IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 August 2024, accepted 16 September 2024, date of publication 7 October 2024, date of current version 24 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3475877

RESEARCH ARTICLE

PolyRes-Net: A Polyhierarchical Residual Network for Decoding Anatomical Complexity in Medical Image Segmentation

AMR MAGDY[1], KHALID N. ISMAIL[2], MARGHNY H. MOHAMED[3,4], MAHMOUD HASSABALLAH[1,5], HAITHAM MAHMOUD[2], AND MOHAMMED M. ABDELSAMEA[6]

[1]Department of Computer Science, Faculty of Computers and Information, South Valley University, Qena 83523, Egypt
[2]College of Computing, Birmingham City University, B5 5JU Birmingham, U.K.
[3]Department of Computer Science and Information Technology, Egypt-Japan University of Science and Technology, New Borg El-Arab City, Alexandria 21934, Egypt
[4]Department of Computer Science, Faculty of Computers and Information, Assiut University, Assiut 71515, Egypt
[5]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
[6]Department of Computer Science, University of Exeter, EX4 4PY Exeter, U.K.

Corresponding author: Khalid N. Ismail (Khalid.ismail@bcu.ac.uk)

ABSTRACT Medical image segmentation entails assigning each pixel in an image to its corresponding class label, a challenging task given the considerable anatomical variations in different cases. The encoder-decoder approach, exemplified by architectures such as U-Net, has emerged as the predominant framework for medical imaging segmentation tasks. In recent years, diverse modifications to the U-Net architecture have been explored, giving rise to distinct models that showcase noteworthy results in comparison to the conventional U-Net design. In this paper, we introduce a novel architectural framework, which we refer to as the Polyhierarchical Residual Network (*PolyRes-Net*). Each encoder step comprises a Multi-Level Residual Block (*MLR-block*) designed to extract local and global feature maps. Furthermore, each decoder step is preceded by an attention gate, aiding in the extraction of the most salient features from the preceding layer, while skip connections correspond to the respective encoder steps. Lastly, the multi-scale feature aggregation (*MSFA*) block consolidates features from various decoder steps. Four benchamar datasets are used for evaluating our model: Krusir-SEG, CVC ClinicDB, 2018 Data Science Bowl, and ISIC-2018 skin lesion segmentation challenge dataset based on two metrics: the Mean Dice Similarity Coefficient (mDSC) and the Mean Intersection Over Union (mIOU). The results of the proposed *PolyRes-Net* outperformed the state-of-the-art segmentation methods. Specifically, *PolyRes-Net* achieves the highest *mDSC* scores of 91.02%, 91.80%, and 89.25% on CVC ClinicDB, 2018 Data Science Bowl, and ISIC-2018 skin lesion segmentation challenge dataset, respectively. Additionally, the highest *mIOU* scores are 85.60%, 85.32%, and 82.14% for the same datasets, further underscoring the efficacy of the proposed model.

INDEX TERMS Medical image segmentation, medical imaging, deep learning.

I. INTRODUCTION

Medical image segmentation is vital for extracting regions of interest from image data. This enables more precise interpretation of medical images for doctors. The main goal of segmenting this data is to identify areas of the anatomy required for diagnosing and treating various diseases which benefit a large number of patients. However, manual annotation of medical image pixels is a costly process due to the vast number of patients with huge details across different modalities of medical images, and the limited number of specialists available [1]. On the other hand, leveraging precise and effective processing of medical images can significantly reduce the time, cost, and potential errors associated with human-based processing [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

Given the complexity and variability of medical imaging data, the utilisation of deep learning networks has emerged as the necessary approach. Deep learning networks are used as state-of-the-art methods in medical imaging segmentation and outperform non-deep state-of-the-art methods. These networks require a large amount of data to train and powerful computation resources to process train data and justify the huge number of network parameters. One of the first deep networks applied to image segmentation is a fully convolutional neural network (FCN) [3].

Nowadays the majority of the medical image segmentation algorithms follow encoder-decoder top-down structures. These algorithms initially encode input images into a latent space using different convolution layers. Subsequently, decoders utilise convolution layers to learn the locations of regions of interest within images. The horizontal propagation of dense feature maps from the encoder to the corresponding decoder layer that applies the spatial information to the deeper layer. This architectural design has significantly produced a more accurate output segmentation map (skip connection), known as UNet [4]. U-Net and its variants have arguably been the most influential leap forward in segmentation algorithms in the recent past. Attention U-Net [5] and MultiResUNet [6] are two examples of variants derived from the original U-Net architecture. Attention UNet use the attention mechanism between skip connections and output from the previous decoder layer to aggregate low-level features from the encoder to high-level features from the decoder, whereas MultiResUNet use MultiRes Block in encoding and decoding layers with a modified path for skip connection from the encoder to decoder layers consist of a chain of convolutional layers with residual connections to alleviate the disparity between the encoder-decoder features.

In image segmentation, the aggregation of multi-scale features holds particular importance for capturing context and spatial information at different scales, which can lead to more accurate and robust segmentation results. The U-Net architecture, commonly used for biomedical image segmentation, employs skip connections to combine features from the encoder and decoder at different resolutions. This enables the model to maintain fine-grained details while also incorporating high-level semantic information. The *PolyRes-Network* adopts the U-Net architecture, employing two branches: an encoder branch and a decoder branch. At each level within these branches, it consists of an *MLR-block*. The *MLR-block* architecture enables the model to identify more useful local and global spatial information in the encoding phase. Meanwhile, the decoding phase utilizes an attention mechanism between the skip connection and the corresponding output from the *MLR-block* of the decoding layer to generate more accurate segmentation feature maps at different scales in each level of the decoder branch. Finally, the multi-scale feature aggregation *MSFA* block aggregates the output from different attention gates along with the output of the final *MLR-block* in the network to generate the final segmentation feature maps of the network. However, while many studies have tackled the segmentation problem, they often lack a seamless transfer of gradient features from the encoder to the decoder. Specifically, there is a deficiency in incorporating skip-connections with attention blocks, as well as transferring features from different decoder layers to a multi-scale feature block for generating final segmentation feature maps. Experimentally this architecture gives our model the ability to output more accurate segmentation maps compared with state-of-the-art methods.

The main contributions of this work are:

- We propose a novel architecture, PolyRes-Net, for semantic image segmentation. The proposed architecture introduces the MLR block in each layer of the encoder-decoder branches of the network, incorporating an attention mechanism in the skip connection, and utilizing a multi-scale features aggregation layer to output the final segmentation feature maps.
- PolyRes-Net relies on preserving low-level features along with high-level features throughout the network's final output. This is achieved by establishing connections at two levels: from the encoder to the decoder through skip connections and from various decoding layers to the final segmentation feature maps using a multi-scale feature aggregation layer.
- To demonstrate the efficacy of the suggested method performs better than alternative algorithms through conducting experiments on a variety of datasets. Utilizing Dice Coefficient (DSC) and Jaccard Index as assessment metrics, we have tested with four distinct medical imaging datasets: Kvasir-SEG, CVC ClinicDB, 2018 Data Science Bowl, and ISIC-2018 skin lesion segmentation challenge dataset.
- An extensive evaluation of PolyRes-Net across four datasets shows a significant improvement over most of the state-of-the-art models that exist nowadays. Therefore, PolyRes-Net can be a new baseline for medical image segmentation tasks.

The remainder of the paper is organised as follows: Section II discusses the previous studies that applied deep learning in the medical domain. Section III discusses and proposes the *PolyRes-Network* system. Section IV evaluates and demonstrates the efficiency of *PolyRes-Network* against other state-of-art methods. Finally, section V wraps up the paper, discussing its conclusion, limitations and future development.

II. RELATED WORK

Systems for computer-aided diagnosis (CAD) have drawn interest as an essential tool for making quick clinical decisions about specific diseases [7]. Among the most challenging tasks for these systems is medical image segmentation, which involves extracting features from medical images to aid in the diagnosis of patient cases. FCN [3] and U-Net [4] have gained significant popularity among semantic segmentation approaches for medical images.

The U-shaped architecture is commonly used in the segmentation task, consisting of an encoder-decoder path. in the setup, the encoder utilizes the channels (filters) and reduces the spatial dimensions in each layer. On the other hand, the decoder reduces the channels while increasing the spatial dimensions. In the end, the spatial dimensions are restored to predict each pixel in the input image [8].

Training Deep Neural Networks is an a challenge process. Accuracy can be increased by training a deep neural network with a deeper network. On the other hand, it could interfere with training and lead to a deterioration issue. UNet++ [9] established a densely linked encoder-decoder network with deep supervision by adding a number of layered, dense skip paths. To help with training and tackle deterioration, a deep residual learning framework [10] is suggested. An architecture called Deep Residual U-Net [11] makes use of U-Net [4] and deep residual learning [10]. An enhanced version of the ResUnet architecture called ResUnet++ [12] makes use of squeezing and excitation blocks, residual blocks, attention blocks, and Atrous Spatial Pyramidal Pooling (ASPP). In order to emphasize pertinent aspects and suppress irrelevant ones, the Squeeze and Excitation (S&E) block was able to model relationships between the various channels and produce a global information map. These S&E blocks were included by FED-Net [13] in their altered U-Net architecture.

In the U-shaped design, a high-level semantic feature maps from the decoder and matching low-level feature maps from the encoder are combined via skip connections. This might lead to a semantic divide between high-level and low-level characteristics, with various strategies being employed to bridge this divide. A suggested architecture called Attention U-Net [5] modifies feature maps transmitted over skip-connections by means of an attention block. In order to exclude irrelevant spatial characteristics that are flowing from the skip connections and retain just the pertinent data, a gating mechanism is formed using the output of the previous decoder block. Furthermore, MultiResUnet [6] suggested designing each U-shaped block to have three 3×3 convolutional layers with 1×1 residual convolutional layer, which may allow comprehending some additional spatial information. This will help to reduce the semantic gap between low-level and high-level feature maps. To improve segmentation results on both natural and medical pictures, there are a number of extensions and modifications based on the U-Net architecture [14], [15], [16], [17], [18].

Atrous convolutions, or dilated convolutions, are introduced by the DeepLab architecture [14] to extract more dense features where information is better kept given objects of different scale. Then, compared to their earlier DeepLab versions, DeepLabV3 [19] exhibits impressive progress. In comparison to the synthesis paths of FCN and UNet, the DeepLabV3 architecture employs a synthesis path with fewer convolutional layers. A skip connection, akin to UNet design, is used by DeepLabV3 between the analysis and synthesis paths. The polyp segmentation architecture introduced by Pranet [18] uses a parallel partial decoder (PPD) to aggregate features in high-level layers and creates a global map that serves as the first guiding area for subsequent components. Features maps can be related to one another using the reverse Attention (RA) module.

HRENet [20] developed an adaptive feature aggregation (AFA) module to dynamically aggregate the multi-level features and send them to the decoder block, as well as an informative context enhancement (ICE) module to intensify the lower-level encoder features under the direction of hard-region attention. Multiple-scale information is included into modules used in other efforts, such as MSRFNet [21] and PolypSeg [22]. In PolypSeg [22], an adaptive scale context module (ASCM) and a semantic global context module (SGCM) were employed. While SGCM strengthens feature fusion between high-level and low-level features and eliminates noise from the low-level features to increase segmentation accuracy, ASCM addresses size fluctuations within polyps and boosts feature representation capabilities. In a comparable manner, MSRFNet [21] combined an additional shape stream network to refine polyp boundaries with cross-scale fusion modules to convey both high-resolution and low-resolution characteristics.

TGANet [23] suggested a text-guided attention architecture that made use of several feature augmentation modules coupled with various encoder blocks to address polyps' fluctuating size and quantity for robust polyp segmentation. To supplement the size-based and number-based feature representations, an auxiliary work is learnt alongside the primary task and utilized as label attention in the decoder blocks. An encoder–decoder architecture is proposed by DCSAU-Net [24]. The primary feature conservation (PFC) approach in the encoder section by integrating the long-range geographical information of the network in the low-level semantic layer, tracking lowers the number of parameters and computation required. The rich primary feature derived from this layer is then provided to the compact split-attention (CSA) block. The CSA module uses a multi-path attention structure to improve the feature representation of various channels. In order to efficiently extract features from the combined features, the decoder concatenates encoded features from each downsampling layer with matching upsampled features using skip connections. Then, the same CSA block is applied.

EANet [25] introduces three modules; the dynamic scale-aware context module (DSC), the edge attention preservation module (EAP), and the multi-level pairwise regression module (MPR) to apply the complementary relationship between edge detection and object segmentation. The dynamic scale-aware context module (DSC) learns the relevant receptive fields dynamically and adaptively based on the scale of the segmented target object, therefore capturing multi-scale contextual information. The segmented object's edge information is extracted using the Edge Attention Preservation module (EAP), which suppresses low-level background noise while keeping edge-related

information intact. By modeling edge and area information, the multi-level pairwise regression module (MPR) is presented to enhance features retrieved from various depth-wise layers, including deeper and shallower layers.

The resulting segmentation task may be affected by encoding the input image as a low-resolution representation by joining high-to-low resolution convolutions in series, and then recovering the high-resolution representation from the encoded low-resolution representation without taking the high-resolution representation into consideration. To ensure that segmentation architectures produce reliable segmentation maps (such as skip connections), it is essential to maintain tracking of a high-resolution representation of the picture. According to [26], [27], [28], and [29], multi-scale fusion may also assist in exchanging high- and low-resolution features during the segmentation job rather than obtaining segmentation maps from low-level representations.

Based on previous work, this paper introduces a novel U-Net segmentation architecture named the Multilevel Residual Network (*PolyRes-Net*), consisting of three main modules: the Multi-Level Residual module (*MLR*), the Attention module, and the Multi-Scale Feature Aggregation module (*MSFA*). The Multi-Level Residual module (*MLR*) is utilized to extract global and local feature maps using two main branches and multiple levels of residual layers providing the model with more accurate and missing features. An attention mechanism is employed between the up-sampled features in the decoder and the corresponding encoder features through skip connections. The multi-scale Feature Aggregation module (*MSFA*) plays a pivotal role in generating the output segmentation feature maps. It aggregates features from different attention gates with the output of the last (*MLR*) block features, ultimately producing the final segmentation feature maps.

III. PROPOSED METHOD

Our proposed method as shown in Fig. 1 composed of two paths **encoder** path consist of five *MLR-blocks* and **decoder** path consist of four *MLR-blocks*. In the decoding stage attention gates are used between skip connection and output from the corresponding decoder layer to combine high-resolution representation and low-resolution representation from encoding to decoding path and get attention to the most important features before going up to the next decoder layer with more rich features. A detailed description of our proposed model *PolyRes-Network* is presented in this section.

Our model depends on multi-scale feature aggregation at different levels. Firstly, from the encoding stage to the decoder stage through skip connection and attention blocks. Secondly from different decoding stages through *MSFA* block to output the final segmentation feature map. This architecture enables the model to transfer and keep more important features from the encoder to decoder layers through skip connection and attention gates, also from different decoder layers at different scales to result in the final segmentation feature maps. Features aggregation that's done through skip connections, attention gates and *MSFA* block enables the model to calibrate some of the misaligned predictions and improve the resulting segmentation features maps.

TABLE 1. Comparison of various segmentation architectures.

Architecture	Methodology
FCN [3] U-Net [4]	- Encoder-decoder architecture with skip connections - Training with cross-entropy loss and Adam optimizer
UNet++ [9]	- Nested skip pathways with deep supervision - Incorporation of residual connections
Deep Residual U-Net [11] ResUnet++ [12]	- Integration of deep residual learning with U-Net architecture - Utilization of squeeze and excitation blocks, ASPP, and attention blocks
Attention U-Net [5]	- Incorporation of attention block for feature refinement - Training with dice loss
Pranet [18]	- Utilization of parallel partial decoder and reverse Attention (RA) module - Multi-level feature aggregation
PolyRes-Net	- Introduction of *MLR* module for feature extraction - Employment of attention mechanism and *MSFA* module for feature refinement and aggregation

Algorithm 1 PolyRes-Network Model

```
Input: original image X to be segmented
Output: segmented images Y for original image X
skip_connections = []
Y = X
//encoder branch
for i from 1 to 4:
    Y = skip_connections[i] = MLR(Y)
    Y = maxpool(Y)
endfor

attentions = []
Y = MLR(Y)
Y = up_convolution(Y)
//decoder branch
for i from 5 to 8:
    Y = attention(Y, skip_connections[9 - i])
    Y = attentions[i - 4]
    Y = MLR(Y)
    if i not equal 8
        Y = up_convolution(Y)
endfor

Y = MSFA(attentions[0], attentions[1], attentions
    [2], attentions[3], Y)
```

A. ENCODER

Encoder branch consists of five *MLR-blocks* as mentioned previously with a gradually increasing number of filters F, that controls the number of feature maps generated from each block.

$$F = U \times \alpha \tag{1}$$

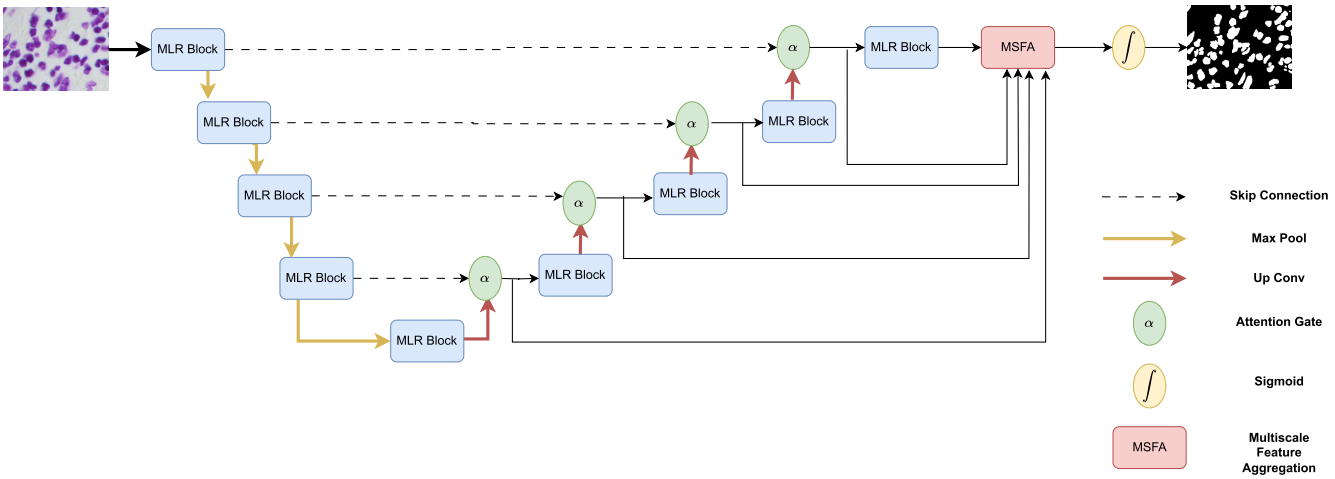


FIGURE 1. PolyRes network architecture which consists of three main components MLR-block, Attention-block and multi-scale feature aggregation block.

TABLE 2. Multi-level residual network architecture, depicting the configuration for each MLR-block in the network. This includes parameters such as the number of input filters, the number of output filters, and the operations performed after each MLR-block.

MLR blocks	#InputFilters	#OutputFilters	FollowedOperations
MLR-block1	3	43	MaxPool
MLR-block2	43	88	MaxPool
MLR-block3	88	177	MaxPool
MLR-block4	177	355	MaxPool
MLR-block5	355	711	UpConv
MLR-block6	256	355	UpConv
MLR-block7	128	177	UpConv
MLR-block8	64	88	UpConv
MLR-block9	32	43	MSFA,Sigmoid

where U is the initial number of filters in the corresponding layer of the network as {32,64,128,256,512} consecutively and α is a scalar coefficient set with value 1.67 as mentioned in MultiResUnet [6]. After each block maxpool operation is conducted to downsample feature maps to capture sufficiently large receptive field and thus, semantic contextual information. This allows features on the coarse spatial grid level to represent pixel position and relationships on a global scale.

MLR-block is the main unit of our network as shown in Fig. 2 consisting of two main branches left and right branch. Each one mainly consists of two consecutive 3×3 convolutional layers with more additional residual paths adding more shapes of information that the network can learn and helping prevent network degradation. Each 3×3 convolutional layer exists in each branch having different sizes of filters computed as follows:

$$FLFilters = F \times 0.333 \quad (2)$$

$$SLFilters = F \times 0.5 \quad (3)$$

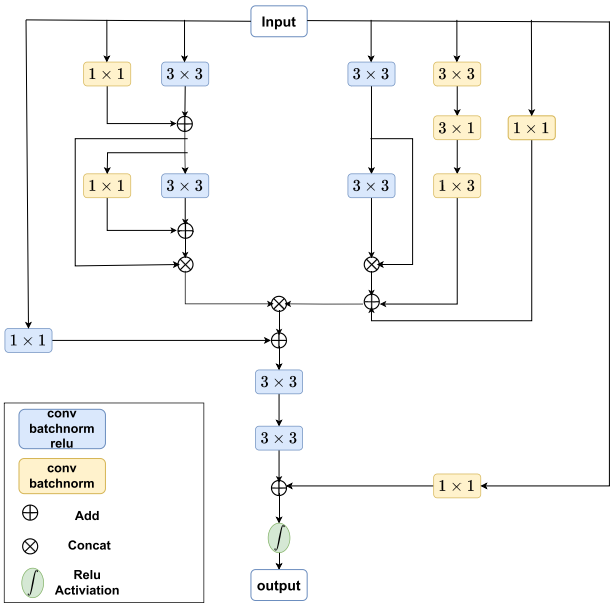


FIGURE 2. MLR-block architecture, illustrating two main branches: left and right branches. Each branch comprises two 3 × 3 convolutional layers, accompanied by additional residual layers for enhanced feature extraction. The primary two branches are concatenated into a unified branch, producing the final set of features.

where *FLFilters*, *SLFilters* are first-layer and second-layer filters consecutively for each layer of the two main branches. As shown in Table 2 each block outputs almost 85% of *F*.

The two main branches try to collect different shapes of data. *LBranch* consist of two consecutive 3×3 convolutional layers as mentioned previously with additional two 1×1 residual layers added locally to each 3×3 layer, then output from each 3×3 layer are concatenated. *RBranch* also consist of two consecutive 3×3 convolutional layers where the output from each layer are concatenated, then additional two residual paths are added to concatenated vector, one residual path consisting of one 1×1 convolutional layer and the other residual path consist of three consecutive convolutional layers as 3×3, 3×1 and 1×3.

MLR-block depend on transfer knowledge on different levels: Firstly, At the level of each branch of the two main branches features computed from *FLFilters* and *SLFilters* are concatenated. Secondly, features computed from each branch are concatenated and fed forward in two consecutive 3×3 convolutional layers with a decreasing number of filters to get the most important features from the concatenated vector then apply RelU activation function and output block features. The left branch tends to extract local features and the right branch extracts more general features with multiple levels of residual paths adding more different shapes of information and preventing network degradation.

B. DECODER

Decoder branch consists of four *MLR-blocks* as mentioned previously with a gradually decreasing number of filters and attention layer applied before each decoding block. By identifying which aspects of the features need greater attention to improve the quality of characteristics that improve the results, the attention mechanism directs its attention to a subset of its input features.

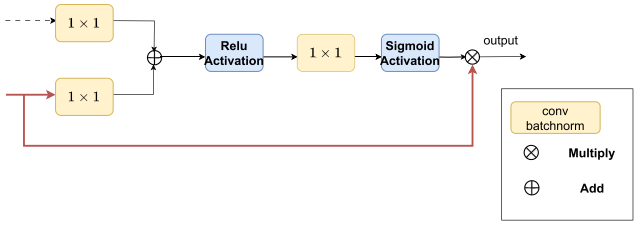


FIGURE 3. The attention block outputs the most important features by capturing the attention between the skip connection and the MLR-block output.

Inspired by the success of attention mechanism with unet [5], our architecture used attention layer in the decoder part between skip connection and previous *MLR-block* output to be able to focus on the essential areas of the feature maps as shown in Fig. 3, so the input for each *MLR-block* in the decoder branch computed as following

$$input = atten(skipconnection, previousblockoutput) \quad (4)$$

The Attention block takes two inputs: the skip connection and the previous *MLR-block* output. Each input is fed into a 1×1 convolutional layer to equalize the number of filters in each input. The outputs from the convolutional layers are then added together, followed by a ReLU activation. Subsequently, another 1×1 convolutional layer is applied. The final decision on where the model should focus its attention, based on the previous *MLR-block*, is made using a sigmoid activation function.

C. MULTI-SCALE FEATURE AGGREGATION

Our model utilizes the multi-scale feature aggregation (*MSFA*) layer to make segmentation decisions, incorporating information not only from the output of the last decoding layer but also considering outputs from each attention gate. Aggregating the outputs from attention gates, rather than from decoding layers alone, provides the network with more accurate features. Attention gates leverage skip connections and the previous decoder layer output to merge high and low-level representations, focusing on the most important features at each scale. The *MSFA* layer then aggregates these crucial features from each scale, in addition to the final decoder layer output, enhancing the resulting segmentation maps of the network.

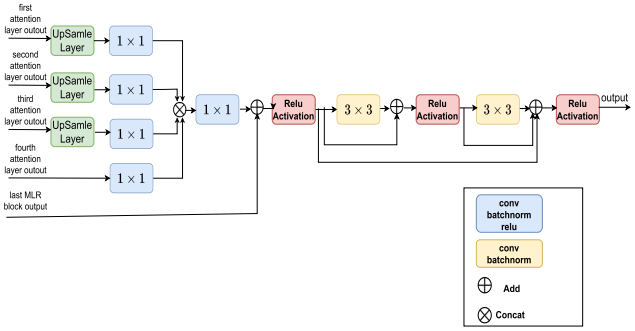


FIGURE 4. Multiscale feature aggregation involves the aggregation of features from different scales obtained from various attention blocks. This aggregation is performed in conjunction with the final MLR-block output, resulting in the generation of final segmentation feature maps.

The *MSFA*, as illustrated in Fig. 4, takes five inputs. Four of them are features from different attention gates at each scale, denoted as $\{atten_i, i = 1, \ldots, 4\}$, and one input is obtained from the final *MLR-block* output in the decoder path. The features from different attention gates with varying scales are unified into a single scale using an upsample layer: $upsample(\{atten_i, i = 1, \ldots, 3\})$. After unifying the attention gates' output to one scale, a 1×1 convolutional layer is used to standardize the number of filters. Subsequently, concatenation is performed between the outputs of the four attention gates (uniformed to one scale), resulting in a concatenated vector. The concatenated vector undergoes a 1×1 layer to downsample the number of filters and extract the most important features. Following the downsampling, an addition operation is performed between this result and the output of the final *MLR-block* in the decoder path. A ReLU activation function is applied to the output of the addition operation, transforming the features into another latent space and capturing non-linear relationships. Finally, the features pass through two 3×3 convolutional layers with a residual path to generate the final segmentation feature maps of the model.

IV. RESULTS

We used four publicly available biomedical imaging datasets to evaluate our model; KvasirSEG [31], CVC-ClinicDB [30], 2018 Data Science Bowl [32], and ISIC-2018 Challenge [33], [34]. Each of these datasets has a different number of images with varying sizes, consisting of the images and their corresponding ground truth masks. Firstly, we resize the images and their masks to 256×256 pixels, then divide

TABLE 3. Quantitative results on the experimented datasets CVC-ClinicDB, Bowl-2018 and KvasirSeg.

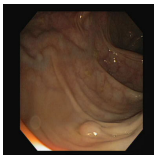
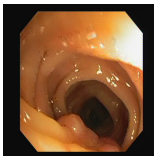
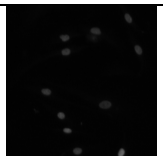
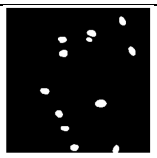
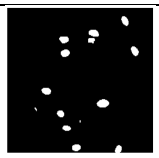
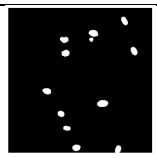
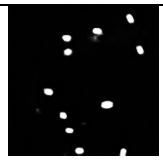
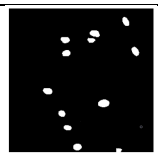
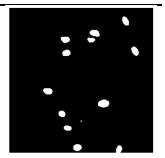
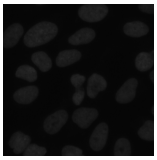
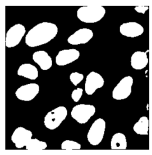
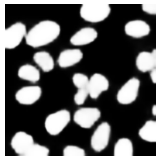
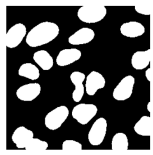
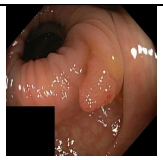
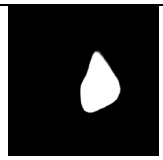
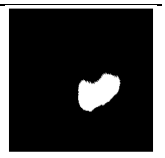
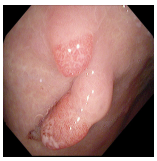
Dataset	Input Image	Ground Truth	MultiReUnet	ReUnet++	Pranet	DCSAU-Net	PolyRes-Net
CVC-ClinicDB							
							
BOWL-2018							
							
KvasirSEG							
							

TABLE 4. Quantitative results on the experimented dataset ISIC-2018.

Dataset	Input Image	Ground Truth	MultiReUnet	ReUnet++	Pranet	MS-Net	PolyRes-Net
ISIC-2018							

each dataset into 80% for training and 20% for testing except ISIC-2018 dataset it's already divided into train and test. The selected datasets are well-known and are commonly used in biomedical image segmentation tasks. We have chosen diverse imaging modalities datasets to evaluate the performance and robustness of the proposed method.

Further preprocessing was applied to optimize the data for use with the PolyRes-Net model. The pixel values were normalized to a range of [0, 1], which facilitated faster convergence during training. To enhance the model's ability to generalize and reduce the risk of overfitting, various data augmentation techniques were used, including random

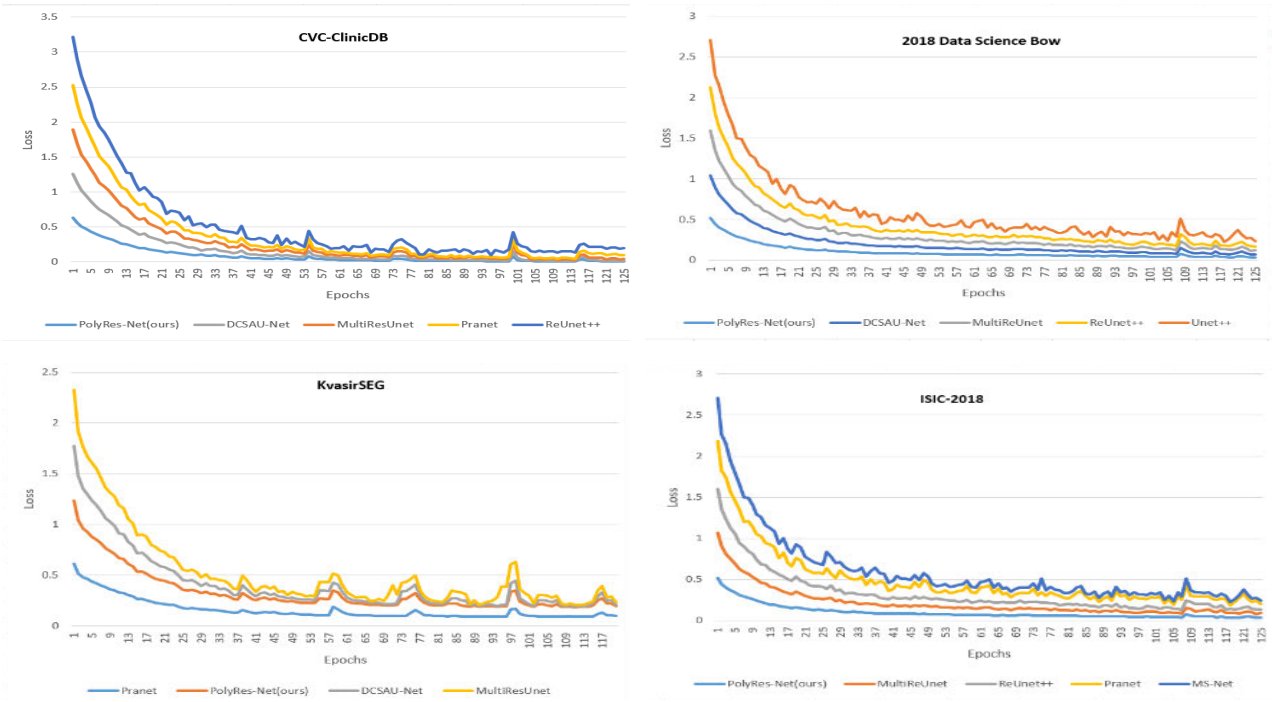


FIGURE 5. Comparison graph based on loss for the experimented datasets CVC-ClinicDB, Bowl-2018, KvasirSeg and ISIC-2018.

rotations, flips, cropping, zooming, and the addition of Gaussian noise. These techniques helped to simulate different conditions and perspectives, making the model more robust. In datasets with class imbalances, a weighted loss function was utilized to ensure that the model focused adequately on less represented classes. Additionally, image enhancement methods like histogram equalization and CLAHE were applied as needed, particularly in cases where medical images had low contrast, to improve the clarity of key features. These preprocessing actions were essential in preparing the data, ultimately contributing to the strong performance of the model across different datasets.

PolyRes-Net implemented using pytorch running on tesla t4 GPU, trained using binary cross entropy (BCE) loss function and evaluated using some of standard computer vision metrics for medical image segmentation such as Mean Dice Coefficient (mDSC), Mean Intersection over Union (mIOU), recall and precision.

$$BCE(y, \hat{y}) = -[y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})] \quad (5)$$

$$mDSC = \frac{1}{N} \sum_{1}^{N} \frac{2 \times TP}{2 \times TP + FP + FN} \quad (6)$$

$$mIoU = \frac{1}{N} \sum_{1}^{N} \frac{TP}{TP + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

where TP, TN, FP, and FN represent true-positive, true-negative, false-positive, and false-negative respectively, for each sample in dataset while N represent number of samples in dataset.

Model trained for 150 epoch using *adam* optimizer [36] and *cosine annealing warm restarts* scheduler [37] with initial learning rate is 0.001, etamin is 0.0008, T0 is 8 and Tmult is 1. The *cosine annealing warm restarts* scheduler consists of two parts: *cosine annealing* and *warm restarts*. The term *cosine annealing* refers to the use of the cosine function as the learning rate annealing function, while *warm restarts* involve periodically resetting the learning rate to the same or higher value to escape local minima and encourage exploration during training. In practical applications, the cosine function has been demonstrated to outperform alternatives such as simple linear annealing. *Warm restarts* is the interesting part: it means that every so often, the learning rate is restated, e.g. re-raised back up and you can see it's impact on our model From Fig.6. Our model scheduler configured with an initial Learning Rate (LR) of 0.001 decreasing it until 0.0008 and warm restart again after eight epochs to start again from 0.001.

Results on CVC-ClinicDB: For the CVC-ClinicDB dataset as shown in Tables 3, 5 and Fig. 5, MLR-Net outperforms all SOTA methods reporting the highest mIoU and mDSC of 0.8560 and 0.9102, respectively. Our method outperformed the most competitive DCSAU-Net [24] with a mIoU of 0.30% and mDSC of 0.40%. This improved performance is due to MLR-Net's unique architecture, which makes use of the MLR-block's dual-branch structure to efficiently extract

TABLE 5. Quantitative results on the experimented datasets.

	mIOU	mDSC	Recall	Precision
Dataset: CVC-ClinicDB [30]				
Unet [4]	0.7537	0.8453	0.8620	0.8587
Unet++ [9]	0.7737	0.8553	0.8720	0.8687
ReUnet++ [12]	0.8037	0.8753	0.8730	0.8725
Pranet [18]	0.8510	0.9095	0.9025	**0.9095**
MultiResUnet [6]	0.8497	0.9010	0.9012	0.9085
DCSAU-Net [24]	0.8530	0.9063	0.9020	0.9090
PolyRes-Net(ours)	**0.8560**	**0.9102**	**0.9077**	0.9080
Dataset: KvasirSEG [31]				
Unet [4]	0.7037	0.7855	0.7115	0.7089
Unet++ [9]	0.6337	0.7453	0.6560	0.6595
ReUnet++ [12]	0.6517	0.7472	0.6720	0.6790
Pranet [18]	**0.8330**	**0.8920**	0.9025	**0.9095**
MultiReUnet [6]	0.7102	0.7905	0.7150	0.7166
DCSAU-Net [24]	0.7420	0.8136	0.7789	0.7850
PolyRes-Net(ours)	0.7560	0.8258	0.7887	0.7899
Dataset: 2018 Data Science Bowl [32]				
Unet [4]	0.8310	0.8995	0.8500	0.8515
Unet++ [9]	0.8127	0.8451	0.8286	0.8297
ReUnet++ [12]	0.8384	0.9078	0.8595	0.8645
Pranet [18]	0.6880	0.7970	0.7762	0.7699
MultiReUnet [6]	0.8352	0.9063	0.8525	0.8595
DCSAU-Net [24]	0.8509	0.9130	0.8908	0.9010
PolyRes-Net(ours)	**0.8532**	**0.9180**	**0.8977**	**0.9091**
Dataset: ISIC-2018 Challenge [33], [34]				
Unet [4]	0.7812	0.8575	0.7998	0.8020
Unet++ [9]	0.7237	0.8059	0.7775	0.7789
ReUnet++ [12]	0.8089	0.8817	0.8200	0.8195
MS-Net [35]	0.7809	0.8630	0.8105	0.8010
Pranet [18]	0.8060	0.8850	0.8149	0.8199
MultiReUnet [6]	0.8119	0.8830	0.8228	0.8298
PolyRes-Net(ours)	**0.8214**	**0.8925**	**0.8578**	**0.8681**

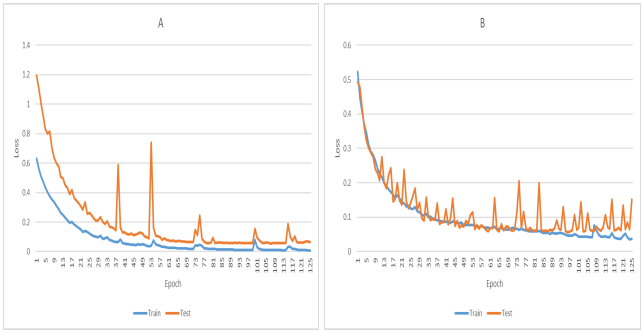


FIGURE 6. Learning curve of PolyRes-Net loss according epochs, A) with CVC ClinicDB dataset. B) with 2018 Data Science Bowl dataset.

a variety of relevant characteristics that facilitate more precise segmentation.

Results on KvasirSEG: As shown in Tables 3, 5 and Fig. 5, PraNet [18] outperforms MLR-Net reporting the highest mIoU and mDSC of 0.8330 and 0.8920, respectively. This result demonstrates that PraNet might have a more effective attention mechanism compared to MLR-Net. However, MLR-Net still receives competitive performance on this dataset which highlights its robustness across different datasets.

Results on 2018 Data Science Bowl: As shown in Tables 3, 5 and Fig. 5, MLR-Net outperforms all SOTA methods reporting the highest mIoU and mDSC of 0.8532 and 0.9180, respectively. Our method outperformed the most competitive DCSAU-Net [24] with a mIoU of 0.20% and mDSC of 0.50%. This enhancement may be attributed to MLR-Net's multi-scale feature aggregation and efficient feature extraction skills, which allow it to extract both fine-grained features and global context from the images.

Results on ISIC-2018 Challenge: As shown in Tables 4, 5 and Fig. 5, MLR-Net outperforms all SOTA methods reporting the highest mIoU and mDSC of 0.8214 and 0.8925, respectively. Our method outperformed the most competitive MultiReUnet [6] with a mIoU of 1.00% and mDSC of 0.75%. Because MLR-Net integrates skip connections, attention mechanisms, and multi-scale feature aggregation effectively, it can collect both local and global information in the images, producing significantly better segmentation results.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduce *PolyRes-Network*, which adopts a U-shape architecture featuring encoder-decoder branches. The *MLR-block*, is the core component and it is employed within these branches to extract features. It leverages a dual-branch structure, employing distinct architectures of residual layers for each branch. This unique design facilitates the extraction of different shapes of important features by allowing each branch to focus on capturing specific characteristics. The model employs skip connections to concatenate feature maps from various encoder and decoder layers. This integration allows the combination of both

low-level and high-level features, facilitating the network's ability to capture detailed information and global context. An attention mechanism is implemented in the decoder branch, where each *MLR-block* receives its input from the output of attention between the skip connection and the previous *MLR-block*. The model constructs its final feature segmentation output using the multi-scale feature aggregation (MSFA) layer. The MSFA layer aggregates the outputs of attention layers at different scales, enhancing the accuracy of the segmentation feature maps.

The experimental results demonstrate the promising performance of our proposed method, surpassing several commonly used methods. However, challenges arise in achieving optimal results with the KvasirSEG dataset. The results highlight the superior performance of the proposed *PolyRes-Net* compared to state-of-the-art segmentation methods. Specifically, *PolyRes-Net* achieves the highest *DSC* scores of 91.02%, 91.80%, and 89.25% on CVC ClinicDB, 2018 Data Science Bowl, and ISIC-2018 skin lesion segmentation challenge dataset, respectively. Additionally, the highest *JC* scores are 85.60%, 85.32%, and 82.14% for the same datasets, further underscoring the efficacy of the proposed model.

We identify opportunities for improvement, particularly in the (MSFA) layer with different aggregation algorithms. We believe changes to the skip connection construction, incorporating CNN layers to bridge the gap between encoder and decoder branch features. Additionally, considering the use of a transformer in the attention layer may yield more accurate results than CNN. Furthermore, we believe that exploring the integration of a backbone model may prove beneficial, providing an avenue to further improve the model's segmentation performance, especially when dealing with diverse modalities of medical images.

REFERENCES

[1] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101693.

[2] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[5] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[6] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[7] S. A. Hicks, S. Eskeland, M. Lux, T. de Lange, K. R. Randel, M. Jeppsson, K. Pogorelov, P. Halvorsen, and M. Riegler, "Mimir: An automatic reporting and reasoning system for deep learning based analysis in the medical domain," in *Proc. 9th ACM Multimedia Syst. Conf.*, Jun. 2018, pp. 369–374.

[8] I. Qureshi, J. Yan, Q. Abbas, K. Shaheed, A. B. Riaz, A. Wahid, M. W. J. Khan, and P. Szczuko, "Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends," *Inf. Fusion*, vol. 90, pp. 316–352, Feb. 2023.

[9] 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. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

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

[11] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[12] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2019, pp. 225–2255.

[13] X. Chen, R. Zhang, and P. Yan, "Feature fusion encoder decoder network for automatic liver lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 430–433.

[14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[15] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, vol. 11045, Granada, Spain. Springer, 2018, pp. 3–11.

[16] C. Li, Y. Tan, W. Chen, X. Luo, Y. Gao, X. Jia, and Z. Wang, "Attention UNet++: A nested attention-aware U-Net for liver CT image segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 345–349.

[17] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "DoubleU-Net: A deep convolutional neural network for medical image segmentation," in *Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2020, pp. 558–564.

[18] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "PraNet: Parallel reverse attention network for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, Lima, Peru. Springer, 2020, pp. 263–273.

[19] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[20] Y. Shen, X. Jia, and M. Q-H. Meng, "HRENet: A hard region enhancement network for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Strasbourg, France. Springer, 2021, pp. 559–568.

[21] A. Srivastava, D. Jha, S. Chanda, U. Pal, H. D. Johansen, D. Johansen, M. A. Riegler, S. Ali, and P. Halvorsen, "MSRF-Net: A multi-scale residual fusion network for biomedical image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2252–2263, May 2022.

[22] J. Zhong, W. Wang, H. Wu, Z. Wen, and J. Qin, "PolypSeg: An efficient context-aware network for polyp segmentation from colonoscopy videos," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Lima, Peru. Springer, Oct. 2020, pp. 285–294.

[23] N. K. Tomar, D. Jha, U. Bagci, and S. Ali, "TGANet: Text-guided attention for improved polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, Sep. 2022, pp. 151–160.

[24] Q. Xu, Z. Ma, N. He, and W. Duan, "DCSAU-net: A deeper and more compact split-attention U-Net for medical image segmentation," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106626.

[25] K. Wang, X. Zhang, X. Zhang, Y. Lu, S. Huang, and D. Yang, "EANet: Iterative edge attention network for medical image segmentation," *Pattern Recognit.*, vol. 12, 2022, Art. no. 108636.

[26] D. Lin, D. Shen, S. Shen, Y. Ji, D. Lischinski, D. Cohen-Or, and H. Huang, "ZigZagNet: Fusing top-down and bottom-up context for object segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 7482–7491.

[27] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[28] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3684–3692.
[29] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.
[30] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.
[31] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-SEG: A segmented polyp dataset," in *Proc. Int. Conf. Multimedia Model.*, Daejeon, South Korea. Springer, 2020, pp. 451–462.
[32] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, and A. E. Carpenter, "Nucleus segmentation across imaging experiments: The 2018 data science bowl," *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, Dec. 2019.
[33] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.
[34] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1–9, Aug. 2018.
[35] B. Zhang, Y. Wang, C. Ding, Z. Deng, L. Li, Z. Qin, Z. Ding, L. Bian, and C. Yang, "Multi-scale feature pyramid fusion network for medical image segmentation," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 18, no. 2, pp. 353–365, Aug. 2022.
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[37] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 2016, *arXiv:1608.03983*.

AMR MAGDY received the bachelor's degree in computer science from the Faculty of Computers and Information, Assiut University, in 2010, and the master's degree in steganalysis from Assiut University, in 2018. His current research interests include deep learning and image processing.

KHALID N. ISMAIL is a Senior Lecturer in computer science with Birmingham City University, where he is also the Course Director for the B.Sc. Computer and Data Science and the B.Sc. Computer Science with AI Programs. Before joining Birmingham City University, in 2021, he was a Senior Research Associate with Durham University and held a postdoctoral research fellow position with the University of Nottingham, the University of Bath, and the University of Leeds. Currently, he supervises several AI-based intelligent project developments and actively participates in industry-based collaborative projects. His primary research interests include artificial intelligence, computer vision, advanced machine learning, image processing, and deep learning, with a focus on solving complex real-world challenges.

MARGHNY H. MOHAMED received the B.Sc. degree in mathematics and the M.Sc. degree in the computer science from the Faculty of Science, Assiut University, Egypt, in 1988 and 1993, respectively, and the PhD. degree in computer science from Kyushu University, in 2001. He is a full-time Professor with Egypt-Japan University of Science and Technology (E-JUST), Egypt; and a Professor with the Computer Science Department, Faculty of Computers and Information, Assiut University. He is the former Vice Dean of Education and Student Affairs; and the Community Service and Environmental Development, Faculty of Computers and Information, Assiut University. He has been working for more than 33 years in the field of teaching, scientific research, and administration. He has supervised about 100 master's and doctoral dissertations for national and international students. He has published about 120 research papers in international journals and conferences. He judged many theses, research papers, and promotions. He was a principal investigator of many advanced research and applied projects. His areas of specialization and research interests include data mining, text mining, information retrieval, web mining, machine learning, natural language processing, artificial intelligence, pattern recognition, neural networks, information security, evolutionary computation, and fuzzy systems. He is a member of the Egyptian Mathematics Society.

MAHMOUD HASSABALLAH received the B.Sc. degree in mathematics and the M.Sc. degree in computer science from South Valley University, Egypt, in 1997 and 2003, respectively, and the D.Eng. degree in computer science from Ehime University, Japan, in 2011. He was a Visiting Scholar with the Department of Computer and Communication Science, Wakayama University, Japan, in 2013; and the GREAH Laboratory, Le Havre Normandie University, France, in 2019. He is currently a Professor of computer science with the Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Saudi Arabia. Also, he is a Full Professor of computer science with the Department of Computer Science, Faculty of Computers and Information, South Valley University. He has published five books and over 100 research papers in refereed international journals and conferences. His research interests include human-centred artificial intelligence, machine learning, computer vision, biometrics, image processing, feature extraction, object detection/recognition, and data security. He is a TPC Member of many conferences, such as ICIAR 2020, IVPAI 2020, AICS 2020, AICS 2021, IVPAI 2021, AVSS 2022, PRIS 2023, ICFIP 2023, ICCCV2024, CCVPR2024, and CIPCV 2024. He serves as a Reviewer for several journals, such as IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Neurocomputing*, *Pattern Recognition*, *Pattern Recognition Letters*, *IET Image Processing*, *IET Computer Vision*, *IET Biometrics*, *Journal of Real-Time Image Processing*, *The Computer Journal*, *Journal of Electronic Imaging*, and *Optical Engineering*. He is an editorial board member of *Pattern Analysis and Applications*, *Journal of Real-Time Image Processing*, *IET Image Processing*, *The Imaging Science Journal*, and *Wireless Communications and Mobile Computing*.

HAITHAM MAHMOUD received the dual B.Sc. and M.Sc. degrees from the Arab Academy for Science and Technology, British University in Egypt (BUE) and Loughborough University, and the Ph.D. degree in electrical engineering from Birmingham City University (BCU), in 2022. He was a Marie-Curie Fellow on the IoT4win Project, BCU, from 2018 to 2021. As a Research Fellow with the Future Communication Research Cluster, he is actively coordinating research activities and overseeing major projects in BCU, including intelligent B5G/6G radio access networks, radio access network resource optimisation, cognitive radio networks, massive MIMO, and future open networks. His research has been supported by H2020, U.K. DSIT, UKRI, NGI, BC, and others in Egypt (STDF and NTI). His main research interests and expertise span a broad range of areas in intelligent networks and utilising AI on digital systems. He has been the Work Package Leader of H2020 IoT4win, an Advisory Board Member of DSIT 5G-Regions, and a Co-PI of BC TSNE with Egypt and AKT. He was selected to be a Reviewer of BC ISPF call 2024. He is a Leading Editor of SI in ACM.

MOHAMMED M. ABDELSAMEA received the Ph.D. (with Doctor Europaeus) degree in computer science and engineering from the IMT-Institute for Advanced Studies, Lucca, Italy, in 2015. He is currently a Senior Lecturer in computer science (machine learning and computer vision) with the Computer Science Department, University of Exeter; and a fellow of the British Higher Education Academy (HEA). Before joining Exeter University, he was a Senior Lecturer in data and information science with the School of Computing and Digital Technology, Birmingham City University, where he was also a Leading Member of the Computer Vision Theme within the DAAI Research Group. He has been working on developing artificial intelligence tools for healthcare and life science applications, since 2005. He has been with several groups of biologists, pathologists, and computer scientists in different research projects in Egypt, Singapore, Italy, and the U.K. He was with the School of Computer Science, Nottingham University; Mechanochemical Cell Biology, Warwick University; Nottingham Molecular Pathology Node (NMPN) and the Division of Cancer and Stem Cells, Nottingham Medical School, as a Research Fellow. In 2016, he received the prestigious Marie-Cuire Research Fellowship from the School of Computer Science, Nottingham University. His current research interests include the development of statistical/classical machine learning and deep learning approaches in the areas of medical image segmentation, medical image classification, semantic/instance segmentation, feature extraction, and data mining and machine learning, with the overall ambition to assist human investigation.