Mathematical Biosciences and Engineering

MBE, 20(7): 12039–12055.
DOI: 10.3934/mbe.2023535
Received: 08 March 2023
Revised: 27 April 2023
Accepted: 05 May 2023
Published: 15 May 2023

http://www.aimspress.com/journal/MBE

Research article

FM-Unet: Biomedical image segmentation based on feedback mechanism Unet

Lei Yuan[1], Jianhua Song[1,2,*] and Yazhuo Fan[2]

[1] The Key Laboratory of Intelligent Optimization and Information Processing, Minnan Normal University, Zhangzhou 363000, China
[2] College of Physics and Information Engineering, Minnan Normal University, Zhangzhou 363000, China

* **Correspondence:** Email: songjianhua@mnnu.edu.cn; Tel: +8605962591443.

Abstract: With the development of deep learning, medical image segmentation technology has made significant progress in the field of computer vision. The Unet is a pioneering work, and many researchers have conducted further research based on this architecture. However, we found that most of these architectures are improvements in the backward propagation and integration of the network, and few changes are made to the forward propagation and information integration of the network. Therefore, we propose a feedback mechanism Unet (FM-Unet) model, which adds feedback paths to the encoder and decoder paths of the network, respectively, to help the network fuse the information of the next step in the current encoder and decoder. The problem of encoder information loss and decoder information shortage can be well solved. The proposed model has more moderate network parameters, and the simultaneous multi-node information fusion can alleviate the gradient disappearance. We have conducted experiments on two public datasets, and the results show that FM-Unet achieves satisfactory results.

Keywords: deep learning; medical image segmentation; convolution neural network; Unet; feedback path

1. Introduction

In recent years, with the improvement of computer hardware performance, deep learning (DL) has been applied in many industrial fields and has demonstrated excellent performance. One of the most important areas is the application of medical image segmentation and classification [1–3].

Compared with traditional machine learning and computer vision methods, DL has more significant advantages in segmentation accuracy [4–6].

Over the past decade, medical image segmentation technology based on deep learning has focused on developing efficient and robust segmentation methods [7]. Unet is a milestone work [8]. It establishes an encoder-decoder convolutional network structure with a skip connection, which is simple and efficient for medical image segmentation with small required datasets. In recent years, the Unet-like structures have become the backbone of almost all leading medical image segmentation methods. Following Unet, many important extension networks have emerged, such as Unet++ [9], Res-Unet [10], AttentionUnet [11] and Trans-Unet [12].

Unet draws on the experience of the Fully convolutional network (FCN), and its network structure consists of two parts. The shrinking network on the left side captures the contextual information in the image, and the extended network on the right side achieves the purpose of accurate positioning of the required segmentation part of the image. Unet also uses skip connection for feature fusion, which combines the down-sampling features of the first half and the up-sampling features of the second half to obtain more accurate context information and a better segmentation effect.

Zhou et al. proposed Unet++ [9], which is composed of a set of different depths Unet and decoders. These decoders are connected intensively with the same resolution through redesigned skip connections. Despite the improved performance, the Unet++ model is very complex, requires additional learnable parameters, and some of its components are redundant for specific tasks [13]. Inspired by the deep residual network (ResNet) [14] and Unet, Zhang et al. proposed the deep residual Unet (Res-Unet) [10]. Res-Unet is still based on the Unet architecture. A series of stacked residual units replace ordinary neural units as basic blocks to build deep Res-Unet, which effectively deepens the number of network training layers. But with the increase of network depth, the training time becomes very long. Researchers also consider introducing self-attention mechanisms into CNNs to improve network performance [15–17]. Ozan Oktay et al. integrated the skip connection of additional focus gates into the U-shaped structure for medical image segmentation [11]. The attention gates (AGs) mechanism implicitly generates soft region suggestions, highlighting salient features useful for specific tasks. The sensitivity and accuracy of the model for dense label prediction are improved by suppressing the features of irrelevant regions. Some researchers have attempted to minimize interference from extraneous regions by preprocessing data prior to network training. Rani et al. [18] discovered that bone structures in chest X-rays could interfere with feature extraction from lung regions, thereby reducing the accuracy of models in detecting, localizing, and visualizing infections during COVID-19 screening. To improve the overall accuracy of their model, they applied bone suppression and lung segmentation preprocessing methods. By preprocessing, the model can minimize the visibility of bones within the lung region while preserving maximum spatial information and resolution [19]. Subsequently, Rani et al. [20] used data augmentation, histogram equalization, and pre-segmentation of L1 vertebrae to calculate the vertebral center as a reference for kidney and ureter localization. The proposed KUB-UNet network was used to verify the effectiveness of this method in enhancing the segmentation of urinary organs in KUB X-ray images.

Currently, Transformer, designed for the sequence-to-sequence prediction, has become an alternative architecture with a global self-attention mechanism [21–25]. Chen et al. proposed Trans-Unet [12], which has the advantages of both Transformers and Unet. On the one hand, the transformer encoders the tokenized image blocks from the CNN feature map into an input sequence to extract the global context. On the other hand, the decoder up-samples the encoded feature and then combines

them with a high-resolution CNN feature map for accurate positioning.

The encoder-decoder structure and the skip connection of Unet have been proved to be efficient and stable network structures [26–29]. As mentioned above, many novel network structures based on Unet structure have been proposed. However, these are improvements proposed on the backward propagation and fusion of the network, with few changes on the forward propagation of the network and forward fusion of information. In order to improve receptive field and pixel level prediction in Unet network, there must be a series of up-sampling and down-sampling operations. However, these operations will inevitably cause information loss and underutilization. Based on this disadvantage, we design a feedback mechanism Unet (FM-Unet) in this paper, adding a feedback path to the encoder and decoder paths of the network to help the network integrate the following step information in the current encoder and decoder. The main contributions of this paper are summarized as follows:

1) A feedback mechanism Unet model for semantic segmentation of medical images is proposed. A feedback path is introduced into the Unet, which integrates the context information of the convolutional blocks and can compensate for the loss of information to improve the segmentation accuracy.

2) Compared with most of the improved networks based on Unet, the FM-Unet model has smaller parameters, which can reduce the cost of computing time and space to a certain extent.

3) In FM-Unet, the concatenation of the feedback path context feature map, the concatenation of the encoder-decoder primary path feature map and the feedback path feature map, and the concatenation of the same node at different time points can better fuse information at each scale and alleviate the problem of gradient disappearance.

The rest of this paper is organized as follows: Section 2 reviews Unet-based segmentation networks and related techniques. Section 3 describes the proposed method. Section 4 gives the experimental results and analysis. Finally, a summary of the proposed model is presented.

2. Related work

2.1. Unet architecture

Unet, whose network structure is shaped like the letter 'U', is composed of convolution, down-sampling, up-sampling, skip connection and other operations, including the down-sampling contraction path on the left and the up-sampling expansion path on the right. Unet contraction path extracts image semantic information, reduces image resolution and expands the receptive field. The network consists of five blocks, each containing two 3 × 3 convolutional layers. After each convolution layer, there is a RELU activation function and a down-sampling operation for the input of the next block. The expansion path predicts pixel by pixel, accurately locates the target position, and restores the image to the size similar to the input image. The extension path contains four blocks, and each block also contains 3 × 3 convolution layers, RELU and up-sampling operation. The skip connection is added between the contraction path and the expansion path. Unet uses concatenation to the crop feature map of the contraction path to the same size as the expansion path and then performs the concatenation operation, which can help the network learn some details lost before the contraction path.

2.2. Residual idea

He et al. proposed the residual network [14], which introduces a constant mapping design to solve

the degradation and gradient disappearance problem in multilayer neural networks. For a stacking layer structure, when the input is x, the learned feature is $H(X)$, and the residual network function can be obtained as $F(X) = H(X) - X$. Currently, the learning feature can be expressed as $H(X) = F(X) + X$. In this way, the optimal solution of the network can be obtained by adjusting the residual function $F(X)$, which is easier than directly learning the original feature $H(X)$. The residual structure in the residual network uses a shortcut connection method, which can also be understood as a quick connection channel so that the feature matrix is added by the interlayers. It is important to note that the $F(X)$ and X shapes should be the same, so the input X is often dimensioned with a 1×1 convolution kernel on the Shortcut path. And here is done by adding the numbers in the same position of the feature matrix.

2.3. Dense structure

Huang et al. proposed the DenseNet model [30], its basic idea is consistent with ResNet, but it establishes a dense connection between all the preceding layers and the following layers, and its name comes from this. Another major feature of DenseNet is feature map reuse through the connecting features on the channel. These features enable DenseNet to achieve better performance than ResNet with fewer parameters and computing costs. Compared with ResNet, DenseNet proposes a more radical mechanism of dense connectivity: connecting all layers. In DenseNet, each layer is connected with all previous layers in the channel dimension (the size of the feature map of each layer is the same here) and used as the input of the next layer.

The output of traditional network at the layer l is $x_l = H_l(x_l\text{-}1)$, that is, only the previous layer is used as input. In DenseNet, all previous layers will be connected as input $x_l = H_l([x_0, x_1, x_2, \cdots, x_{l\text{-}1}])$, which $x_l = H_l(\cdot)$ represents a nonlinear transformation function, and it is a combined operation and includes a series of Batch Normalization (BN), ReLU, Pooling and Conv operations.

2.4. Feedback mechanism

The concept of feedback in cybernetics involves adjusting the input based on the change in the output. Researchers have incorporated this idea into Unet networks to enhance the accuracy of medical image segmentation. Shibuya et al. [31] proposed a Feedback U-Net with convolutional LSTM, where the second round of input is fed back through the first round of Unet output, and convolutional LSTM is used to extract features based on those obtained in the first round. However, this approach only implements feedback between two stages. This kind of long-distance feedback inevitably has a semantic gap, limiting the transfer of features learned in the first stage to the second stage. Lin et al. [32] proposed Refine U-Net. In order to alleviate the semantic gap in the skip connection, the global refinement module of the middle layer was added to the Unet skip connection. To this end, the encoder output is progressively upsampled as feedback features, and they are fused with the corresponding decoder-side output features. However, this work only jump-connects the feedback feature map of the encoder to the decoder, and does not consider the feedback of the decoder information. Furthermore, these works seldom consider the feedback of encoder-inside and decoder-inside information.

2.5. Variants of Unet architecture

Zhou et al. proposed Unet++ [9], which is composed of a set of Unet with different depths and

decoders. These decoders are connected intensively with the same resolution through redesigned skip connections. Inspired by the Inception module for CNNs to achieve more efficient computing, Nabil et al. introduced MultiResUnet [33], an enhanced Unet architecture that uses a convolutional layer chain with residual connections, rather than simply connecting features from the encoder path to the decoder path. These residual connections not only reduce the semantic gap between encoder and decoder features, but also make learning easier while segmenting images from various modalities robustly at different scales. Valanarasu et al. argued that Unet has poor performance in detecting small anatomical structures with fuzzy noise boundaries and proposed Ki-Net [34], an overcomplete network architecture that can project data onto high dimensions. When combined with Unet, Ki-Net can capture details and perform target segmentation better than Unet. The introduction of the self-attentive mechanism also improves the performance of the network [35,36]. Oktay et al. integrated the skip connection of additional focus gates into the U-shaped structure for medical image segmentation [11]. At the same time, people are trying to combine CNNs and Transformer. Chen et al. combined Transformer and CNNs to form a powerful encoder for 2D medical image segmentation [12]. The complementarity of Transformer and CNN is used to improve the segmentation capability of the model.

Although these variants of Unet architecture show good results in biomedical image segmentation, the problem of information loss due to convolution operation in the encoder and decoder paths still exists, which affects the final segmentation accuracy.

3. Methods

3.1. Proposed network architecture

FM-Unet utilizes the classic U-shaped network as the basic network architecture. By introducing a feedback mechanism, the output feature map of the next convolution block is updated through the feedback convolution block, as shown in Figure 1(a). At the same time, the feedback output in the feedback path will also be output to the next feedback convolution block through down-sampling or up-sampling. The feedback mechanism of the coder-decoder designed in FM-Unet is shown in Figure 1(b),(c).

The mechanism integrates not only the output feature map of the next convolution block in the primary path, but also the output feature map of the last convolution block in the feedback path. The structure is well fused with the information of the context and less information can be lost. The complete structure of the FM-Unet includes a basic Unet architecture primary path (yellow block in Figure 1) and a feedback path (green block in Figure 1). In the encoder of Figure 1(b), layer I of the primary path receives $FX_{i\text{-}1\_0}$, down-sampling feature map, and the output X_{i_0} of this layer is obtained. At the same time, X_{i_0} down-sampling is used for the next layer operation to get X_{i+1_0} on the primary path. The feedback path convolution block receives the output X_{i+1_0} from the next layer of the primary path and the output $X_{i\text{-}1\_1}$ from the previous layer of the feedback path. According to a short skip connection, the feedback path output X_{i_1} is concatenated with the original primary path convolutional block output X_{i_0} in the channel dimension (green dashed line in Figure 1). This short skip connection can allow the concatenated two feature maps to have a smaller semantic gap and low training difficulty [37]. The feature map after feedback concatenation is the FX_{i_0}, the feedback output of the primary path will completely replace the previous X_{i_0}, which will be used for the next operation. The feedback structure of the decoder is shown in Figure 1(c). The basic implementation process is similar

to the feedback structure of the encoder. The detailed implementation diagram of the designed feedback mechanism can refer to Figures 2 and 3.

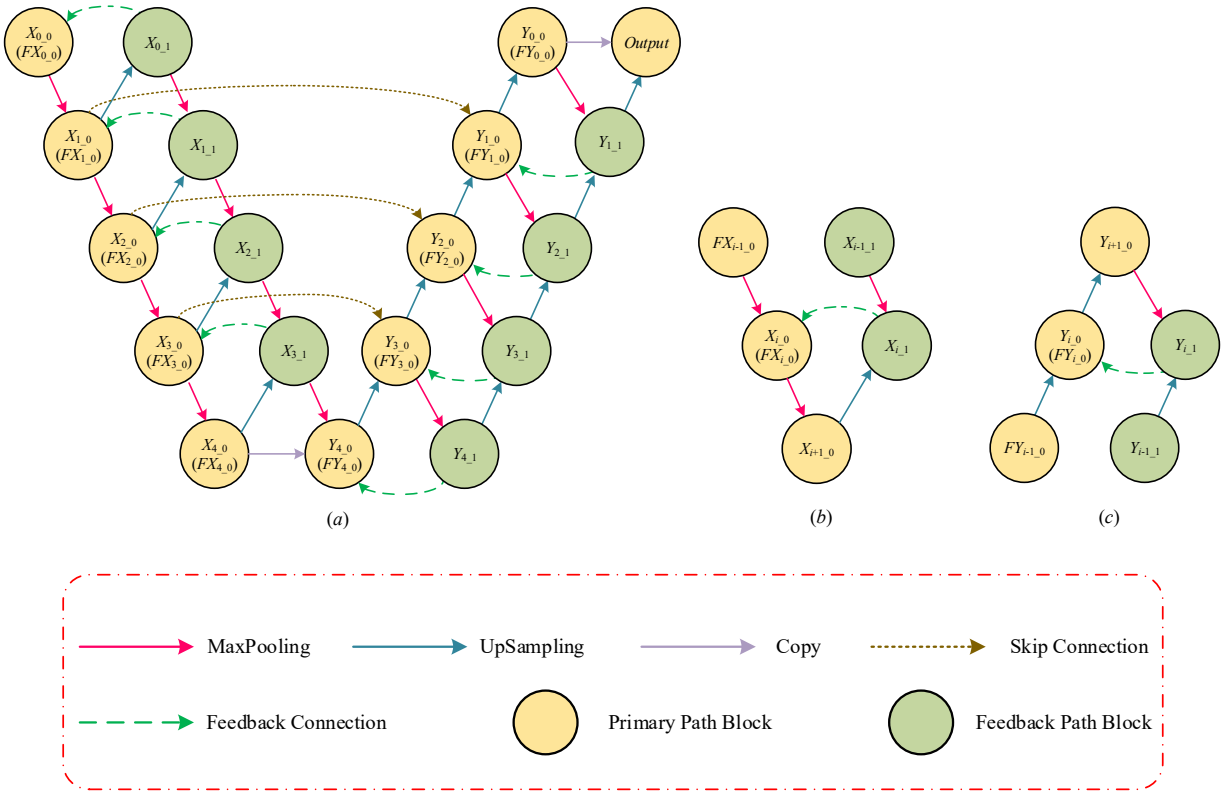


Figure 1. Schematic diagram of overall architecture and each module. (a) Architecture of FM-Unet, (b) Encoder feedback mechanism, (c) Decoder feedback mechanism.

3.2. Encoder feedback mechanism

In the encoder path, the detailed implementation diagram of the feedback mechanism is shown in Figure 2.

First, the primary path receives the input H×W×C feature map, which is obtained by 2 × 2 max polling of the previous convolution block $FX_{i\text{-}1\_0}$ (step 1). It should be noted that if the feature map is the initial input image, then the received input feature map is not obtained from the down-sampling of the previous module, it is the original input image with 1 channel or 3 channels. Every convolutional block consists of two convolutional layers and the corresponding activation function. At this node, we get the primary path output X_{i_0}. The output of the convolution block passes through 2 × 2 max polling as the input of the next layer in the primary path (step 2), the convolution block output of this layer will be up-sampling into the feedback path (step 3), The input to the feedback path may also come from a layer above the feedback path (step 4). This output of the feedback convolution block gives a feature map incorporating contextual information. The feedback feature map obtained here will be concatenated with the previously output feature map of the primary path of the current layer in the channel dimension (steps 5 and 6), then the concatenated feature map is used as input of the current primary path convolution block. The feature map from this convolution block is the updated feature map

of the primary path layer. We get the output FX_{i_0} after the primary path is updated. And this feature map will completely replace the previous output feature map of the layer for subsequent operations.

The Encoder Feedback Mechanism that we designed enables the first layer node in the encoder to perform feedback updates on its original output, X_{i_0}. The feedback output, X_{i_1}, is obtained by capturing the information X_{i+1_0} of the next layer in the primary path, and the information X_{i-1_1} of the upper layer in the feedback path at the feedback node. The feedback output, X_{i_1}, is concatenated with the original output, X_{i_0}, and updated after passing through the convolution block, culminating in the primary path returning the output FX_{i_0}. This operation achieves the shrinking network on the left that captures the image's contextual information, thus mitigating information loss.

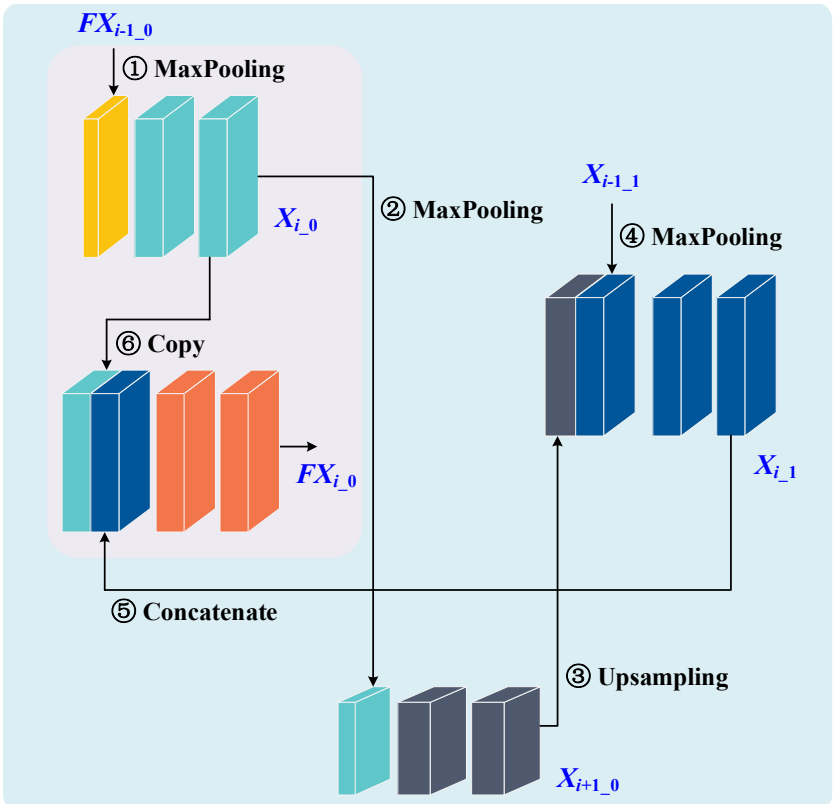


Figure 2. Detailed implementation diagram of encoder feedback mechanism.

3.3. Decoder feedback mechanism

In the decoder path, the detailed implementation diagram of the feedback mechanism is shown in Figure 3.

First, the primary path receives the bilinear interpolation up-sampling feature map of the decoder (step 1). Meanwhile, the feedback feature map after the decoder master path update is stitched by skip connection (step 2). The skip connection follows the standard Unet architecture. It concatenated two feature maps in the channel dimension. The output of this convolution block is up-sampled to the next layer of the decoder primary path (step 3), and this layer is similarly stitched together with the encoder primary path feature map (step 4) and the up-sampled feature map. The feature map in this layer will be down-sampled into the feedback path of the decoder (step 5), and the input in this feedback path

may also have the feature map from the previous layer of the decoder feedback path (step 6). The output of this feedback convolutional block will be connected to the primary path of the current layer via a short ship connection (step 7). Currently, the primary path needs to concatenate the feedback convolution block output feature map, the previous output feature map of the primary path of the current layer (step 8) and the skip connection encoder primary path feedback feature map (step 9). We get the output FY_{i_0} after the primary path is updated. The output of this input to the convolutional block is the current feedback feature map of that layer of the primary path and, like the encoder, completely replaces the previous output of that layer for subsequent operations.

We designed the Decoder feedback mechanism to enable the original output Y_{i_0} of the i-th layer node in the decoder to be updated with feedback. The feedback output Y_{i_1} is obtained by capturing the information Y_{i+1_0} of the next layer of the primary path, and the information $Y_{i\text{-}1\_1}$ of the upper layer of the feedback path in the feedback node. The feedback path output Y_{i_1}, the original output Y_{i_0} and the encoder primary path feedback output FX_{i_0} are concatenated after the convolution block to obtain the updated primary path feedback output FY_{i_0}. This operation allows the extended network on the right to accurately locate the segmented part of the image required and provide more detailed information.

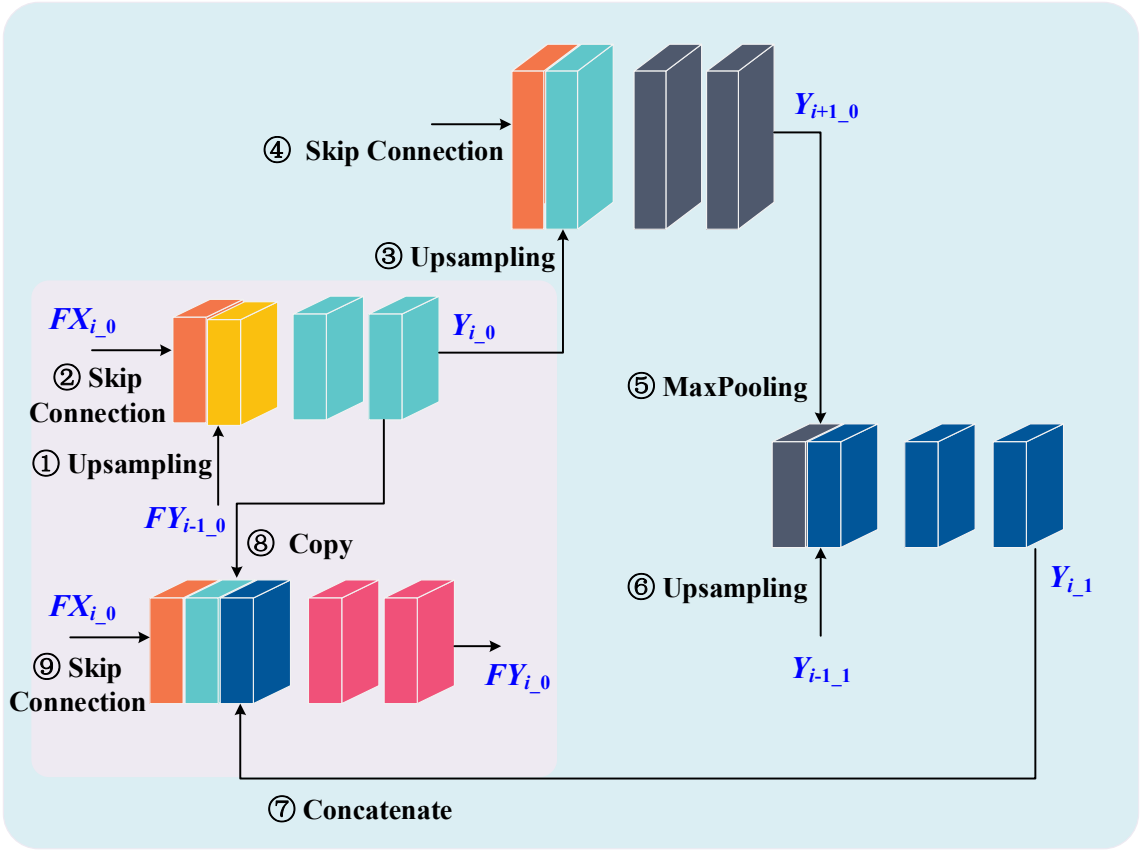


Figure 3. Detailed implementation diagram of decoder feedback mechanism.

4. Experiments

In this section, we conduct extensive experiments to evaluate the performance of the proposed image segmentation framework and compare it with the baseline model on several benchmark datasets.

4.1. Datasets

Breast Ultra Sound Images (BUSI) [38] is a medical images dataset of breast cancer by ultrasound scans. The BUSI dataset collected included breast ultrasound images of women aged 25 to 75 collected in 2018. The number of patients is 600 women. The dataset consists of 780 images, all of which are cropped to different sizes to remove unused and unimportant borders from the images. The images are in PNG format and divided into three categories: normal, benign and malignant. Each image has its ground truth (mask image). Both benign and malignant images are used in the experiments. To standardize network input sizes and take advantage of GPU parallelization, we resized 647 images to 256 × 256 RGB.

The International Skin Imaging Collaboration 2018 (ISIC 2018) [39] is the world's largest skin image analysis challenge and has organized the world's largest public dermoscope image library. This challenge was divided into three image analysis tasks: lesion segmentation, lesion attribute detection and disease classification. We performed only the lesion segmentation task. There are 2594 images in the dataset containing three different categories, including 20.0% melanoma, 72.0% nevus, and 8.0% seborrheic keratosis. The dataset consists of images of various resolutions, and we adapt all images to 512 × 512 RGB images for the same reason.

The STARE [40] dataset was first introduced by Michael Goldbaum in 1975 as a color fundus image database designed for retinal vessel segmentation. This dataset comprises 20 fundus images, among which 10 have lesions and the remaining 10 do not have any lesions. The images in the dataset have a resolution of 605 × 700. To avoid overfitting the model, we randomly cropped each picture to a size of 256 × 256 four times, and introduced random noise to the images. This approach not only enhances the number of datasets, but also satisfies standardized network input sizes and GPU parallel processing of data requirements.

4.2. Implementation details

All experiments were run on a Tesla P40 (24 GB) graphics card. FM-Unet is developed based on Python 3.8 framework using an SGD optimizer with learning rate of 0.001, momentum of 0.9, and weight decay of 0.0001. SGD optimizer was bound to a cosine annealing decay learning rate controller with a minimum learning rate of 0.00001 and a cosine function period (T_max) of one epoch. The batch size is set to 4 and a total of 100 epochs are performed. We split the dataset by a random factor according to the training set of 70% and validation set of 30%. In order to increase the diversity of training samples, the training model has a stronger generalization ability. We also expanded the data with random rotations and random adjustments of hue, brightness and cropping of the images.

FM-Unet uses a combination of binary cross-entropy and dice coefficients as the loss function for all of the above dataset training. The cross-entropy loss function is described as shown in Eq (1).

$$BCE(Y, \hat{Y}) = -\sum_{i=1}^{N} Y(x_i) \cdot log\hat{Y}(x_i) \tag{1}$$

where Y and $\hat{Y}$ are the true image label and the predicted image respectively, and x_i is the i-th pixel of the image. The calculation formula of the *Dice* coefficient is shown in Eq (2).

$$Dice = \frac{2 \cdot |Y \cap \hat{Y}|}{|Y| + |\hat{Y}|} \tag{2}$$

where $Y \cap \hat{Y}$ denotes the intersection of the sets Y and $\hat{Y}$, $|Y|$ and $\left|\hat{Y}\right|$ denote the number of their elements. For the segmentation task, Y and $\hat{Y}$ denote the Ground Truth (GT) and predict mask of the segmentation. The *DiceLoss* is *DiceLoss* = 1-*Dice*, which is calculated as shown in Eq (3).

$$DiceLoss(Y, \hat{Y}) = 1 - \frac{2 \cdot |Y \cap \hat{Y}|}{|Y| + |\hat{Y}|} \tag{3}$$

The combination of binary cross entropy and dice coefficients between Y and $\hat{Y}$ is used as the final loss function. The loss function is expressed as Eq (4).

$$L = 0.5BCE(Y, \hat{Y}) + DiceLoss(Y, \hat{Y}) \tag{4}$$

In order to evaluate the performance of the proposed framework relative to the baseline approach, we use F_1_score and Intersection-Over-Union (IOU) as evaluation metrics. F_1-score is a measure of classification problems. F_1-score is often used as the final measure in some multi-classification and binary classification problems. It is the harmonic average of the accuracy rate and recall rate, with a maximum of 1 and a minimum of F_1-score is defined as shown in Eq (5).

$$F_1 = 2 \cdot \frac{Y \cdot \hat{Y}}{Y + \hat{Y}} \tag{5}$$

The IOU is a standard performance measure for object segmentation problems, and its definition is shown in Eq (6). Given a set of images, IOU gives the similarity between the predicted regions and GT of the objects present in the set.

$$\text{IOU} = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} \tag{6}$$

where $|\cdot|$ denotes the base of the set. IOU is the area of overlapping between the predicted segmentation and GT divided by the joint area between the predicted segmentation and GT. For binary or multi-class segmentation, the average IOU is calculated by taking the IOUs of all classes and averaging them. The IOU index ranges from 0 to 1, where 1 represents a perfect match between GT and predicted segmentation, and 0 indicates a complete mismatch between them.

4.3. Comparisons with state-of-the-art methods

In this section, FM-Unet is compared with existing models to verify the effectiveness of the method. We choose the classical and more popular models for medical image segmentation: Unet [8], Unet++ [9], ResUnet [10], Attention Unet [11], TransUnet [12], MedFormer [41] and UNeXt [42].

4.3.1. Validation of model performance

Figure 4 shows the experimental results of FM-Unet and other models on the BUSI dataset, the parameter of loss function on the training set and the evaluation index on the validation set within 100 epochs are compared. We performed a random split in the BUSI dataset by training set 70% and validation set 30%. We also augmented the dataset with random rotation, adjustment of hue, brightness and cropping of the images.

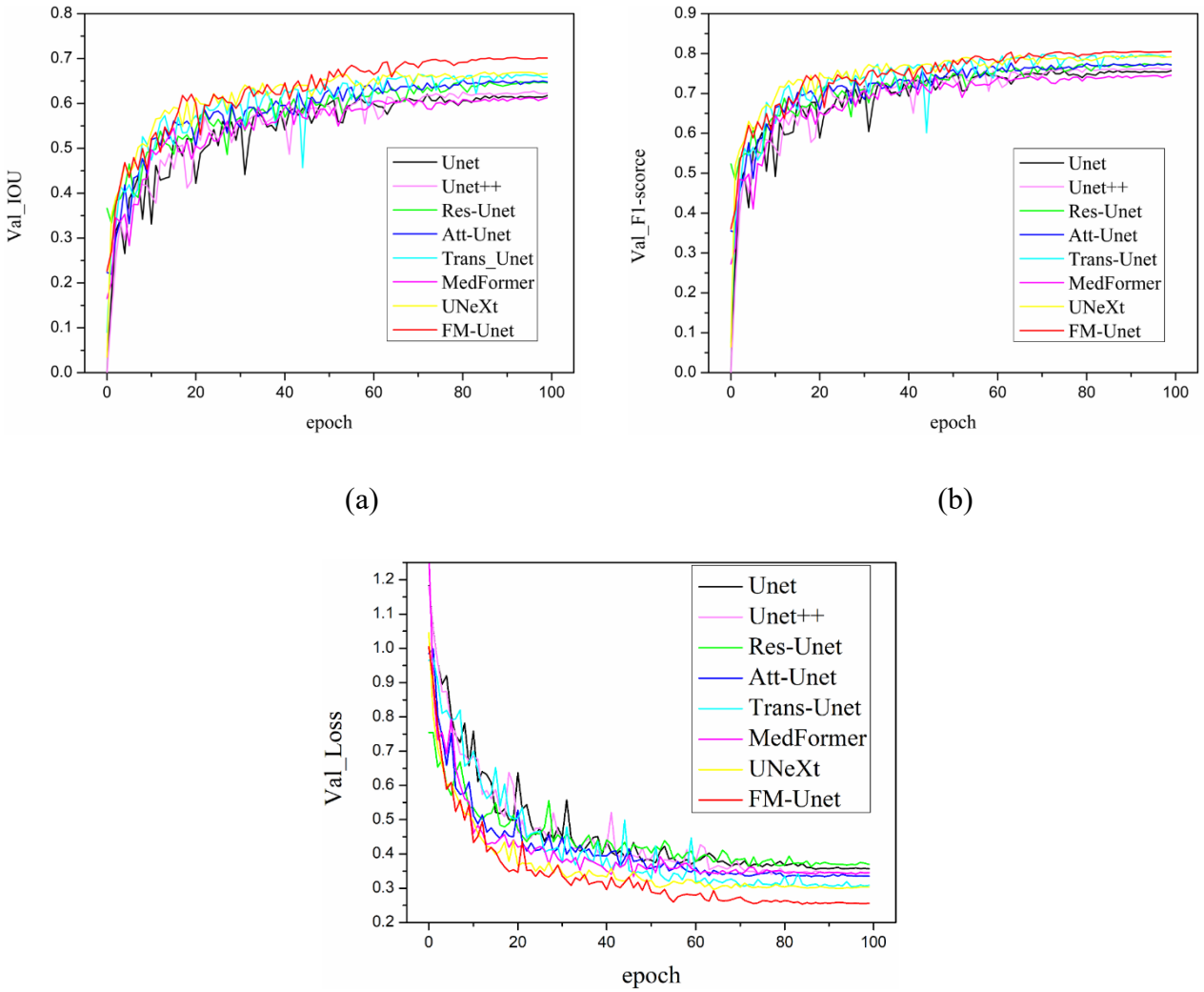


Figure 4. Demonstration of validation performance within 100 epochs: (a) IOU, (b) F_1-score and (c) DiceLoss. We record the value of the IOU and F_1-score on the validation set after each epoch.

Figure 4(a)–(c) show the changes in the loss function, IOU coefficients and F_1-score of each model on the validation set with the increase of iterations in 100 epochs. In Figure 4(a), we can see that the loss function of FM-Unet decreases the fastest and tends to be stable around 70 epochs. At the same time, in Figure 4(b),(c), it can be seen that FM-Unet performs best in both the IOU coefficient and F_1-score, and achieves good results in less iterations. FM-Unet has achieved a better IOU coefficient and F_1-score at 50 epochs. Compared with TransUnet and AttentionUnet, FM-Unet can achieve better segmentation results on a smaller number of iterations. Compared with Unet, Unet++ and ResUnet, IOU coefficients and F_1-score of FM-Unet are steadily increasing in the early training period, while the remaining networks have very large jumps in IOU coefficients and F_1-score, especially Unet network, which has the most drastic oscillation. It proves that FM-Unet is more adaptable, robust and effective, and the whole network design is more scientific.

4.3.2. Validation of model performance

Table 1 shows the Params data for FM-Unet, Unet and its improved architecture model, which shows the spatial complexity and relative computation time of the model. We obtained the data in Table 1 on an input tensor of 256 × 256 × 3. From Table 1, we can find that the Params of the Unet are about 31.13 MB. UNet++ halves the output channels of the Unet network and consolidates Unet structures of different sizes into one network. Its Params are about 9.16 MB, more than three times smaller than Unet. The Res-Unet and Att-Unet are also improvements based on Unet, but there are relatively large increases on Params, 62.74 MB and 51.99 MB, respectively. With the explosion of Transformers in the last year, Trans-Unet, a fusion of Transformers and Unet networks, has seen a huge increase in Params. Nowadays, most medical image segmentation research focuses on attention mechanisms and Transformers, but this neglects computation time and graphics capacity in the pursuit of better segmentation performance. FM-Unet has a little more than the Params of Unet, and the Params data is still in a small range. However, the params are still in the smaller range of 48.03 MB than the more popular attention-based and Transformers-related networks, 3.96 MB smaller than Att-Unet and 57.29 MB smaller than Trans-Unet.

Table 1. Comparison of model parameters.

Networks	Params (in MB)
Unet	31.13
Unet++	**9.16**
Res-Unet	62.74
Att-Unet	51.99
Trans-Unet	105.32
MedFormer	28.07
UNeXt	1.47
FM-Unet	48.03

4.3.3. Quantitative comparisons on different datasets

a) Results on Breast Ultra Sound Images (BUSI) dataset

Table 2 shows the experimental results of several of the most popular segmentation networks on the BUSI segmentation dataset. As shown in Table 2, FM-Unet obtained the highest IOU and F_1-score scores with 70.21% and 80.53%, respectively. Compared with the results of the other seven models, IOU coefficients of FM-Unet improved by 8.16%, 7.49%, 5.39%, 4.96%, 3.29%, 8.95% and 3.26%, respectively, F_1-score improved by 4.51%, 4.10%, 3.06%, 3.03, 1.23%, 6.05% and 1.16%, respectively. It shows that FM-Unet achieves better results than other models and achieves state-of-the-art.

b) Results on International Skin Imaging Collaboration (ISIC 2018) dataset

Table 2 also shows the experimental results of several advanced segmentation networks on the ISIC 2018 segmentation dataset. I experimental results show that FM-Unet obtained IOU the highest IOU and F_1-score with 82.14% and 89.95%, respectively. Compared with the results of the other seven models, IOU coefficients of FM-Unet improved by 9.45%, 7.90%, 8.98%, 7.04%, 1.63%, 1.00% and 0.44%, respectively, F_1-score improved by 6.22%, 5.51%, 5.88%, 4.57, 1.04%, 0.97% and 0.25%,

respectively. It indicates that FM-Unet achieves better results than other models on ISIC 2018 dataset and achieves state-of-the-art.

c) Results on Structured Analysis of the Retina (STARE) dataset

Table 2 also shows the experimental results of several advanced segmentation networks on the STARE segmentation dataset. The experimental results show that FM-Unet obtained IOU the highest IOU and F_1-score with 66.13% and 79.47%, respectively. Compared with the results of the other seven models, IOU coefficients of FM-Unet improved by 2.41%, 1.51%, 1.71%, 2.61%, 0.69%, 1.63% and 1.86%, respectively, F_1-score improved by 1.68%, 1.01%, 0.64%, 1.94%, 0.41%, 1.09% and 1.26%, respectively. It indicates that FM-Unet achieves better results than other models on STARE dataset and achieves state-of-the-art.

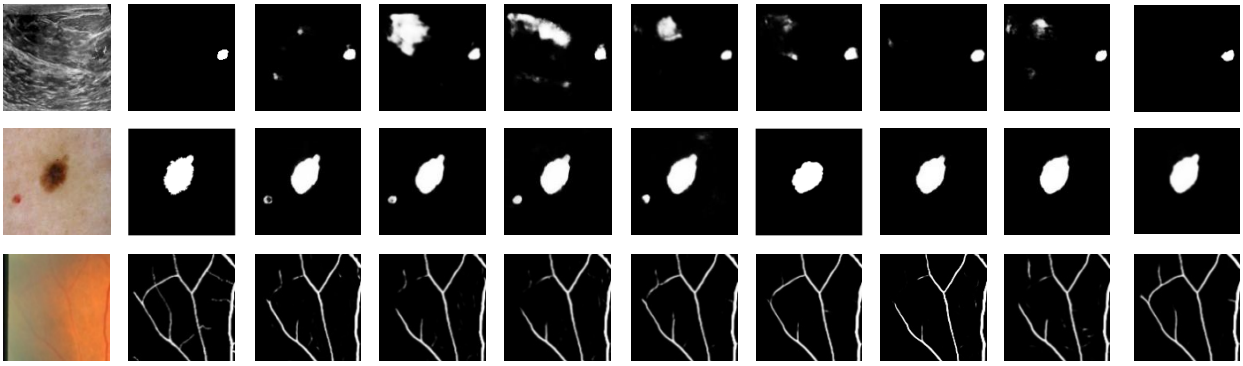
Table 2. Segmentation accuracy of different methods on the BUSI dataset, ISIC (2018) and STARE dataset.

Networks	BUSI		ISIC 2018		STARE	
	IOU (%)	F_1-score (%)	IOU (%)	F_1-score (%)	IOU (%)	F_1-score (%)
Unet	62.05	76.02	72.69	83.72	63.72	77.79
Unet++	62.72	76.43	74.24	84.76	64.62	78.46
Res-Unet	64.82	77.47	73.16	84.07	64.42	78.83
Att-Unet	65.25	77.50	75.10	85.38	63.52	77.53
Trans-Unet	66.92	79.30	80.51	88.91	65.44	79.06
MedFormer	61.26	74.48	81.14	88.98	64.50	78.38
UNeXt	66.95	79.37	81.70	89.70	64.27	78.21
FM-Unet	**70.21**	**80.53**	**82.14**	**89.95**	**66.13**	**79.47**

4.3.4. Qualitative comparative analysis

Quantitative evaluation to show performance differences may not be sufficient to fully understand the advantages of the proposed model. As shown in the evaluation index results in Table 2, FM-Unet achieves the best results, but visual observation is required to determine whether the proposed model works as expected. To this end, in Figure 5, we also give some visual comparison examples of the segmentation in BUSI dataset, ISIC 2018 dataset and STARE dataset.

FM-Unet employs a feedback mechanism and achieves better results than other state-of-the-art segmentation networks. These visual segmentation results show that FM-Unet can recover finer segmentation details successfully, and unexpected segmentation results are less likely to occur for complex backgrounds.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 5. Qualitative comparisons. Row 1-ISIC dataset, Row 2-BUSI dataset, Row 3-STARE dataset. (a) Input, (b) Ground Truth, (c) Unet, (d) UNet++, IRes-Unet, (f) Att-Unet, (g) Trans-UNet, (h) MedFormer, (i) UNeXt and (j) FM-Unet.

5. Conclusions

In this paper, we introduce a novel feedback encoder-decoder depth convolution network architecture based on the U-shaped structure for medical image segmentation. The core idea is to add an encoder feedback path and a decoder feedback path to the basic Unet framework. The proposed feedback path focuses on the information loss of up-sampling and down-sampling, which is well integrated with contextual information. And FM-Unet is efficiently modeled with less complexity and parameters in improved Unet-based networks. Experimental results show that our proposed architecture outperforms the state-of-the-art baselines on various benchmarks. In addition, our network achieves very excellent segmentation results in complex backgrounds. Our work has some limitations. We have developed FM-Unet, a network that relies entirely on convolutional operations. However, the localization and weight sharing of the receiver domain in convolutional operations make it difficult for our network to learn global information. For our future work, we plan to enhance the feedback mechanism's information extraction mechanism so that the feedback module can have a global sensory field. Additionally, we aim to explore a more effective information fusion architecture between the primary and feedback paths.

Acknowledgments

This work is partially supported by the Natural Science Foundation of Fujian Province (Grant Nos. 2020J01816 and 2022J01916), the National Natural Sciences Foundation of China (Grant No. 62106092) and the Principal Foundation of Minnan Normal University (Grant No. KJ18010).

Conflict of interest

The authors have no conflict of interest.

References

1. A. Sinha, J. Dolz, Multi-scale self-guided attention for medical image segmentation, *IEEE J. Biomed. Health. Inf.*, **25** (2021), 121–130. https://doi.org/10.1109/JBHI.2020.2986926
2. X. Zhang, K. Liu, K. Zhang, X. Li, Z. Sun, B. Wei, SAMS-Net: Fusion of attention mechanism and multi-scale features network for tumor infiltrating lymphocytes segmentation, *Math. Biosci. Eng.*, **20** (2023), 2964–2979. https://doi.org/10.3934/mbe.2023140
3. J. Cheng, S. Tian, L. Yu, C. Gao, X. Kang, X. Ma, et al., ResGANet: Residual group attention network for medical image classification and segmentation, *Med. Image Anal.*, **76** (2022), 102313. https://doi.org/10.1016/j.media.2021.102313
4. M. Moghbel, S. Mashohor, R. Mahmud, M I. B. Saripan, Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography, *Artif. Intell. Rev.*, **50** (2018), 497–537. https://doi.org/10.1007/s10462-017-9550-x
5. B. Dourthe, N. Shaikh, S. A. Pai, S. Fels, S. H. M. Brown, D. R. Wilson, et al., Automated segmentation of spinal muscles from upright open MRI using a multiscale pyramid 2D convolutional neural network, *Spine*, **47** (2022), 1179–1186. https://doi.org/10.1097/BRS.0000000000004308
6. T. Zhou, L. Li, G. Bredell, J. Li, E. Konukoglu, Volumetric memory network for interactive medical image segmentation, *Med. Image Anal.*, **83** (2023), 102599. https://doi.org/10.1016/j.media.2022.102599
7. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683
8. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networksfor biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, Springer, **9351** (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
9. Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. M. Liang, UNet++: A Nested U-Net Architecture for Medical Image Segmentation, in *Deep Learning in Medical Image Anylysis and Multimodal Learning for Clinical Decision Support*, Springer, **11045** (2018), 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
10. Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual Unet, *IEEE Geosci. Remote Sens. Lett.*, **15** (2018), 749–753. https://doi.org/10.1109/LGRS.2018.2802944
11. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention U-net: learning where to look for the pancreas, *arXiv preprint*, 2018, arXiv:1804.03999v3. https://doi.org/10.48550/arXiv.1804.03999
12. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, et al., Transunet: Transformers make strong encoders for medical image segmentation, *arXiv preprint*, 2021, arXiv:2102.04306. https://doi.org/10.48550/arXiv.2102.04306
13. Y. Chen, B. Ma, Y. Xia, α-UNet++: A data-driven neural network architecture for medical image segmentation, in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*, Springer, (2020), 3–12. https://doi.org/10.1007/978-3-030-60548-3_1
14. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770–778. https://doi.org/10.1109/CVPR.2016.90

15. Z. Li, H. Zhang, Z. Li, Z. Ren, Residual-attention UNet++: a nested residual-attention U-Net for medical image segmentation, *Appl. Sci.*, **12** (2022), 7149. https://doi.org/10.3390/app12147149
16. J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, et al., Dual attention network for scene segmentation, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 3146–3154. https://doi.org/10.1109/CVPR.2019.00326
17. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., An image is worth 16x16 words: transformers for image recognition at scale, *arXiv preprint*, 2021, arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929
18. G. Rani, A. Misra, V. S. Dhaka, D. Buddhi, R. Sharma, E. Zumpano, et al., A multi-modal bone suppression, lung segmentation, and classification approach for accurate COVID-19 detection using chest radiographs, *Intell. Syst. Appl.*, **16** (2022), 200148. https://doi.org/10.1016/j.iswa.2022.200148
19. G. Rani, A. Misra, V. S. Dhaka, E. Zumpano, E. Vocaturo, Spatial feature and resolution maximization GAN for bone suppression in chest radiographs, *Comput. Methods Programs Biomed.*, **224** (2022), 107024. https://doi.org/10.1016/j.cmpb.2022.107024
20. G. Rani, P. Thakkar, A. Verma, V. Mehta, R. Chavan, V. Dhaka, et al., KUB-UNet: segmentation of organs of urinary system from a KUB X-ray image, *Comput. Methods Programs Biomed.*, **224** (2022), 107031. https://doi.org/10.1016/j.cmpb.2022.107031
21. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, et al., Attention is all you need, *arXiv preprint*, 2017, arXiv:1706.03762. https://doi.org/10.48550/arXiv.1706.03762
22. J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, V. M. Patel, Medical transformer: Gated axial-attention for medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention–MICCAI*, Springer, (2021), 36–46. https://doi.org/10.1007/978-3-030-87193-2_4
23. H. Huang, S. Xie, L. Lin, Y. Iwamoto, X. Han, Y. W. Chen, et al., ScaleFormer: Revisiting the transformer-based backbones from a scale-wise perspective for medical image segmentation, *arXiv preprint*, 2022, arXiv:2207.14552. https://doi.org/10.48550/arXiv.2207.14552
24. Y. Zhang, H. Liu, Q. Hu, Transfuse: Fusing transformers and cnns for medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021*, Springer, (2021), 14–24. https://doi.org/10.1007/978-3-030-87193-2_2
25. H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, et al., Swinunet: Unet-like pure transformer for medical image segmentation, *arXiv preprint*, arXiv:2105.05537. https://doi.org/10.48550/arXiv.2105.05537
26. Y. Liu, N. Qi, Q. Zhu, W. Li, CR-U-Net: Cascaded U-net with residual mapping for liver segmentation in CT images, in *IEEE Visual Communications and Image Processing (VCIP)*, (2019), 1–4. https://doi.org/10.1109/VCIP47243.2019.8966072
27. L. Hong, R. Wang, T. Lei, X. Du, Y. Wan, Qau-Net: Quartet attention U-net for liver and liver-tumor segmentation, in *IEEE International Conference on Multimedia and Expo (ICME)*, (2021), 1–6. https://doi.org/10.1109/ICME51207.2021.9428427
28. J. You, P. L. Yu, A. C. Tsang, E. L. Tsui, P. P. Woo, C. S. Lui, et al., 3D dissimilar-siamese-U-Net for hyperdense middle cerebral artery sign segmentation, *Comput. Med. Imaging Graphics*, **90** (2021), 101898. https://doi.org/10.1016/j.compmedimag.2021.101898
29. M. Jiang, F. Zhai, J. Kong, A novel deep learning model DDU-net using edge features to enhance brain tumor segmentation on MR images, *Artif. Intell. Med.*, **121** (2021), 102180. https://doi.org/10.1016/j.artmed.2021.102180

30. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 4700–4708. https://doi.org/10.1109/CVPR.2017.243
31. E. Shibuya, K. Hotta, Feedback U-Net for cell image segmentation, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2020), 4195–4203. https://doi.org/974-975.10.1109/CVPRW50498.2020.00495
32. D. Lin, Y. Li, T. L. Nwe, S. Dong, Z. Oo, RefineU-Net: Improved U-Net with progressive global feedbacks and residual attention guided local refinement for medical image segmentation, *Pattern Recognit. Lett.*, **138** (2020), 267–275. https://doi.org/10.1016/j.patrec.2020.07.013
33. N. Ibtehaz, M. S. Rahman, MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation, *Neural Networks*, **121** (2020), 74–87. https://doi.org/10.1016/j.neunet.2019.08.025
34. J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, V. M. Patel, KiU-Net: Towards accurate segmentation of biomedical images using over-complete representations, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*, Springer, (2020), 363–373. https://doi.org/10.1007/978-3-030-59719-1_36
35. S. Woo, J. Park, J. Lee, I. Kweon, CBAM: Convolutional block attention module, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
36. H. Zhao, H. Zhang, X. Zheng, A multiscale attention-guided UNet++ with edge constraint for building extraction from high spatial resolution imagery, *Appl. Sci.*, **12** (2022), 5960. https://doi.org/10.3390/app12125960
37. Y. Pang, Y. Li, J. Shen, L. Shao, Towards bridging semantic gap to improve semantic segmentation, in *2019 IEEE/CVF International Conference on Computer Vision*, (2019), 4230–4239. https://doi.org/10.1109/ICCV.2019.00433
38. W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, 2020. Available from: https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset.
39. The International Skin Imaging Collaboration (ISIC 2018). Available from: https://challenge.isic-archive.com/landing/2018/.
40. A. D. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, *IEEE Trans. Med. Imaging*, **19** (2000), 203–210. https://doi.org/10.1109/42.845178
41. Y Gao, M Zhou, D Liu, Z. Yan, S. Zhang, D. Metaxas, A data-scalable transformer for medical image segmentation: architecture, model efficiency, and benchmark, *arXiv preprint*, 2023, arXiv:2203.00131, 2022. https://doi.org/10.48550/arXiv.2203.00131
42. J. M. J. Valanarasu, V. M. Patel. Unext: Mlp-based rapid medical image segmentation network, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022*, Springer, (2022), 23–33. https:/doi.org/10.1007/978-3-031-16443-9_3