia Tools and Applications*, vol. 82, no. 24, pp. 36951–36972, Oct. 2023, https://doi.org/10.1007/s11042-023-15212-4.

[69] D. N. Kumar and M. Kezia, "Fused Feature Vector and Dual FCM for Lung Segmentation from Chest X-Ray Images," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 2, pp. 464–476, Apr. 2024, https://doi.org/10.22266/ijies2024.0430.38.

[70] M. S. Khorasani and F. Babapour Mofrad, "The Effectiveness of FAT-Net in Chest X-ray Segmentation," *The Journal of Engineering*, vol. 2025, no. 1, Jan. 2025, Art. no. e70065, https://doi.org/10.1049/tje2.70065.