

Article

LPE-Unet: An Improved UNet Network Based on Perceptual Enhancement

Suwei Wang , Chenxun Yuan  and Caiming Zhang * 

School of Software, Shandong University, Jinan 250101, China

* Correspondence: czhang@sdu.edu.cn

Abstract: In Computed Tomography (CT) images of the coronary arteries, the segmentation of calcified plaques is extremely important for the examination, diagnosis, and treatment of coronary heart disease. However, one characteristic of the lesion is that it has a small size, which brings two difficulties. One is the class imbalance when computing loss function and the other is that small-scale targets are prone to losing details in the continuous downsampling process, and the blurred boundary makes the segmentation accuracy less satisfactory. Therefore, the segmentation of calcified plaques is a very challenging task. To address the above problems, in this paper, we design a framework named LPE-UNet, which adopts an encoder–decoder structure similar to UNet. The framework includes two powerful modules named the low-rank perception enhancement module and the noise filtering module. The low-rank perception enhancement module extracts multi-scale context features by increasing the receptive field size to aid target detection and then uses an attention mechanism to filter out redundant features. The noise filtering module suppresses noise interference in shallow features to high-level features in the process of multi-scale feature fusion. It computes a pixel-wise weight map of low-level features and filters out useless and harmful information. To alleviate the problem of class imbalance caused by small-sized lesions, we use a weighted cross-entropy loss function and Dice loss to perform mixed supervised training on the network. The proposed method was evaluated on the calcified plaque segmentation dataset, achieving a high F1 score of 0.941, IoU of 0.895, and Dice of 0.944. This result verifies the effectiveness and superiority of our approach for accurately segmenting calcified plaques. As there is currently no authoritative publicly available calcified plaque segmentation dataset, we have constructed a new dataset for coronary artery calcified plaque segmentation (Calcified Plaque Segmentation Dataset, CPS Dataset).

Keywords: calcified plaque segmentation; attention mechanism; low rank; CNN; gating mechanism



Citation: Wang, S.; Yuan, C.; Zhang, C. LPE-Unet: An Improved UNet Network Based on Perceptual Enhancement. *Electronics* **2023**, *12*, 2750. <https://doi.org/10.3390/electronics12122750>

Academic Editors: Luca Mesin and Gemma Piella

Received: 11 May 2023

Revised: 7 June 2023

Accepted: 19 June 2023

Published: 20 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Coronary artery disease (CAD) has become one of the leading causes of human mortality [1] and accounts for the highest proportion of all cardiac deaths. It is characterized by sudden onset and high mortality. The main cause of CAD is atherosclerosis-induced coronary artery stenosis or obstruction. In the clinical diagnosis of CAD, the coronary artery calcification score reflects an important criterion for the degree of coronary atherosclerosis and is commonly used to assess the overall calcification status of the coronary arteries [2]. Therefore, there is an urgent need to develop an efficient and accurate calcified plaque segmentation algorithm, as this is a critical step in calculating the plaque calcification score. Traditional manual segmentation methods are time-consuming and labor-intensive, and difficult to apply to CT images with exploding data volumes. Automatic segmentation algorithms based on chest CT images can effectively reduce the workload of physicians and assist in the detection and diagnosis of diseases and have some research value. In the past, it has been explored for a long time. Existing conventional algorithms and DL-based segmentation algorithms usually divide the segmentation process into two stages, first locating or segmenting larger tissues or organs, such as coronary arteries or the heart,

and then using further algorithms to segment calcified plaques in the coronary arteries. However, these methods still have significant drawbacks, leaving room for improvement in terms of accuracy and difficulty in ensuring the structural integrity of the lesion margins. Therefore, there is still important research value in the field of calcified plaque segmentation.

In recent years, DL has demonstrated powerful performance in various medical image detection and segmentation tasks, such as tumor, polyp, and ulcer detection [3]; organ segmentation [4]; and COVID-19 infection segmentation [5]. Compared to conventional natural images and medical images, calcified plaques are characterized by small scale, irregular shape, and unclear boundaries. Accurate segmentation of too small targets has always been a difficult problem in computer vision. There is often a trade-off between enlarging the receptive field and losing detail information in conventional CNN. A larger receptive field means that each neuron in the network can extract features over a larger region during inference, resulting in more accurate segmentation results. CNN usually enlarges the receptive field by down-sampling through consecutive pooling layers to extract high-level semantic features. With the increase in network layers, the proportion of the actual receptive field to the theoretical receptive field decreases step by step [6]. Therefore, to obtain a larger receptive field, a very deep network is usually used, including many pooling layers. However, more pooling layers mean more loss of detail information, which is unfavorable for pixel-wise image segmentation tasks. When the segmentation target is too small, too many pooling operations will make the network lose sight of the target, making the image difficult to reconstruct; if there are too few pooling layers, extracting the desired high-level semantic features is difficult. Therefore, the number of pooling layers used by the network is a matter of balance.

In segmenting calcified plaques in coronary CT angiography images, contextual information plays a significant role in addition to the inherent morphological features of the plaques. Contextual information at different scales can provide different receptive fields, which can be achieved by pyramid structures that include convolution kernels of different sizes or dilation rates. Moreover, not all feature maps contribute equally to the segmentation task, and only task-related channels and regions are of concern. The redundancy problem inside convolutional filters is also prominent.

With the increase in network depth, the extracted features have higher-level semantic features, more accurate localization ability, and less detailed information. On the other hand, shallow features contain rich detailed information. In order to compensate for the loss of detailed information caused by downsampling, shallow features are used to fuse with high-level features to predict the target jointly. However, for the segmentation task of small-scale calcified plaques, directly concatenating shallow features and high-level features will reduce the localization ability of the features, since shallow features contain much irrelevant information that is not useful for prediction. To improve the prediction accuracy of the model, it is necessary to filter out the noise in shallow features before performing multi-scale feature fusion. Another issue worth noting is that the small-scale characteristics of calcified plaques can bring about a serious class imbalance problem between foreground and background, which is not conducive to the convergence of the model.

This paper proposes an end-to-end calcium scoring network, LPE-UNet, to accurately segment calcium deposits in coronary artery CT images. To address the problem of lost detail information during downsampling, this study first reduces the number of down-sampling operations while maintaining segmentation performance. Before inputting the extracted features into the decoder, a low-rank perceptual enhancement module is used to enhance and purify the features. The low-rank perceptual enhancement module consists of a multi-scale feature extraction module and a feature filtering module based on attention mechanisms. Multi-scale feature extraction not only provides the model with contextual information at different scales for prediction but also expands the receptive field of the model without losing detailed information, allowing the model to collect information more accurately over a larger range for segmentation. The attention module consists of two

sub-modules, channel attention and low-rank spatial attention, which enable the model to focus only on relevant features and regions for the task.

In addition, to suppress the interference of irrelevant features in the fusion of deep and shallow features, a gate mechanism for filtering redundant information was constructed on the skip connection in this paper. The highly correlated spatial positions of the output were clustered, and the obtained results were used to mask the features, blocking out task-irrelevant regions. Finally, to address the problem of class imbalance, we proposed a novel weighted loss function that combines the weighted cross-entropy loss function and the Dice loss for network supervision training. This further improves the accuracy of calcium deposit segmentation and avoids the problem of network convergence difficulty when using Dice loss alone. We conducted comprehensive experiments on the coronary artery calcium deposit dataset and achieved optimal performance. The experimental results confirm the feasibility and effectiveness of the proposed algorithm in this paper. In summary, the contributions of this paper are mainly focused on four aspects:

1. We propose a new end-to-end segmentation network, LPE-UNet, which accurately segments calcium plaques in the coronary artery.
2. We design a low-rank perceptual enhancement module (LPE), which adds the ability to capture and process multi-scale information to the model and establish global dependencies between local features. It can improve the segmentation accuracy.
3. By adding noise optimization modules to suppress noise information in skip connections, which use attention based gates to filter low-level features, We further improved the accuracy of network inference.
4. Due to the small-scale characteristics of calcium plaques, class imbalance can easily occur during network training. We proposed a new weighting method that uses a combination of weighted cross-entropy and DiceLoss for mixed supervision training of the network, which effectively improves the stability of model training and prediction accuracy.

2. Related Work

2.1. Calcified Plaque Segmentation Algorithm

In recent years, many CT image-based algorithms have been proposed for calcium deposit segmentation. Existing segmentation methods typically divide the segmentation process into two stages, first, segmenting or locating larger tissues such as coronary arteries or the heart, and then segmenting the calcium deposits in the coronary arteries. Gao et al. [7] developed a method to automatically detect the range of the calcified plaque with acoustic shadowing in IVUS images by combining of the Gaussian mixture model and Markov random field. Yoshida et al. [8] used linear enhancement filters and region-growing algorithms to extract the coronary arteries, and then used threshold segmentation to extract calcium deposits from the enhanced images. Sun et al. [9] proposed a coronary artery plaque detection method based on automatic scale selection and fuzzy c-means. They first estimated the vessel diameter by obtaining the response of DoG operation in each slice center of the artery and cropped the slices based on the features. Then, they used fuzzy c-means to classify the cropped images and extracted calcium deposits based on their intensity, quantifying arterial stenosis rate. In recent years, with the rapid development of computer hardware and DL theory, many excellent DL-based algorithms for calcium deposit segmentation have been proposed one after another.

DL-based algorithms for calcified plaque segmentation can be broadly categorized into patch-based scoring and pixel-wise segmentation. The first type involves chunking the input image into patches and scoring each patch to determine the presence of calcified plaques. In this approach, Jelmer et al. [10] cropped three orthogonal CT patches as input for a CNN and predicted the probability of calcified plaques within the image blocks. Lessmann et al. [11] used three parallel CNNs to process the three orthogonal CT patches, and the fused features are used to score automatic coronary. In a follow-up work [2], they

employed two sequential CNN, where the first network identified potential calcifications based on location and the second network identified true calcified plaques.

The second type of approach involves pixel-wise segmentation of calcified plaques. Lekadir et al. [12] cropped b-ultrasound pattern maps into small image blocks and classified them using a CNN for plaque detection. Santini et al. [13] used axial projections centered around candidate pixels to perform segmentation using the U-Net architecture [14]. Shadmi et al. [15] considered both U-Net and FC-DenseNet architectures for coronary calcium segmentation and Agatston score prediction. Ma et al. [16] proposed DenseRAUnet, a novel network that takes advantage of dense connections. Zhang et al. [17] introduced CAC-net, which utilizes 2D U-DenseNet for intra-layer calcification features and 3D U-Net for inter-slice calcification features. Lee et al. [18] employed a 3D neural network to detect primary calcified tissue and then used SegNet for plaque segmentation. Li et al. [19] used two branches to detect media-adventitia and lumen, and combined their results with the original image to enhance plaque detection. We have summarized the calcification plaque segmentation algorithm in Table 1.

Table 1. Table of calcified plaque segmentation algorithms.

Non DL-Based Algorithms	DL-Based Algorithms	
	Patch-Based Scoring	Pixel-Wise Segmentation
Gao et al. [7]	Jelmer et al. [10]	Lekadir et al. [12]
Yoshida et al. [8]	Lessmann et al. [11]	Santini et al. [13]
Sun et al. [9]	Lessmann et al. [2]	Shadmi et al. [15]
		Ma et al. [16]
		Zhang et al. [17]
		Lee et al. [18]
		Li et al. [19]

2.2. Medical Image Segmentation

The segmentation of calcified plaques in coronary arteries, as a specific task within medical image segmentation, is closely intertwined with other related segmentation tasks. In the field of medical image segmentation, these tasks mutually reinforce each other, leading to shared advancements and progress.

Ronneberger et al. [14] introduced the U-Net structure for medical image segmentation, which has since become the benchmark for many segmentation tasks. U-Net utilizes a symmetric architecture with skip connections that effectively combine low-resolution and high-resolution feature maps, enabling the fusion of low-level and high-level features. Since its inception, U-Net has been widely adopted in the medical imaging field, leading to numerous meaningful improvements. Res-UNet [20] enhances U-Net by replacing each submodule with residual connected modules, resulting in improved performance for retinal vessel segmentation. Unet++ [21] incorporates long and short connections, where each decoder layer fuses connections from lower- and same-level feature maps in the encoder, as well as high-level feature maps from the decoder. This approach has shown advancements in medical segmentation tasks. To capture a larger perceptual field, Lopez et al. [22] applied a superposition of multi-scale atrous convolutions to brain tumor segmentation, achieving notable improvements in accuracy. Zhao et al. [23] integrate a feature pyramid with a U-Net++ model to segment coronary arteries in ICAs, and proposed a compound loss function that contains Dice loss and dilated Dice loss. Mu et al. [24] introduced the long skip connections of the attention gate at each layer to emphasize the field of view (FOV) for IAs and embed residual-based short skip connections in each layer to implement in-depth supervision to help the network converge.

The rise of attention has brought great attention-based work. The attention U-Net [25] introduces a self-attentive mechanism in each module of U-Net, enabling the suppression of irrelevant regions in input images while highlighting salient features in specific local regions. This approach has been applied to pancreas image segmentation. Kaul et al. [26] proposed

the FocusNet, which combines spatial attention and channel attention for medical image segmentation. This approach effectively emphasizes informative image regions during the segmentation process. Additionally, the Transformer model [27], known for its global attention mechanism, has been adapted for medical image segmentation. Researchers, such as [4], have added a transformer branch to the original convolutional U-Net architecture to leverage the benefits of global attention.

3. Method

3.1. Overview of LPE-Unet

Like [21,25,28–31], the proposed LPE-UNet network in this paper is based on the U-shaped architecture of the UNet network commonly used in medical image segmentation tasks. As shown in Figure 1, the model consists of four parts, encoders, decoders, low-rank perception enhancement modules, and noise optimization modules. The encoders and decoders form the backbone network. The encoders take low-level features as an input and use successive convolution and a downsampling pool layers to extract more abstract semantic features, and the decoder takes the upsampled high-level features and the same level features as inputs and gradually recovers the detail through successive convolution to obtain the final segmentation results. Due to the small scale of calcified plaques, the method in this paper reduces the number of layers of the downsampling pool in order to avoid losing the feature information of the object to be segmented. The configurations of encoders and decoders are shown in Table 2. Let F_{ei} and F_{di} represent the output feature map of the i th encoder and decoder separately. To supplement the lost reception field of view, a low-level perceptual enhancement module is built into the network to expand the perceptual field. Furthermore, the global attention is computed to weight the extracted multiscale features by establishing long-distance dependencies between features to highlight the important information in the features.

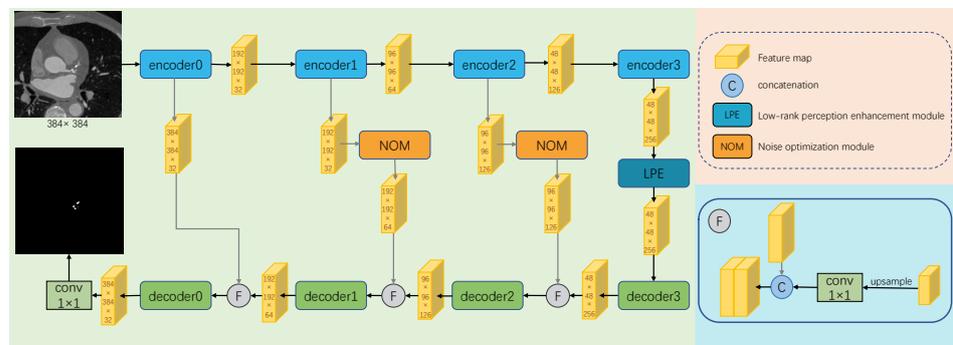


Figure 1. An overview of the proposed LPE-Unet. It takes an encoder–decoder format.

Traditional networks based on the UNet architecture use skip connections to supplement the decoder with detailed information lost due to continuous downsampling. However, the shallow feature maps have higher resolution and contain more complex information, and a large number of irrelevant features have a negative impact on the subsequent inference of the network. To address this issue, this paper proposes a noise optimization mechanism to filter out noisy information in low-level features. By learning a global weight distribution, this mechanism suppresses unimportant features in the feature map and highlights task-relevant features.

Finally, to address the severe inter-class imbalance problem caused by the small size of calcified plaques, this paper designs and implements a new cross-entropy weighting method, which focuses the optimization of the network on the regions of interest. During training, we use a combination of weighted cross-entropy and DiceLoss for mixed supervised training of the network, which effectively improves the stability of model training and the accuracy of predictions.

Table 2. Parameter table of the backbone of our model. Each convolution is followed by a group normalization and a ReLU activation, which is omitted for readability.

Layer Name	Parameters	Output Size
encoder0	conv3 × 3, 16 conv3 × 3, 32 maxpool2 × 2, stride2	32, 192, 192
encoder1	conv3 × 3, 32 conv3 × 3, 64 MaxPool2 × 2, stride2	64, 96, 96
encoder2	conv3 × 3, 64 conv3 × 3, 128 MaxPool2 × 2, stride2	128, 48, 48
encoder3	conv3 × 3, 128 conv3 × 3, 256	256, 48, 48
decoder3	conv3 × 3, 256 conv3 × 3, 256	256, 48, 48
decoder2	conv3 × 3, 128 conv3 × 3, 128	128, 96, 96
decoder1	conv3 × 3, 64 conv3 × 3, 64	64, 192, 192
decoder0	conv3 × 3, 32 conv3 × 3, 32	32, 384, 384

3.1.1. Low-Rank Perception Enhancement Module

The low-rank perception enhancement module consists of a multi-scale feature extraction module and a hybrid attention module to enhance features.

Extracting semantic information in different receptive fields can effectively improve segmentation accuracy. However, the calculation and model training difficulty increases with the increase in the receptive field, which limits the size of the receptive field. Inspired by Atrous Spatial Pyramid Pooling [32] and inception modules [33], we use a multi-scale feature module to extract multi-scale features, which increases the size of the reception fields without significantly increasing the computation and losing resolution. Multi-scale features can provide different context information. As shown in the top left corner in Figure 2, multi-scale feature extraction includes four branches. These four branches include one branch with a 1 × 1 convolution kernel and three branches with 3 × 3 dilated convolutions with dilation rates of 1, 2, and 3. The feature maps of different branches are concatenated to obtain multi-scale features containing rich semantic information. Finally, the concatenated features undergo a convolution operation to reduce channels. Let F_{aspp} represent the output of the multi-scale feature extraction module, and the calculation process is as follows:

$$F_{msf} = \text{conv}_1(\varphi(\text{conv}_1(F_{e3}), C_{a1}(F_{e3}), C_{a2}(F_{e3}), C_{a3}(F_{e3}))) \tag{1}$$

where $\text{conv}_1(\cdot)$ represents 1 × 1 convolutions. φ represents the feature fusion function, here the concatenation operation is used. $C_{ai}(\cdot)$ represent 3 × 3 dilated convolutions with atrous rate i .

The hybrid attention module consists of two sub-modules, the channel attention module and the low-rank based spatial attention module, which enable the model to focus only on relevant features and regions for the task at hand. We use a SE module to model channel attention, which computes the relationships between different channels and learns a weight for each channel. In order to extract more important and critical feature information. The multi-scale feature F_{msf} enters the channel attention module, first undergoes global average pooling and then passes through two consecutive fully connected layers. The output channels of the first fully connected layer are less than the

input channels, and the output channels of the second fully connected layer are equal to C . The calculation process of the channel weights is shown as follows.

$$W_c = \theta \left(f_{l2} \left(\theta \left(f_{l1} \left(P_{Gavg}(F_{msf}) \right) \right) \right) \right) \tag{2}$$

where $\theta(\cdot)$ is the sigmoid function, f_{l1} and f_{l2} are fully connected layers, and $P_{Gavg}(\cdot)$ represent global average pooling. After channel attention, the features become:

$$F_c = W_c \times F_{msf} \tag{3}$$

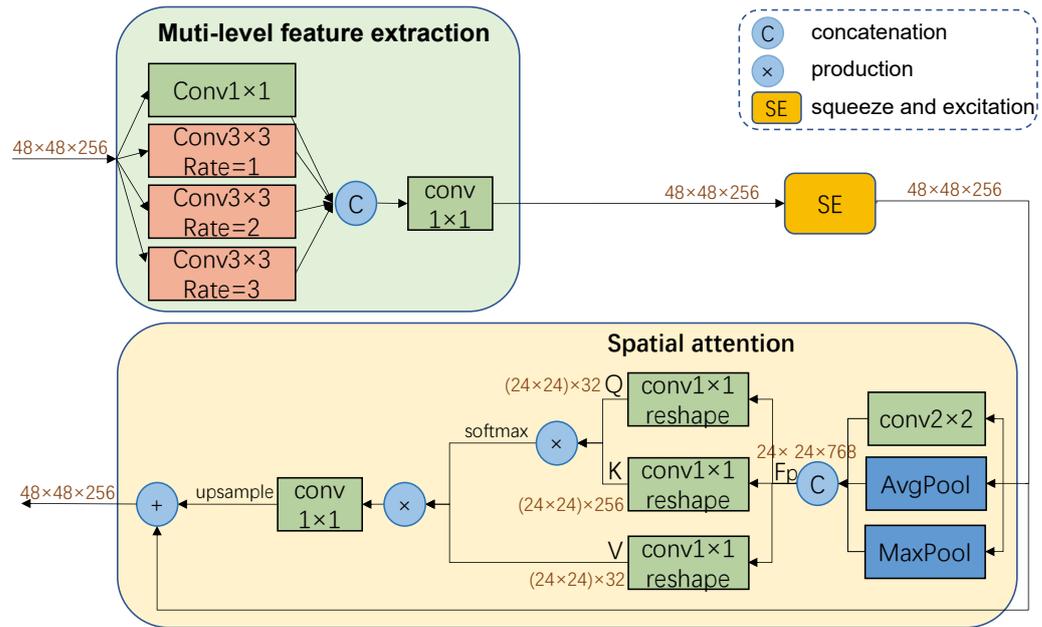


Figure 2. Low-rank perception enhancement module structure diagram.

To model the importance of different positions, we added a spatial attention module to the network which can establish global dependencies between different local features and introduce low rank in non-local self-attention. This module is inspired by the non-local self-attention mechanism. Since lesion areas and background areas contain different features and the edge areas of lesions have different contexts from the internal areas, we use a mixed pooling strategy to help distinguish between lesion areas and background areas. Specifically, the mixed pooling module includes three branches, an average pooling, a maximum pooling, and a 2×2 convolution. The maximum pooling can extract the most prominent features, the average pooling can extract flat background information, and a 2×2 convolutional branch is added to reduce information loss. After passing through the mixed pooling part, F_c is transformed into three sets of feature maps, which are concatenated to obtain F_{mp} . the size of F_{mp} is reduced to one-quarter of F_c , thereby reducing the computational complexity of the model. The formula for computing F_{mp} is as follows:

$$F_{mp} = \varphi(P_{avg}(F_c), P_{max}(F_c), conv_2(F_c)) \tag{4}$$

Among them, P_{avg} and P_{max} represent the average pooling layer and the maximum pooling layer, respectively, and $conv_2$ represents 2×2 convolution. φ represents the feature fusion function, here the concatenation operation is used. Then F_{mp} is used to calculate the global dependencies between different local features.

When computing spatial attention, instead of computing correlations between all locations, we map pixels to a compact low-dimensional space, and compute the correlation matrix between two feature spaces as global descriptors, then reconstruct the features using global descriptors. By mapping features to a low-dimensional space, we hope to discover

semantic information with a higher degree of abstraction. Let C denote the dimensionality of the F_{mp} and K denote the dimensionality of the low-dimensional space, where K is significantly smaller than C . By mapping the original features to the low-dimensional space, the value V and query Q of each position are obtained. Then the C dimensional key K is computed in the original feature space. By multiplying the key and query, the correlation matrix $A \in R^{KC}$ of the two feature spaces is obtained. Then the reconstructed features are obtained by multiplying the value and the feature space similarity matrix A . Because K of the low-dimensional space is significantly smaller than C , using K global descriptors to reconstruct the C dimensional features led to a low rank in the reconstructed features. F_{mp} first passes through two 1×1 convolution branches to obtain Q and K , then we reshape Q and K to two-dimensional matrixes by merging spatial dimensions. The correlation matrix $A \in R^{KC}$ is calculated as follows:

$$A = \delta \left(\frac{\text{conv}_1^Q(F_{mp})^T \text{conv}_1^K(F_{mp})}{\sqrt{d_k}} \right) \quad (5)$$

Then we reconstruct features of each location with global descriptors, and the calculation process is as follows

$$W_s = \text{conv}_1(A^T \text{conv}_1^V(F_{mp})) \quad (6)$$

where conv_1^Q , conv_1^K , and conv_1^V are all 1×1 convolution kernels, δ represents the softmax function. d_k is a regularization term that ensures that the sum of each row is equal to 1. By multiplying the value and A to obtain the reconstructed features, each location can receive the information of the whole space. This operation allows the model to establish feature dependencies at a global level, facilitating the understanding of how each pixel relates to the overall context. The output of the spatial attention module is:

$$F_s = W_s + F_c \quad (7)$$

3.1.2. Noise Optimization Module

High-level features capture more abstract semantic information and have the ability of precise localization, as images pass through successive downsampling layers, spatial details are lost. However, for accurate image segmentation tasks, spatial details are crucial alongside semantic information. To address this, UNet architecture connects the encoder and decoder at the same level through skip connections and gradually incorporates shallow detail information from the encoder to the decoder. Nonetheless, shallow features with higher resolution contain more generic low-level semantic information, they may be irrelevant or even intrusive to our task. This information, which can negatively impact the accuracy of calcified plaque segmentation, is referred to as noise. Directly fusing shallow and deep features will introduce this noise, thereby damaging the network's precise localization ability and reducing the accuracy of segmentation. Based on the assumption that the feature maps in each channel have similar compatible weight distributions, we use spatial attention to filter out noise information in the shallow features. In this paper, we construct a spatial attention-based gating mechanism on the skip connections to suppress the noise in low-level features. The detailed structure of the module is shown in Figure 3. To reduce computation, we only added the module in the second and third layers of the network. The module consists of two branches, which calculate the global and local spatial attention maps, respectively, and then fuse the results of the two branches to obtain the final attention map. First, a channel information aggregation function θ is used to calculate the feature information of the two-dimensional spatial position, and then the input channel aggregation information F_θ is expanded into one-dimensional data, and then two consecutive fully connected layers are used to learn the feature information of each position in the matrix. Then the learned one-dimensional global attention tensor is restored to a

two-dimensional matrix. Let W_{Gi} represent the global attention which takes the output of i th encoder as input, it is calculated as follows

$$W_{Gi} = f_{f2}(f_{f1}(\theta(F_{ei}))). \tag{8}$$

where θ is the channel information aggregation function, and a 1×1 convolution is used in this paper. f_{f1} and f_{f2} represent two consecutive fully connected layers. Since W_{Gi} did not consider the variability of features between channels, we add another path as a complement to the first path. In this approach, we use a probabilistic learning function T to simultaneously compare the importance between different channels and locations. The function T here is implemented using a 3×3 convolutional layer.

$$W_{li} = T(F_{ei}) \tag{9}$$

We fuse the two attention features together, and then use the sigmoid function to compute the global weight distribution matrix. In order to avoid the complete loss of features due to the weight being too small, a threshold ϵ is added to the calculated probability map, here we take ϵ as 0.01. The residual connection is used here, and the calculation process is shown in the formula, and F_{Gi} is the output feature matrix.

$$F_{Gi} = \left(\sigma \left(\frac{1}{C(X)} (W_{Gi} + W_{li}) \right) + \epsilon \right) * F_{ei} + F_{ei} \tag{10}$$

$C(X)$ is a regularization term so that the sum of each row of the weight matrix is 1.

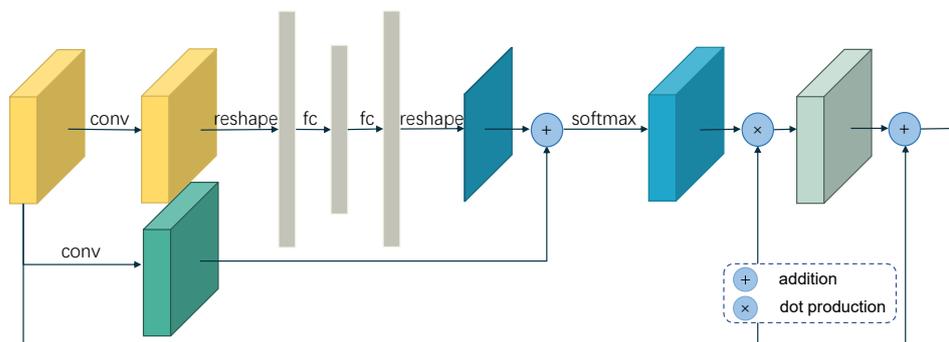


Figure 3. Block diagram of noise optimization module.

3.2. Mixed Training

The calcium plaque segmentation dataset has two main characteristics, high image resolution and small and irregular lesion size. Therefore, there is a significant class imbalance problem in the task of calcium plaque segmentation in coronary artery CT images. The cross-entropy loss function focuses more on the segmentation quality of global segmentation results, and using the cross-entropy loss function in class-imbalanced scenarios often fails to achieve good results. Since the true goal of segmentation is to maximize the dice coefficient, the Dice loss function performs well in scenarios with severe positive and negative sample imbalances. However, when the batch size is small and the segmentation target is too small, even a small prediction error may cause significant changes in the loss value, leading to oscillations in the loss curve and difficulty in network convergence. The high image resolution of the calcium plaque segmentation dataset leads to a small batch size, and the small lesions make it difficult to perform image scaling operations. Therefore, using the Dice loss function directly for network training is not suitable for this task. This paper combines the advantages of weighted cross-entropy loss function and Dice loss function and uses a hybrid loss function to train the network. During training, the network is first trained with the iterative weighted cross-entropy loss function (DWCE Loss) for multiple rounds, and then trained with the Dice loss function until the network converges.

Based on the problems raised above, this method proposes a weighted cross-entropy loss function based on the distance from the pixel point in the image to the lesion. The traditional binary cross entropy loss function formula is as follows

$$l(p) = -(y_p \log(x_p) + (1 - y_p) \log(1 - x_p)) \quad (11)$$

where p represents the current pixel, y is the label, and x is the predicted value. During the calculation of the cross-entropy loss function, each pixel contributes the same to the final loss value. In the case of unbalanced samples in this task, the model optimization direction will be dominated by a large number of negative sample losses, which will make the model difficult to train or converge to non-optimal results. In order to solve this problem, this paper proposes a relatively novel weighting method.

In order to facilitate the demonstration of the calculation process, we use ω to represent the weight matrix. The initial value is set to the label value of the ground truth label Y , that is, $\omega_0 = Y$. The process of one iteration is as follows

$$\omega_t = \rho(\omega) = \omega_{t-1} + y_t * \alpha \quad (12)$$

$$y_t = \eta(y_0) \quad (13)$$

where y_0 is the label, η represents the expansion operation, each iteration, the label is expanded by one pixel, and α is selected based on experience, indicating the weight of the expanded label matrix. t is the number of iterations of the calculation process. After t times of operations, the final weight matrix ω appears as a stepped matrix with the label as the center and the weights gradually decrease outward. In this way, the role of positive examples is improved, the role of negative classes is weakened, and the phenomenon of category imbalance is alleviated. Since the weight of the boundary negative class in the obtained ω matrix is 0, in order to make the negative class away from calcified plaque can also contribute to the optimization of the network, we do the following transformation on ω :

$$\omega = \omega + \delta\omega_0 \quad (14)$$

Among them, δ is used to balance the relationship between ω and marginal negative classes. The setting of parameter values can be modified according to the proportion of the foreground class in the ground truth. In this experiment, δ is set to 0.1. We use the symbol L_{DWCE} to represent the weighted cross-entropy loss (DWCE Loss). The calculation process is as follows

$$L_{DWCE} = \frac{1}{N} \sum \omega * l(p) \quad (15)$$

where $l(p)$ represents the calculation result of the conventional cross-entropy loss function, ω is the weight distribution map calculated above, and N is the sum of the weight matrix ω . Compared with DiceLoss, which ignores a large number of background elements, DWCE Loss takes into account the value of each pixel. Therefore, when training the network, we sequentially use ω_{30} , ω_{10} , ω_5 to pre-train the network, and finally use DiceLoss to train the network to obtain the final convergence result. The whole training process can be understood as the process of gradually concentrating the focus of the network on the lesion area.

4. Experiments

4.1. Experiment Setting

The batch size of the network is set to 16, an initial learning rate of 0.0001, and weight decay of 0.0001. To increase the amount of data and reduce GPU memory usage, we use a sliding window to extract patches of size 384×384 with a step size of 35 from the original images as input to the network. Finally, we obtain the segmentation results by averaging the results of multiple patches.

4.2. Dataset

The experimental data were all from Qianfoshan Hospital in Shandong Province. We selected representative data from 130 coronary heart disease patients from a large number of CT image data for annotation. The annotation process was completed by two experienced radiologists and checked by another physician to ensure the accuracy of the dataset. Each patient's data consisted of about 60 to 100 DCM files (each file represents a slice of CT images). Due to differences in physician operating habits and equipment models, the pixel distance and slice spacing of the CT images also varied. Generally, the pixel distance is 0.5 or 0.35, and the slice interval is 0.5 to 0.7. To reduce the impact of equipment, the patient's CT image data are resampled, with a distance of 0.5 between adjacent slices and a pixel distance of 0.35 within the same slice. As there were some areas with high HU values in the data, to avoid the image data distribution being too concentrated after normalization, which may lead to unclear feature differences, we set a threshold to suppress the high HU values. To increase the amount of data and the model's generalization ability, we performed routine data augmentation operations, including image normalization, flipping and rotation, adding Gaussian noise, adding elastic deformation, and randomly adjusting contrast. A sample image of the calcified plaque segmentation dataset is shown in Figure 4.

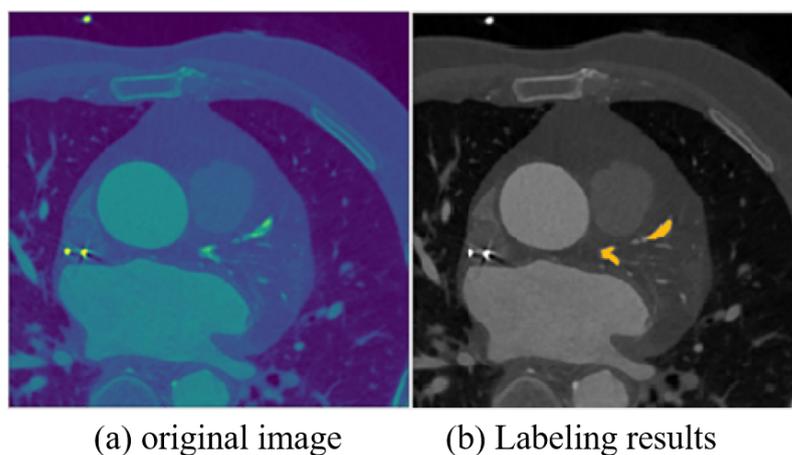


Figure 4. Calcified plaque segmentation dataset sample figure (The yellow mark in (b) represents Ground Truth).

4.3. Evaluation Index

Intersection over Union: Iou is used to represent the mean value of the ratio between the overlapping part of ground truth and the prediction and the union of Ground Truth and the segmentation. This avoids the interference of background pixels in the image when calculating the segmentation effect, and the representation of the segmentation result is more accurate. Iou can be expressed as:

$$\text{Iou} = \frac{TP}{(FP + FN + TP)} \quad (16)$$

Iou is widely used in various semantic segmentation tasks and target detection tasks.

Dice score: Dice can be defined as twice the ratio between the overlapping area of the prediction and the ground truth between the sum area of both, which is popular in medical image segmentation. The Dice score is defined as:

$$\text{Dice} = \frac{2TP}{(FP + FN + 2TP)} \quad (17)$$

F1_score: The F1 score takes both Precision and Recall into consideration. When we analyze the experimental results, Precision is usually used to measure whether the

segmentation result has a misjudgment, and Recall is usually used to measure whether the result has a false positive error.

$$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{(\text{Recall} + \text{Precision})} \tag{18}$$

F1_score is usually used when the recall rate and precision rate need to be controlled at the same time, and it is also one of the most commonly used evaluation indicators in segmentation tasks.

4.4. Results

We conducted comprehensive experiments to verify the effectiveness and superiority of the proposed method. First, we compared the performance of the LPE-UNet model with several classic networks (UNet [14], DeepLab v3+ [34], FCN [35], Attention-UNet [25], nnUNet [28], UNet++ [21], and R2U_UNet [29]) on the coronary artery calcification plaque dataset. Secondly, we conducted sufficient ablation experiments to evaluate the effectiveness of each proposed module. All experiments were conducted using 4-fold cross-validation, where the dataset was randomly divided into training set (60%), validation set (20%), and test set (20%).

Table 3 shows the experimental results of our method compared with other methods. It can be seen that our method outperforms other existing models, which is consistent with our expectations. In order to better compare and evaluate the overall performance of the models, we used F1_score as the evaluation metric to further qualitatively evaluate the segmentation results. We defined F1 scores greater than 0.95 as very accurate, between 0.93 and 0.95 as accurate, between 0.90 and 0.93 as moderately accurate, between 0.85 and 0.90 as slightly inaccurate, and less than 0.85 as very inaccurate, in order to evaluate the accuracy of all segmentation results. We conducted the statistical analysis of all test set outputs, and the results are shown in Figure 5. As can be seen from the figure, the number of samples that are very accurate in the computed results of our method model is much higher than that of other models, while the number of samples that are very inaccurate and slightly inaccurate is much lower than that of other models. Therefore, we believe that our method has better overall performance and stability.

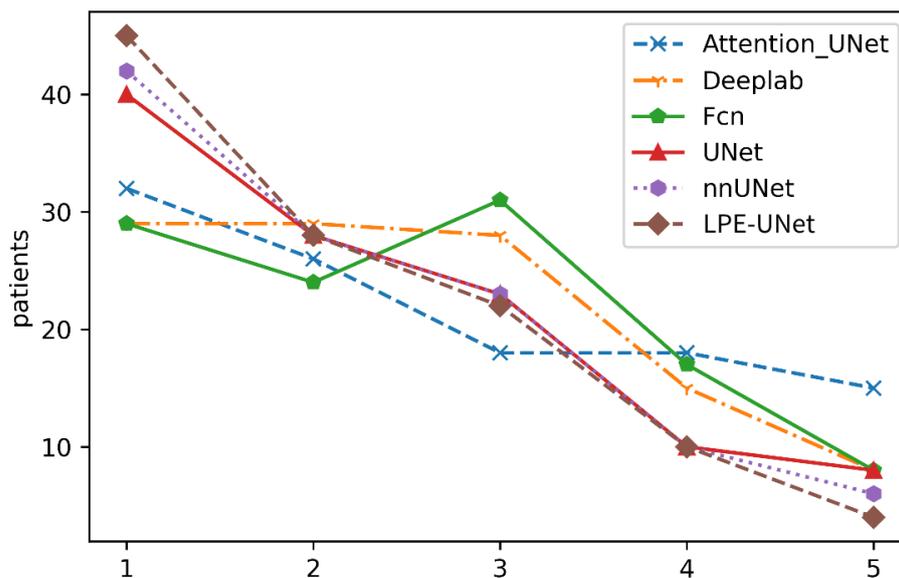


Figure 5. Qualitative evaluation results of test set CT images. The vertical axis is the number of samples, and the horizontal axis is the evaluation level. From 1 to 5 on the horizontal axis represent very accurate, accurate, moderately accurate, slightly inaccurate, and very inaccurate, respectively.

To further validate the effectiveness of the proposed method, we visually compared the segmentation results of our method with other methods. As shown in Figure 6, we selected eight different test images, including two images with small targets, three images with multiple small lesion targets, and three images with larger targets. Overall, compared with other models, the method proposed in this paper is more consistent with the Ground Truth for the segmentation results of image lesion edge pixels, with better lesion edge integrity, more accurate segmentation, and better visual effects, which verifies the effectiveness and superiority of the method proposed in this paper. Compared with other methods for lesion segmentation, due to the role of the noise optimization module and mixed loss function, our method has a more accurate division of image contours and lesion boundaries, while effectively reducing the possibility of false positives. Our method has improved the segmentation ability for smaller targets of calcified plaques, with more accurate segmentation boundaries and effectively suppressing false positive errors.

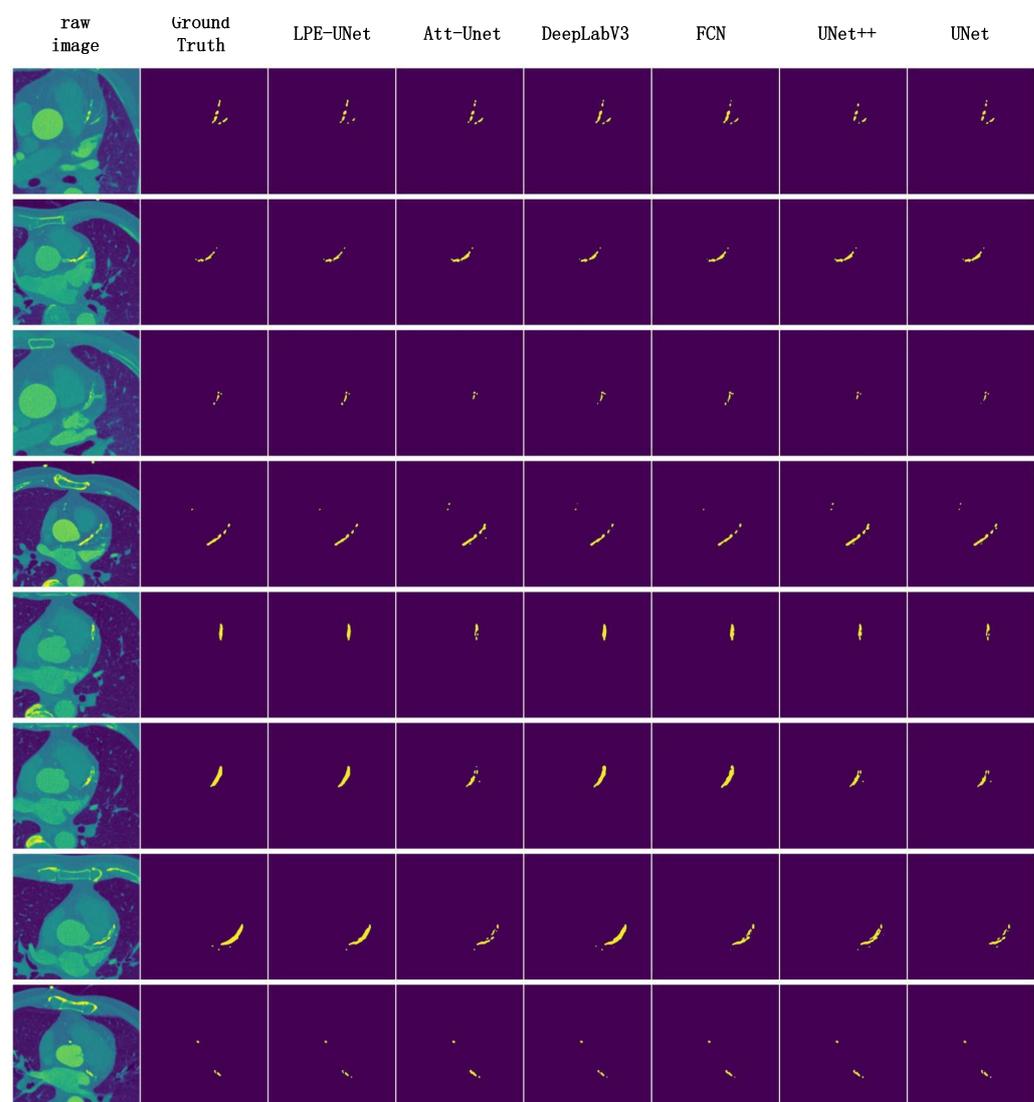


Figure 6. Comparison chart of experimental results.

We perform ablation experiments on the proposed model to verify the effectiveness of the algorithm improvements. In our experiments, we use the UNet network with one reduced downsampling operation as our baseline network and gradually add the proposed modules and training strategy to verify the effectiveness of the low-rank perception enhancement module, noise optimization module, and mixed training strategy, respectively. The results of the ablation experiments are shown in Table 4. From the experimental results,

it is clear that each time we add a new part to the network, our model further improves the accuracy of the segmentation results. In Experiment 2, we added the multi-scale feature extraction module (MSF) to the network to expand the perceptual field and aggregate different level features, while avoiding feature loss caused by the pooling layer, and the experimental result of the Iou metric improved by 0.7% (F1_score improved by 0.46% and Dice score improved by 0.43%), which shows that our insight works. In Experiment 3 we added the noise optimization modules (NOM) at the base of Experiment 2 to suppress the effect of irrelevant features in the low-level features passed through the skip connections, and the experimental result of Iou was further improved by 0.21% (F1_score was improved by 0.35% and Dice score improved by 0.20%). In Experiment 4, a hybrid attention mechanism is added to Experiment 3. The multi-scale feature extraction module and the hybrid attention mechanism together form the low-rank perception enhancement module (LPE), the network is trained using Dice Loss, and the experimental effect F1_score is further improved by 0.25%, Iou score is improved by 0.15%, and Dice score is improved by 0.14%. It can be verified that the global dependencies and inter-channel dependencies established by the attention mechanism play an important role in achieving accurate segmentation of small target objects in medical images. In Experiment 5, we first use multiple iterations of DWCE Loss to adequately pre-train the network and perform preliminary optimization on the easy-to-learn parts of the network to avoid the oscillation phenomenon when training directly with Dice Loss in the case of small batch and then use Dice Loss to live the final training results, and the experimental results prove the effectiveness of the hybrid loss function. DWCE Loss has a stronger learning ability compared with the Dice Loss function, which easily leads to the transitional fitting of the training set. In the experiment, we added an extra dropout layer at the end of the network to reduce the degree of overfitting. The results of the ablation experiments fully verify the effectiveness of each module and the loss function.

Table 3. Comparative test results (LPE-UNet is the method of this paper).

Model	F1	Iou	Dice
UNet	0.9272	0.8740	0.9328
FCN	0.9200	0.8673	0.9289
DeepLabV3+	0.9278	0.8744	0.9330
Attention-Unet	0.9277	0.8747	0.9332
Unet++	0.9207	0.8657	0.9280
R2U_UNet	0.8034	0.7400	0.8506
nnUNet	0.9310	0.8790	0.9356
LPE-UNet	0.9410	0.8950	0.9446

Table 4. Ablation experiment results, where MSF, LPE, Nom, Dice represent the multi-scale feature extraction module, the low-rank perception enhancement module, the noise optimization module, and Dice Loss. LPE-UNet is the work we proposed.

Model	F1	Iou	Dice
Baseline	0.9319	0.8808	0.9366
Baseline + MSF	0.9365	0.8878	0.9406
Baseline + MSF + AGate	0.9386	0.8913	0.9425
Baseline + LPE + AGate + Dice	0.9401	0.8938	0.9439
LPE-UNet	0.9410	0.8950	0.9446

5. Discussion

In this work, we introduce a novel network framework called LPE-UNet, which is specifically designed for segmenting small-scale calcified plaques. Given the limited area occupied by calcified plaques, it becomes crucial to leverage rich and clean contextual information for accurate segmentation. To address this challenge, our framework focuses on extracting multi-scale contextual information and incorporating attention mechanisms

to weigh the extracted features, enhancing the model's ability to locate the target accurately. By combining shallow and deep features at each level of the network, we aim to enhance the detection accuracy of the framework. To achieve this, we calculate an attention matrix for the shallow features, which serves to supplement filtered spatial detail information to the deep features. However, the level-by-level fusion of shallow and deep features is not the optimal choice, as the expressive power of the features can gradually dilute during the process of feature transmission. To address this limitation, our future work aims to explore alternative approaches by incorporating more skip connections that combine high-level semantics from different-scale feature maps with low-level semantics. We believe that this integration will help preserve and enhance the expressive power of the features throughout the network, leading to improved segmentation results. Furthermore, in the context of CT images, calcified plaques exhibit noticeable continuity; however, treating each CT image independently discards valuable localization information. To tackle this issue, we propose treating adjacent CT images as a single sample, enabling more effective utilization of the inherent continuity of calcified plaques. This approach is also part of our future work, as we anticipate that it will further enhance the segmentation performance. Through these enhancements, we expect to achieve more accurate and robust segmentation results in our future research endeavors.

6. Conclusions

In this paper, we propose an improved UNet network, called LPE-UNet, to achieve the segmentation of small-scale calcified plaques. For the model, on the one hand, we construct a filter gate mechanism on the skip connection to suppress irrelevant feature information from interfering with the segmentation result. On the other hand, we solve the contradiction between the use of pooling layers to enlarge the receptive field and the loss of features in traditional networks by building a dependency enhancement module. This also provides the model with the ability to capture and fuse multi-scale information and model long-term dependencies and channel dependencies in the model. In addition to improving the model, we propose a new weighted cross-entropy loss function based on the distance between pixels in the image and pixels between lesions to weight the loss function, which suppresses the imbalance between foreground and background classes caused by the small segmentation targets. We use this with Dice Loss to conduct mixed training on the model to obtain a more accurate segmentation result. The experimental results on the calcified plaque segmentation dataset verify the feasibility and effectiveness of the proposed network and loss function in this paper.

Author Contributions: Conceptualization, S.W. and C.Z.; Methodology, S.W. and C.Y.; Software, S.W. and C.Y.; Validation, S.W. and C.Z.; Formal analysis, S.W. and C.Y.; Investigation, C.Y.; Resources, C.Z.; Data curation, S.W.; Writing—original draft preparation, S.W. and C.Y.; Writing—review and editing, C.Z.; Visualization, S.W.; Supervision, C.Y. and C.Z.; Project administration, C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the National Natural Science Foundation of China (NSFC) Joint Fund with Zhejiang Integration of Informatization and Industrialization under Key Project (Grant No. U22A2033) and NSFC (Grant No. 62072281).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source of the data is stated in the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Virani, S.S.; Alonso, A.; Benjamin, E.J.; Bittencourt, M.S.; Callaway, C.W.; Carson, A.P.; Chamberlain, A.M.; Chang, A.R.; Cheng, S.; Delling, F.N.; et al. Heart disease and stroke statistics—2020 update: A report from the American Heart Association. *Circulation* **2020**, *141*, e139–e596. [[CrossRef](#)] [[PubMed](#)]
2. Lessmann, N.; van Ginneken, B.; Zreik, M.; de Jong, P.A.; de Vos, B.D.; Viergever, M.A.; Išgum, I. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Trans. Med. Imaging* **2017**, *37*, 615–625. [[CrossRef](#)] [[PubMed](#)]
3. Rahim, T.; Usman, M.A.; Shin, S.Y. A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging. *Comput. Med. Imaging Graph.* **2020**, *85*, 101767. [[CrossRef](#)] [[PubMed](#)]
4. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
5. He, J.; Zhu, Q.; Zhang, K.; Yu, P.; Tang, J. An evolvable adversarial network with gradient penalty for COVID-19 infection segmentation. *Appl. Soft Comput.* **2021**, *113*, 107947. [[CrossRef](#)] [[PubMed](#)]
6. Luo, W.; Li, Y.; Urtasun, R.; Zemel, R. Understanding the effective receptive field in deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; p. 29.
7. Gao, Z.; Hau, W.K.; Zhang, H.; Zhang, Y.T. Automatic Detection of Calcified Plaque with Acoustic Shadowing. In Proceedings of the International Conference on Health Informatics, Verona, Italy, 15–17 September 2014; Zhang, Y.T., Ed.; pp. 197–199.
8. Yoshida, Y.; Fujisaku, K.; Sasaki, K.; Yuasa, T.; Shibuya, K. Semi-automatic detection of calcified plaque in coronary CT angiograms with 320-MSCT. In Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016; pp. 1703–1707.
9. Sun, Q.; Yang, G.; Shu, H. Calcified coronary plaques detection in CTA based on automatic scale selection and fuzzy C means. In Proceedings of the 2016 International Conference on Machine Learning and Cybernetics (ICMLC), Jeju Island, Republic of Korea, 10–13 July 2016; Volume 2, pp. 807–813.
10. Wolterink, J.M.; Leiner, T.; Viergever, M.A.; Išgum, I. An automatic deep learning approach for coronary artery calcium segmentation. In Proceedings of the MICCAI, Munich, Germany, 5–9 October 2015; Volume 9349.
11. Lessmann, N.; Išgum, I.; Setio, A.A.; de Vos, B.D.; Ciompi, F.; de Jong, P.A.; Oudkerk, M.; Willem, P.T.M.; Viergever, M.A.; van Ginneken, B. Deep convolutional neural networks for automatic coronary calcium scoring in a screening study with low-dose chest CT. In Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis, San Diego, CA, USA, 27 February–3 March 2016; Volume 9785, pp. 255–260.
12. Lekadir, K.; Galimzianova, A.; Betriu, A.; del Mar Vila, M.; Igual, L.; Rubin, D.L.; Fernández, E.; Radeva, P.; Napel, S. A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 48–55. [[CrossRef](#)] [[PubMed](#)]
13. Santini, G.; Latta, D.D.; Martini, N.; Valvano, G.; Gori, A.; Ripoli, A.; Susini, C.L.; Landini, L.; Chiappino, D. An automatic deep learning approach for coronary artery calcium segmentation. In *EMBECE & NBC 2017*; Springer: Singapore, 2018; pp. 374–377.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; pp. 234–241.
15. Shadmi, R.; Mazo, V.; Bregman-Amitai, O.; Elnekave, E. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 24–28.
16. Ma, J.; Zhang, R. Automatic calcium scoring in cardiac and chest CT using DenseRAUnet. *arXiv* **2019**, arXiv:1907.11392.
17. Zhang, W.; Zhang, J.; Du, X.; Zhang, Y.; Li, S. An end-to-end joint learning framework of artery-specific coronary calcium scoring in non-contrast cardiac CT. *Computing* **2019**, *101*, 225581–225593. [[CrossRef](#)]
18. Lee, J.; Gharaibeh, Y.; Kolluru, C.; Zimin, V.N.; Dallan, L.A.P.; Kim, J.N.; Bezerra, H.G.; Wilson, D.L. Segmentation of Coronary Calcified Plaque in Intravascular OCT Images Using a Two-Step Deep Learning Approach. *IEEE Access* **2020**, *8*, 225581–225593. [[CrossRef](#)] [[PubMed](#)]
19. Li, Y.C.; Shen, T.Y.; Chen, C.C.; Chang, W.T.; Lee, P.Y.; Huang, C.C.J. Automatic Detection of Atherosclerotic Plaque and Calcification From Intravascular Ultrasound Images by Using Deep Convolutional Neural Networks. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* **2021**, *68*, 1762–1772. [[CrossRef](#)] [[PubMed](#)]
20. Xiao, X.; Lian, S.; Luo, Z.; Li, S. Weighted Res-UNet for High-Quality Retina Vessel Segmentation. In Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 327–331.
21. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018; pp. 3–11.
22. Xiao, X.; Lian, S.; Luo, Z.; Li, S. Dilated convolutions for brain tumor segmentation in mri scans. In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI), Quebec City, QC, Canada, 10–14 September 2017; pp. 253–262.

23. Zhao, C.; Vij, A.; Malhotra, S.; Tang, J.; Tang, H.; Pienta, D.; Xu, Z.; Zhou, W. Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Comput. Biol. Med.* **2021**, *1323*, 104667. [[CrossRef](#)] [[PubMed](#)]
24. Mu, N.; Lyu, Z.; Rezaeitalshmahalleh, M.; Tang, J.; Jiang, J. An attention residual u-net with differential preprocessing and geometric postprocessing: Learning how to segment vasculature including intracranial aneurysms. *Med. Image Anal.* **2023**, *84*, 102697. [[CrossRef](#)] [[PubMed](#)]
25. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
26. Kaul, C.; Manandhar, S.; Pears, N. Focusnet: An Attention-Based Fully Convolutional Network for Medical Image Segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 455–458.
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017.
28. Isensee, F.; Jaeger, P.F.; Kohl, S.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [[CrossRef](#)] [[PubMed](#)]
29. Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv* **2018**, arXiv:1802.06955.
30. Islam, M.T.; Khan, H.A.; Naveed, K.; Nauman, A.; Gulfam, S.M.; Kim, S.W. LUVS-Net: A Lightweight U-Net Vessel Segmentor for Retinal Vasculature Detection in Fundus Images. *Electronics* **2023**, *12*, 1786. [[CrossRef](#)]
31. Peng, J.; Li, Y.; Liu, C.; Gao, X. The Circular U-Net with Attention Gate for Image Splicing Forgery Detection. *Electronics* **2023**, *12*, 1451. [[CrossRef](#)]
32. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
33. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
34. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
35. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.