frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 02 November 2023
DOI 10.3389/fphys.2023.1209659

Check for updates

OPEN ACCESS

EDITED BY
Guangyu Zhu,
Xi'an Jiaotong University, China

REVIEWED BY
Yong Wang,
Max Planck Society, Germany
Helton Bíscaro,
University of São Paulo, Brazil

*CORRESPONDENCE
Jingfeng Jiang,
✉ jjiang1@mtu.edu

†PRESENT ADDRESS
Nan Mu,
School of Computer Science, Sichuan Normal University, Chengdu, Sichuan, China

RECEIVED 21 April 2023
ACCEPTED 25 September 2023
PUBLISHED 02 November 2023

CITATION
Mu N, Lyu Z, Rezaeitaleshmahalleh M, Bonifas C, Gosnell J, Haw M, Vettukattil J and Jiang J (2023), S-Net: a multiple cross aggregation convolutional architecture for automatic segmentation of small/thin structures for cardiovascular applications.
Front. Physiol. 14:1209659.
doi: 10.3389/fphys.2023.1209659

S-Net: a multiple cross aggregation convolutional architecture for automatic segmentation of small/thin structures for cardiovascular applications

Nan Mu[1,2†], Zonghan Lyu[1,2], Mostafa Rezaeitaleshmahalleh[1,2], Cassie Bonifas[1,2], Jordan Gosnell[3], Marcus Haw[3], Joseph Vettukattil[1,3] and Jingfeng Jiang[1,2]*

[1]Department of Biomedical Engineering, Michigan Technological University, Houghton, MI, United States, [2]Center for Biocomputing and Digital Health, Health Research Institute, Institute of Computing and Cybernetics, Michigan Technological University, Houghton, MI, United States, [3]Betz Congenital Health Center, Helen DeVos Children's Hospital, Grand Rapids, MI, United States

With the success of U-Net or its variants in automatic medical image segmentation, building a fully convolutional network (FCN) based on an encoder-decoder structure has become an effective end-to-end learning approach. However, the intrinsic property of FCNs is that as the encoder deepens, higher-level features are learned, and the receptive field size of the network increases, which results in unsatisfactory performance for detecting low-level small/thin structures such as atrial walls and small arteries. To address this issue, we propose to keep the different encoding layer features at their original sizes to constrain the receptive field from increasing as the network goes deeper. Accordingly, we develop a novel S-shaped multiple cross-aggregation segmentation architecture named S-Net, which has two branches in the encoding stage, i.e., a resampling branch to capture low-level fine-grained details and thin/small structures and a downsampling branch to learn high-level discriminative knowledge. In particular, these two branches learn complementary features by residual cross-aggregation; the fusion of the complementary features from different decoding layers can be effectively accomplished through lateral connections. Meanwhile, we perform supervised prediction at all decoding layers to incorporate coarse-level features with high semantic meaning and fine-level features with high localization capability to detect multi-scale structures, especially for small/thin volumes fully. To validate the effectiveness of our S-Net, we conducted extensive experiments on the segmentation of cardiac wall and intracranial aneurysm (IA) vasculature, and quantitative and qualitative evaluations demonstrated the superior performance of our method for predicting small/thin structures in medical images.

KEYWORDS

automatic segmentation, cardiac wall, intracranial aneurysm (IA), small/thin structure, fully convolutional network (FCN)

1 Introduction

Cardiovascular disease is one of the leading causes of death, accounting for one death every 34 s in the United States[1]. A comprehensive understanding of cardiac function and brain vessel integrity is essential for preventing, diagnosing, and treating this life-threatening disease. Automatic medical image segmentation of cardiac imaging data plays a crucial role in medical 3D printing (Lindquist et al., 2021), computer modeling (Jou et al., 2003; Cebral et al., 2005), and computer-aided diagnosis of cardiovascular systems (Lindquist Liljeqvist et al., 2021; Sunderland et al., 2021; Rezaeitaleshmahalleh et al., 2023c), assisting cardiologists, radiologists, and surgeons in making clinical decisions efficiently.

In the last 2 decades, considerable research efforts have been devoted to computational hemodynamics (Jou et al., 2003; Cebral et al., 2005). Accurate "patient-specific" vasculature segmentation from 3D imaging data is critical for subsequent numerical evaluations of each patient's hemodynamic environment. However, an automated workflow for model creation in computational hemodynamics still requires considerable attention (Mu et al., 2023a; Rezaeitaleshmahalleh et al., 2023b; Lyu et al., 2023), particularly in delineating small brain arteries (approximately 0.5 mm in diameter). Similarly, although considerable research has been devoted to segmenting cardiac imaging data (Suri, 2000; Petitjean and Dacher, 2011), whole heart wall automatic segmentation remains challenging because the thickness of the atrial wall is particularly thin (approximately 1–4 mm, and down to 0.5 mm in pediatric populations). Existing studies have only focused on myocardial wall segmentation of the left and right ventricles (Zhu et al., 2013; Yang et al., 2018). Those unmet needs motivate the work presented in this study.

Early medical image segmentation methods include active contours (Xu and Prince, 1998), template matching (Lalonde et al., 2001), edge detection (Zhao et al., 2006), shape modeling (Tsai et al., 2003), machine learning (Zhang et al., 2004), etc. More recently, deep-learning-based methods have been developed to extract abundant and powerful data-specific features from cardiac and brain imaging data. Typically, most CNN models developed for medical image segmentation are of the encoder-decoder type, which is one of the most popular end-to-end architectures, e.g., fully convolutional network (FCN) (Yuan et al., 2017), U-Net (Ronneberger et al., 2015), and their variants (Bhalerao and Thakur, 2019; Isensee et al., 2021). Among these structures, the encoder is usually deployed to progressively extract higher-level medical image features. At the same time, the decoder is generally employed for recovering and integrating the extracted (multi-scale) features back to the original image size. Eventually, this end-to-end configuration generates the final segmentation result. Theoretically, as the network goes deeper, more high-level features are extracted at the expense of losing low-level detail information. Although skip connections generally help propagate local features from the encoder to the decoder, they still fail to adequately capture small/thin anatomical structures in cardiovascular application. This shortcoming is well noted in the literature (Mu et al., 2023a).

Currently, the U-shaped architecture in the classic U-Net model (Ronneberger et al., 2015) achieves satisfactory performance in segmenting large structures but suffers from significant limitations. Recall that the U-shaped architecture contains both top-down and bottom-up paths (see Figure 1A). First, a large amount of spatial detail information is lost during the downsampling of the bottom-up path and cannot be easily recovered. Second, the top-down path may gradually weaken the global context information in high-level features, resulting in incomplete segmentation results. Third, making predictions only for high-resolution feature maps with weaker semantics inevitably limits the expressiveness of small/thin object recognition. Existing approaches address these problems by introducing an attention mechanism (Mu et al., 2023a) or self-attention (Valanarasu et al., 2021a) into the U-shaped structures, refining the feature map in a recursive (Alom et al., 2019) or cascading (Mu et al., 2023b) manner, etc. However, the abovementioned newer networks are configured to focus only on high-level features and cannot detect some low-level detail structures.

In this paper, we investigate an alternative approach, i.e., controlling the receptive field size of the convolutional filters in the encoder. Thus, the new configuration can effectively guide a CNN model to capture fine-grained details and small/thin structures even in deeper layers. Specifically, we discard the traditional U-shaped framework and propose constructing two resampling and downsampling branches with an S-shaped cross aggregation to control receptive field size. As a result, we name our new network configuration S-Net, which enables learning of low-level detail information and high-level discriminative knowledge in the encoding stage. As shown in Figure 1B, compared with the bottom-up encoder (see Figure 1A), the proposed S-shaped network adds resampling layers to the encoder to avoid spatial information loss while ensuring a small receptive field for capturing low-level details during convolution.

Furthermore, considering the characteristics and complementarity of downsampling and resampling branches in our S-Net, we design a simple multiple cross aggregation module (MCAM) to efficiently integrate multilevel high-level and low-level features efficiently, ensuring robust and comprehensive feature representation. In particular, we also configure dilated convolutions of different specifications in the encoder to extract rich global context information and propagate it to the top-down decoder through lateral connections to strengthen the global dependence of decoding features. Last, we perform supervised prediction of features on all decoding layers, which combines semantically strong, low-resolution features with semantically weak, high-resolution features along the top-down path. The new S-Net design is anticipated to enhance the segmentation of large and small anatomical targets, and its architecture is illustrated in Figure 2.

In this study, we perform segmentation experiments of the heart wall and intracranial aneurysm (IA) vasculature to demonstrate the merits of the proposed S-Net. The whole heart wall segmentation requires dealing with thin structures, while segmenting brain vasculature with IAs involves small vessel annotations. The success of both applications is highly dependent on the ability of CNNs to discriminate fine-grained structures. The performance of the proposed S-Net is compared with five other state-of-the-art medical segmentation models.

[1] Centers for Disease Control and Prevention, National Center for Health Statistics (NCHS). About Multiple Cause of Death, 1999–2020. CDC WONDER Online Database website. Atlanta, GA: Centers for Disease Control and Prevention; 2022. Accessed February 21, 2022.

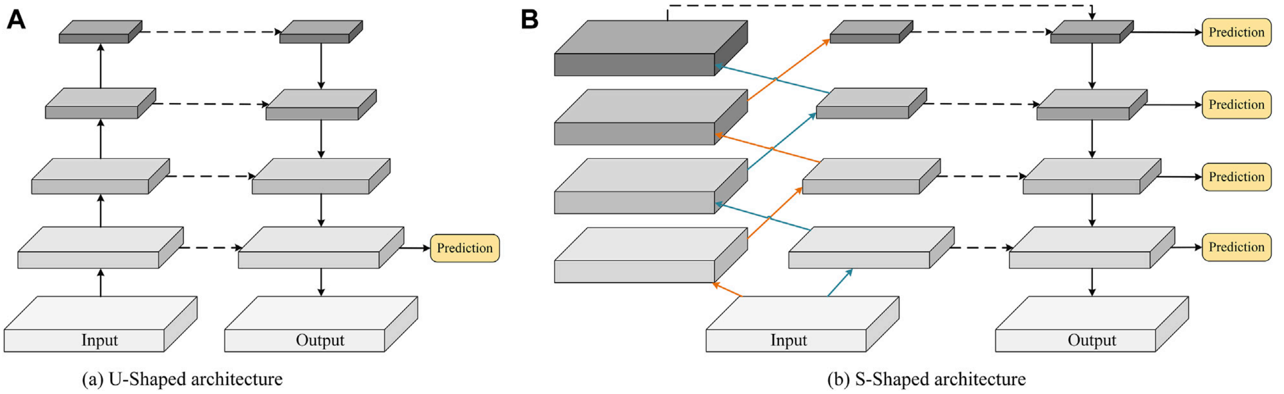

FIGURE 1
Schematic comparison of **(A)** the conventional U-shaped structure and **(B)** the proposed S-shaped structure.

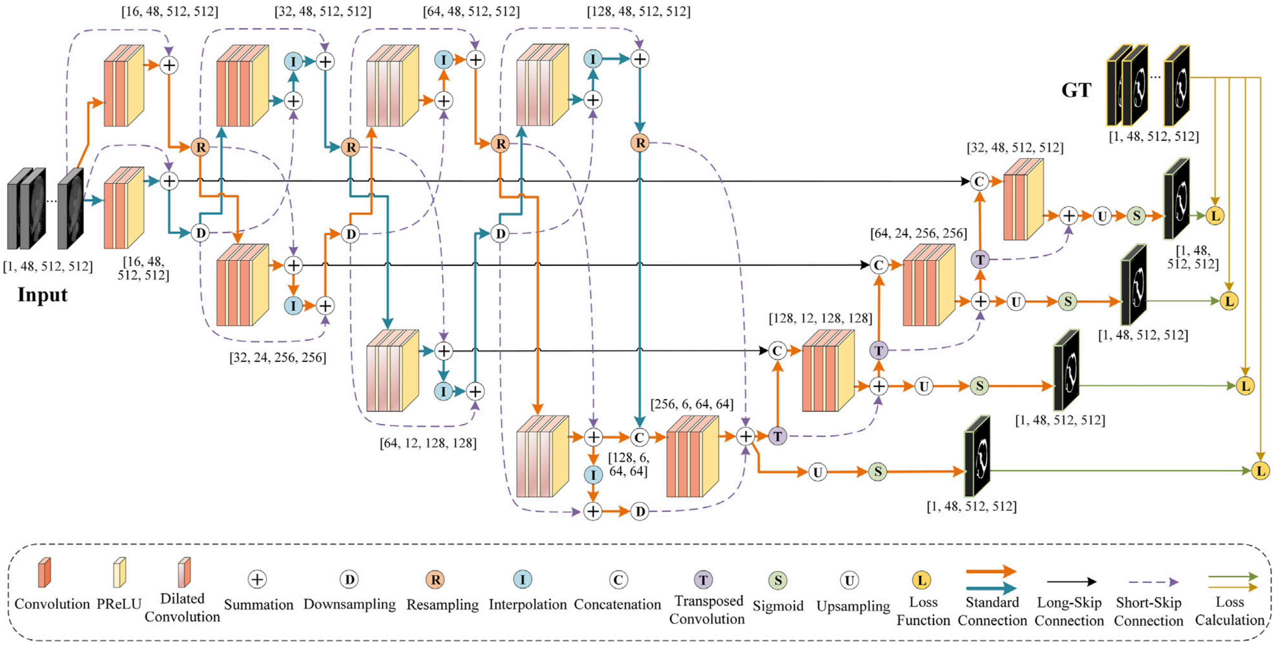

FIGURE 2
An illustrative diagram showing the architecture of the proposed S-Net.

Overall, our contributions are summarized as follows. First, we propose an efficient dual-branch encoder to explore spatial details and global contextual information. Moreover, we design a multiple cross-aggregation module to exploit the complementarity of these two branches for improving the feature representation capability on small structures. Second, we improve the learning effectiveness of S-Net by optimizing semantic and localization knowledge based on top-down multilevel supervision. Third, we experimentally explore the proposed S-Net for image segmentation of the heart wall and IA vasculature, demonstrating that our model achieves state-of-the-art performance both quantitatively and qualitatively, especially for small volume prediction.

2 Related works

This section briefly reviews U-Net-based medical image segmentation architectures and cardiac wall segmentation models.

2.1 U-Net architecture

Since its proposal, U-Net (Ronneberger et al., 2015), with an encoder-decoder structure, has inspired substantial further developments. Its variants have become widely adopted tools for medical image segmentation tasks. By extending the dimensionality of the basic U-Net framework, 3D U-Net (Çiçek et al., 2016) enables 3D volume segmentation, which has been of significant effect in many biomedical applications. Some attention-based U-Nets (Oktay et al., 2018; Mu et al., 2023a) emphasize local information by employing attention units to allow the network to focus on specific objects of importance while ignoring unnecessary regions. To segment targets of various sizes and shapes, inception U-Nets (Chen et al., 2018; Zhang et al., 2020) utilize convolutional filters of multiple sizes in the same network layer to analyze images with different salient regions efficiently. To overcome the training difficulty of convergence caused by the degradation of deep CNN features, residual U-Nets (Bhalerao and Thakur, 2019; Yu et al., 2019) acquire feature maps from one network layer and add them to a deeper layer to improve the performance. In addition, recurrent U-Nets (Alom et al., 2019; Jiang et al., 2020) optimize the expressiveness of the feature maps by merging recurrent feedback loops into the convolutional layer to obtain the context of adjacent units. To compensate for the information loss in the deeper layers of CNNs, dense U-Nets (Li et al., 2018; Wang et al., 2019) construct identity mappings for each layer, which depend not only on the former layer but also on all previous layers.

In summary, those U-Net variants have been designed to optimize the segmentation results by tuning the network structure, introducing new modules, etc. However, those variants largely follow the encoder-decoder design constructed by downsampling and upsampling convolutional layers (See Figure 1A). Recall that the existing U-shaped configuration is prone to spatial information loss and insufficient knowledge acquisition of small/thin targets. Thus, in this study, our work is fundamentally different than that of prior publications.

2.2 Cardiac wall segmentation

Cardiac wall segmentation faces two main challenges. First, the mixture of contrast, blood, and dynamic myocardial structures causes blurred boundaries of the heart wall, making them difficult to be distinguished. Second, the atrial wall, interatrial septum, and cardiac valves are thin and irregularly shaped, which makes them highly unrecognizable. Although there have been a wealth of studies addressing segmentation of the whole heart (Zhuang et al., 2010; Xu et al., 2019), four chambers (Zheng et al., 2008), left and/or right ventricle (Ringenberg et al., 2014; Lu et al., 2019), and left and/or right atrium (Tobon-Gomez et al., 2013; Tobon-Gomez et al., 2015), there appear to be no studies of whole heart wall segmentation.

Notably, Zhu et al. proposed an automated method based on variational region growth to segment the myocardial walls of right and left ventricles using cardiac computed tomography (CT) images (Zhu et al., 2013). Yang et al. presented a multi-component deformable model combined with 2D-3D U-Net for segmenting the ventricular walls from cardiac magnetic resonance imaging (MRI) (Yang et al., 2018). Ye et al. applied a PC-UNet to segment the left ventricle myocardium wall in conjunction with CT data (Ye et al., 2021). However, none of the early research involved segmenting the whole heart wall. This is because the segmentation of the whole heart wall is challenging for the following reasons. First, the heart organ has multiple chambers and large vessels with complex geometry. The shape of the heart wall varies considerably between different subjects or in the same subject with different cardiac conditions, and this variation in shape is particularly pronounced when pathological conditions are involved. Therefore, it is difficult to accurately capture the complex shape of the whole heart wall using priori models trained from a limited training dataset. Second, depending on the intensity distribution (i.e., texture pattern) of the medical images, some boundaries between anatomical substructures are visually indistinct, e.g., the valve planes that separate the atria and great vessels from the ventricles, the boundaries between the left atrium and the pulmonary veins and between the right atrium and the superior/inferior vena cava, and the thin walls of the atria and vessels. These ill-defined boundaries make fully automated whole heart wall segmentation challenging to achieve. Segmenting thin cardiac structures like atrial walls is also a technically challenging problem. Third, the intensity distribution between some adjacent tissues or substructures is highly similar; e.g., the intensity of the myocardium is analogous to the neighboring papillary muscles, liver, and body muscles. Therefore, segmentation models relying only on image intensities have difficulty separating the heart wall from similar tissues. Finally, due to the complex motion within the heart, the imaging data may contain severe motion artifacts, interference noise, and intensity inhomogeneities, leading to unsmooth and undesirable delineation of the heart wall.

3 Proposed method

The overall pipeline of our S-Net is depicted in Figure 2. It consists of eight coding blocks, four decoding blocks, and four supervision layers, where every two coding blocks form a densely connected MCAM. In this section, we first describe the structure of our S-Net model. Later, we elaborate on the proposed MCAM, and finally, we present the specific implementation details.

3.1 S-Net architecture

The essence of the proposed S-Net is a fully convolutional network similar to the classic U-Net (Ronneberger et al., 2015), consisting of encoding and decoding, but unlike the symmetric U-shaped structure of U-Net, the information propagation path of our S-Net in the encoding stage is interleaved, approximately following an S shape. Specifically, two strategies are used at the encoding stage to encode CT scans' small/thin structures. First, we use convolution and downsampling to acquire four-layer features by gradually halving 3D sizes and increasing the number of channels. Second, we construct four-layer features with increasing channel numbers, but the 3D sizes remain the same as the input image patch for perceiving small targets. At the decode stage, we perform multiple cross-aggregation on these encoded features: Long-skip connections are used to fuse them to complement the four-layer upsampled convolutional features. Hence, abundant structural information is decoded for considerable performance gains in delineating small/thin objects.

Furthermore, we leverage the Sigmoid activation function to implement layer-by-layer prediction on the multi-scale features generated by different decoding layers and calculate the errors between the predicted results and the ground truths. Errors are

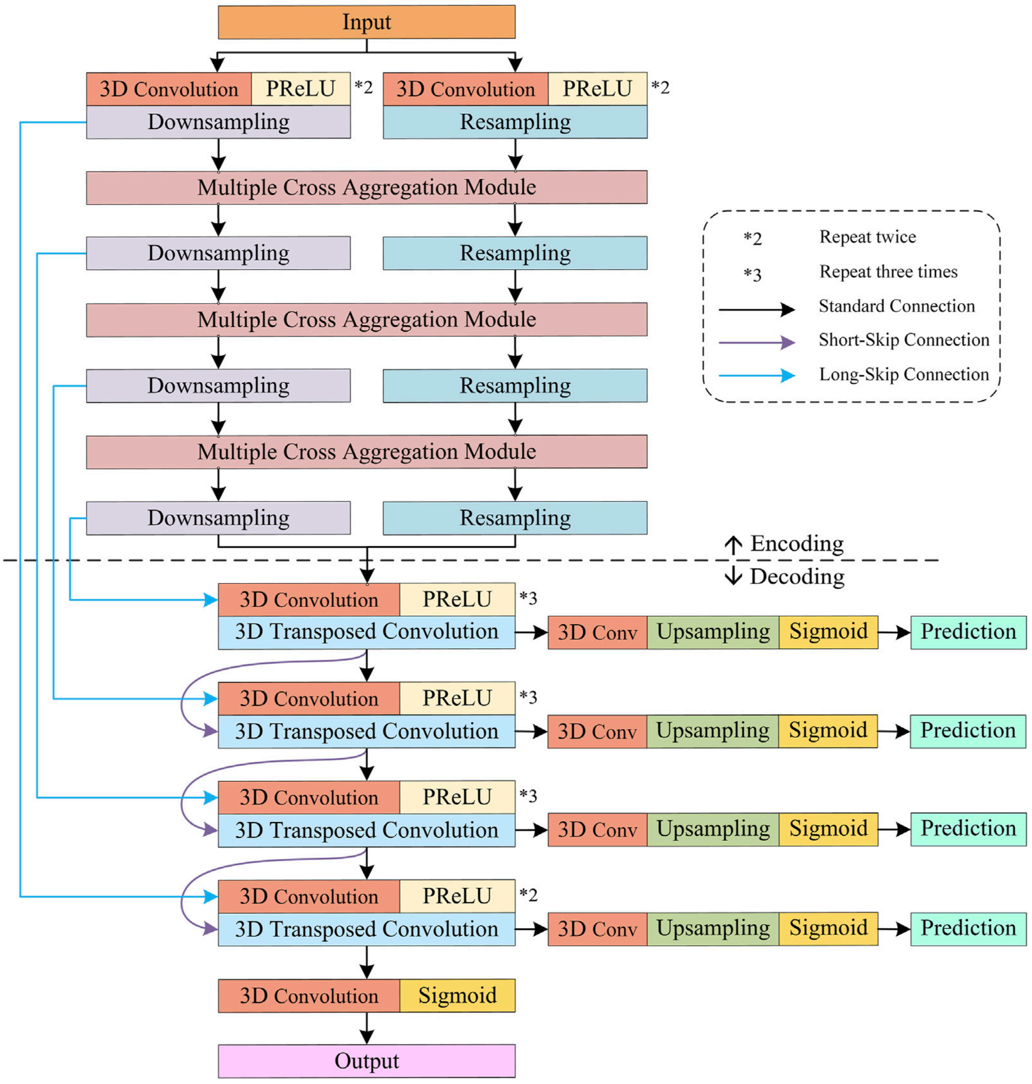

FIGURE 3
Encoding and decoding configurations of the proposed S-Net.

then back-propagated to update the model training parameters. The encoding and decoding structures of the proposed S-Net are illustrated in Figure 3.

3.1.1 Rationale

Regarding the conventional U-Net and its variants, the feature size in the encoding stage is gradually decreasing. For an input 3D image of size $D \times W \times H$, after 3D convolution and 3D pooling operations, the size of the ith layer feature map is reduced to $\frac{D}{2^i} \times \frac{W}{2^i} \times \frac{H}{2^i}$; thus, for a 3D convolution with a kernel size of 3^3, its relative receptive field with respect to the features of the ith layer becomes $(2^i \times 3)^3$. As a result, the receptive field of the convolutional layers in the encoder increases as the network deepens, making the deeper layers focus on high-level (semantic) features and, therefore, it cannot extract the knowledge for segmenting small/thin objects, fine details, etc. In light of this finding, we added a series of non-downsampled convolutional layers to the decoder, attempting to maintain a similar resolution to the input image at each layer, and interacting with the original downsampled convolutional layers to encode objects of different scales. Since the dimensions of the added convolutional layers are all kept as $D \times W \times H$, the 3^3 convolution operations maintain the same small 3^3 receptive field for each layer, enabling the network to perceive tiny structures.

3.1.2 Encoder and decoder

Encoder. Dedicated to fine-grained 3D segmentation, the encoder of the proposed S-Net contains two main branches, i.e., a downsampled branch of four convolutional blocks with decreasing resolution and a resampled branch of four convolutional blocks preserving the original resolution. In particular, each convolution block includes two or three 3D convolutions followed by a parametric rectification linear unit (PReLU) activation function. Similarly, the channel numbers in the convolutional blocks of both branches increase as the network goes deeper. In addition, the first two convolution blocks have a kernel size of 3^3, with a stride of 1 and a padding of 1. The last two convolution blocks are mainly based on four different dilated convolutions with kernel size 3^3 and stride 1, but with padding and dilation of {2, 2}, {3, 3}, {4, 4}, and {5, 5}, respectively, and their corresponding receptive fields are 3^3, 7^3, 9^3, and 11^3. Such a convolution setting undoubtedly allows the encoder to have a diverse range of receptive fields, which helps capture multi-scale targets.

Moreover, it is important to mention that for the two encoding branches, the downsampling relies on a 3D convolution with a kernel size of 2^3 and a stride of 2 for halving the 3D feature size, while the resampling is based on a 3D convolution with a kernel size of 3^3, a stride of 1, and a padding of 1 to retain the original 3D size. More importantly, the two encoding branches interact with information through MCAM to learn multi-scale objective knowledge and propagate discriminative information from the encoding layer to the decoding layer of the identical resolution through long-skip connections, contributing to recovering the lost spatial information from downsampling encoding.

Decoder. As shown in Figure 3, the decoder comprises four convolution blocks with doubled resolution and halved channel numbers, each containing two to three 3D convolutions, followed by PReLU. Specifically, all the 3D convolutions have a kernel size of 3^3, a stride of 1, and a padding of 1. Also, all four 3D transposed convolution operators have a kernel size of 2^3 and a stride of 2 to progressively recover the full spatial resolution of the network output. Subsequently, the feature maps generated by transposition convolution are 1) concatenated with the feature maps in the encoding path by long-skip connections to integrate more accurate pixel localization information and 2) provided with high-level semantic information for the next layer through short-skip connections (He et al., 2016). In the last layer, a 3D convolution with a kernel size of 1 × 1 × 1 is added to reduce the channel number of the output image to the label number, and the result is forwarded to the Sigmoid activation function to obtain the predicted voxels.

3.1.3 Loss function

Typically, the semantic information of higher-level features in the decoder is gradually diluted along the top-down path; thus, there are significant semantic gaps in the feature maps generated by multiple convolutional layers of different depths. Generally, global contextual information is gradually ignored as the feature resolution of additional decoding layers increases. Therefore, the traditional U-Net structure using only high-resolution features with weak semantics for prediction will inevitably omit small/thin targets. Given this, the proposed S-Net enriches the semantic and location information by predicting the features of different decoding layers separately, thus enhancing the representation ability for small/thin object segmentation.

Specifically, as shown in Figure 3, for the features generated by the four decoding blocks, a 1 × 1 × 1 3 D convolution is first applied to reduce the channel number of these features to 1. The feature size is adjusted to the original image size by trilinear upsampling. Finally, the prediction results are obtained by the Sigmoid activation function, which will be utilized to calculate the back-propagation errors for updating the model parameters.

Let P_i^0 and P_i^1 denote the predicted background and foreground voxels, respectively, where $i \in \{1, 2, 3, 4\}$ indicates the predictions of the four decoding layers. Meanwhile, G^0 and G^1 denote the voxels of the two labels corresponding to the ground truth (GT), respectively. Correspondingly, the proposed loss function between the predictions and the ground truth is defined as follows:

$$
\begin{aligned}
L &= \sum_{i=1}^{4} \frac{\left|G^1 \cap P_i^1\right|}{\left|G^1 \cap P_i^1\right| + \alpha\left|G^0 \cap P_i^1\right| + \beta\left|G^1 \cap P_i^0\right|} \\
&= \sum_{i=1}^{4} \frac{\mathrm{TP}}{TP + \alpha \times FP + \beta \times FN}
\end{aligned}
$$

where $|G^1 \cap P_i^1|$, $|G^0 \cap P_i^1|$, and $|G^1 \cap P_i^0|$ indicate the True Positive (TP), False Positive (FP), and False Negative (FN), respectively. The hyperparameters α and β are exploited to control the trade-off between FPs and FNs. It is worth noting that in this paper, we set α to 0.3 and β to 0.7 to emphasize FN over FP, i.e., giving more weight to Recall ($\frac{\mathrm{TP}}{\mathrm{TP+FN}}$) than to Precision ($\frac{\mathrm{TP}}{\mathrm{TP+FP}}$). Focusing more on Recall than Precision increases the probability of false detection of non-heart walls as heart walls to some extent but avoids the probability of missing detection of true heart walls. We experimentally verified that the weighting configurations of 0.3 and 0.7 for FP and FN are most effective for trading off the miss detection and false detection rates. Such a setup avoids the missed detection of small targets to a certain extent and improves the generalization ability to imbalanced data during training.

3.2 Multiple cross aggregation module

To fully exploit the feature representation capability of the down-sampling and resampling branches in the encoder, we propose a multiple cross-aggregation module to integrate the features of both branches on each encoding layer. Since the feature scales of the two branches are different and the encoded discriminative information also differs, we attempt to capture complementary features from the two branches to further improve the quality of the features learned by the deep network. The structural details of the proposed multiple cross-aggregation module can be seen in Figure 4.

As shown in Figure 4, let the feature maps from the downsampling and resampling branches be denoted as F_i^D and F_i^R, respectively, where $i \in \{1, 2, 3, 4\}$ represents the ith encoding layer. For the features F_i^D and F_i^R generated from the previous (i-1)th layer, three 3D convolutions followed by PReLUs are first performed to extract deeper features. Next, the features before and after convolution are combined based on short-skip connections for facilitating network convergence. Then, trilinear interpolation is employed to upsample/downsample the convolutional features to a specific size (i.e., the feature size of the other branch) to obtain the cross-aggregated features. Those cross-aggregated features are finally added to features of the other branches through the residual short-skip connections to generate new complementary features F_{i+1}^R and F_{i+1}^D that are forwarded to the next layer.

Formally, the processing of multiple cross-aggregation can be expressed as:

$$
\begin{aligned}
F_{i+1}^R &= F_i^R + \mathrm{Interp}\left(F_i^D + \mathrm{PReLU}\left(\mathrm{Conv}\left(F_i^D\right)\right)\right), \\
F_{i+1}^D &= F_i^D + \mathrm{Interp}\left(F_i^R + \mathrm{PReLU}\left(\mathrm{Conv}\left(F_i^R\right)\right)\right),
\end{aligned}
$$

where Conv, PReLU, and Interp denote the 3D convolution, PReLU, and trilinear interpolation operations, respectively. By extracting the complementary features from two branches of different scales, it will be beneficial to improve the network's segmentation performance, making it capable of capturing fine-grained targets.

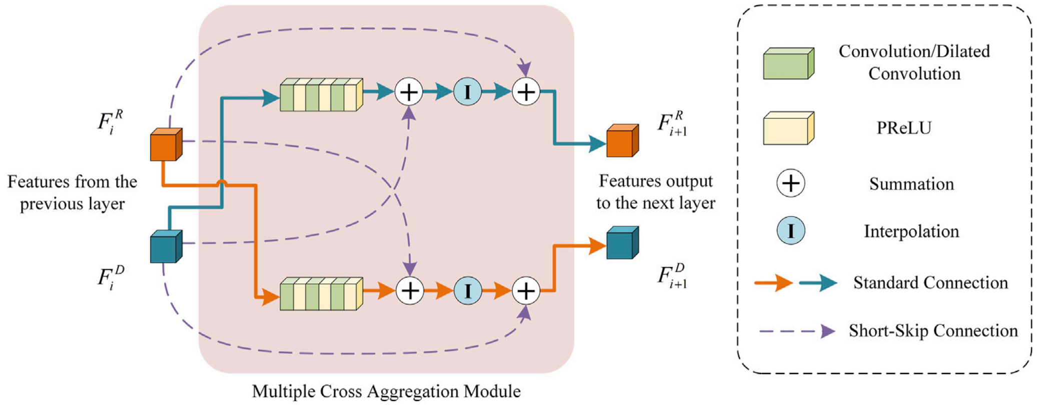
F_i^R
Features from the previous layer
F_i^D
I
I
Multiple Cross Aggregation Module
F_{i+1}^R
Features output to the next layer
F_{i+1}^D
Convolution/Dilated Convolution
PReLU
Summation
I Interpolation
Standard Connection
Short-Skip Connection

FIGURE 4
An Illustration of the proposed multiple cross-aggregation module.

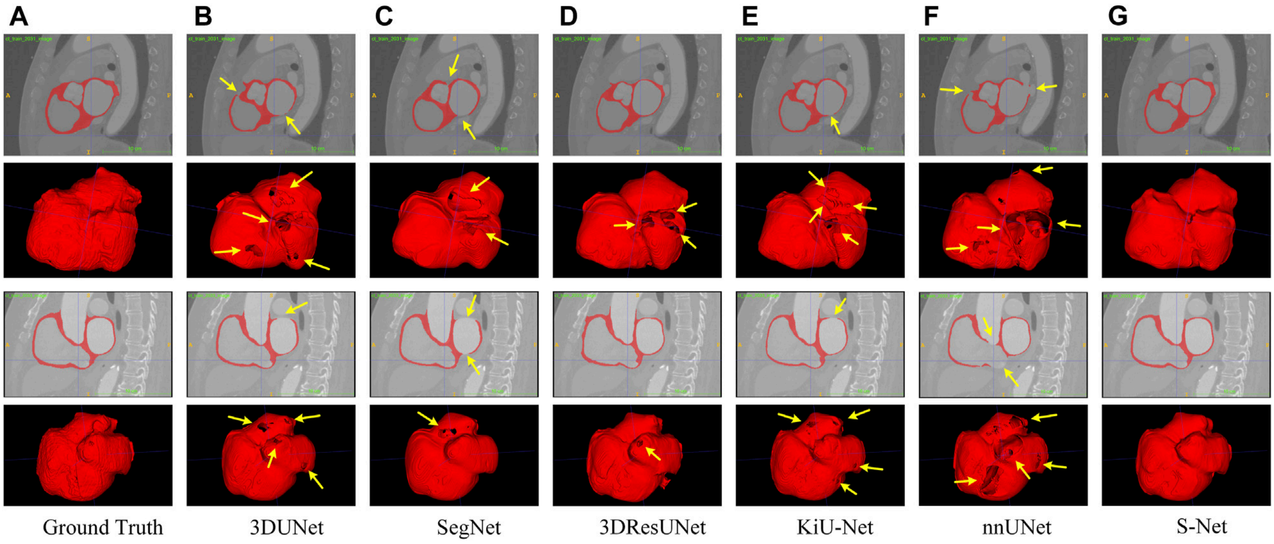

FIGURE 5
Qualitative comparison of six different models for segmenting the whole heart wall. Sagittal and 3D views are provided for each case. Arrows point to holes in the arterial wall. Images in column **(A)** are the Ground Truth and columns **(B–G)** represent the prediction results of 3DUNet, SegNet, 3DResUNet, KiU-Net, nnUNet, and our S-Net, respectively.

3.3 Implementation details

In this study, the proposed S-Net model is implemented by PyTorch framework (Version 2.1). To train and test the S-Net, we deploy patch-based learning by randomly cropping the original image along its vertical axis into a series of 48 × 512 × 512 voxel patches as the input to the network. In particular, the feature maps generated by the resampling branch in the encoder maintain the original size to cross-aggregate with the features whose size is gradually halved in the downsampling branch. Both training and testing tasks are accelerated by dual Tesla V100 PCIe GPUs with 32 GB memory. It is worth noting that when out-of-memory occurs during training, the feature size in the resampling branch will be appropriately scaled to fit the limited GPU RAM. For the specific setup, our model is trained for 1,000 epochs utilizing the Adam optimizer with a batch size of 2 and an initial learning rate of 1×10^{-4}. In addition, during training, a dropout operation is performed after each encoding and decoding layer to reduce overfitting; i.e., some elements of the output features are randomly zeroed with a probability of 0.3 using samples from the Bernoulli distribution.

4 Experiments

In this section, we first describe the experimental datasets and evaluation metrics used and then validate the superiority of the proposed S-Net through comparative experiments and ablation studies.

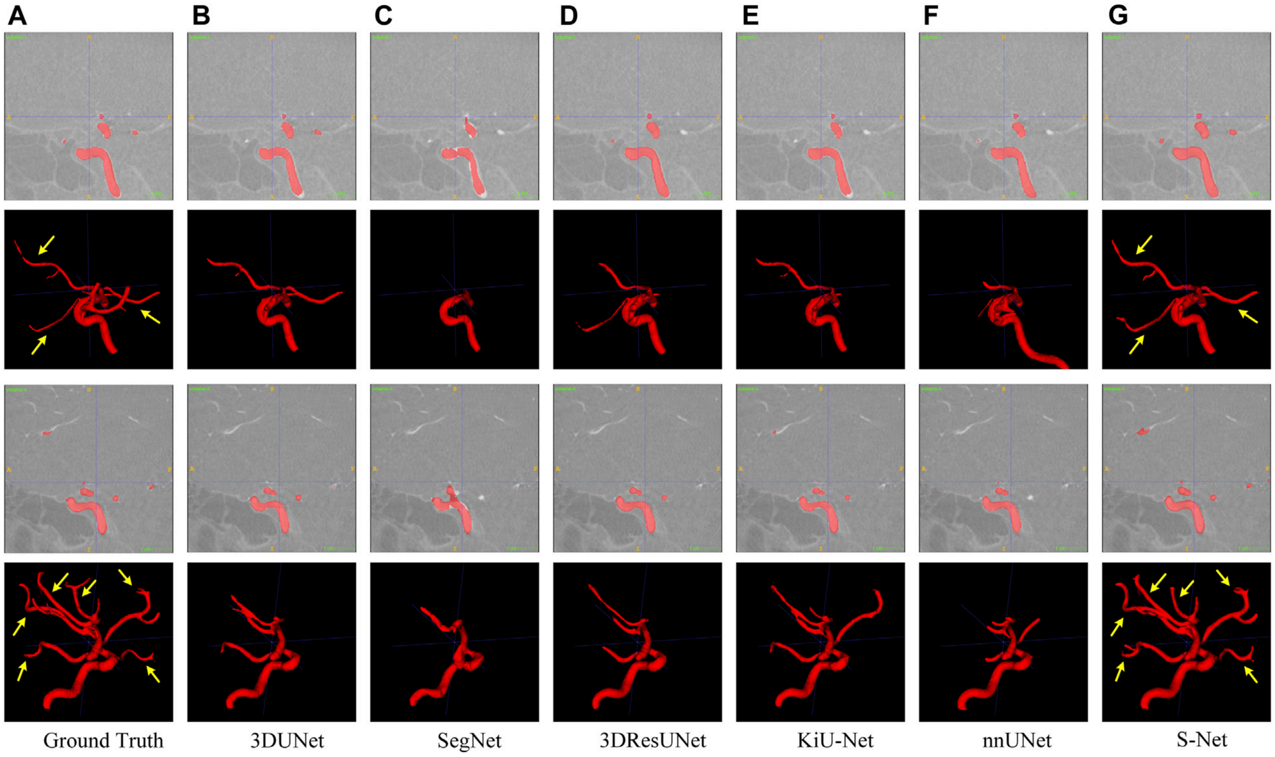

FIGURE 6
Qualitative comparison of six different models for segmenting the IA vasculature. Sagittal and 3D views are provided for each case. Arrows point to long segments of small vessels. Images in column **(A)** are the Ground Truth and columns **(B–G)** represent the prediction results of 3DUNet, SegNet, 3DResUNet, KiU-Net, nnUNet, and our S-Net, respectively.

4.1 Experimental datasets

We have fully demonstrated the application of our S-Net in two different segmentation tasks. The first task was to segment the whole heart wall in the cardiac CT data (Zhuang and Shen, 2016) provided by the Multi-Modality Whole Heart Segmentation (MM-WHS) challenge[2]. We selected 40 CT images in NIFTI format having 512 × 512 pixels with 177–363 slices, and two experienced operators manually annotated the heart walls to obtain the GT labels. Specifically, we divided the data into 30 training images and ten testing images. The second task was to segment the IA and its vasculature in the IA dataset (Mu et al., 2023a), containing 23 3D rotational angiography (3DRA) images with a resolution of 256 × 256 × 256. We used 15 cases for training and 8 cases for testing.

4.2 Evaluation metrics

We employed six widely used metrics to assess the performance of our S-Net and five other state-of-the-art deep-learning segmentation methods, including four volume measures, i.e., dice similarity coefficient (DICE), relative volume error (RVE), sensitivity, and specificity, and two surface measures in mm, i.e., 95% Hausdorff distance (HD95) and average symmetric surface distance (ASSD). A more detailed description of these evaluation metrics can be found in our previous publication (Mu et al., 2023a). All metrics were computed by comparing the predicted segmentation maps of all test data with their corresponding GTs. The mean and 95% confidence level of the calculated metrics are provided in the later experimental section.

4.3 Comparison results

To verify the effectiveness of the proposed S-Net, we compared it with five state-of-the-art 3D medical image segmentation models, including 3DUNet (Çiçek et al., 2016), SegNet (Badrinarayanan et al., 2017), 3DResUNet (Bhalerao and Thakur, 2019), KiU-Net (Valanarasu et al., 2021b), and nnUNet (Isensee et al., 2021). All algorithms are provided by respective authors of methods mentioned above, and for a fair comparison, we utilized the same experimental setup for training and testing. The average training time for our S-Net to complete each epoch was 1.227 min, while the other five comparison models consumed 0.697, 0.689, 0.855, 0.796, and 3.717 min, respectively.

Qualitative Comparison. To subjectively demonstrate the advantages of our S-Net, we provide some visual examples of the various models for heart wall and IA vasculature segmentation, as

[2] For details of the MM-WHS dataset, please refer to https://zmiclab.github.io/zxh/0/mmwhs/data.html

TABLE 1 Quantitative comparison (mean ±95% confidence level) of different models for the whole heart wall segmentation regarding six evaluation metrics. The smaller the RVE, HD95, and ASSD values, the better the segmentation effect. The best results are highlighted in bold.

Models	DICE	Sensitivity	Specificity	RVE	HD95	ASSD
3DUNet	0.8018 ± 0.0271	0.8099 ± 0.0442	0.9918 ± 0.0024	0.0806 ± 0.0373	8.8262 ± 1.5267	2.4516 ± 0.3366
SegNet	0.7856 ± 0.0251	0.7914 ± 0.0400	0.9914 ± 0.0023	**0.0759 ± 0.0385**	9.1680 ± 1.4390	2.6365 ± 0.3613
3DResUNet	0.8297 ± 0.0253	0.8234 ± 0.0461	0.9935 ± 0.0026	0.0832 ± 0.0378	7.8312 ± 1.8458	2.1058 ± 0.3469
KiU-Net	0.8126 ± 0.0252	0.7899 ± 0.0416	0.9943 ± 0.0015	0.0905 ± 0.0442	7.7801 ± 0.9845	2.1789 ± 0.2700
nnUNet	0.8381 ± 0.0270	0.7770 ± 0.0426	**0.9971 ± 0.0010**	0.1485 ± 0.0497	9.1379 ± 2.0217	1.9339 ± 0.2751
S-Net	**0.9012 ± 0.0145**	**0.8947 ± 0.0388**	0.9966 ± 0.0010	0.0762 ± 0.0280	**4.6806 ± 1.7503**	**1.2233 ± 0.2146**

TABLE 2 Quantitative comparison of different models for IA vasculature segmentation.

Models	DICE	Sensitivity	Specificity	RVE	HD95	ASSD
3DUNet	0.8294 ± 0.0292	0.7437 ± 0.0477	0.9998 ± 0.0001	0.2098 ± 0.0566	48.3566 ± 15.4985	4.9263 ± 1.9197
SegNet	0.7081 ± 0.0345	0.5998 ± 0.0477	0.9995 ± 0.0001	0.3103 ± 0.0557	36.6319 ± 12.5793	4.8530 ± 1.1163
3DResUNet	0.8050 ± 0.1050	0.8173 ± 0.0655	0.9983 ± 0.0025	0.3467 ± 0.4517	48.2625 ± 14.2790	6.1627 ± 3.9610
KiU-Net	0.8213 ± 0.0273	0.7233 ± 0.0420	**0.9998 ± 0.0001**	0.2413 ± 0.0459	45.7190 ± 13.7946	4.7272 ± 1.2201
nnUNet	0.8342 ± 0.0501	0.7920 ± 0.0725	0.9995 ± 0.0003	0.1542 ± 0.0765	56.1059 ± 18.9685	6.6677 ± 2.9604
S-Net	**0.8735 ± 0.0288**	**0.9329 ± 0.0390**	0.9990 ± 0.0003	**0.1363 ± 0.0664**	**24.7471 ± 17.5222**	**2.2070 ± 1.0808**

The best results are shown in bold fonts.

shown in Figure 5 and Figure 6. It can be easily seen that our model can generate more accurate and complete segmentation results than other compared methods. As observed in Figure 5, the surfaces of the heart walls segmented by the other five methods have large or small holes (indicated by the yellow arrows). Since quantitative assessment of cardiac function is primarily achieved by analyzing the shape attributes like heart wall thickness, enclosed area, or shape variation of the heart wall boundaries, it is crucial to completely and accurately determine the heart wall's internal (endocardial) and external (epicardial) boundaries. Although other existing models have the ability to segment the heart wall, holes on the surface and internal mis-segmentation greatly affect the evaluation of cardiac function.

In contrast, our S-Net can segment the intact heart wall. These results (see Figure 5G) indicate that our model has a more robust characterization capability to capture the structural content and surface details, which greatly contribute to the subsequent analysis of cardiac function. The comparison example illustrated in Figure 6 shows that the other models perform well for segmenting large arteries (e.g., internal carotid artery) but are prone to miss some small vessels. This observation suggests that our S-Net is more effective in modeling long-range dependencies and small structures than other CNN models, achieving segmentation maps closest to the ground truth masks. Meanwhile, the multilayer supervision also drives the network to efficiently integrate semantic and localization knowledge, which is crucial for generating high-quality prediction results.

Quantitative Comparison. Table 1 and Table 2 list the quantitative results of the cardiac wall and IA vasculature segmentation in terms of six quantitative metrics. From Table 1, we can observe that the average scores of our S-Net for almost all metrics on the testing set outperform all state-of-the-art comparative models. In terms of DICE and Sensitivity scores, our S-Net achieved gains of 0.0631 and 0.0848 compared with the second-ranked nnUNet and 3DUNet, respectively. For HD95 and ASSD scores, the surface errors between our results and GT are reduced by 4.4573 mm and 0.7106 mm, respectively, compared with the suboptimal nnUNet. For the RVE and Specificity scores, we have only 0.0003 and 0.0005 differences compared with the first-ranked SegNet and nnUNet, respectively. However, as seen in Figures 5C,E, there are conspicuous holes in the heart wall segmented by SegNet and nnUNet. Those results demonstrate that our S-Net can capture richer global context and local detail information than its counterparts (i.e., the other five CNN-based models). To verify that our S-Net can perform prediction more effectively with the Sigmoid activation function, we compared the results of training and testing using five other activation functions, namely LeakyReLU, Softmax, ReLU, PReLu, and Hyperbolic Tangent (Tanh). The results show that when the LeakyReLU, Softmax, and Tanh activation functions are used, the loss function (see Section 3.1.3) struggles to converge. The average DICE of the whole heart wall segmentation results are only 0.1348 and 0.2398, respectively, when ReLU and PReLU activation functions are used. In addition, it can be viewed from Table 2 that our S-Net is optimal for the segmentation of IA vasculature, except for the Specificity score, which is also attributed to the dual-branch cross-aggregation structure that effectively detects more complete small vessels.

4.4 Ablation studies

In this subsection, we perform a series of ablation experiments based on four variants of the proposed S-Net to validate the potential

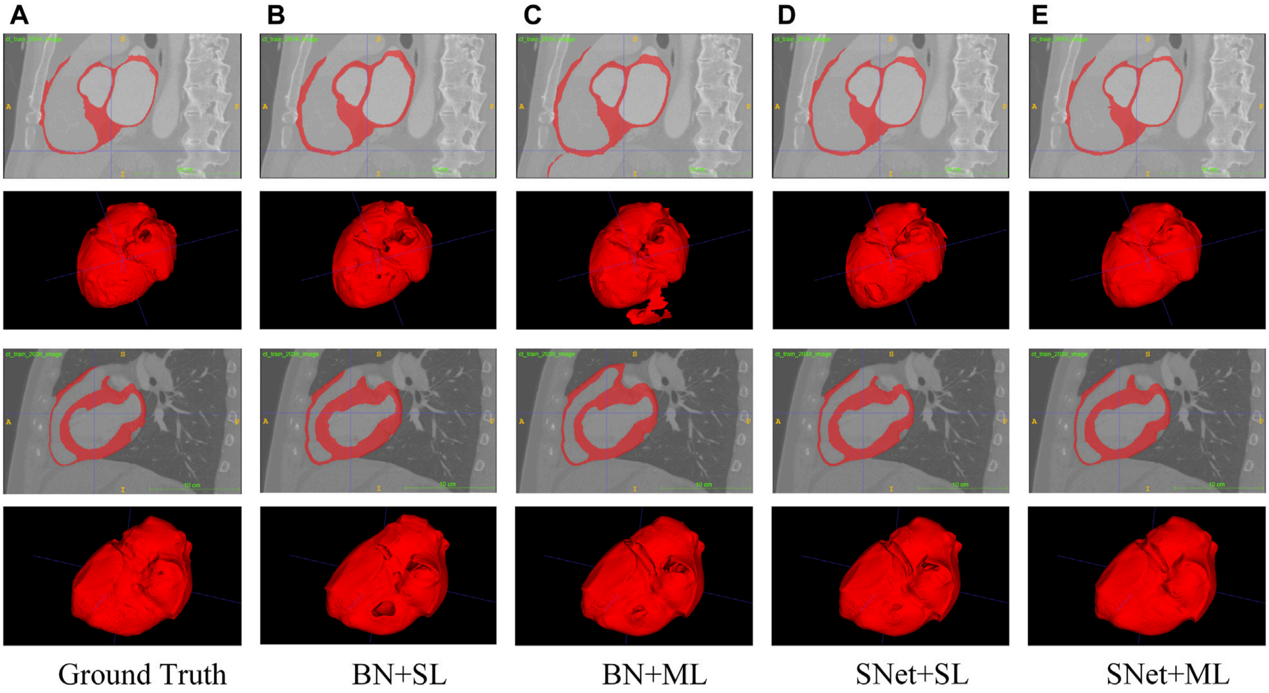

FIGURE 7
Visualization of the heart wall segmentation results generated by the four variants of the proposed S-Net. Images in column **(A)** are the Ground Truth and columns **(B–E)** represent the prediction results for baseline BN+SL, BN+ML, SNet+SL, and SNet+ML, respectively.

of our network architecture, including 1) a backbone network based on residual U-Net, which only uses the loss of a single decoding layer as supervision (denoted as BN + SL); 2) a backbone network with multiple decoding layer losses as supervision (denoted as BN + ML); 3) a dual-branch S-Net based on the loss of a single decoding layer (denoted as SNet + SL); 3) a dual branch S-Net utilizing multiple decoding layer loss (denoted as SNet + ML). All experiments were performed on the heart wall and IA vasculature datasets. Figures 7, 8 depict the subjective quality improvement of the configuration with a dual-branch encoder and a multi-supervised decoder. Tables 3 and 4 illustrate the objective results of these ablation experiments. We found that the performance improves as dual branch encoding and multi-supervised decoding are added to the network.

Effectiveness of the dual-branch encoder. From the results shown in Figures 7, 8, the two S-shaped structures (see d) and e)) are able to produce significant segmentation improvement compared to the two U-shaped baselines (see b) and c)). This implies that the crossed dual-stream encoding structure formed by adding the proposed resampling branch, attributed to the effective aggregation of global and local information, can remarkably optimize the overall segmentation performance. Specifically, the configuration of the dual-branch encoding effectively reduces the holes caused by inadequate heart wall segmentation (see b) vs. d) and c) vs. e) in Figure 7). Also, the dual-branch encoding avoids the noise (indicated by yellow arrows in Figure 8) generated by IA vasculature segmentation and achieves delineating more complete small vessels (see b) vs. d) and c) vs. e) in Figure 8). In addition, the performance of the objective evaluation metrics is also improved by introducing the dual-branch encoding; see 1) vs. 3) and 2) vs. 4) in Table 3, where the DICE scores are increased by 0.0533 and 0.0516, respectively, and similarly, see 1) vs. 3) and 2) vs. 4) in Table 4, where the DICE scores achieve gains of 0.0476 and 0.0632, respectively. These experimental results validate the necessity of the proposed dual-branch encoder.

Effectiveness of multi-supervised decoder. As shown in b) vs. c) and d) vs. e) of Figures 7, 8, replacing the single supervision (i.e., BN + SL or SNet + SL) in the decoder with multiple supervision (i.e., BN + ML or SNet + ML) also helps to improve the quality of heart wall and IA vasculature segmentation results. The improved results stem from the fact that the multi-supervision effectively integrates the high-level semantic and low-level localization knowledge, allowing the network to focus on the integrity of the target and the exactness of the details. This stipulation can also be confirmed by 1) vs. 2) and 3) vs. 4) in Table 3, where the DICE scores are improved by 0.0193 and 0.0176, respectively, and also, similarly, for 1) vs. 2) and 3) vs. 4) in Table 4, where the DICE scores gain 0.0078 and 0.0234, respectively. The results of these evaluation metrics presented in Tables 3 and 4 objectively demonstrate that the performance of the predictions is improved with the adoption of multi-supervision. Clearly, our comparison results show that the multi-supervised decoder optimizes the segmentation results to some extent.

5 Discussions

This study investigates the segmentation performance of the proposed network structure based on S-shaped multiple cross-aggregation for the whole heart wall and IA vasculature. The

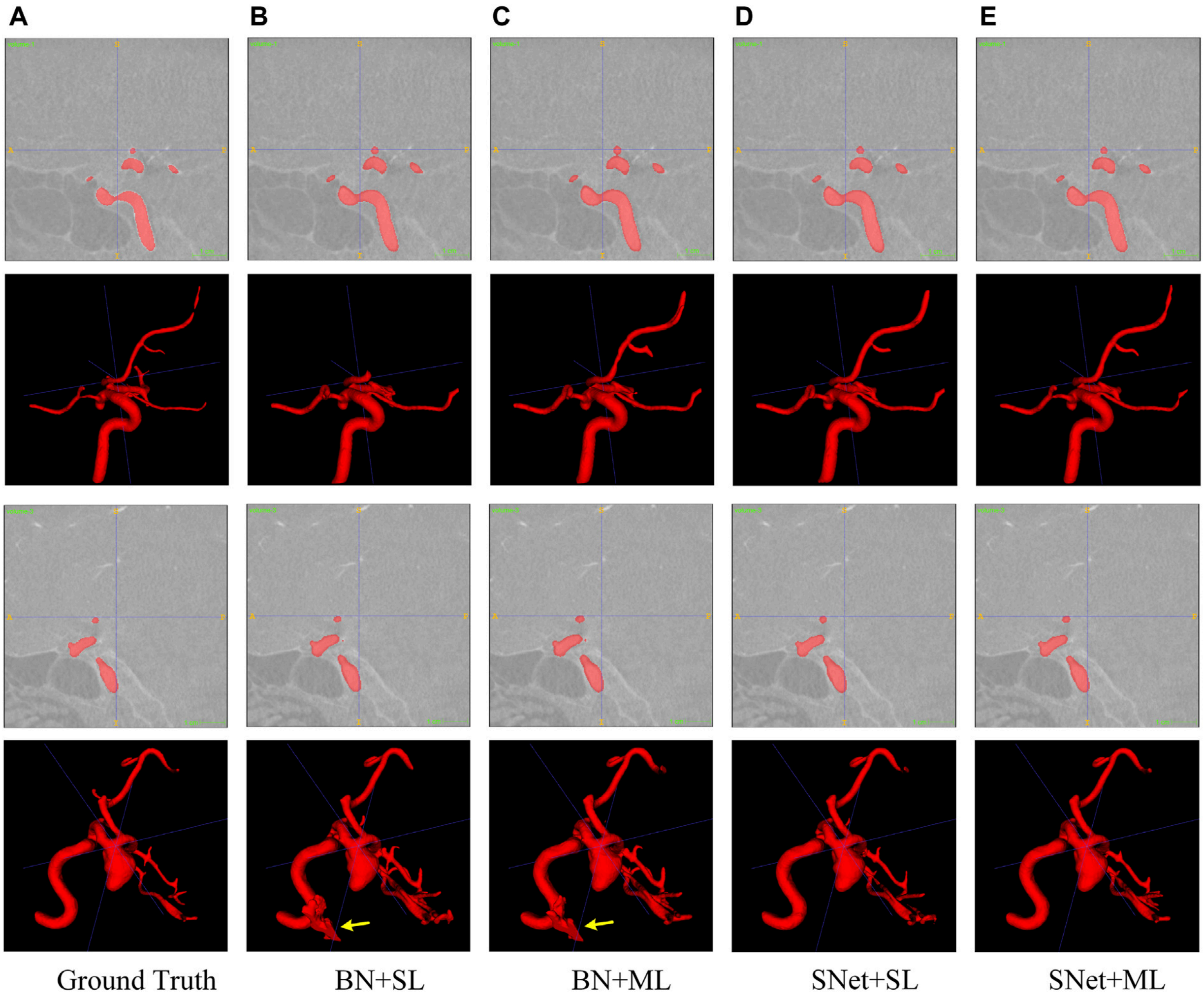

FIGURE 8
Visualization of the IA vasculature segmentation results generated by the four variants of the proposed S-Net. Images in column **(A)** are the Ground Truth and columns **(B–E)** represent the prediction results for baseline BN+SL, BN+ML, SNet+SL, and SNet+ML, respectively.

TABLE 3 Objective evaluation metric results of the four variants of the proposed S-Net for heart wall segmentation. The best results are highlighted by bold fonts.

Models	DICE	Sensitivity	Specificity	RVE	HD95	ASSD
1) BN + SL	0.8303 ± 0.0129	0.9286 ± 0.0208	0.9880 ± 0.0028	0.2381 ± 0.0594	6.8191 ± 0.5995	2.2405 ± 0.1257
2) BN + ML	0.8496 ± 0.0274	0.9147 ± 0.0262	0.9906 ± 0.0032	0.1586 ± 0.0956	25.7313 ± 18.5340	3.4594 ± 1.8512
3) SNet + SL	0.8836 ± 0.0121	0.9008 ± 0.0336	0.9947 ± 0.0013	0.0858 ± 0.0206	5.3262 ± 1.3749	1.4512 ± 0.1763
4) SNet + ML	**0.9012 ± 0.0145**	**0.8947 ± 0.0388**	**0.9966 ± 0.0010**	**0.0762 ± 0.0280**	**4.6806 ± 1.7503**	**1.2233 ± 0.2146**

The best results are shown in bold fonts.

TABLE 4 Objective evaluation metric results of the four variants of the proposed S-Net for IA vasculature segmentation. The best results are highlighted by bold fonts.

Models	DICE	Sensitivity	Specificity	RVE	HD95	ASSD
1) BN + SL	0.8025 ± 0.0575	0.9217 ± 0.0493	0.9979 ± 0.0012	0.3212 ± 0.2349	38.5696 ± 19.6079	4.9117 ± 3.4622
2) BN + ML	0.8103 ± 0.0319	0.9330 ± 0.0378	0.9981 ± 0.0005	0.3067 ± 0.1055	27.0335 ± 15.5266	2.8468 ± 0.9589
3) SNet + SL	0.8501 ± 0.0380	**0.9363 ± 0.0343**	0.9986 ± 0.0003	0.2055 ± 0.0569	**21.8773 ± 13.7142**	2.2607 ± 0.9994
4) SNet + ML	**0.8735 ± 0.0288**	0.9329 ± 0.0390	**0.9990 ± 0.0003**	**0.1363 ± 0.0664**	24.7471 ± 17.5222	**2.2070 ± 1.0808**

The best results are shown in bold fonts.

superior predictive ability of the proposed S-Net model is validated by comparison and ablation experiments, mostly outperforming the five compared state-of-the-art models in six evaluation metrics, where the DICE measured on two datasets (whole heart wall and IA) are 0.9012 and 0.8735, respectively. More importantly, qualitative experimental comparisons adequately demonstrate that our S-Net is capable of segmenting small/tiny structures, i.e., thin heart walls, e.g., the holes revealed by other segmentation models (indicated by the arrows in Figures 5B–F), and tiny arteries (indicated by the arrows in a) and g) of Figure 6).

It is clear from the experimental results that the proposed S-Net is a good backbone architecture for small-volume segmentation. Three key strategies attributed to our success. First, we design an efficient two-branch encoder, i.e., a regular downsampling encoding branch with progressively halved resolution network layers and a resampling encoding branch with fixed resolution network layers, to explore spatial details and global contextual information. This configuration allows the encoder to consider large and small receptive fields, which can efficiently guide the CNN model to capture small/thin structures and fine-grained details. Our intuitive explanation of the abovementioned dual-branch encoder can be seen in Figures 7, 8. Second, our multiple cross-aggregation module also plays a vital role in guaranteeing the comprehensiveness and robustness of encoded features. In other words, the module effectively integrates multilayer high-level and low-level features. Since high-level features contain rich semantic and global context knowledge, while low-level features have plenty of details and localization information, the propagation of the fused features to the decoder through the hierarchical horizontal connection strengthens the global dependency and local details of the decoded features, facilitating the detection of small volumes. Third, we perform multilevel supervised prediction for all decoding layers, effectively combining high-level and low-level features with different semantics and resolutions along a top-down path, optimizing the network's learning efficiency for targets of various sizes and thus improving the completeness of prediction for tiny structures. The improvement in segmentation performance by the multi-supervised decoder can be seen in Tables 3 and 4. Collectively, the three proposed strategies allow our model to achieve state-of-the-art performance and have the ability to segment small volumes effectively.

Besides cardiovascular applications, the proposed S-Net could also be applied to segment vasculature in other organs, e.g., hepatic veins/arteries and retinal vessels. We also stipulate that the proposed S-Net could also be used in oncological applications (e.g., brain tumors, colon cancer, breast cancer, lung nodules). For instance, the improved detection of small targets allows us to identify small, hard-to-detect tumors and complex tumor compositions. As a result, the proposed S-Net can be integrated into an artificial intelligence (AI) system to enhance the prediction and detection of disease progression, thereby elevating the clinical management of cancer patients. It is worth noting successes of such predictive modeling have been achieved in predicting the growth of abdominal aortic aneurysms (Rezaeitaleshmahalleh et al., 2023a; Rezaeitaleshmahalleh et al., 2023c) and the rupture status of intracranial aneurysms (Sunderland et al., 2021; Jiang et al., 2023). In the future, we will expand the application of the proposed S-Net to oncological applications.

Although our two-branch coding structure facilitates the extraction of detailed features, it has the drawbacks of high computational complexity and long training time, and its training process requires high-performance hardware with large amounts of memory and is time-consuming. Our future work will optimize the training model by reducing the parameters through regularization and model pruning. Furthermore, the training time will be reduced by performing batch normalization and adding pooling layers to allow fast convergence.

6 Conclusion

In this paper, motivated to overcome the drawbacks of existing U-shaped segmentation architectures, we propose an S-Net framework for small/thin structure segmentation of medical images. Our novelty lies in exploring a dual-branch encoder consisting of resampling and downsampling convolutional layers to capture the information from large and small receptive fields for more accurate learning of small targets and finer details. These two branches are efficiently integrated by leveraging a novel S-shaped multiple cross-aggregation approach for effective training. Meanwhile, we enhance the global context and local detail knowledge in the decoding stage by propagating complementary features from the encoding layers through lateral connections. We also supervise features from all decoding layers in the top-down path to fully optimize the semantics and localization of the prediction results. To verify the superiority of the proposed S-Net for small/thin structure predictions, we performed segmentation experiments on the heart wall and IA vasculature, and the results demonstrated that the proposed model outperforms state-of-the-art CNN methods, significantly improving the structural accuracy and surface quality of segmented volumes and having the ability to adequately capture small structures, which possesses the potential to facilitate clinical applications.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://zmiclab.github.io/zxh/0/mmwhs/ and http://ecm2.mathcs.emory.edu/aneuriskweb/about.

Ethics statement

Institutional Review Boards at Michigan Technological University and Helen DeVos Children's Hospital approved the study. Publicly available imaging data were used in this study, and no patient consent is needed.

Author contributions

NM and JJ: conceptualization and methodology. NM, MR, CB, JG, and ZL: formal analysis, data processing, visualization, and investigation. JV, MH, and JJ: resources and supervision. NM, JV, and JJ: writing–original draft preparation. NM, CB, ZL, MR, JG, MH, JV, and JJ: writing–review and editing. NM, JG, MH, JV, and JJ: project administration and funding acquisition. All authors contributed to the article and approved the submitted version..

Funding

The study is partly supported by a research contract from Spectrum Health Foundation (Corewell Health) and a research grant from the National Institutes of Health (R01-EB029570A1). NM is supported by a post-doctoral fellowship from the American Heart Association (23Post1022454). MR receives partial funding from the Health Research Institute at Michigan Technological University.

Acknowledgments

We want to thank the Aneurisk and Multi-Modality Whole Heart Segmentation (MM-WHS) projects for providing brain aneurysm and heart data used in this project, respectively.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., and Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* 6 (1), 014006. doi:10.1117/1.JMI.6.1.014006

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Analysis Mach. Intell.* 39 (12), 2481–2495. doi:10.1109/TPAMI.2016.2644615

Bhalerao, M., and Thakur, S. (2019). "Brain tumor segmentation based on 3D residual U-Net," in *International MICCAI brainlesion workshop* (USA: MICCAI).

Cebral, J. R., Castro, M. A., Burgess, J. E., Pergolizzi, R. S., Sheridan, M. J., and Putman, C. M. (2005). Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models. *Am. J. Neuroradiol.* 26 (10), 2550–2559.

Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., and Rueckert, D. (2018). DRINet for medical image segmentation. *IEEE Trans. Med. Imaging* 37 (11), 2453–2462. doi:10.1109/TMI.2018.2835303

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation, International Conference on Medical Image Computing and Computer-Assisted Intervention, 8-12 October, Canada IEEE.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *IEEE conference on computer vision and pattern recognition* (IEEE).8-12 October, Canada.

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18 (2), 203–211. doi:10.1038/s41592-020-01008-z

Jiang, J., Rezaeitaleshmahalleh, M., Lyu, Z., Mu, N., Ahmed, A. S., Md, C. M. S., et al. (2023). Augmenting prediction of intracranial aneurysms' risk status using velocity-informatics: initial experience. *J. Cardiovasc. Transl. Res.* doi:10.1007/s12265-023-10394-6

Jiang, Y., Wang, F., Gao, J., and Cao, S. (2020). Multi-path recurrent U-Net segmentation of retinal fundus image. *Appl. Sci.*, 10(11), 3771–3717. doi:10.3390/app10113777

Jou, L. D., Quick, C. M., Young, W. L., Lawton, M. T., Higashida, R., Martin, A., et al. (2003). Computational approach to quantifying hemodynamic forces in giant cerebral aneurysms. *AJNR Am. J. Neuroradiol.* 24 (9), 1804–1810.

Lalonde, M., Beaulieu, M., and Gagnon, L. (2001). Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching. *IEEE Trans. Med. Imaging* 20 (11), 1193–1200. doi:10.1109/42.963823

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., and Heng, P.-A. (2018). H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* 37 (12), 2663–2674. doi:10.1109/TMI.2018.2845918

Lindquist, E. M., Gosnell, J. M., Khan, S. K., Byl, J. L., Zhou, W., Jiang, J., et al. (2021). 3D printing in cardiology: a review of applications and roles for advanced cardiac imaging. *Ann. 3D Print. Med.* 4, 100034. doi:10.1016/j.stlm.2021.100034

Lindquist Liljeqvist, M., Bogdanovic, M., Siika, A., Gasser, T. C., Hultgren, R., and Roy, J. (2021). Geometric and biomechanical modeling aided by machine learning improves the prediction of growth and rupture of small abdominal aortic aneurysms. *Sci. Rep.* 11 (1), 18040. doi:10.1038/s41598-021-96512-3

Lu, X., Chen, X., Li, W., and Qiao, Y. (2019). Graph cut segmentation of the right ventricle in cardiac MRI using multi-scale feature learning, Proceedings of 3rd International Conference on Cryptography, Security and Privacy, January 19-21, 2019, Malayasia IEEE.

Lyu, Z., King, K., Rezaeitaleshmahalleh, M., Pienta, D., Mu, N., Zhao, C., et al. (2023). Deep-learning-based image segmentation for image-based computational hemodynamic analysis of abdominal aortic aneurysms: a comparison study. *Biomed. Phys. Eng. Express* 9 (6), 067001. doi:10.1088/2057-1976/acf3ed

Mu, N., Lyu, Z., Rezaeitaleshmahalleh, M., Tang, J., and Jiang, J. (2023a). An attention residual U-Net with differential preprocessing and geometric postprocessing: learning how to segment vasculature including intracranial aneurysms. *Med. Image Anal.* 84, 102697–102718. doi:10.1016/j.media.2022.102697

Mu, N., Lyu, Z., Rezaeitaleshmahalleh, M., Zhang, X., Rasmussen, T., McBane, R., et al. (2023b). Automatic segmentation of abdominal aortic aneurysms from CT angiography using a context-aware cascaded U-Net. *Comput. Biol. Med.* 158, 106569. doi:10.1016/j.compbiomed.2023.106569

Petitjean, C., and Dacher, J.-N. (2011). A review of segmentation methods in short axis cardiac MR images. *Med. Image Anal.* 15 (2), 169–184. doi:10.1016/j.media.2010.12.004

Rezaeitaleshmahalleh, M., Lyu, Z., Mu, N. A. N., and Jiang, J. (2023b). Using convolutional neural network-based segmentation for image-based computational fluid dynamics simulations of brain aneurysms: initial experience in automated model creation. *J. Mech. Med. Biol.* 23 (04), 2340055. doi:10.1142/S0219519423400559

Rezaeitaleshmahalleh, M., Lyu, Z., Mu, N., Zhang, X., Rasmussen, T. E., McBane, R. D., et al. (2023a). Characterization of small abdominal aortic aneurysms' growth status using spatial pattern analysis of aneurismal hemodynamics. *Sci. Rep.* 13 (1), 13832. doi:10.1038/s41598-023-40139-z

Rezaeitaleshmahalleh, M., Sunderland, K. W., Lyu, Z., Johnson, T., King, K., Liedl, D. A., et al. (2023c). Computerized differentiation of growth status for abdominal aortic aneurysms: a feasibility study. *J. Cardiovasc. Transl. Res.* 16, 874–885. doi:10.1007/s12265-022-10352-8

Ringenberg, J., Deo, M., Devabhaktuni, V., Berenfeld, O., Boyers, P., and Gold, J. (2014). Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI. *Comput. Med. Imaging Graph.* 38 (3), 190–201. doi:10.1016/j.compmedimag.2013.12.011

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention* (Canada: MICCAI).

Sunderland, K., Wang, M., Pandey, A. S., Gemmete, J., Huang, Q., Goudge, A., et al. (2021). Quantitative analysis of flow vortices: differentiation of unruptured and ruptured medium-sized middle cerebral artery aneurysms. *Acta Neurochir.* 163 (8), 2339–2349. doi:10.1007/s00701-020-04616-y

Suri, J. S. (2000). Computer vision, pattern recognition and image processing in left ventricle segmentation: the last 50 years. *Pattern Analysis Appl.* 3 (3), 209–242. doi:10.1007/s100440070008

Tobon-Gomez, C., Geers, A. J., Peters, J., Weese, J., Pinto, K., Karim, R., et al. (2015). Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *IEEE Trans. Med. Imaging* 34 (7), 1460–1473. doi:10.1109/TMI.2015.2398818

Tobon-Gomez, C., Peters, J., Weese, J., Pinto, K., Karim, R., Schaeffter, T., et al. (2013). "Left atrial segmentation challenge: a unified benchmarking framework," in *Statistical atlases and computational models of the heart. Imaging and modelling challenges* (China: 4th International Workshop).

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., et al. (2003). A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Med. Imaging* 22 (2), 137–154. doi:10.1109/TMI.2002.808355

Valanarasu, J. M. J., Oza, P., Hacihaliloglu, I., and Patel, V. M. (2021a). “Medical transformer: gated axial-attention for medical image segmentation,” in *Medical image computing and computer assisted intervention* (Canada: MICCAI).

Valanarasu, J. M. J., Sindagi, V. A., Hacihaliloglu, I., and Patel, V. M. (2021b). KiU-Net: overcomplete convolutional architectures for biomedical image and volumetric segmentation. *IEEE Trans. Med. Imaging*, 1–13. doi:10.1109/TMI.2021.3130469

Wang, Z.-H., Liu, Z., Song, Y.-Q., and Zhu, Y. (2019). Densely connected deep U-Net for abdominal multi-organ segmentation, IEEE International Conference on Image Processing, october 8 11 2023, Germany IEEE.

Xu, C., and Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Process.* 7 (3), 359–369. doi:10.1109/83.661186

Xu, X., Wang, T., Shi, Y., Yuan, H., Jia, Q., Huang, M., et al. (2019). “Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching,” in *Medical image computing and computer assisted intervention* (Canada: MICCAI).

Yang, D., Huang, Q., Axel, L., and Metaxas, D. (2018). “Multi-component deformable models coupled with 2D-3D U-Net for automated probabilistic segmentation of cardiac walls and blood,” in *IEEE 15th international symposium on biomedical imaging* (USA: IEEE).

Ye, M., Huang, Q., Yang, D., Wu, P., Yi, J., Axel, L., et al. (2021).PC-U net: learning to jointly reconstruct and segment the cardiac walls in 3D from CT data, *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges: 11th International Workshop*, Canada: MICCAI.

Yu, W., Fang, B., Liu, Y., Gao, M., Zheng, S., and Wang, Y. (2019). “Liver vessels segmentation based on 3d residual U-Net,” in IEEE international conference on image processing (IEEE).January 19-21, 2019, Malayasia.

Yuan, Y., Chao, M., and Lo, Y.-C. (2017). Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Trans. Med. Imaging* 36 (9), 1876–1886. doi:10.1109/TMI.2017.2695227

Zhang, J., Ma, K.-K., Er, M.-H., and Chong, V. (2004). Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine, International Workshop on Advanced Image Technology, January 19-21, 2019, Malayasia IEEE.

Zhang, Z., Wu, C., Coleman, S., and Kerr, D. (2020). DENSE-INception U-net for medical image segmentation. *Comput. Methods Programs Biomed.* 192, 105395. doi:10.1016/j.cmpb.2020.105395

Zhao, Y.-q., Gui, W.-h., Chen, Z.-c., Tang, J.-t., and Li, L.-y. (2006). "Medical images edge detection based on mathematical morphology", in IEEE Engineering in Medicine and Biology 27th Annual Conference.January 19-21, 2019, Malayasia, IEEE,).

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., and Comaniciu, D. (2008). Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Med. Imaging* 27 (11), 1668–1681. doi:10.1109/TMI.2008.2004421

Zhu, L., Gao, Y., Appia, V., Yezzi, A., Arepalli, C., Faber, T., et al. (2013). Automatic delineation of the myocardial wall from CT images via shape segmentation and variational region growing. *IEEE Trans. Biomed. Eng.* 60 (10), 2887–2895. doi:10.1109/TBME.2013.2266118

Zhuang, X., Rhode, K. S., Razavi, R. S., Hawkes, D. J., and Ourselin, S. (2010). A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Trans. Med. Imaging* 29 (9), 1612–1625. doi:10.1109/TMI.2010.2047112

Zhuang, X., and Shen, J. (2016). Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Med. Image Anal.* 31, 77–87. doi:10.1016/j.media.2016.02.006