LGPS: A Lightweight GAN-Based Approach for Polyp Segmentation in Colonoscopy Images

Fiseha B. Tesema, *Member, IEEE*, Alejandro Guerra Manzanares ,Tianxiang Cui, Qian Zhang, Moses Solomon, Sean He

***Abstract*—Colorectal cancer (CRC) is a major global cause of cancer-related deaths, with early polyp detection and removal during colonoscopy being crucial for prevention. While deep learning methods have shown promise in polyp segmentation, challenges such as high computational costs, difficulty in segmenting small or low-contrast polyps, and limited generalizability across datasets persist. To address these issues, we propose LGPS, a lightweight GAN-based framework for polyp segmentation. LGPS incorporates three key innovations: (1) a MobileNetV2 backbone enhanced with modified residual blocks and Squeeze-and-Excitation (ResE) modules for efficient feature extraction; (2) Convolutional Conditional Random Fields (ConvCRF) for precise boundary refinement; and (3) a hybrid loss function combining Binary Cross-Entropy, Weighted IoU Loss, and Dice Loss to address class imbalance and enhance segmentation accuracy. LGPS is validated on five benchmark datasets and compared with state-of-the-art(SOTA) methods. On the largest and challenging PolypGen test dataset, LGPS achieves a Dice of 0.7299 and an IoU of 0.7867, outperformed all SOTA works and demonstrating robust generalization. With only 1.07 million parameters, LGPS is 17 times smaller than the smallest existing model, making it highly suitable for real-time clinical applications. Its lightweight design and strong performance underscore its potential for improving early CRC diagnosis. Code is available at https://github.com/Falmi/LGPS/.**

***Index Terms*—Deep Learning, Image Segmentation, Polyp Segmentation, Medical Image Analysis, Generative Adversarial Networks, GAN**

I. Introduction

Colorectal cancer (CRC) is one of the most prevalent and deadly forms of cancer worldwide, accounting for 10% of all cancer-related deaths [2]. Early detection and removal of polyps during colonoscopy are critical to preventing the progression of CRC. However, manually identifying and segmenting polyps in colonoscopy images is challenging due to their significant variability in size, shape, color, and texture, along with the presence of image artifacts such as motion blur, reflectance, and low contrast with surrounding tissues [3], [22]. These challenges often lead to false negatives or inaccurate segmentations, underscoring the need for robust and reliable Computer-Aided Diagnosis (CAD) systems.

In recent years, deep learning-based approaches, particularly Convolutional Neural Networks (CNNs), have shown remarkable success in automating polyp segmentation in colonoscopy images [5], [6]. Models such as U-Net [10], PraNet [6], HarDNet-MSEG [9], and WDFF-Net [24] have achieved remarkable segmentation accuracy by leveraging encoder-decoder architectures, attention mechanisms, and multi-scale feature fusion. Despite their effectiveness, these methods face three significant limitations: (i) high computational cost, making them unsuitable for real-time applications [4]; (ii) difficulty in segmenting small or low-contrast polyps [18]; and (iii) poor generalization across datasets due to variations in imaging conditions [22]. These limitations hinder the adoption of these segmentation methods in clinical practice.

To address these challenges, we propose the Lightweight GAN-based framework for Polyp Segmentation (LGPS). As shown in Figure 1, LGPS achieves state-of-the-art (SOTA) segmentation accuracy with only 1.07 million parameters, making it 17 times smaller than the smallest existing SOTA method. The framework introduces three key innovations: (1) a GAN architecture with a generator for image segmentation and a discriminator enhanced with Convolutional Conditional Random Fields (ConvCRF) to refine spatial coherence and boundary details; (2) a custom hybrid loss function combining Binary Cross-Entropy (BCE), weighted IoU, and Dice losses to address class imbalance and ensure precise segmentation; and (3) a MobileNetV2-based generator with modified residual Squeeze-and-Excitation (ReSE) blocks, enabling SOTA performance while maintaining a lightweight design suitable for deployment on resource-constrained devices. LGPS also demonstrates robust generalization across internal and external validation datasets, including the challenging PolypGen dataset, outperforming larger and more complex models. Its lightweight design and superior accuracy underscore its potential for real-time clinical applications.

II. Related Work

The rapid development of deep learning methods has significantly advanced computer vision, particularly in medical image segmentation. In this section, we provide an overview of major progress in medical image segmentation methods,

This work was supported by RKE Internal Seed Funding for a Small Research and Knowledge Exchange Project [105231000022] and Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute[I01240100007], University of Nottingham Ningbo China(UNNC), Ningbo, China. (Corresponding author: Fiseha B. Tesema)

Fiseha B. Tesema is with the School of Computer Science and the Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute, UNNC, Ningbo, China. (e-mail:Fiseha-Berhanu.Tesema@nottinham.edu.cn)

Alejandro Guerra Manzanares is with School of Computer Science, UNNC, Ningbo, China (e-mail:alejandro.guerra-manzanares@nottingham.edu.cn)

Tianxiang Cui, is with School of Computer Science, UNNC, Ningbo, China (e-mail:tianxiang.cui@nottingham.edu.cn)

Qian Zhang, is with School of Computer Science, UNNC, Ningbo, China (e-mail:qian.zhang@nottingham.edu.cn)

Moses Solomon, is with Department of Chemical and Environmental Engineering, and the Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute, UNNC, Ningbo, China.(e-mail:moses.solomon@nottingham.edu.cn)

Sean He, is with School of Computer Science, UNNC, Ningbo, China (e-mail:sean.he@nottingham.edu.cn)

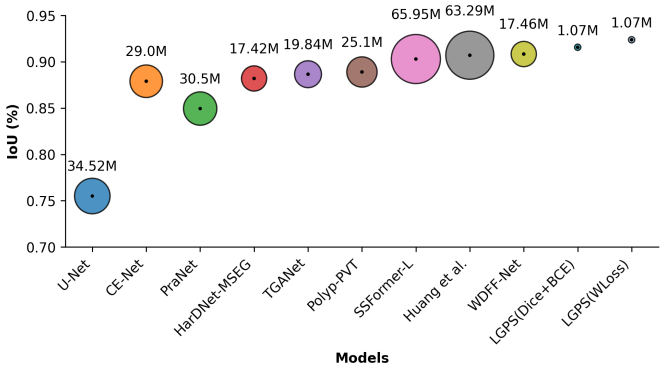

Fig. 1. **Comparison of model size and performance.** The area of the circles relates to the size of the model in terms of the number of parameters, while the left axis reports the IoU value of each model on the CVC-ClinicDB dataset. The proposed model, LGPS, outperforms all state-of-the-art models with 17 times fewer parameters.

focusing on polyp segmentation research. We highlight the strengths and limitations of existing approaches, which emphasize the need for lightweight, efficient, and generalizable models suitable for real-time clinical applications.

A. U-Net Architectures

The domain of medical image segmentation has made remarkable progress with the advent of deep learning. U-Net [28], introduced in 2015, revolutionized the field with its encoder-decoder architecture and skip connections, enabling precise localization and segmentation of medical structures. Since then, U-Net has become a benchmark for many segmentation tasks due to its simplicity and effectiveness. However, it struggles with complex visual patterns, such as polyps with blurry edges or low contrast, often leading to under-segmentation or over-segmentation [45].

To address these limitations, subsequent works have introduced advanced architectures and techniques. CE-Net [27] enhances U-Net by incorporating a context encoder module to capture global context information, improving segmentation accuracy in challenging regions. Similarly, PraNet [6] incorporates parallel reverse attention modules to focus on boundary cues and region relationships, achieving SOTA performance in polyp segmentation tasks. Despite their improvements, these models often require significant computational resources, making them unsuitable for real-time applications or deployment on resource-constrained environments [9].

B. Attention Mechanisms and Feature Aggregation

Attention mechanisms have emerged as a powerful tool for improving segmentation accuracy by enabling models to focus on diagnostically relevant regions [8]. HarDNet-MSEG [9] uses a cascaded partial decoder and the HarDNet68 [15] backbone to achieve high accuracy and inference speed, making it suitable for real-time applications. SANet [16] introduces a shallow attention module to address pixel imbalance in small polyps, effectively reducing background noise and improving segmentation accuracy. TGANet [18] leverages text-based embeddings and auxiliary classification tasks to handle drastic scale variations in polyp size. While these methods demonstrate remarkable results, their computational complexity remains a significant drawback [19].

C. Transformer-Based Approaches

The emergence of transformer-based models has further advanced the field of medical image segmentation. Transformers excel at capturing long-range dependencies, making them particularly effective for complex segmentation tasks [7]. Polyp-PVT [4] employs a Pyramid Vision Transformer (PVT) to learn robust feature representations, achieving SOTA performance in polyp segmentation. SSFormer [19] introduces a progressive local decoder to refine segmentation results, while WDFF-Net [24] combines dual-branch feature fusion with progressive and scale-aware strategies to address under-segmentation and size variation. Despite their effectiveness, transformer-based models are often computationally expensive, with large numbers of parameters and slow inference times, limiting their practicality for real-time applications [25].

D. Lightweight Models for Real-Time Deployment

Given the computational challenges of existing methods, there is a growing demand for lightweight models that can deliver competitive performance while maintaining low memory and computational requirements, especially in resource-constrained environments [23]. Several works in the literature have proposed lightweight architectures for image segmentation, demonstrating the feasibility of efficient and real-time solutions.

Ni et al. [36] introduced a bilinear attention network with an adaptive receptive field for the segmentation of surgical instruments. Their approach leverages bilinear attention mechanisms to capture fine-grained details while maintaining computational efficiency. Similarly, Wang et al. [37] proposed LEDNet, a lightweight encoder-decoder network that uses ResNet50 in the encoder block and an attention pyramidal network in the decoder block. LEDNet achieves real-time semantic segmentation with a significant reduction in computational complexity, making it suitable for resource-constrained environments. Another notable contribution is Squeeze U-Net [38], which is inspired by the U-Net architecture [28]. Squeeze U-Net achieves a 12$\times$ reduction in model size compared to traditional U-Net while maintaining efficient performance in terms of multiplication-accumulation (MAC) operations and memory usage. This makes it particularly suitable for deployment on devices with limited computational resources. ERFNet [39] introduced an efficient residual factorized convolutional network for real-time semantic segmentation. ERFNet achieves a balance between accuracy and speed, making it a strong candidate for real-time applications.

These works highlight the potential of lightweight architectures for real-time image segmentation. However, most existing lightweight models are designed for general-purpose segmentation tasks and have not been specifically optimized

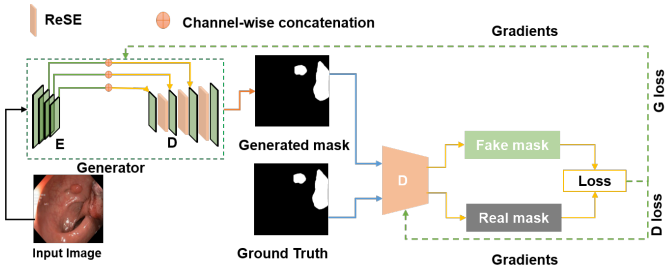

Fig. 2. a) Overview of the LGPS architecture, showing the generator and discriminator components.

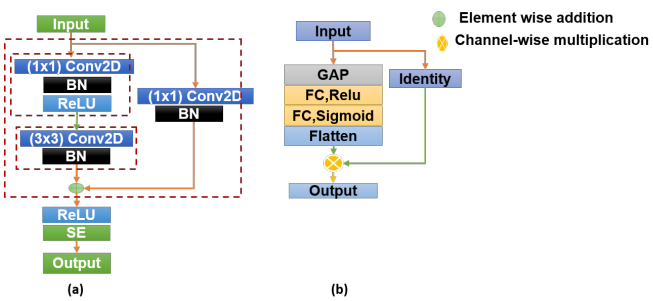

Fig. 3. a) ReSE block b) SE block

for polyp segmentation in colonoscopy images. Polyp segmentation presents unique challenges, such as the need to handle small and irregularly shaped polyps, blurry boundaries, and low contrast with surrounding tissues. These challenges require the development of specialized lightweight models tailored to the particularities of polyp segmentation.

Addressing this significant gap, we propose LGPS, an efficient lightweight GAN-based framework designed specifically for polyp segmentation. LGPS achieves SOTA segmentation accuracy without sacrificing computational efficiency, making it suitable for real-time clinical applications.

III. Proposed Model Architecture

LGPS is a novel GAN-based architecture designed for efficient and precise polyp segmentation in medical images. The model leverages a lightweight backbone, modified Residual Blocks with Squeeze-and-Excitation (ReSE) [30] mechanisms, and a refinement module to achieve SOTA segmentation accuracy while maintaining computational efficiency. As shown in Figure 2(a), the LGPS architecture consists of two primary components: a generator (G) that produces segmentation masks from input images and a discriminator (D) that evaluates the quality of these masks. In the following, we provide a detailed description of each component, along with theoretical justifications for their design.

A. Generator Architecture

The generator follows an encoder-decoder architecture with a modified MobileNetV2 backbone, chosen for its efficiency and lightweight design. The encoder (E) extracts multi-scale features from the input image, while the decoder (D) refines and upsamples these features to produce precise segmentation masks. We perform key modifications to the MobileNetV2 [29] backbone to reduce its size to 1.07 million parameters, ensuring computational efficiency for real-time applications. The components of the generator are as follows:

1) Encoder: The encoder is built using a pre-trained MobileNetV2 model, which employs depthwise separable convolutions to reduce computational complexity while maintaining performance. To further optimize the model, we made the following modifications to MobileNetV2: (i) reduce the number of filters in each layer by a factor of 2 to decrease the number of parameters while retaining essential feature extraction capabilities; (ii) remove the final classification layer and redundant intermediate layers to keep only the essential feature extraction layers for segmentation tasks; and (iii) add depthwise separable convolutions with reduced expansion factors to minimize computational overhead.

These modifications resulted in a lightweight MobileNetV2 backbone with only 1.07 million parameters. The encoder extracts feature maps from four intermediate layers, which are used as skip connections to preserve spatial details during decoding. This multi-scale feature extraction enables the model to handle polyps of varying sizes and shapes, including small polyps that are often missed by other methods.

2) Modified Residual with Squeeze-and-Excitation block (ReSE): The ReSE block is an essential component of the generator architecture, aiming to enhance feature extraction and recalibration. As shown in Figure 3 (a), it combines traditional residual connections with the SE mechanism, enabling the model to dynamically recalibrate feature maps and focus on diagnostically relevant regions.

The ReSE block consists of several components that work together to improve feature representation and segmentation performance. They are described as follows.

The first component is the bottleneck layer, which reduces the number of channels by a factor of 4 using a 1x1 convolution layer. This step reduces computational complexity while preserving essential features. The output of the 1x1 convolution is passed through a Batch Normalization layer and a ReLU activation function, expressed as:

$$x_{\text{bottleneck}} = \text{ReLU}(\text{BatchNorm}(\text{Conv2D}_{1\times1}(x))), \quad (1)$$

where x is the input tensor, and $\text{Conv2D}_{1\times1}$ denotes a 1x1 convolution.

Next, the bottleneck output is passed through a 3x3 convolution layer to extract spatial features. This layer captures local patterns and structures in the feature maps, which are critical for accurate polyp segmentation. The output of the 3x3 convolution is normalized using Batch Normalization:

$$x_{\text{spatial}} = \text{BatchNorm}(\text{Conv2D}_{3\times3}(x_{\text{bottleneck}})). \quad (2)$$

A residual connection is then added between the input and the output of the 3x3 convolution to facilitate gradient flow and improve training stability[46]. If the number of channels in the input tensor does not match the output tensor,

a 1x1 convolution is applied to the input tensor to adjust its dimensions:

$$x_{\text{shortcut}} = \text{BatchNorm}(\text{Conv2D}_{1\times1}(x)). \tag{3}$$

The residual connection is implemented using an Add operation, followed by a ReLU activation:

$$x_{\text{residual}} = \text{ReLU}(\text{Add}([x_{\text{shortcut}}, x_{\text{spatial}}])). \tag{4}$$

Finally, the Squeeze-and-Excitation (SE) block recalibrates the feature maps by modeling channel-wise interdependencies[30]. As show in Figure 3 (b), the SE mechanism consists of three steps. First, the squeeze block uses global average pooling to aggregate spatial information into channel-wise descriptors:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j), \tag{5}$$

where $x_c(i, j)$ is the feature map at spatial location (i, j) for channel c, and $H \times W$ is the spatial dimension. Second, the excitation block uses two fully connected (FC) layers to model non-linear channel interdependencies, generating attention weights:

$$s_c = \sigma(W_2\delta(W_1 z_c)), \tag{6}$$

where W_1 and W_2 are learnable weights, δ is the ReLU activation, and σ is the sigmoid activation. Third, the recalibration block applies the attention weights s_c to the input feature maps using a Multiply operation, emphasizing diagnostically relevant features:

$$x_{\text{se}} = \text{Multiply}([x_{\text{residual}}, s]). \tag{7}$$

This ReSE block, with its combination of bottleneck layers, spatial feature extraction, residual connections, and SE mechanisms, significantly enhances the generator's ability to extract and recalibrate features for accurate polyp segmentation.

3) Decoder: The decoder progressively upsamples the feature maps and concatenates them with skip connections from the encoder to recover spatial details. It consists of four upsampling stages, each followed by the ReSE. The upsampling is performed using bilinear interpolation, and the skip connections are concatenated with the upsampled feature maps to preserve spatial information. The final output of the decoder is a binary segmentation map, obtained by applying a 1x1 convolution with a sigmoid activation:

$$M_{\text{pred}} = \sigma(\text{Conv2D}(1, (1, 1))(F_d^4)), \tag{8}$$

where F_d^4 is the final feature map from the decoder. The use of skip connections and progressive upsampling ensures that the model preserves fine-grained spatial details, a key strength for accurate polyp segmentation.

B. Discriminator Architecture

The discriminator employs a patch-based adversarial framework with Convolutional Conditional Random Fields (ConvCRF) refinement to improve spatial consistency in polyp segmentation. It processes concatenated pairs of input images and predicted masks $(I_{\text{in}}, M_{\text{pred}}) \in \mathbb{R}^{256\times256\times4}$ through five convolutional layers. Each layer uses a kernel size of $(3, 3)$, stride 2, and LeakyReLU activation ($\alpha = 0.2$), progressively increasing the number of filters from 64 to 512. The final layer produces a patch-wise real/fake probability map, providing fine-grained feedback to the generator.

To address spatial inconsistency in GAN-based segmentation, we introduce ConvCRF layers, which refine local spatial coherence through learnable 3×3 convolutions. A ConvCRF layer consists of a 3×3 convolutional operation followed by a sigmoid activation:

$$\text{ConvCRF}(F_d^i) = \sigma(\text{Conv}_{3\times3}(F_d^i)), \tag{9}$$

where F_d^i represents the feature maps from the previous layer, and σ is the sigmoid activation function. This operation enforces smoothness in predicted masks while preserving edges.

In our implementation, four ConvCRF layers are applied sequentially after the final convolutional layer for refinement. The refined feature maps are computed as:

$$F_{\text{refined}} = \text{ConvCRF}(F_d^i), \tag{10}$$

where F_d^i represents the feature maps from the final convolutional layer. This ensures smoothness and edge preservation in predicted masks.

The final output is a patch-wise real/fake probability map:

$$D(x) = \sigma(F_{\text{refined}}), \tag{11}$$

where $D(x)$ represents the discriminator's output probability. The discriminator is trained to distinguish between real (ground truth) and generated masks, providing adversarial feedback to the generator.

C. Adversarial Training and Loss Functions

The generator and discriminator are trained in an adversarial manner, where the generator aims to minimize the difference between real and generated masks, while the discriminator attempts to correctly classify real and fake masks. The training process follows a minimax game, defined as:

$$\min_G \max_D \mathcal{L}_{\text{total}}(G, D), \tag{12}$$

where G and D represent the generator and discriminator, respectively. The generator's loss function is a weighted combination of Binary Crossentropy Loss (BCE), Weighted Intersection over Union (IoU) Loss, and Dice Loss:

$$\mathcal{L}_{\text{total}} = \lambda_1\mathcal{L}_{\text{BCE}} + \lambda_2\mathcal{L}_{\text{IoU}} + \lambda_3\mathcal{L}_{\text{Dice}}. \tag{13}$$

These losses guide the generator to produce accurate and realistic segmentation masks. Below, we describe each component of the hybrid loss function in detail.

1) Binary Crossentropy Loss (BCE): The BCE measures the pixel-wise difference between the predicted mask M_{pred} and the ground truth mask M_{true}. It is defined as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} \Big[M_{\text{true}}^i \log(M_{\text{pred}}^i) + (1 - M_{\text{true}}^i) \log(1 - M_{\text{pred}}^i) \Big], \tag{14}$$

where N is the total number of pixels, and M^i_{true} and M^i_{pred} are the ground truth and predicted values for the i-th pixel, respectively.

2) Weighted IoU(WIoU) Loss: The Weighted IoU Loss addresses class imbalance by assigning different weights to the foreground (polyps) and background regions. It is defined as:

$$\mathcal{L}_{\text{IoU}} = 1 - \text{WIoU}(M_{\text{true}}, M_{\text{pred}}), \tag{15}$$

where the Weighted IoU is computed as:

$$\text{WIoU}(M_{\text{true}}, M_{\text{pred}}) = \alpha \cdot \text{IoU}_{\text{fg}} + (1-\alpha) \cdot \text{IoU}_{\text{bg}}. \tag{16}$$

Here, α is the weight for the foreground (typically set to 0.7), and IoU_{fg} and IoU_{bg} are the IoU values for the foreground and background, respectively. These are computed as:

$$\text{IoU}_{\text{fg}} = \frac{\sum(M_{\text{true}} \cdot M_{\text{pred}}) + \epsilon}{\sum M_{\text{true}} + \sum M_{\text{pred}} - \sum(M_{\text{true}} \cdot M_{\text{pred}}) + \epsilon}, \tag{17}$$

$$\text{IoU}_{\text{bg}} = \frac{A + \epsilon}{B + C - A + \epsilon}, \tag{18}$$

where $A = \sum\big((1 - M_{\text{true}}) \cdot (1 - M_{\text{pred}})\big)$, $B = \sum(1 - M_{\text{true}})$, $C = \sum(1 - M_{\text{pred}})$, and ϵ is a small constant (e.g., 10^{-6}) to avoid division by zero.

The weighted IoU loss ensures that the model focuses on both foreground and background regions, addressing the challenge of class imbalance.

3) Dice Loss: The Dice Loss measures the overlap between the predicted mask and the ground truth mask. It is defined as:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum(M_{\text{true}} \cdot M_{\text{pred}}) + \epsilon}{\sum M_{\text{true}} + \sum M_{\text{pred}} + \epsilon}, \tag{19}$$

where ϵ is a small constant to ensure numerical stability. The Dice Loss is particularly effective for segmentation tasks with imbalanced classes, as it emphasizes the overlap between the predicted and ground truth masks.

4) Total Loss: The total loss for the generator is a weighted combination of the BCE, Weighted IoU, and Dice Losses:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{BCE}} + \lambda_2 \mathcal{L}_{\text{IoU}} + \lambda_3 \mathcal{L}_{\text{Dice}}. \tag{20}$$

The weights λ_1, λ_2, and λ_3 are hyperparameters that balance the contributions of each loss term. In our experiments, we set $\lambda_1 = 0.4$, $\lambda_2 = 0.3$, and $\lambda_3 = 0.3$ to achieve a balance between pixel-wise accuracy, segmentation overlap quality, and boundary precision.

5) Discriminator Loss: The discriminator is trained using BCE to classify real and fake masks:

$$\mathcal{L}_D(y_{\text{true}}, y_{\text{pred}}) = -\frac{1}{N}\sum_{i=1}^{N}\Big(y^i_{\text{true}} \log(D(y^i_{\text{pred}})) + (1 - y^i_{\text{true}}) \log(1 - D(y^i_{\text{pred}}))\Big), \tag{21}$$

where y_{true} and y_{pred} are the ground truth and predicted labels, respectively, and D is the discriminator's output probability. The adversarial training framework encourages the generator to produce precise and realistic segmentation masks.

TABLE I
PERFORMANCE EVALUATION METRICS

Metrics	Description
Dice	Dice = (2×TP)/(2×TP+FP+FN)
IoU	IoU = (TP)/(TP+FP+FN)
Recall	Recall = (TP)/(TP+FN)
Precision	Precision = (TP)/(TP+FP)
Accuracy	Accuracy = (TP+TN)/(TP+TN+FP+FN)
F2	F2 = (5 × P ×R)/(4 × P + R)

IV. EXPERIMENTAL RESULT AND ANALYSIS

A. Dataset and Evaluation Metrics

The experiments utilize six public polyp segmentation datasets: Kvasir-SEG [1], CVC-ClinicDB [34], ETIS [35], CVC-300 [45], and PolypGen [22]. These datasets vary in terms of the number of images and their resolutions. Kvasir-SEG contains 1,000 images with variable sizes, while CVC-ClinicDB provides 612 images at a fixed resolution of 384×288. CVC-ColonDB includes 380 images with a resolution of 574×500, and ETIS consists of 196 images at a higher resolution of 1225×966. CVC-300 offers 60 images with the same resolution as CVC-ColonDB (574×500), and PolypGen, the largest dataset, contains 1,537 images with variable sizes. These datasets collectively provide a diverse and comprehensive foundation for the experiments.

The performance of the LGPS model was evaluated using a Dice coefficient (Dice), Intersection over Union (IoU), Recall, Precision, F2 score, and Accuracy. The formulas for calculating each metric are shown in Table I.

B. Implementation Details

The proposed LGPS model is implemented using the TensorFlow and Keras frameworks. The model is trained on a NVIDIA RTX A6000 GPU. The Adam optimizer is used with a learning rate of 1×10^{-4} and a batch size of 16. The Adam optimizer is also used for the discriminator with a learning rate of 1×10^{-4} and a batch size of 16. Input images are preprocessed by resizing them to a fixed resolution of 256×256 pixels and normalizing pixel values to the range $[0, 1]$. To improve the robustness of the model, several data augmentation techniques are applied during training. These include random horizontal and vertical flips with a probability of 0.5, random rotation by an angle between $-10°$ and $10°$, random brightness adjustment by a factor between 0.9 and 1.1, and random contrast adjustment by a factor between 0.9 and 1.1. The testing set is not augmented and is directly resized into 256 × 256. Following the PraNet [6] 900 and 550 images from the Kvasir-SEG and CVC-ClinicDB datasets, respectively, are used as the training set, while the remaining 100 and 62 images are used as the testing set.

C. Ablation Experiments

1) Ablation Experiment on Loss Function: To evaluate the impact of different loss functions, we conducted an ablation study using the Kvasir-SEG dataset. We tested various combinations of Binary Cross-Entropy (BCE), Intersection over Union (IoU), Weighted IoU (WIoU), and Dice Loss, as

TABLE II
PERFORMANCE OF DIFFERENT LOSS FUNCTION COMBINATIONS IN THE ABLATION EXPERIMENTS.

Loss Fun.	Dice	IoU	Recall	Pre.	F2	Acc.
WIoU	0.8530	0.8436	0.7900	0.9123	0.8118	0.9498
BCE_only	0.8552	0.8464	0.7866	0.9202	0.8101	0.9506
Dice_only	0.8494	0.8407	0.7722	0.9494	0.7990	0.9494
BWIoU	0.8515	0.8396	0.8145	0.8816	0.8271	0.9477
BIoU	0.8529	0.8431	0.7954	0.9052	0.8152	0.9494
BDice	**0.8582**	**0.8478**	**0.8049**	0.9063	**0.8233**	0.9508
3Loss A	0.8575	0.8477	0.7905	0.9217	0.8136	**0.9512**
3Loss B	0.8431	0.8331	0.7880	0.8816	0.8141	0.9457
3Loss C	0.8377	0.8289	0.7775	0.8952	0.7986	0.9450

summarized in Table II. The results reveal that the combination of BCE and Dice Loss (BDice) achieved the highest Dice score (0.8582) and IoU (0.8478), outperforming other combinations. Below, we discuss the performance of standalone and combined loss functions.

Standalone loss functions address one specific aspect of the segmentation task. Weighted IoU (WIoU) achieved a Dice score of 0.8530 and IoU of 0.8436. WIoU balances foreground and background regions, addressing class imbalance but lacks pixel-wise accuracy and boundary precision. BCE Only achieved a Dice score of 0.8552 and IoU of 0.8464. While BCE Loss ensures pixel-wise classification accuracy, it struggles with class imbalance and boundary precision. Dice Only achieved a Dice score of 0.8494 and IoU of 0.8407. Dice Loss handles class imbalance and optimizes overlap but lacks pixel-wise precision.

Combined loss functions address multiple aspects of the segmentation task by integrating two or more losses. The BCE + Dice Loss (BDice) combination achieved the highest Dice score (0.8582) and IoU (0.8478). This combination balances pixel-wise accuracy (BCE) with overlap quality and boundary precision (Dice), addressing class imbalance and producing well-defined boundaries. BCE + WIoU (BWIoU) achieved a Dice score of 0.8515 and IoU of 0.8396. While BWIoU improves over standalone WIoU by balancing pixel-wise accuracy and foreground-background balancing, it does not explicitly optimize for boundary precision. BCE + IoU (BIoU) achieved a Dice score of 0.8529 and IoU of 0.8431. This combination balances pixel-wise accuracy with overlap metrics but lacks the boundary refinement provided by Dice Loss.

Hybrid loss combinations, such as 3Loss A ($0.4 \cdot \text{BCE} + 0.3 \cdot \text{WIoU} + 0.3 \cdot \text{Dice}$), achieved competitive results, with a Dice score of 0.8575 and IoU of 0.8477. While this hybrid loss balances pixel-wise accuracy, foreground-background balancing, and overlap quality, it is slightly outperformed by BDice, suggesting that the additional complexity of combining three losses does not always translate to better performance. Similarly, 3Loss B ($0.3 \cdot \text{BCE} + 0.4 \cdot \text{WIoU} + 0.3 \cdot \text{Dice}$) achieved a Dice score of 0.8431 and IoU of 0.8331. This combination places more emphasis on WIoU, reducing its effectiveness in handling boundary precision and pixel-wise accuracy. 3Loss C ($\text{BCE} + \text{IoU} + \text{Dice}$) achieved a Dice score of 0.8377 and IoU of 0.8289. This combination lacks the weighted balancing of foreground and background regions, which reduces its effectiveness in handling class imbalance.

In conclusion, the BCE + Dice Loss (BDice) combination is the most effective for polyp segmentation, as it addresses the key challenges of class imbalance, boundary precision, and pixel-wise accuracy without introducing unnecessary complexity. Standalone losses and hybrid combinations, while useful in specific scenarios, do not outperform the simpler BCE + Dice combination.

TABLE III
ABLATION EXPERIMENT ON THE CONTRIBUTION OF EACH MODULES.

Model	Dice	IoU
Baseline	0.8575	0.8477
W/o ReSE	0.8445	0.8366
W/o ConvCRF	0.8519	0.8436
MRB w/o SE	0.8475	0.8378
W/o ConvCRF and ReSE	0.8415	0.8376

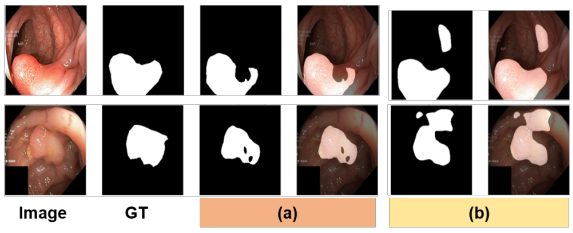

Fig. 4. Visualized heat maps (a) with ConvCRF and ReSE and (b) without ConvCRF and ReSE

2) Ablation Experiment on the Contribution of Each Module: To evaluate the contribution of each component in LGPS, we conducted an ablation study using the 3Loss A as the benchmark, we named it LGPS Weighted Loss (LGPS WLoss). The study systematically removed key components and analyzed their impact on segmentation performance. The results, presented in Table III, are evaluated using the Dice and IoU.

The baseline model, which includes all components (ReSE, ConvCRF layers, and the WLoss function), achieved the highest performance with a Dice of 0.8575 and IoU of 0.8477. This result demonstrates the effectiveness of the complete model configuration, where each component contributes to improving segmentation accuracy and spatial coherence.

When the ReSE was removed, the Dice dropped to 0.8445 (a reduction of 1.30%), and the IoU decreased to 0.8366 (a reduction of 1.11%). This performance degradation highlights the importance of the MRB in capturing hierarchical features and enhancing the model's ability to handle complex polyp structures. The residual connections within these blocks facilitate gradient flow during training, enabling the model to learn more robust representations.

Removing the SE mechanism from the ReSE resulted in a Dice of 0.8475 (a reduction of 1.00%) and an IoU of 0.8378 (a reduction of 0.99%). The SE mechanism dynamically recalibrates channel-wise feature responses, emphasizing diagnostically relevant features while suppressing less useful ones. Its removal leads to a noticeable drop in performance, particularly in scenarios where fine-grained feature discrimination is critical.

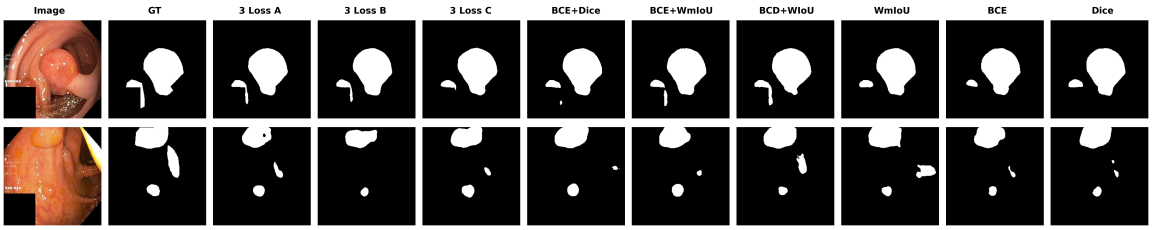

Fig. 5. Ablation experiment on different loss functions.

When the ConvCRF layer was removed, the Dice decreased to 0.8519 (a reduction of 0.56%), and the IoU dropped to 0.8436 (a reduction of 0.41%). The ConvCRF layer plays a crucial role in refining segmentation masks by enforcing spatial coherence and preserving boundary details. Its removal results in slightly less precise segmentation, particularly around polyp edges and small structures.

Removing both the ConvCRF layer and the ReSE led to the most significant performance degradation, with the Dice dropping to 0.8415 (a reduction of 1.60%) and the IoU decreasing to 0.8376 (a reduction of 1.01%). This result underscores the complementary roles of these components: the ReSE block enhances feature extraction, while the ConvCRF layer refines the final segmentation output. Their combined removal significantly impacts the model's ability to accurately segment polyps, even with the WLoss function in place.

The ablation study demonstrates that each component of LGPS contributes meaningfully to its overall performance. The ReSE mechanism are critical for robust feature extraction, while the ConvCRF layer ensures precise boundary preservation and spatial coherence. The baseline model, which includes all components, achieves the best performance, highlighting the importance of their synergistic integration.

3) Qualitative Ablation Study: Impact of Key Components on Segmentation Performance: The heat map of the features, both with and without the ConvCRF and ReSE, is shown in 4. It is evident that the network focuses more on the object areas when both modules are introduced. Without these modules, the network activates non-polyp regions and fails to precisely localize the polyp region and its shape. However, with the inclusion of both modules, the polyp regions are accurately activated. This indicates that the modules enhance the object regions while suppressing the background, thereby improving segmentation accuracy.

4) Qualitative Analysis of Segmentation Masks: Evaluating the Impact of Different Loss Functions: To complement the quantitative findings of the ablation study, as shown in Figure 5, we performed a qualitative analysis of the segmentation masks generated by the different loss functions. This analysis focused on visual inspection of the segmentation results, particularly for challenging cases such as small polyps, boundary regions, and areas with class imbalance.

The qualitative analysis revealed several key observations. The segmentation masks produced by the combination 3Loss A exhibited the highest precision in boundary localization. The edges of the polyps were well-defined, and the masks closely aligned with the ground truth annotations, even in regions with complex shapes or irregular boundaries. For small polyp regions, the baseline 3Loss A and 3Loss B demonstrated superior performance. The segmentation masks generated by these loss functions accurately captured small polyps, with minimal false positives or missed regions. This aligns with the quantitative results from the ablation study, confirming their effectiveness in handling small and underrepresented structures.

The WLoss functions, particularly the baseline and BCD + WIoU, showed a remarkable ability to handle class imbalance. In images with a high background-to-polyp ratio, these loss functions produced segmentation masks that effectively prioritized polyp regions without over-segmenting the background. However, standalone loss functions such as WIoU, BCE, and Dice loss, while effective in segmenting large polyp regions, exhibited limitations in generalizing across diverse cases. For instance, they occasionally produced fragmented masks for small polyps or failed to predict small polyp regions altogether. Additionally, these standalone loss functions struggled with boundary precision in regions of low contrast.

Some loss functions, particularly standalone and binary loss functions, exhibited tendencies toward over-segmentation or under-segmentation. Over-segmentation was observed in regions with ambiguous boundaries, while under-segmentation occurred in cases where the polyp regions were small or poorly contrasted against the background. These challenges highlight the limitations of using single loss functions in complex segmentation tasks.

The qualitative analysis underscores the strengths and limitations of the evaluated loss functions in polyp segmentation. The combination 3Loss A consistently demonstrated superior performance in boundary precision, small polyp localization, and handling class imbalance. Standalone loss functions, while effective for large polyps, struggled with small polyp regions and boundary precision. These findings reinforce the importance of combining multiple loss functions to address the diverse challenges in polyp segmentation and provide valuable insights for future improvements in segmentation models.

D. Qualitative Assessment of LGPS and Existing Methods

To qualitatively evaluate the performance of different state-of-the-art (SOTA) segmentation methods, we visualize the segmentation results on the Kvasir-SEG and CVC-ColonDB datasets, as shown in Fig. 6. The visualization highlights the strengths and limitations of existing methods compared to the proposed LGPS.

Method	Backbone Network	Parameters (M)	Dice	IoU	Recall	Precision	F2
U-Net [14]	ResNet34	34.52	0.8762	0.7550	0.8732	0.8999	0.8784
CE-Net [27]	ResNet34	29.00	0.9280	0.8790	0.9080	0.9150	0.8990
PraNet [6]	Res2Net	30.50	0.8995	0.8495	0.9500	0.9450	0.9490
HarDNet-MSEG [33]	HardNet68	17.42	0.9320	0.8820	0.9200	0.9460	0.9290
TGANet [18]	ResNet50	19.84	0.9457	0.8866	0.9437	0.9519	0.9439
Polyp-PVT [4]	PVT	25.10	0.9370	0.8890	0.9490	0.9280	0.9360
SSFormer-L [19]	PVT	65.95	0.9470	0.9030	0.9560	0.9420	0.9530
Huang et al. [17]	ResNet50	63.29	0.9492	0.9071	0.9534	0.9483	0.9511
WDFF-Net [24]	HardNet68	17.46	**0.9521**	0.9084	**0.9702**	**0.9711**	**0.969**
Ours (Weighted Loss)	MobileNet-V2	1.07	0.9261	**0.9238**	0.8607	0.9683	0.8802
Ours (Dice + BCE)	MobileNet-V2	1.07	0.9117	0.9157	0.8473	0.9655	0.8686

TABLE IV
COMPARISON WITH SOTA METHODS ON THE CVC-CLINICDB DATASET.

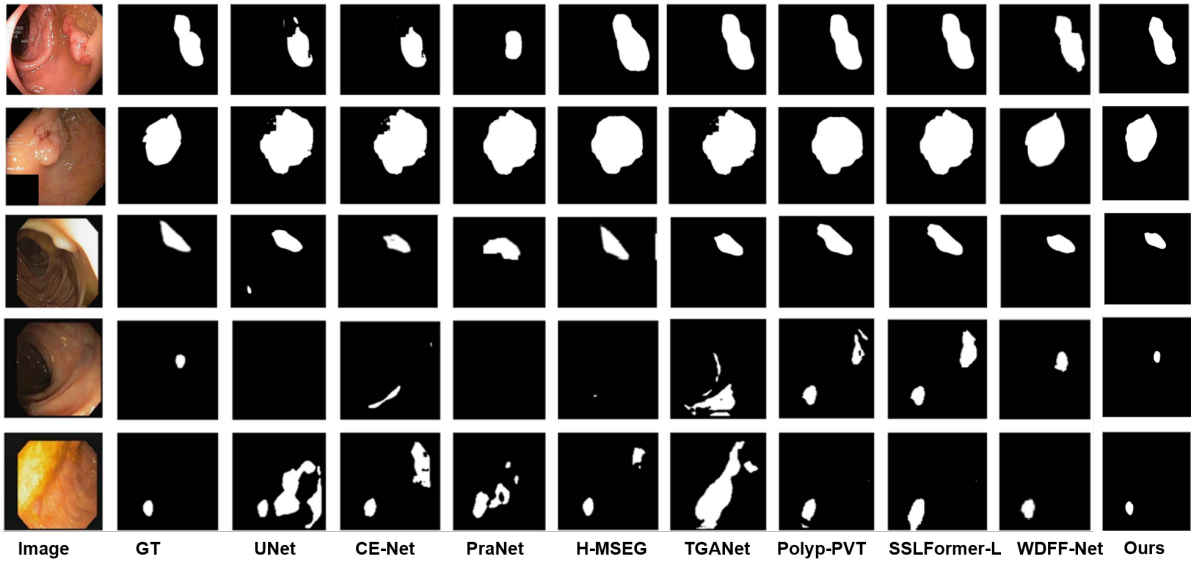

Fig. 6. Visualization of segmentation results on the Kvasir-SEG and CVC-ColonDB datasets. Rows 1–2 shows large polyps. Rows 3–4 show cases with blurry polyp boundaries and low contrast, while rows 4–5 depict cases with significant variations in polyp size, particularly small polyps.

TABLE V
COMPARISON WITH SOTA ON ETIS, AND CVC-300 DATASETS.

Method	ETIS		CVC-300	
	Dice	**IoU**	**Dice**	**IoU**
U-Net [14]	0.3980	0.3350	0.7100	0.6270
CE-Net [27]	0.5859	0.5700	0.8706	0.7970
PraNet [6]	0.6280	0.5670	0.8710	0.7970
HarDNet-MSEG [33]	0.6770	0.6630	0.8870	0.8210
TGANet [18]	0.6630	0.5860	0.8850	0.8190
Polyp-PVT [4]	0.7870	0.7600	0.9000	0.8330
Huang et al. [17]	0.7510	0.6800	0.9110	0.8490
Su et al. [26]	**0.8160**	0.7330	0.9120	0.8490
WDFF-Net(2024) [24]	0.7581	0.7241	**0.9161**	0.8533
Ours (Weighted Loss)	0.7447	0.7742	0.8502	0.8648
Ours (Dice + BCE)	0.7451	**0.7746**	0.8556	**0.8690**

TABLE VI
COMPARISON WITH SOTA METHODS ON THE POLYPGEN DATASET.

Method	Dice	IoU	Recall	Pre.	F2
U-Net (2015) [14]	0.5995	0.5347	0.6829	0.7523	0.6105
U-Net++(2018) [41]	0.5964	0.5310	0.6765	0.7546	0.6089
ResU-Net++(2019) [42]	0.3982	0.3149	0.5887	0.4444	0.4314
HarDNet-MSEG(2021) [33]	0.6089	0.5376	0.7116	0.7124	0.6246
ColonSegNet (2021)[6]	0.5486	0.4718	0.6554	0.6687	0.5617
UACANet(2021) [43]	0.6531	0.5777	0.7493	0.7531	0.6678
UNeXt (2022) [44]	0.4552	0.3761	0.6135	0.5600	0.4805
TransNetR (2023)[40]	0.6668	0.6058	0.6135	0.5600	0.6706
WDFF-Net (2024) [24]	0.6687	0.6102	0.6893	0.7602	0.6723
Ours (WLoss)	**0.7299**	**0.7867**	0.6807	**0.8233**	0.6958
Ours (Dice+BCE)	0.7276	0.7835	**0.6997**	0.7948	**0.7061**

Existing methods often struggle with challenges such as blurry boundaries and low contrast. As shown in rows 1-2 existing methods over segmented polyp region, however, LGSP precisely segment the polyp region. As shown in rows 3–4 of Fig. 6, existing methods frequently fail to achieve complete segmentation when polyp boundaries are blurry or exhibit low contrast with surrounding tissues. This results in incomplete or inaccurate segmentation masks. Additionally, existing methods face difficulties with variable polyp sizes. Rows 4–5 illustrate that existing methods tend to miss small polyps or produce fragmented segmentation results when there are significant variations in polyp size. This is particularly problematic for small polyps, which are often overlooked or inaccurately segmented. In contrast, the proposed LGPS demonstrates superior performance in addressing these challenges. LGPS effectively segments flat polyps with low contrast, as shown in rows 3–4. The model's ability to capture subtle boundary details ensures complete and accurate segmentation, even in challenging cases. Furthermore, LGPS accurately segments polyps with

large variations in size, including small polyps, as depicted in rows 4–5. This robustness is attributed to the model's multi-scale feature extraction and boundary refinement mechanisms. However, in the fourth row, LGPS struggles slightly to precisely segment the polyp region under conditions of poor visibility. The visualization experiment demonstrates that LGPS outperforms existing methods in accurately segmenting polyps with blurry boundaries, low contrast, and significant size variations. These results underscore the model's ability to handle real-world challenges in polyp segmentation, making it a reliable tool for clinical applications.

E. State-of-the-Art (SOTA) Analysis and Discussion

To validate the effectiveness of LGPS, we compared it against nine SOTA polyp segmentation methods, including both universal segmentation networks (e.g., U-Net [14], CE-Net [27]) and dedicated polyp segmentation networks (e.g., PraNet [6], HarDNet-MSEG [33], TGANet [18], SSFormer-L [19], Polyp-PVT [4], and WDFF-Net [24]). The comparison was conducted on four public datasets: CVC-ClinicDB, ETIS, CVC-300, and PolypGen.

1) Segmentation Accuracy: As shown in Table IV, LGPS demonstrates competitive segmentation accuracy on the CVC-ClinicDB dataset. The WLoss variant achieves a Dice score of 0.9261 and an IoU of 0.9238, outperforming PraNet (Dice = 0.8995) and achieving results comparable to several SOTA methods. Notably, LGPS achieves the highest IoU (0.9238) among all methods, surpassing even the recent WDFF-Net (IoU = 0.9084). This highlights the effectiveness of LGPS in achieving high segmentation accuracy, particularly when leveraging the WLoss function, which balances multiple loss terms to improve performance.

2) Generalization Ability: One of the key strengths of LGPS is its exceptional generalization capability, particularly on unseen and challenging datasets. To evaluate this, we conducted experiments on three unseen datasets: PolypGen, ETIS, and CVC-300. As shown in Table V and VI, LGPS achieves strong performance on ETIS (IoU = 0.7746) and CVC-300 (IoU = 0.8690), demonstrating its robustness to diverse imaging conditions and unseen data. In terms of Dice, LGPS shows competitive results on both datasets, with 0.7447 on ETIS and 0.8502 on CVC-300.

Notably, LGPS achieves SOTA performance on the PolypGen dataset, the largest and most challenging test set, with a Dice score of 0.7299 and an IoU of 0.7867. This is a significant achievement, as PolypGen contains diverse polyp types and imaging conditions, making it a rigorous benchmark for evaluating generalization. To the best of our knowledge, LGPS is the first polyp segmentation model to demonstrate such strong generalization performance on unseen datasets. The adversarial training framework, combined with the WLoss function, enables the model to learn robust features that generalize well across different datasets and imaging conditions.

3) Model Efficiency: A key advantage of LGPS is its lightweight design, enabled by the MobileNet-V2 backbone. With only 1.07 million parameters, LGPS is significantly more efficient than SOTA methods such as WDFF-Net (17.46M parameters), SSFormer-L (65.95M parameters), and Huang et al. (63.29M parameters). Despite its compact architecture, LGPS achieves competitive or superior performance on multiple datasets, making it suitable for real-time applications in clinical settings. This efficiency is particularly important for deploying the model in resource-constrained environments, such as endoscopy suites, where computational resources are limited.

4) Discussion: The results demonstrate that LGPS achieves a compelling balance of accuracy, efficiency, and generalization. The WLoss variant, which combines BCE, WIoU, and Dice Loss, consistently outperforms the Dice+BCE variant, particularly on unseen datasets. This highlights the importance of balancing multiple loss terms to improve segmentation performance and generalization.

The strong performance of LGPS on unseen datasets, particularly PolypGen, can be attributed to its GAN-based architecture. The adversarial training framework encourages the generator to produce realistic and accurate segmentation masks, while the discriminator provides fine-grained feedback to improve boundary preservation and spatial coherence. This makes LGPS particularly effective in handling diverse and challenging imaging conditions, which are common in real-world clinical settings.

While LGPS achieves SOTA performance on PolypGen and SOTA IoU on other datasets, there is room for improvement in terms of Dice, where it slightly underperforms compared to some methods. Future work could explore integrating additional attention mechanisms or leveraging transformer-based backbones to further enhance performance. These improvements could address the current limitations and extend the applicability of LGPS to a wider range of medical imaging tasks.

V. Conclusion

In this paper, we introduced LGPS, a lightweight GAN-based framework for polyp segmentation in colonoscopy images. LGPS addresses critical challenges such as blurry boundaries, small polyp detection, and computational inefficiency, making it suitable for real-time clinical applications. The framework integrates a MobileNetV2 backbone with ReSE, and ConvCRF to achieve SOTA performance with only 1.07 million parameters. A hybrid loss function combining Binary Cross-Entropy (BCE), Weighted IoU Loss, and Dice Loss further enhances segmentation accuracy by addressing class imbalance.

Extensive experiments on five public datasets demonstrate the effectiveness of LGPS. On the challenging PolypGen dataset, LGPS achieves a Dice of 0.7299 and a mean IoU of 0.7867, outperforming existing methods in both accuracy and efficiency. The model also exhibits strong generalization capabilities on unseen datasets, such as ETIS and CVC-300 , highlighting its robustness to diverse imaging conditions. Its lightweight design makes it highly suitable for deployment on resource-constrained devices, offering significant potential for real-time clinical use.

Future work will explore extending LGPS to other medical imaging tasks, such as lesion detection and organ segmenta-

tion, and integrating transformer-based architectures to further enhance performance. By addressing key challenges in polyp segmentation, LGPS sets a new benchmark for efficient and accurate medical image analysis, paving the way for improved clinical outcomes.

REFERENCES

[1] D. Jha et al., "Kvasir-seg: A segmented polyp dataset," in *Proc. Int. Conf. MultiMedia Modeling*, 2020, pp. 451–462.

[2] H. Yao et al., "Motion-based camera localization system in colonoscopy videos," *Med. Image Anal.*, vol. 73, 2021, Art. no. 102180.

[3] O. S. Kayhan and J. C. van Gemert, "On translation invariance in CNNs: Convolutional layers can exploit absolute spatial location," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14274–14285.

[4] B. Dong et al., "Polyp-PVT: Polyp segmentation with pyramid vision transformers," 2021, arXiv:2108.06932.

[5] Y. Fang et al., "Selective feature aggregation network with area-boundary constraints for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2019, pp. 302–310.

[6] D.-P. Fan et al., "PraNet: Parallel reverse attention network for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2020, pp. 263–273.

[7] Z. Li et al., "Scribformer: Transformer makes CNN work better for scribble-based medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 43, no. 6, pp. 2254–2265, 2024.

[8] Y. Xie et al., "Attention mechanisms in medical image segmentation: A survey," 2023, arXiv:2305.17937.

[9] C.-H. Huang et al., "HarDNet-MSEG: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 FPS," 2021, arXiv:2101.07172.

[10] N. Siddique et al., "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[11] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 1, pp. 121–130, Jan. 2021.

[12] J. M. J. Valanarasu et al., "Medical transformer: Gated axial-attention for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2021, pp. 36–46.

[13] H. Cao et al., "Dual-branch residual network for lung nodule segmentation," *Appl. Soft Comput.*, vol. 86, 2020, Art. no. 105934.

[14] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2019, arXiv:1810.07810.

[15] P. Chao et al., "HarDNet: A low memory traffic network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3552–3561.

[16] H. Fan and H. Ling, "SANet: Structure-aware network for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 42–49.

[17] X. Huang et al., "Polyp segmentation network with hybrid channel-spatial attention and pyramid global context guided feature fusion," *Comput. Med. Imag. Graph.*, vol. 98, 2022, Art. no. 102072.

[18] N. K. Tomar et al., "TGANet: Text-guided attention for improved polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2022, pp. 151–160.

[19] J. Wang et al., "Stepwise feature fusion: Local guides global," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2022, pp. 110–120.

[20] Q. Wang et al., "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11534–11542.

[21] X. Chen et al., "Learning Euler's elastica model for medical image segmentation," 2020, arXiv:2011.00526.

[22] S. Ali et al., "A multi-centre polyp detection and segmentation dataset for generalisability assessment," *Sci. Data*, vol. 10, no. 1, 2023, Art. no. 75.

[23] H. I. Liu et al., "Lightweight deep learning for resource-constrained environments: A survey," *ACM Comput. Surv.*, vol. 56, no. 10, pp. 1–42, 2024.

[24] J. Cao et al., "WDFF-Net: Weighted dual-branch feature fusion network for polyp segmentation with object-aware attention mechanism," *IEEE J. Biomed. Health Informat.*, 2024.

[25] N. K. Tomar et al., "TransResU-Net: Transformer based ResU-Net for real-time colonoscopy polyp segmentation," 2022, arXiv:2206.08985.

[26] Y. Su et al., "Accurate polyp segmentation through enhancing feature fusion and boosting boundary performance," *Neurocomputing*, vol. 545, p. 126233, 2023.

[27] Z. Gu et al., "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[28] O. Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[29] M. Sandler et al., "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. CVPR*, 2018, pp. 4510–4520.

[30] J. Hu et al., "Squeeze-and-excitation networks," in *Proc. CVPR*, 2018, pp. 7132–7141.

[31] P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 1125–1134.

[32] L.-C. Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, 2018, pp. 801–818.

[33] C.-H. Huang et al., "HarDNet-MSEG: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 FPS," 2021, arXiv:2101.07172.

[34] N. Tajbakhsh et al., "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Trans. Med. Imaging*, vol. 35, no. 2, pp. 630–644, 2015.

[35] J. Silva et al., "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, pp. 283–293, 2014.

[36] Z.-L. Ni et al., "BARNet: Bilinear attention network with adaptive receptive field for surgical instrument segmentation," 2020, arXiv:2001.07093.

[37] Y. Wang et al., "LEDNet: A lightweight encoder-decoder network for real-time semantic segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 1860–1864.

[38] N. Beheshti and L. Johnsson, "Squeeze U-Net: A memory and energy efficient image segmentation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 364–365.

[39] E. Romera et al., "ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 263–272, 2017.

[40] D. Jha et al., "TransNetR: Transformer-based residual network for polyp segmentation with multi-center out-of-distribution testing," in *Proc. Med. Imaging Deep Learn.*, 2024, pp. 1372–1384.

[41] Z. Zhou et al., "Unet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, 2018, pp. 3–11.

[42] D. Jha et al., "Resunet++: An advanced architecture for medical image segmentation," in *Proc. IEEE Int. Symp. Multimedia*, 2019, pp. 225–2255.

[43] T. Kim et al., "UACANet: Uncertainty augmented context attention for polyp segmentation," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 2167–2175.

[44] J. M. J. Valanarasu and V. M. Patel, "UNeXt: MLP-based rapid medical image segmentation network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2022, pp. 23–33.

[45] D. Vázquez et al., "A benchmark for endoluminal scene segmentation of colonoscopy images," *J. Healthcare Eng.*, vol. 2017, p. 4037190, 2017.

[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.