Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Polyp Segmentation with Deep Learning: Utilizing DeeplabV3+ Architecture and Various CNN Backbones

Yaren AKGÖL[1*], Buket TOPTAŞ[1]

[1] Bandırma Onyedi Eylül University, Software Eng. Dept, yarenakgol@ogr.bandirma.edu.tr, Orcid No: 0009-0004-5987-0171
[1] Bandırma Onyedi Eylül University, Software Eng. Dept, btoptas@bandirma.edu.tr, Orcid No: 0000-0003-2556-8199

ARTICLE INFO

Article history:

Received 16 July 2024
Received in revised form 8 Oct 2024
Accepted 17 Oct. 2024
Available online 23 December 2024

Keywords:

polyp segmentation, deeplabv3+, colorectal cancer, colon cancer

Doi: 10.24012/dumf.1517112

ABSTRACT

Polyps are abnormal tissue growths that often serve as early indicators for various types of cancer. Early detection is crucial in the treatment of diseases like colorectal cancer, which has a high mortality rate. There is a significant need for automated diagnostic systems to detect these cancers efficiently. This article introduces a deep learning-based diagnostic system using the Deeplabv3+ architecture, which is enhanced by integrating four different backbone networks: Model 1 (DeeplabV3+), Model 2 (ResNet50), Model 3 (SqueezeNet), and Model 4 (VGG16). The enhanced architectures have been tested on the publicly available Kvasir-SEG and CVC-ClinicDB datasets for the task of polyp segmentation. Experimental results indicate that the best segmentation performance on the Kvasir-SEG dataset was achieved with Model 1, showing a mean Dice coefficient (mDice) of 0.858, a mean Intersection over Union (mIoU) of 0.850, an accuracy (Acc) of 0.948, a recall of 0.824, and a precision (Pr) of 0.896. For the CVC-ClinicDB dataset, the highest metrics were observed with Model 2 for mDice (0.914) and mIoU (0.912), and Model 1 for specificity (Sp) at 0.996 and precision at 0.959, whereas Model 4 exhibited the highest accuracy of 0.974. These results demonstrate the effectiveness of our models in automating the detection of polyps, potentially aiding in the early diagnosis of colorectal cancer.

Introduction

Medical image segmentation represents a highly sensitive and challenging field with significant impacts on human health. Challenges such as overlapping cells, differentiation of retinal blood vessels into arteries and veins, skin lesions obscured by hair, and polyps of various shapes and sizes exemplify the difficulties faced in this domain. However, recent years have witnessed rapid progress and significantly improved success rates in various areas of medical image segmentation, including retina [1], cell [2], mammogram [3], skin [4], and polyp segmentation [5], driven by advancements in deep learning methods. These developments constitute a significant leap forward in the field of medical image analysis, enhancing not only diagnostic accuracy but also streamlining workflows within medical practices. The integration of deep learning into medical imaging paves the way for more precise, efficient, and potentially life-saving diagnostic procedures.

Polyps, which are abnormal tissue growths developing on surfaces such as the colon, rectum, and stomach, represent one of the primary areas where deep learning architectures have been extensively applied in medical fields. Although polyps may appear benign, they increase the risk of cancer. According to cancer statistics published in 2022, colon and rectum cancer ranks third worldwide in cancer-related deaths [6].

The early detection of polyps can facilitate timely treatment options and potentially reduce mortality rates associated with polyp-related complications. To achieve early detection, automated segmentation systems are essential. However, the segmentation of polyps is challenged by their irregular shapes, size variations, and differences in location and color. To address these challenges, a significant number of studies have been conducted in the literature. These studies and advanced methodologies are discussed in greater detail under the 'Related Work' section.

Related work

Traditional methods of polyp segmentation rely on manual feature extraction processes [7],[8]. With the advent of deep neural network architectures that minimize manual intervention by extracting semantic information from high-dimensional images, researchers have increasingly focused on leveraging these architectures.

Gangrade et al.[9] proposed a modified DeepLabV3+ architecture for polyp segmentation from colonoscopy images. This architecture is comprised of encoder and decoder layers. The encoder utilizes a pre-trained dilated

convolutional residual network to optimally achieve feature map resolution. The encoder-decoder structures include the Atrous Spatial Pyramid Pooling (ASPP) module and the ResNet101 backbone used in the proposed model. ASPP generates a multi-scale feature map by employing atrous convolution and global average pooling. Zhang et al. [5] have proposed the Dual-Branch Multi-İnformation Aggregation Network (DBMIANet) method to segment the same type of polyps reliably and effectively. DBMIA-Net employs three auxiliary modules to enhance its feature extraction and segmentation capabilities. These are the Adaptive Channel-Wise Graph Convolution (ACGC), Global Information Aggregation (GIA), and Edge Information Aggregation (EIA) modules. These modules are utilized by a transformer encoder and a Convolutional Neural Networks (CNN) encoder. The GIA module is used in aggregating global information, whereas the EIA module is employed for edge information aggregation. The ACGC module has been developed to improve the capability of learning channel feature associations representation. Li et al. [10] have proposed a model to address the challenge of detecting small polyps. The proposed model utilizes a two-stage transfer learning approach. In the first stage, the network is trained to identify specific areas of polyp lesions and to save initial weights. The second stage employs transfer learning to segment the relevant area in more detail. A Pyramid Vision Transformer (PVT) has been used as the feature backbone.

Li et al. [11] proposed a cross-level information fusion and guidance-oriented approach to polyp segmentation networks. The method employs a transformer encoder to build a robust feature representation. Modules such as the Edge Feature Processing Module (EFPM) and Crosslevel İnformation Processing Module (CIPM) are utilized to enhance the feature information coming from the encoders. EFPM focuses on the boundary information of polyps and is used to gather and process multi-scale features transmitted by various encoder layers. An İnformation Guidance Module (IGM) has been suggested to combine the processed features of EFPM and CIPM, maximizing the segmentation effect. Jia et al. [12] have proposed a semi-supervised framework named PolypMixNet, aimed at achieving colorectal polyp segmentation. The framework utilizes a mean teacher architecture and novel augmentation techniques within its model architecture. PolypMixNet includes a Polyp-Aware Mix-Up Algorithm (PolypMix) and a dual-level consistency regularization. PolypMix enhances the diversity of training data and addresses class imbalance in colonoscopy datasets. He et al. [13] introduced the Boundary-Guided Filter Network (BGF-Net), known for achieving enhanced medical image segmentation. In the proposed model, DeepLabV3+ has been selected as the backbone for BGF-Net. BGF-Net is composed of four main components: ResNet-101, Channel Boundary Guided (CBG), Spatial Boundary Guided (SBG), and Boundary Guided Filter (BGF). During the encoding process, CBG is connected to ResNet-101 to extract channel weights about boundary features. SBG is designed to capture spatial weights and to guide and optimize low-level features. BGF directs and preserves appropriate segment boundaries through refined boundary resolution. Liu et al. [14] have proposed a novel Feature Combination Network (FCA-Net) for accurately detecting polyp sizes and locations. The proposed model comprises three modules: the Edge Perception Module (EPM), the Boundary-Guided Feature Aggregation Module (BFAM), and the Iterative Context Aggregation Module (ICAM). EPM is capable of simultaneously extracting initial boundary guidance maps from both low and high-level features. BFAM enhances hierarchical features, better preserves boundary details, and recalibrates positioned objects by integrating boundary information into the segmentation network. ICAM employs a contextaware approach to better leverage dependencies between features at different scales. Wang et al. [15] introduced CPSNet, a novel model for concealed polyp segmentation. CPSNet consists of three main modules: the Deep Multi-Scale Feature Fusion Module (DMF), the Camouflaged Object Detection Module (CDM), and the Multi-Scale Feature Enhancement Module (MFEM). These modules work collaboratively to enhance the segmentation process, increasing both resilience and accuracy. The model employs the DMF module for the progressive fusion of features, gathering structural and semantic information of polyps from deep features. CDM is used in shallow features to effectively identify camouflaged and concealed polyps. Furthermore, MFEM has been developed to seamlessly combine shallow and deep features, considering both local and global perspectives.

Liu et al. [16] detail the components of the proposed CAFE-Net architecture, which includes the PVT as its backbone, the Feature Completion and Exploration Module (FSEM), the Cross-Attention Decoder Module (CADM), and the Multi-Scale Feature Aggregation (MFA). During the decoder phase, CADM has been utilized to successfully amalgamate high and low-level features. Shao et al. [17] highlight that the variability in shapes and sizes of polyps poses challenges in early diagnosis. To address this issue, the Adaptive Feature Aggregation Network (AFANet) has been proposed. The proposed model is composed of the Multi-modal Balancing Attention Module (MMBA) and the Global Context Module (GCM). The MMBA module facilitates the extraction of enhanced local features by utilizing contextual information in the foreground, background, and edge regions of images, paying special attention to these areas. The GCM module, on the other hand, captures the global contextual features from the top of the encoder to examine the pathological image's global contextual characteristics in greater detail, and then transfers these features to the decoder layer. Muhammad et al. [18] have proposed a novel polyp segmentation method known as MMFIL-Net. The proposed model incorporates the Hierarchical Multi-Source Feature Interaction Module (HMFIM) and Multi-Source Feature Interaction Blocks (MFIB). MFIB aims to achieve generalized performance by manipulating multi-level and multi-source features to minimize the differences between low and high-level feature maps. Additionally, the Multiple Receptive Field Feature Interaction Block (MRFFIB) addresses segmentation issues of polyps of various sizes. To tackle the challenge of detecting and segmenting early-stage polyps

with ambiguous boundary information, the Dual Source Attention Fusion Block (DSAFB) was developed. The model utilizes EfficientNet-B0 as its encoder block. Yue et al. [19] have introduced the Boundary Uncertainty Aware Network (BUNet) to enhance polyp detection. Emphasizing the awareness of the shapes and sizes of polyps, a PVT encoder has been used to learn multi-scale feature representations. For low-level features, the Boundary Exploration Module (BEM) was preferred. Utilizing boundary information from BEM, the Boundary Uncertainty Aware Module (BUM) was proposed for detecting error-prone areas in high-level features. The pyramid image transformer PVT-V2 was employed to extract multi-scale and robust features. BUM consists of two parallel convolutional branches, which are supervised by the ground truths of the polyp and background. By taking the difference between feature maps generated from these branches, boundary uncertainty regions are identified, incorporating boundary cues from the BEM module.

Ahmed and Hasan [20] proposed a Twin Segmentation Network (Twin-SegNet) to enhance segmentation performance by merging polyp and background reconstructions. The model is structured into three main components: polyp and background segmentation models, the Partial Channel Recalibration (PCR) section, and the merging section. To ensure accurate segmentation of polyp regions and the background, mean squared error has been utilized. The Wavelet Convolutional Block (WCB) is recommended for edge information, while the Partial Channel Recalibration (PCR) block is proposed to allow for mutual feature exchange. In the initial part, the final convolution layer and sigmoid activation are extracted. The concluding part contains a final convolution layer followed by a sigmoid to generate the foreground and background segmentation maps. Fan et al. [21] have proposed Super-Resolution-Assisted Small Targets Polyp Segmentation Network (SRSegNet), emphasizing unified learning and multi-task learning. The proposed model consists of two main components. Firstly, a method for joint learning of high and low resolutions is utilized. It comprises two sub-segmentation branches that process the network's high and low-resolution inputs simultaneously. Each branch extracts features at different resolution levels and collectively they extract the network's entire feature set. Secondly, a multi-task learning approach is employed. This approach includes two sub-branches within the network, each conducting two different tasks simultaneously: low-resolution segmentation and super-resolution. Liu et al. [22] proposed the Multi-level Feature Fusion Network (MLFF-Net) to enhance segmentation performance by integrating multi-level feature fusion and attention mechanisms. The network is comprised of three modules: the Multi-scale Attention Module (MAM), the High-level Feature Enhancement Module (HFEM), and the Global Attention Module (GAM). MAM collects polyp details and information at various scales from the shallow outputs of the encoder. In HFEM, deeper features from the encoders are interlinked, enhancing the overall feature set. Meanwhile, the attention mechanism within HFEM reorganizes the importance of the combined features, dampening irrelevant components while accentuating information critical to the task. GAM merges information from both encoder and decoder features and is used to model dependencies between different regions of an image, ensuring the model accounts for information from more distant. Pan et al. [23] proposed A Global Guided Local Feature Stepwise Aggregation Network for polyp segmentation (GLSNet) to improve performance in polyp segmentation. The model incorporates three modules: the Spatial Feature Enhancement (SFE) module, the Globally Guided Local Feature Enhancement (GLFE) module, and the Feature Stepwise Aggregation (FSA) module. The SFE module enhances the spatial features of polyps, allowing for the acquisition of more detailed information about them. The GLFE module utilizes high-level features to capture noise in low-level features and uncovers polyp information hidden within superficial features. Lastly, the FSA module combines positional and semantic information of polyps across different scales to achieve the final segmentation results.

Lin et al. [24] introduced CSwinDoubleU-Net, a novel dual U-shaped image segmentation network that combines an interlaced convolutional structure with Shifted Windows (Swin) Transformer to address segmentation challenges such as differentiation between polyp regions and backgrounds and motion blur. The model is a CNN-based structure featuring a U-shaped encoder and decoder. The first U-shaped encoder network ensures the precise location of encoded features at each step by considering positional information, achieved through multiple convolutional layers. Subsequently, the second U-shaped encoder network garners additional global feature information using Swin Transformer layers with the shifted window technique. Finally, a Convolution Feature And Self-Attention Feature Fusion Module (CSFFM) has been developed to merge local convolutional features from the first U-shaped structure with global self-attention features from the second U-shaped structure. Liu and Song [25] proposed Attention Combined Pyramid Vision Transformer (Att-PVT), a novel approach that combines CNN and PVT to accurately detect the position and size of polyps. The proposed model consists of three main components: Multidimensional Information Extraction (MIE), Cascaded Context Integration (CCI), and Multilevel Feature Fusion (MFF). Att-PVT utilizes feature maps through the MIE module. CCI aims to learn semantic and spatial information by adaptively combining the top three layers of polyp features. The MFF module integrates boundary information from a higher-level global map with lower-level layers. This module is crucial for the accurate segmentation of colorectal polyps. Nguyen and Nguyen [26] proposed PolyPooling, a model designed for the precise segmentation of polyps. The encoder utilizes PoolFormer, employing a hierarchical structure to encode multilevel features. For decoding, the model leverages the Hamburger module and the Convolutional Block Attention Module (CBAM). The SegFormer decoder processes blocks in parallel with associated MLP modules before combining them. Within the proposed model, the Channel-wise Feature Pyramid (CFP) and refinement module are used in conjunction with the Pooling Reverse Attention module (Pooling-RA). CFP

enables parallel learning while capturing finer details. The Pooling-RA module is suggested to mitigate computational complexity. Huang et al. [27] proposed a U-Net neural network model to enhance segmentation accuracy rates. The traditional U-Net serves as the foundational architecture for Reparameterized Convolutional Network (RCNU-Net). For segmentation, a specific loss calculation method known as CDLoss is employed within the proposed model. To prevent gradient loss, a multi-branched structural model is utilized. The CBAM acts as a crucial bridge by expanding the receptive field of filters in both channel and spatial dimensions. This facilitates a secondary feature extraction that leverages attention benefits for superior contextual information aiding segmentation. This study employs a joint loss calculation method that combines both Cross-Entropy Loss (CELoss) and Dice Loss in the overall architecture.

Despite new methods emerging in the literature for polyp segmentation, researchers have yet to fully overcome the challenges encountered in polyp segmentation. This field remains highly current and exciting. In this article, unlike other literature studies, the performance of the DeeplabV3+ architecture's backbone networks, which have provided quite successful results in segmentation processes, is examined. Four different CNN architectures were applied to the backbone network of the DeepLabV3+ architecture: Model 1 (DeepLabV3+), Model 2 (ResNet50), Model 3 (SqueezeNet), and Model 4 (VGG16); and the results were tested on two significant datasets.

The organization of the rest of the paper is as follows: In the Materials and methods section, we present the Used datasets, DeepLabV3+ architecture, Evaluation metrics and Implementation details. Results section contains the results section, which presents the experimental studies and results. The last section includes Discussion and conclusion.

Material and method

Used datasets

In this article, experiments were conducted on the two most preferred datasets in polyp segmentation studies to compare and evaluate the models used with the latest state-of-the-art (SOTA) methods. These datasets are the publicly available Kvasir-SEG dataset [28] and the CVC-ClinicDB dataset [29].

The Kvasir-SEG dataset, created by expert clinicians in 2020, comprises a total of 1000 colonoscopy images. The resolution of these images varies between 1920x1072 and 332x487. Ground truth (GT) images are available for all 1000 images. The CVC-ClinicDB dataset, introduced in 2015, consists of a total of 612 colorectal images obtained from 31 colorectal sequences. Ground truth images are also available for each image in this dataset.

DeepLabV3+ architecture

The DeepLabV3+ architecture consists of encoder and decoder modules [30]. The encoder is used to reduce feature maps and extract a set of semantic features, while the decoder is used to restore spatial information and generate more explicit segmentation features. The DeepLabV3+ architecture is an enhancement of the DeepLabV3 architecture. This network architecture adds a decoding module based on DeepLabV3 and the output combination of the encoder module becomes the input to the decoder module. The architecture broadly employs a backbone network and an ASPP mechanism. The ASPP mechanism consists of one 1×1 atrous convolution and three 3×3 atrous convolutions with atrous rates of 6, 12, and 18, respectively, along with a global pooling layer. The four convolution operations and one pooling layer are processed in parallel. The backbone network is utilized to extract semantic information of features [31]. The original DeepLabV3+ architecture employs the Xception model Chen et al. [30] as its backbone network. Features from the Backbone Network and ASPP mechanism are subjected to 4 times subsampling. Fig. 1 shows the DeepLabV3+ architecture. In this article, an end-to-end framework using the DeepLabV3+ architecture for the automatic segmentation of polyps has been developed. DenseNet [32], ResNet50 [33], SqueezeNet [34], and VGG16 [35] models were used as the Backbone Network, and the segmentation success of the architecture was tested across different backbone networks.

The DenseNet architecture derives its name from "Densely Connected Convolutional Networks." It is distinguished by its dense connections, allowing each layer to receive inputs from all preceding layers. This structure promotes the reuse of features, increases the model's parameter efficiency, and facilitates more effective gradient propagation through deep networks. As a result, high performance is achieved with fewer parameters, even in deeper networks.

The ResNet50 architecture is a popular CNN architecture that utilizes the residual learning approach for deep learning. Thanks to residual blocks, it enables efficient transmission of gradients to deeper layers, thus overcoming the vanishing gradient problem encountered with very deep networks and facilitating easier model training. The SqueezeNet architecture offers exceptional parameter efficiency among CNN architectures. It uses 'fire' modules in its structure and directly classifies the feature maps of each class using global average pooling instead of fully connected at the last layer, thus significantly reducing the model size.

The VGG16 architecture is a CNN architecture that demonstrates that depth can improve model performance. In its structure, 3x3 convolutional filters and maximum pooling layers are sequentially ordered and complemented by three dense layers. Each convolution block improves in-depth feature extraction by increasing the number of filters, and the ReLU activation function is used.

Pre-trained CNN networks serve as the "backbone" of the model by extracting low-level features from the input image. Using DenseNet architecture as the backbone network for pre-trained DeepLabV3+ architecture is referred to as Model 1, utilizing ResNet 50 architecture is named Model 2, employing SqueezeNet architecture is named Model 3, and utilizing VGG16 architecture is named Model 4.

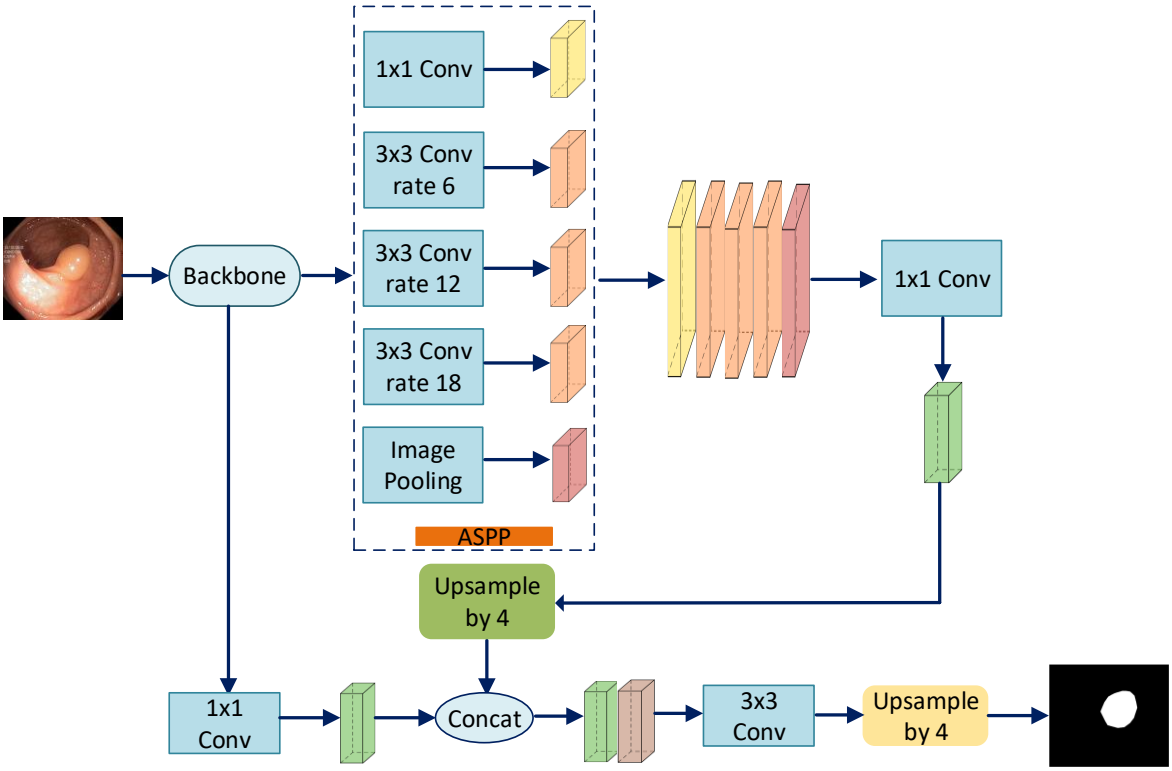

Figure 1. DeepLabV3+ architecture

.

Evaluation metrics

There are many different evaluation metrics commonly used in the field of medical image segmentation. This article has opted for the most frequently used evaluation metrics, and Equations (1)- (7) provide the mathematical expressions for these metrics. In the mathematical expressions, the terms FN, FP, TP, and TN are used to denote the number of false negatives, false positives, true positives, and true negatives, respectively.

$$mDice = \frac{1}{1+k}\sum_{i=0}^{k}\frac{2TP}{2TP+FP+FN} \quad (1)$$

$$mIoU = \frac{1}{1+k}\sum_{i=0}^{k}\frac{TP}{TP+FP+FN} \quad (2)$$

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Sp = \frac{TN}{TN+FP} \quad (5)$$

$$Pr = \frac{TP}{TP+FP} \quad (6)$$

$$MAE = \frac{1}{w\ x\ H}\sum_{i=1}^{w}\sum_{i=1}^{H}|\ S(i,j) - G(i,j) \quad (7)$$

Implementation details

All CNN architectures integrated into the backbone network of the DeepLabV3+ architecture have been trained using the TensorFlow framework on an NVIDIA RTX A4000 GPU. The same set of hyperparameters has been chosen for each model. These parameters are presented in Table 1.

Table 1. Optimal hyper-parameter values.

Hyper-Parameter	Value
Learning rate	0,0001
Batch size	8
Optimizer	Adam
Activation function	ReLU

Results

The datasets used in polyp segmentation with the DeepLabV3+ architecture have been divided into training,

validation, and testing sets. Each dataset was shuffled before being fed into the network and randomly split into 80% for training and 20% for testing. 20% of the data in the training dataset was used for validation. Accordingly, of the Kvasir-SEG dataset, 640 were used for training, 200 for testing, and 160 for validation. For the CVC-ClinicDB dataset, 392 were used for training, 122 for testing, and 98 for validation.

The experimental results obtained from the integration of the DeepLabV3+ architecture with different CNN architectures have been presented in terms of mDice and mIoU evaluation metrics for the Kvasir-SEG and CVC-ClinicDB datasets, respectively, in Fig. 2 and Fig. 3.

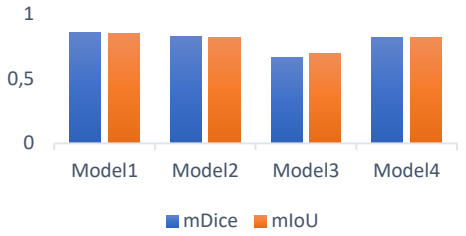

Figure 2. Results from the Kvasir-seg dataset.

According to Fig. 2, the evaluation metrics for the Kvasir-SEG dataset results are displayed, indicating that the best performance in polyp segmentation was achieved with Model 1, while the lowest performance was observed with Model 3.

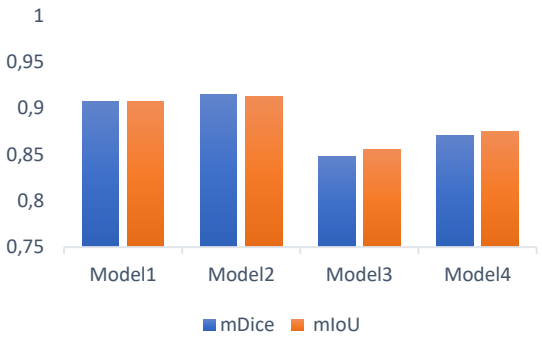

Figure 3. Results from the CVC-ClinicDB dataset.

According to Fig. 3, the evaluation metrics for the results on the CVC-ClinicDB dataset are displayed, indicating that the best performance in polyp segmentation was achieved with Model 2, while the lowest performance was observed with Model 3.

The quantitative findings of the models and their comparison with SOTA methods are presented in Table 2. The performance success of SOTA methods is derived from the article by [9]. Gangrade et al. [9], conducted experiments with a 256x256 image size and training for 25 epochs. Additionally, the learning rate was set at 0.0002, and the Adam optimizer was chosen as the optimization method.

The dataset was divided into 80% training, 10% validation, and 10% testing. For the comparison to be fair, it's necessary to keep the hyper-parameters consistent; however, this article has conducted a comparison with hyper-parameter analysis. Gangrade et al. [9] trained for 25 epochs, but this was not considered sufficient for updating the weights. There might have been a case of overfitting. Choosing a learning rate of 0.0002 can increase the likelihood of the network getting stuck in local minima and extend the time it takes to reach the global minimum. The purpose of this table is to display the performance of methods on datasets. Thus, it demonstrates the level at which models obtained through the integration of different CNN architectures with DeepLabV3+ architecture stand in the literature.

In this article, data augmentation was not performed on the two datasets used, and the data were not subjected to preprocessing. Due to the images in the dataset being of various resolutions, the images were provided to the network architectures at a resolution of 256x256.

Fig. 4 and Fig. 5 display some example qualitative results of experiments conducted on the Kvasir-SEG and CVC-ClinicDB datasets, respectively. In these visuals, the first column represents images from the original dataset. The second column represents the ground truth (GT) images. The last four columns show the segmentation results for Model 1, Model 2, Model 3, and Model 4, respectively.

Looking at the test images provided in Fig. 4, it can be observed that Models 3 and Model 4 have low segmentation success in the third image. This situation indicates that these architectures struggle with images of non-polypoid lesions.

When looking at the test images provided in Fig. 5, it is observed that Model 3 is the most challenged. Upon examining the visual results, it can be concluded that the SqueezeNet architecture performs weaker as a backbone for DeepLabV3+ compared to other architectures.

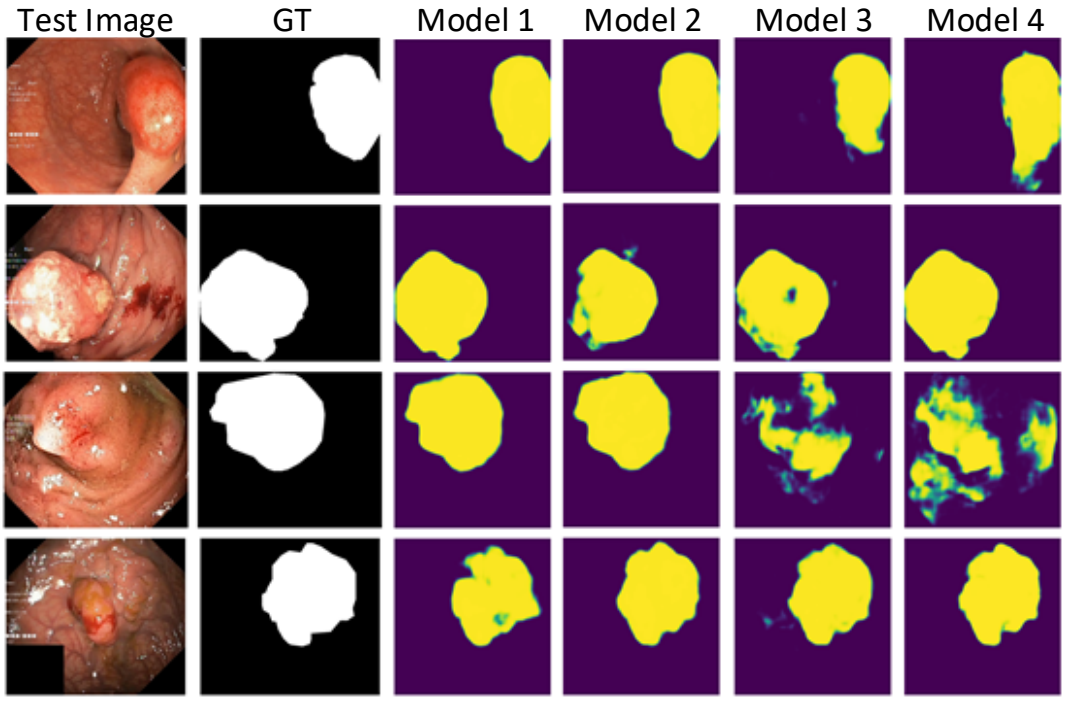

Figure 4. Kvasir-seg dataset segmentation results.

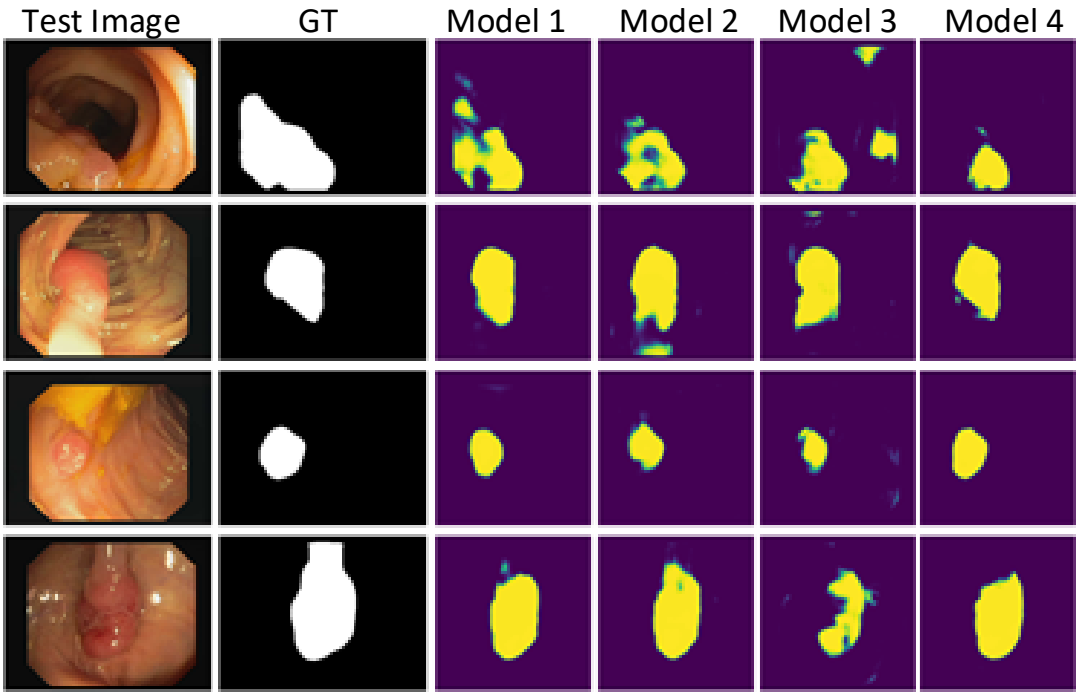

Figure 5. CVC-ClinicDB dataset segmentation results.

Discussion and conclusion

Polyps, although generally benign structures, can be considered precancerous lesions in some cases. The early diagnosis and treatment of these structures are of great importance, especially in preventing serious health issues like colorectal cancer.

This article presents polyp segmentation on colonoscopy images using the DeepLabV3+ architecture with various backbone networks. In the method, pre-trained CNN

networks are used as the backbone for the DeepLabV3+ architecture. This approach allows the network to adapt more quickly to datasets by leveraging what it has learned from previous tasks, achieving higher performance with less input.

The performance of the conducted study has been compared with other SOTA methods in the literature. Upon examining Table 2, it is observed that SOTA methods yield different results for each dataset. According to the Table 2, the best result on the Kvasir-SEG dataset was obtained with the DeepLabV3+ architecture's Inception backbone.

However, the best result on the CVC-ClinicDB dataset was achieved with Model 2, which uses the ResNet50 architecture as its backbone network. Experimental studies have been evaluated using metrics such as mDice, mIoU, Acc, Recall, SP, and MAE. According to these metric results, the worst backbone network for both datasets was the SqueezeNet architecture. In the study, experiments were conducted on raw data without any data augmentation or preprocessing. While this approach may reduce the models' generalization ability, it has sped up the analysis of the models' performance on raw data. For the future, it is recommended to modify the DeepLabV3+ architecture with the Resnet50 backbone network to increase the models' robustness and generalizability.

Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared.

There is no conflict of interest with any person / institution in the article prepared.

Authors' Contributions

AKGÖL: Conceptualization, Methodology, Writing – original draft. TOPTAŞ: Supervision, Validation, Writing – editing.

References

[1] B. Toptaş and D. Hanbay, "Separation of arteries and veins in retinal fundus images with a new CNN architecture," Comput. Methods Biomech. Biomed. Eng. Imaging Vis., vol. 11, no. 4, pp. 1512–1522, 2023, doi: 10.1080/21681163.2022.2151066.

[2] M. Toptaş and D. Hanbay, "Mikroskobik Kan Hücre Görüntülerinin Güncel Derin Öğrenme Mimarileri ile Bölütlemesi," *Mühendislik Bilim. ve Araştırmaları Derg.*, vol. 5, no. 1, pp. 135–141, 2023, doi: 10.46387/bjesr.1261689.

[3] C. Özdemir, "Meme Ultrason Görüntülerinde Kanser Hücre Segmentasyonu için Yeni Bir FCN Modeli," *Afyon Kocatepe Univ. J. Sci. Eng.*, vol. 23, no. 5, pp. 1160–1170, 2023, doi: 10.35414/akufemubid.1259253.

[4] N. Şahin, N. Alpaslan, and D. Hanbay, "Robust optimization of SegNet hyperparameters for skin lesion segmentation," *Multimed. Tools Appl.*, vol. 81, no. 25, pp. 36031–36051, 2022, doi: 10.1007/s11042-021-11032-6.

[5] W. Zhang, F. Lu, H. Su, and Y. Hu, "Dual-branch multi-information aggregation network with transformer and convolution for polyp segmentation," *Comput. Biol. Med.*, vol. 168, 2024, doi: 10.1016/j.compbiomed.2023.107760.

[6] A. Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, "Cancer statistics, 2021," *Ca Cancer J Clin*, pp. 7–33, 2021.

[7] O. H. Maghsoudi, "Superpixel based segmentation and classification of polyps in wireless capsule endoscopy," in *2017 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2017 - Proceedings*, 2017, pp. 1–4. doi: 10.1109/SPMB.2017.8257027.

[8] S. Hwang and M. E. Celebi, "Polyp detection in Wireless Capsule Endoscopy videos based on image segmentation and geometric feature," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2010, pp. 678–681. doi: 10.1109/ICASSP.2010.5495103.

[9] S. Gangrade, P. C. Sharma, A. K. Sharma, and Y. P. Singh, "Modified DeeplabV3+ with multi-level context attention mechanism for colonoscopy polyp segmentation," *Comput. Biol. Med.*, vol. 170, 2024, doi: 10.1016/j.compbiomed.2024.108096.

[10] S. Li *et al.*, "Boundary guided network with two-stage transfer learning for gastrointestinal polyps segmentation," *Expert Syst. Appl.*, vol. 240, 2024, doi: 10.1016/j.eswa.2023.122503.

[11] W. Li, Z. Huang, F. Li, Y. Zhao, and H. Zhang, "CIFG-Net: Cross-level information fusion and guidance network for Polyp Segmentation," *Comput. Biol. Med.*, vol. 169, 2024, doi: 10.1016/j.compbiomed.2024.107931.

[12] X. Jia *et al.*, "PolypMixNet: Enhancing semi-supervised polyp segmentation with polyp-aware augmentation," *Comput. Biol. Med.*, vol. 170, 2024, doi: 10.1016/j.compbiomed.2024.108006.

[13] Y. He, Y. Yi, C. Zheng, and J. Kong, "BGF-Net: Boundary guided filter network for medical image segmentation," *Comput. Biol. Med.*, vol. 171, 2024, doi: 10.1016/j.compbiomed.2024.108184.

[14] D. Liu, H. Deng, Z. Huang, and J. Fu, "FCA-Net: Fully context-aware feature aggregation network for medical segmentation," *Biomed. Signal Process. Control*, vol. 91, 2024, doi: 10.1016/j.bspc.2024.106004.

[15] H. Wang *et al.*, "Unveiling camouflaged and partially occluded colorectal polyps: Introducing CPSNet for accurate colon polyp segmentation," *Comput. Biol. Med.*, vol. 171, 2024, doi: 10.1016/j.compbiomed.2024.108186.

[16] G. Liu *et al.*, "CAFE-Net: Cross-Attention and Feature Exploration Network for polyp segmentation," *Expert Syst. Appl.*, vol. 238, 2024, doi: 10.1016/j.eswa.2023.121754.

[17] D. Shao, H. Yang, C. Liu, and L. Ma, "AFANet: Adaptive Feature Aggregation for Polyp Segmentation," *Med. Eng. Phys.*, p. 104118, 2024, doi: 10.1016/j.medengphy.2024.104118.

[18] Z. U. D. Muhammad, U. Muhammad, Z. Huang, and N. Gu, "MMFIL-Net: Multi-level and multi-source feature interactive lightweight network for polyp segmentation," *Displays*, vol. 81, 2024, doi: 10.1016/j.displa.2023.102600.

[19] G. Yue *et al.*, "Boundary uncertainty aware network for automated polyp segmentation," *Neural Networks*, vol. 170, pp. 390–404, 2024, doi: 10.1016/j.neunet.2023.11.050.

[20] S. Ahmed and M. K. Hasan, "Twin-SegNet: Dynamically coupled complementary segmentation networks for generalized medical image segmentation," *Comput. Vis. Image Underst.*, vol. 240, 2024, doi: 10.1016/j.cviu.2023.103910.

[21] 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," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 3, 2024, doi: 10.1016/j.jksuci.2024.101981.

[22] J. Liu, Q. Chen, Y. Zhang, Z. Wang, X. Deng, and J. Wang, "Multi-level feature fusion network combining attention mechanisms for polyp segmentation," *Inf. Fusion*, vol. 104, 2024, doi: 10.1016/j.inffus.2023.102195.

[23] X. Pan, C. Ma, Y. Mu, and M. Bi, "GLSNet: A Global Guided Local Feature Stepwise Aggregation Network for polyp segmentation," *Biomed. Signal Process. Control*, vol. 87, 2024, doi: 10.1016/j.bspc.2023.105528.

[24] Y. Lin, X. Han, K. Chen, W. Zhang, and Q. Liu, "CSwinDoubleU-Net: A double U-shaped network combined with convolution and Swin Transformer for colorectal polyp segmentation," *Biomed. Signal Process. Control*, vol. 89, 2024, doi: 10.1016/j.bspc.2023.105749.

[25] X. Liu and S. Song, "Attention combined pyramid vision transformer for polyp segmentation," *Biomed. Signal Process. Control*, vol. 89, 2024, doi: 10.1016/j.bspc.2023.105792.

[26] D. C. Nguyen and H. L. Nguyen, "PolyPooling: An accurate polyp segmentation from colonoscopy images," *Biomed. Signal Process. Control*, vol. 92, 2024, doi: 10.1016/j.bspc.2024.105979.

[27] B. Huang, T. Huang, J. Xu, J. Min, C. Hu, and Z. Zhang, "RCNU-Net: Reparameterized convolutional network with convolutional block attention module for improved polyp image segmentation," *Biomed. Signal Process. Control*, vol. 93, 2024, doi: 10.1016/j.bspc.2024.106138.

[28] D. Jha *et al.*, "Kvasir-SEG: A Segmented Polyp Dataset," 2020, pp. 451–462. doi: 10.1007/978-3-030-37734-2_37.

[29] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imaging Graph.*, vol. 43, pp. 99–111, Jul. 2015, doi: 10.1016/j.compmedimag.2015.02.007.

[30] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," Feb. 2018, doi: https://doi.org/10.48550/arXiv.1802.02611.

[31] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua. pp. 3462–3471, 2017. doi: 10.1109/CVPR.2017.369.

[32] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," Aug. 2016, [Online]. Available: http://arxiv.org/abs/1608.06993

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[34] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," Feb. 2016, [Online]. Available: http://arxiv.org/abs/1602.07360

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[37] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Trans. Med. Imaging*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609.

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