



AVDNet: Joint Coronary Artery and Vein Segmentation with Topological Consistency

Wenji Wang^a, Qing Xia^{a,*}, Zhennan Yan^b, Zhiqiang Hu^a, Yinan Chen^a, Wen Zheng^c, Xiao Wang^c, Shaoping Nie^c, Dimitris Metaxas^d, Shaoting Zhang^{a,e}

^a*SenseTime Research, Beijing, 100080 China*

^b*SenseBrain Technology, NJ, 08540 USA*

^c*Center for Coronary Artery Disease, Beijing Anzhen Hospital, Capital Medical University, Beijing, 100029 China*

^d*Department of Computer Science, Rutgers University, NJ, 08854 USA*

^e*Shanghai Artificial Intelligence Laboratory, Shanghai, 200032 China*

ARTICLE INFO

Article history:

Received 1 August 2022

Received in final form 1 August 2022

Accepted 1 August 2022

Available online 1 August 2022

Communicated by S. Sarkar

2000 MSC: 41A05, 41A10, 65D05, 65D17

Keywords: coronary CT angiography, vessel segmentation, artery and vein disentanglement, topology consistency, inverse distance weights

ABSTRACT

Coronary CT angiography (CCTA) is an effective and non-invasive method for coronary artery disease diagnosis. Extracting an accurate coronary artery tree from CCTA image is essential for centerline extraction, plaque detection, and stenosis quantification. In practice, data quality varies. Sometimes, the arteries and veins have similar intensities and locate closely, which may confuse segmentation algorithms, even deep learning based ones, to obtain accurate arteries. However, it is not always feasible to re-scan the patient for better image quality. In this paper, we propose an artery and vein disentanglement network (AVDNet) for robust and accurate segmentation by incorporating the coronary vein into the segmentation task. This is the first work to segment coronary artery and vein at the same time. The AVDNet consists of an image based vessel recognition network (IVRN) and a topology based vessel refinement network (TVRN). IVRN learns to segment the arteries and veins, while TVRN learns to correct the segmentation errors based on topology consistency. We also design a novel inverse distance weighted dice (IDD) loss function to recover more thin vessel branches and preserve the vascular boundaries. Extensive experiments are conducted on a multi-center dataset of 700 patients. Quantitative and qualitative results demonstrate the effectiveness of the proposed method by comparing it with state-of-the-art methods and different variants. Prediction results of the AVDNet on the Automated Segmentation of Coronary Artery Challenge dataset are available at <https://github.com/WennyJJ/Coronary-Artery-Vein-Segmentation> for follow-up research.

© 2024 Elsevier B. V. All rights reserved.

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally (WHO, 2021). CVDs refer to a group of disorders of the heart and blood vessels, and coronary artery disease

(CAD) is one of them. CAD happens when the major blood vessels that supply your heart become damaged or diseased due to the buildup of plaques. It often develops over decades without a noticeable problem until patients have a significant blockage or a heart attack. Therefore, early diagnosis is critical and coronary computed tomography angiography (CCTA) is an effective and commonly used imaging method for this purpose.

*Corresponding author. Email: xiaqing@sensetime.com; Tel.: +86-186-0192-0416.

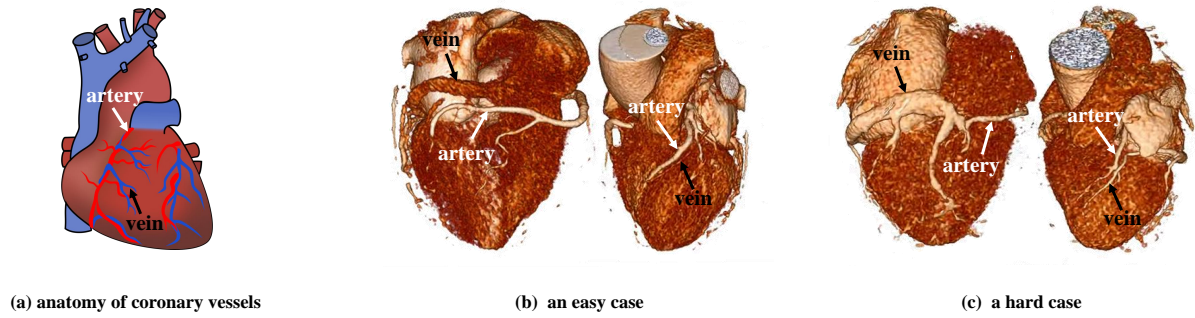


Fig. 1. Anatomy of coronary vessels and two real examples. (a) anatomy of coronary arteries (red) and veins (blue) (b) an easy case characterized by a significant intensity difference between coronary arteries and veins, allowing for easy distinction of the vessels. (c) a challenging case with similar intensity of both coronary arteries and veins, which can result in segmentation errors in local regions.

Segmentation of coronary arteries from CCTA images is a vital step before the centerline extraction, plaque detection, and stenosis quantification in a typical diagnosis pipeline. Research on automatic and objective vessel extraction has been carried out for decades. (Kirbas and Quek, 2004) reviewed and divided vessel segmentation methods into six main categories: pattern recognition techniques (Niki et al., 1993; Sarwal and Dhawan, 1994), model-based approaches (van der Weide et al., 2001), tracking-based approaches (Tolias and Panas, 1998), artificial intelligence-based approaches (Stansfield, 1986), neural network-based approaches (Nekovei and Sun, 1995; Hunter et al., 1995) and miscellaneous tube-like object detection approaches (Kompatsiaris et al., 2000). However, traditional vascular segmentation methods are time-consuming, parameter-sensitive, and thus difficult to achieve the desired accuracy on unseen cases with various diseases and image artifacts. This limits the large-scale data analytics and integration with clinical systems.

The advance in deep learning brings new ideas for segmenting blood vessels in medical images. In particular, researchers have adopted various deep learning methods on coronary artery segmentation from CCTA images in recent years (Kong et al., 2020; Chen et al., 2019, 2018; Lei et al., 2020; Pan et al., 2021; Song et al., 2022). The Automated Segmentation of Coronary Arteries Challenge (ASOCA) ¹ in MICCAI 2020 was held to encourage developing fully automatic segmentation methods of

the full coronary artery tree by establishing a standardized annotated dataset of healthy and diseased coronary arteries from 60 subjects. These studies focused on coronary arteries and ignored the influence of coronary veins in arterial segmentation. Within the human blood circulation system, the coronary arteries and veins of the heart interweave in a complex manner, like in Fig.1 (a). In ideal CCTA images, the intensities of arteries are much higher than those of veins, which makes it easy to distinguish between them even when they are very close, such as the case in Fig.1 (b). However, in some practical cases like Fig.1 (c), it is difficult to tell them apart due to similar intensities caused by the misoperation of technicians or blood circulation differences in patients. In such a challenging scenario, the artery segmentation methods are prone to errors. These errors could be corrected by revising the segmentation results manually or acquiring the CCTA images one more time. However, both solutions cost much longer time in image processing, and the latter increases the risk of radiation for patients. In this paper, we focus on obtaining accurate and reliable coronary artery segmentation results automatically under various conditions.

In order to achieve our goal, we integrate coronary vein segmentation into the task. As far as we know, this is the first study to segment coronary artery and vein vessels simultaneously from the CCTA images. We propose an artery and vein disentanglement network (AVDNet), consisting of two stages. In AVDNet, an image-based vessel recognition network (IVRN) learns to output the initial segmentation results and followed by a topology-based vessel refinement network (TVRN) to re-

¹<https://asoca.grand-challenge.org/>

vise the results. The IVRN has one encoder and two decoders, which are responsible for image feature extraction, coronary vessel segmentation, and artery-vs-vein classification, respectively. The IVRN has the ability to distinguish between arteries and veins. However, when nearby arteries and veins have similar intensities in some local regions, the IVRN learned from image features may produce topologically incorrect results. This motivates us to design an additional network, TVRN, to revise the initial segmentation by taking advantage of structural features to maintain topological consistency. The TVRN consists of two encoders and one decoder. One encoder learns structural features from recognized results of IVRN with different states, and the other one learns image features from varying HU values. Finally, the decoder receives the concatenated features and learns how to reduce the intensity interference and maintain the topology consistency. We also design a novel inverse distance weighted dice (IDD) loss. The IDD loss pays more attention to voxels near edges and vice versa, which can segment more thin structures and protect vascular edges. We employ a multi-center dataset of 700 patients to carry out experiments. Among them, the testing dataset consists of 200 patients, 100 images with strong discrimination of arterial and venous intensity values forming a easy dataset, and the remaining 100 ones forming a hard dataset. We adopt centerline overlap to evaluate the completeness of the extracted vessel tree and present a weighted dice coefficient to address the problem that the traditional dice metric is region-based, being easily affected by thick branches. This weighted dice metric takes more care of thin vessels and reduces the effect of thick vessels on the results. To facilitate follow-up research, we perform inference on the ASOCA dataset by using AVDNet and open-source the prediction results with coronary artery and vein vessels. In summary, the main contributions of this paper include:

1. First study of joint coronary artery and vein segmentation, which aims to produce reliable and accurate vessel tree extraction even from low-quality images.
2. A novel method, AVDNet, is proposed, which consists of two-stage learning. First, IVRN learns from image fea-

tures to obtain an initial segmentation, and then TVRN learns from morphological features to refine the segmentation to ensure topology consistency.

3. A novel loss, IDD, is designed to preserve thin vessels and obtain accurate vascular boundaries.
4. A weighted dice coefficient is designed to better evaluate the segmentation results, paying more attention to any missed vessel segments rather than to ambiguous boundaries of thick vessels.

The rest of the paper is organized as follows. Section 2 introduces some related work. Section 3 provides details of the proposed method. Section 4 reports the experimental settings and results, followed by discussions. The work is concluded in Section 5.

2. Related work

In recent years, deep learning based methods have been successfully applied to medical image segmentation tasks (He et al., 2021; Cui et al., 2022; Zhuang et al., 2022; Wang et al., 2022; Davies et al., 2022; Luo et al., 2022; Chang et al., 2022; Qu et al., 2021; Wang et al., 2019b). These methods are more efficient and generally provide more reliable results than traditional methods. In this section, we will review some deep learning based methods on tubular-structure segmentation, coronary artery segmentation, and artery and vein segmentation, respectively.

2.1. Tubular-structure segmentation methods

For 2D tubular segmentation, Laibacher et al. (Laibacher et al., 2019) proposed a U-Net-like architecture to segment retinal vessels in the fundus image. Since the size of training data in medical applications is usually small, Wu et al. (Wu et al., 2021) proposed a scale and context sensitive network to tackle the problems of large variations of scale and semantics, while Zhao et al. (Zhao et al., 2018) utilized generative adversarial network to synthesize retinal and neuronal images for improving segmentation performance. Xia et al. (Xia et al.,

2019) developed a matrix decomposition model to take advantage of temporal motion information in the x-ray coronary angiographic video data for 2D vessel segmentation. For 2.5D tubular segmentation, Yun et al. (Yun et al., 2019) proposed a novel airway segmentation method in volumetric chest computed tomography, which performed voxel-by-voxel segmentation by a 2.5D convolutional neural network. Cui et al. (Cui et al., 2019) applied a 2.5D segmentation network from three orthogonal axes to automatically segment pulmonary vessels. The 2.5D network has lower network complexity and memory usage compared to 3D networks. For 3D tubular segmentation, Wang et al. (Wang et al., 2019a) proposed a novel radial distance loss to improve segmentation of thin tubular structures for 3D bronchus in CT scans. Furthermore, a novel geometry-aware tubular structure segmentation method was proposed by Wang et al. (Wang et al., 2020), which combines intuitions from the classical distance transform for skeletonization and modern deep segmentation networks. Qin et al. (Qin et al., 2019) presented a voxel-connectivity aware approach named AirwayNet for accurate airway segmentation on CT scans. This approach transformed the conventional binary segmentation task into 26 tasks of connectivity prediction by connectivity modeling.

2.2. Coronary artery segmentation methods

Besides the above tubular segmentation approaches, several recent studies have adopted deep learning with different network architectures for fully automatic coronary artery reconstruction in CCTA data, e.g., by using a 3D multi-channel U-Net (Chen et al., 2019), a 3D attention fully convolution network (Lei et al., 2020), and a 3D Dense-U-Net (Pan et al., 2021). Shen et al. (Shen et al., 2019) combined both the traditional level set method and 3D fully convolutional network to produce more smooth segmentation results. To explicitly learn the anatomical structure of the coronary artery, Kong et al. (Kong et al., 2020) proposed a tree-structured convolutional gated recurrent unit model. In practice, due to the large size of CCTA data and limited memory of GPU, a CCTA scan may be divided into small patches to obtain prediction in the original resolution. Gu et al. (Gu and Cai, 2021) uses both 2D CNN

and 3D CNN to perform aorta and coronary artery segmentation. 2D CNN first extracted large-field-of-view information in a slice-by-slice fashion and generated segmentation results. Then, 3D CNN extracted the inter-slice information to refine the segmentation results in 2D CNN. Song et al. (Song et al., 2022) used a 2D classification network to screen out the non-coronary-artery slices to obtain a smaller ROI for 3D segmentation. Instead of voxel-based segmentation methods, Wolterink et al. (Wolterink et al., 2019) modeled the coronary artery with a tubular surface mesh and utilized graph convolutional networks (GCN) to optimize the location of the surface mesh vertices for lumen segmentation. Despite these previous researches, some challenges are remaining in this particular task, such as false positives of arterial predictions.

2.3. Artery and vein segmentation methods

In pulmonary vessel segmentation and hepatic vessel segmentation tasks, vessel classification, such as artery and vein classification, is clinically required. Nardelli et al. (Nardelli et al., 2018) combined CNN with a graph-cut strategy to classify vessels into arteries and veins on chest CT images. Jimenez et al. (Jimenez-Carretero et al., 2019) presented a framework to approach the separation of tree-like structures and designed graph-cut methodology to ensure connectivity as well as the spatial and directional consistency of the derived sub-trees. This framework was applied to the pulmonary artery and vein classification. Qin et al. (Qin et al., 2021) presented a CNNs-based method to segment airway, artery, and vein simultaneously in non-contrast computed tomography. It enjoys superior sensitivity to tenuous peripheral bronchioles, arterioles, and venules. Keshwani et al. (Keshwani et al., 2020) proposed TopNet to detect voxels on vascular centerlines and estimate connectivity between center-voxels in the tree structure to be reconstructed. With TopNet, the extracted portal and hepatic veins can better preserve their topology. Guo et al. (Guo et al., 2022) proposed a framework to segment portal veins and hepatic veins from multi-phase MR images, which takes the vascular flow into account to improve segmentation performance.

Different from the above studies, existing research on coro-

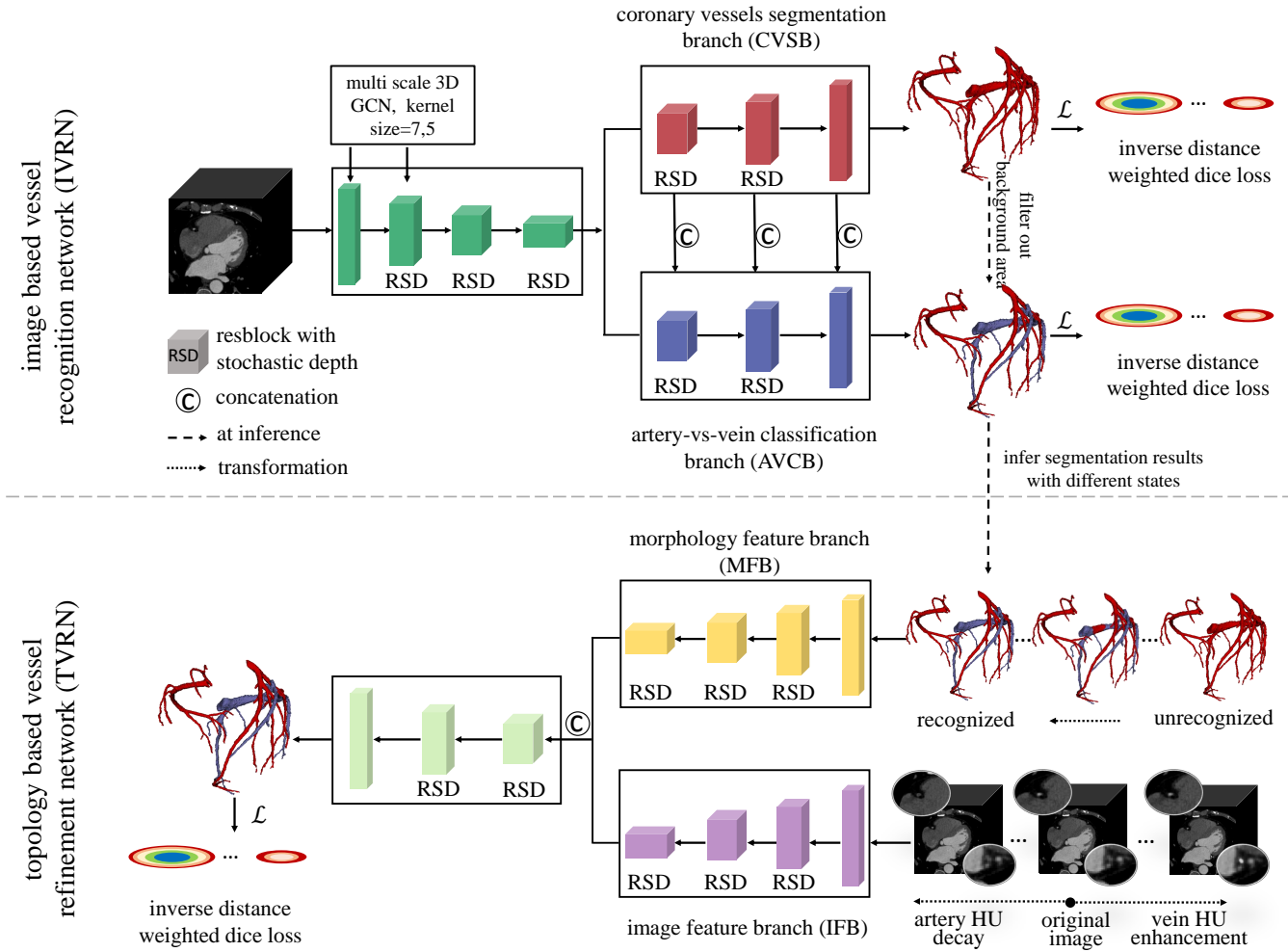


Fig. 2. Overview of the proposed AVDNet. IVRN and TVRN are cascaded in the AVDNet. For IVRN, CVSB learns a 2-class segmentation task to predict vessels, including both artery and vein, as the foreground category; AVCB learns to distinguish between coronary arteries and veins in the foreground. TVRN combines the morphology input from the segmentation results of the AVCB and the image input with varying HU values to output the revised segmentation results.

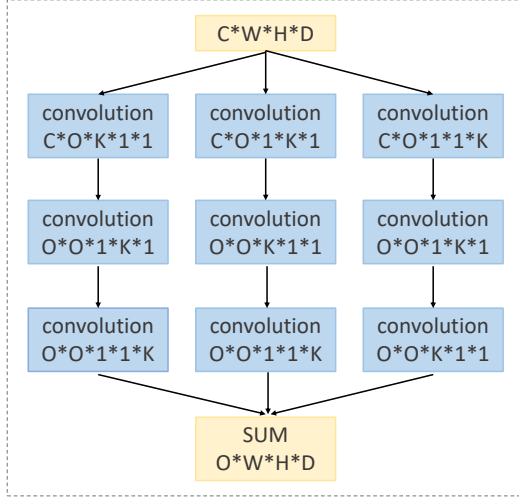
nary artery segmentation ignored the influence of venous vessels. In practical cases, coronary artery and vein in CCTA images have similar HU values and cross each other in close vicinity, prone to incorrect segmentation results. Inspired by this, we integrate coronary vein segmentation into coronary artery segmentation task and propose the AVDNet for artery and vein classification.

3. Methods

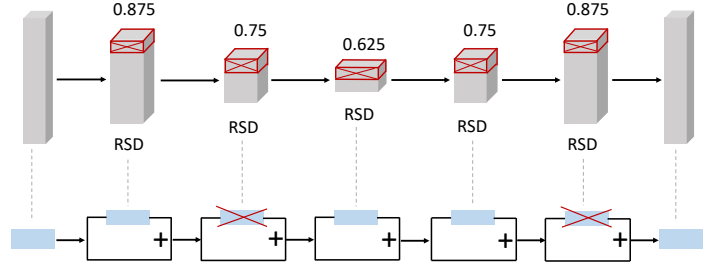
We propose a coronary artery and vein segmentation method, AVDNet, which is illustrated in Fig.2. It consists of two networks IVRN and TVRN. We first use the IVRN to segment both arteries and veins from CCTA image, and then use the TVRN

to revise incorrect predictions regarding the topology consistency. In order to preserve thin vessel structures while obtaining accurate vascular boundaries, we propose an inverse distance weighted dice (IDD) loss which pays more attention to voxels near boundaries. Both IVRN and TVRN are supervised by the IDD loss. The complete names of terms, acronyms, along with their detailed purpose can be found in Table.S1 in the supplementary materials.

In the following subsections, we explain the proposed networks and loss functions in detail. We denote the set of input CCTA ROI images as $X = \{x_i : i = 1, \dots, N\}$, where the size of x_i is $D * W * H$. The set of corresponding ground-truth labels is denoted as $Y = \{y_i\}$, y_i is a three-class label including back-



(a) 3D GCN module in IVRN



(b) RSD module in both IVRN and TVRN

Fig. 3. Details of two modules used in the AVDNet: (a) 3D global convolutional network (GCN) module and (b) random stochastic depth (RSD) module.

ground, coronary artery and vein. $Y_{Bin} = \{y_i^b\}$ denotes the binary version vessel label, which merge artery and vein into the same category. The set of initial segmentation results from IVRN is denoted as S , and the refined results by TVRN as R .

3.1. Image-based vessel recognition network (IVRN)

IVRN splits the multi-class segmentation task into two sub-tasks, one for coronary vessels and background recognition, and the other for artery-vs-vein classification. As shown in the upper part of Fig 2, the IVRN has one encoder and two decoders for multi-task learning. It takes CCTA image as input and outputs two types of segmentation from the two decoders, coronary vessels segmentation branch (CVSB) and artery-vs-vein classification branch (AVCB).

In 3D medical image segmentation tasks, receptive field is important for accurate and robust prediction. We employ a multi-scale 3D global convolutional network (GCN) and a random stochastic depth (RSD) module in the encoder for extracting comprehensive image features for accurate distinguish of coronary arteries and veins. The 3D GCN is illustrated in Fig.3, which is expanded from the traditional 2D GCN (Peng et al., 2017). With 3D GCN, on the one hand, we can set a larger kernel size to increase the receptive field, and on the other hand, using a separate $k * 1 * 1$ convolution in different directions (D, W, H) will have a better performance than using $k * k * k$

convolution directly as reported in (Peng et al., 2017). Due to the small targets of vessels, we just place 3D GCN to the first few layers in the encoder, whose features have higher resolution. We also combine the 3D GCN with multi-scale kernels to increase the diversity of features. Since thin structures may disappear in deep intermediate layers, which could limit the segmentation accuracy. To improve the network robustness, we use tailored RSD (Huang et al., 2016) modules to randomly bypass some layers, which are illustrated in Fig.3.

In the decoder part, CVSB aims to detect all the vessels belonging to the coronary, and provides vascular features for AVCB. The AVCB aims to distinguish arteries from veins. It takes features at different resolutions from the CVSB and concatenates them with its own extracted features to pass forward. In this way, the branch can pay more attention to learning the discriminative features between arteries and veins, without considering the background area. It contributes to accelerating network convergence and improving performance. For this purpose, we only calculate the loss of the foreground area (artery and vein) during the training process in AVCB. Therefore, the objective function of CVSB is defined as:

$$\min_{\theta} \sum_{i=1}^N \mathcal{L}_{\mathcal{D}\mathcal{D}}(f(x_i; \theta), y_i^b), \quad (1)$$

where N is the number of training data. The objective function

of AVCB is defined as:

$$\min_{\theta} \sum_{i=1}^N \mathcal{L}_{\mathcal{DD}}(f(x_i; \theta), y_i). \quad (2)$$

In the inference stage, the CVSB uses the prediction results to help the AVDC filter out the background area, generating the final segmentation results.

On the one hand, the IVRN can improve vascular detection ability, finding as many coronary vessels as possible. On the other hand, it can focus on distinguishing between arteries and veins in the foreground, reducing the difficulty of the multi-label segmentation task.

3.2. Topology-based vessel refinement network (TVRN)

We noticed that the IVRN may produce inconsistent segmentation results along the same vessel in some cases. It is due to the image features are mostly based on intensity and anatomy, ignoring the topology consistency of the vessels themselves. This motivates us to design an additional network, TVRN, to revise incorrect segmentation results by taking advantage of structural features to maintain topological consistency. The TVRN consists of two encoders and one decoder, as shown in the bottom of Fig.2. Two encoders learn morphological features and image features, respectively. All the features from both encoders are concatenated and forwarded to the decoder to output the corrected results. The main contributions lie in the encoder part. The morphology feature branch (MFB) aims to learn robust structural features from various vascular shapes. Specifically, we feed the recognition results of coronary artery and vein from IVRN to MFB. These results are inferred by IVRN using models at different epochs, which increases the variety of input. These various vascular shapes, from completely unrecognized results to fully recognized results, can help the MFB to learn robust structural features.

Meanwhile, the image feature branch (IFB) extract image features as complementary to help the refinement. During training, we augment input images by changing vascular HU values in the original images. There are two methods to augment the original images. The first method involves reducing the HU value of coronary arteries, while the second method focus on

enhancing the HU value of coronary veins. They all aim to simulate imperfect scans that have similar HU values for arteries and veins. Specifically, we first use the original image and its corresponding labels to calculate the mean HU values of coronary arteries and veins. We then obtain the difference between these two mean values. To vary the vascular HU values, we set a range of ratios for the augmentation operation and randomly select a ratio within this range, which is then multiplied by the above difference. Finally, we apply the resulting value to either enhance the HU value of coronary veins or decay the HU value of coronary arteries, with a 50% chance for each option. The implementation follows the formula:

$$HU_{diff} = \text{mean}(HU_{artery}) - \text{mean}(HU_{vein}), \quad (3)$$

$$image_{artery-decay} = image_{ori} - HU_{diff} * ratio * mask_a, \quad (4)$$

$$image_{vein-enhance} = image_{ori} + HU_{diff} * ratio * mask_v, \quad (5)$$

where $image_{ori}$ is the original image; $ratio \in (1,2)$; $mask_a$ represents the coronary artery mask and $mask_v$ means the coronary vein mask.

The objective function of TVRN is defined as:

$$\min_{\theta} \sum_{i=1}^N \mathcal{L}_{\mathcal{DD}}(f(s_i, \hat{x}_i; \theta), y_i), \quad (6)$$

where $s_i \in S$, $\hat{x}_i \in \hat{X}$, and \hat{X} is an augmented set of X by varying HU values.

3.3. Inverse distance weighted dice loss (IDD)

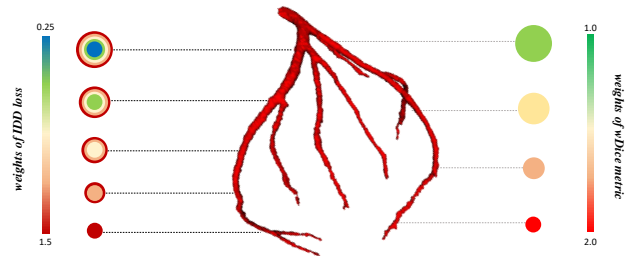


Fig. 4. Illustration of the weights of IDD loss (left) and the weights of wDice metric (right) with respect to different cross-sections of a vascular branch. The two weights are designed differently. The weight of IDD loss for a vascular voxel is based on its distance to the vascular boundary. Thus, the weights can be different in a cross-section, voxels closer to the centerline have smaller weights. While the weights of wDice metric are calculated based on the thickness of the cross-section. Thicker branches have smaller weights. The maximum weights are indicated in red, while the minimum weights are indicated in blue.

As mentioned in the previous subsections, we design an inverse distance weighted dice (IDD) loss to optimize our networks IVRN and TVRN, in order to preserve more thin structures and obtain accurate vascular edges. The left side of Fig.4 illustrates the proposed inverse distance weights. Voxels closer to the vascular edge are set to larger weights, and vice versa. The voxels in the center of the thickest vessel have the smallest weight.

We discuss how to generate these inverse distance weights. First, a 3D euclidean distance transform² is used to generate distance transforms of vessels. Each vascular voxel has a distance to the nearest boundary. Then, the distance map is normalized and transformed into an inverse distance map, so that voxels of thin structures and vascular edges have larger weights. The weights of background voxels are set to zero. We noticed that if the weights on boundaries are much higher than the neighboring background, the learned model tends to oversegment the vessels. Thus, we use a gaussian filter to smooth the gap between the vascular edges and nearby background regions to suppress the thickening of the vessels. Distance maps of the coronary artery and vein are generated separately by performing the above steps. Finally, the two distance maps for the coronary artery and vein of each case are integrated and the background voxels of zero weight are reassigned to a small weight based on the following considerations. In our method, a preprocessed CCTA ROI has dimension of 144*208*208. This ROI has over 6 million voxels, while the number of voxels belonging to the coronary artery and vein is less than 300000. accounting for only 5% of the total volume. Given that this is a typical small-target segmentation task, where both the foreground and background regions need to be identified, we have assigned a small empirical weight of 0.05 for the background. This weight is determined by considering the proportion of foreground and background voxels. Generation process is described in algorithm 1 in detail. Our distance map is different from the one proposed in (Wang et al., 2019a), we believe that setting larger weights near the vascular edge is beneficial to maintaining the shape of

Algorithm 1: The generation process of the inverse distance weights.

Input: Ground-truth labels Y

for $vessel \in (artery, vein)$ **do**

- 1.distance transform: calculate the 3D euclidean distance transform of vessels and obtain distance map D ;
- 2.distance map normalization: $D_N = \frac{D - \min(D)}{\max(D) - \min(D)}$;
- 3.inverse distance map: $D_I = 1.5 - D_N$;
- 4.nonlinear transformation: $D_I = e^{D_I / \max(D_I) - 1}$;
- 5.smoothing: background voxels are set zero weight, then use gaussian filter to smooth D_I to get D_G ;

end

- 6.combine D'_G s for artery and vein, and assign a small weight (0.05) to the background voxels with zero weight to obtain the final distance map D_F ;

return The inverse distance map D_F .

the vessel. If the segmentation branches are too thick, they will be prone to adhesion.

By integrating the inverse distance weight map into the dice loss, the IDD loss is defined as follows:

$$\mathcal{L}_{IDD} = 1 - \left(\sum_{c=1}^C \frac{2 * \sum_{k=1}^K p_{k,c} * y_{k,c} * d_k}{\sum_{k=1}^K p_{k,c}^2 * d_k + \sum_{k=1}^K y_{k,c}^2 * d_k} \right) / C, \quad (7)$$

where $c \in C$ denotes c -th class, $k \in K$ represents k -th voxel in the image, $p \in P$ means predictions by the network, and $d \in D_F$ are the inverse distance weights.

3.4. Implementation details

We employ similar encoder-decoder backbone for both IVRN and TVRN and use resblocks to extract robust features. In the resblock, a convolution layer, batch normalization layer and relu layer form a base layer, and multiple base layers are connected by using residual connections. In the encoder of IVRN, the first group of features is extracted by $3 * 3 * 3$ convolutions and 3D GCN with kernel sizes of 7 and 5, respectively. These features are concatenated to pass forward. The second group of features is extracted by $3 * 3 * 3$ convolutions and 3D GCN with only a reduced kernel size of 5 because it has a halved resolution. All following layers in the encoder and decoders use $3 * 3 * 3$ convolutions. In the training of IVRN, we first fix the parameters of the AVCB to train the encoder and the CVSB. When the CVSB tends to converge, all parameters of the IVRN are updated together. In the TVRN, the inputs of

²<https://github.com/seung-lab/euclidean-distance-transform-3d/>

MFB are randomly sampled from different vessel recognition results by the AVCB, i.e. 20% fully recognized results with minimum errors, 60% results with clear errors, and 20% completely unrecognized results. For the RSD modules in the IVRN and TVRN, 0.875, 0.75, and 0.625 are set for the living rate.

To learn more discriminative features of coronary artery and vein, we feed the network a complete CCTA ROI image instead of some randomly cropped cubes. Specifically, in data preprocessing, the CCTA image (average original resolution $0.41 \times 0.41 \times 0.51 \text{ mm}^3$) is resampled to isotropic resolution of 0.9 mm, and then a heart localization network (Wang et al., 2021) is employed to locate and extract the heart ROI of size $144 \times 208 \times 208$. The intensities are normalized by 2048 and clamped between -1 and 1. During the training process, several data augmentation techniques, including image translation, rotation, flipping, scaling and HU value shifting, are randomly performed to increase data varieties. Segmentation results are post-processed by removing isolated components smaller than a certain threshold (800 mm^3 in the experiments). All models are implemented in Pytorch and trained on 8 Tesla V100 GPUs with 32 GB memory. We use the Adam optimizer with the learning rate of 10^{-3} and the maximum epoch is set to 2000.

4. Experiments

4.1. Dataset and metric

We conduct extensive experiments to evaluate our method by using a multi-center dataset of 700 CCTA images. These images were acquired by different CT scanners, including the Siemens SOMATOM Definition, the Siemens SOMATOM Force, the GE Discovery, the GE Revolution, and the Philips IQon-Spectral CT. The in-plane resolution ranges from 0.3 mm to 0.6 mm and the slice thickness ranges from 0.5 mm to 0.7 mm. The in-plane image size is 512×512 , while the number of slices in z-axis varies from 200 to 500. Of these 700 CCTA images, 400 cases are used for training, 100 cases for validation and 200 cases for testing. Each data was annotated on the SenseCare platform (Duan et al., 2020) by two experienced doctors, one doctor with 7 years of clinical experience and one

specialist with 13 years of experience in cardiology. Two doctors worked together to discuss any disagreement and agreed to any changes to update the annotation. The annotations include both coronary arteries and veins, with detailed labeling of branches with diameters as small as 0.5 mm (equivalent to just one pixel). The samples were divided into "easy cases" and "hard cases" based on the difference between coronary arterial and venous intensities. Specifically, cases with a difference of less than 150 HU were classified as hard cases, while those with a larger difference were classified as easy cases. Based on this threshold, there were 494 easy cases and 206 hard cases. We randomly selected 100 hard cases and 100 easy cases to form the testing dataset and the training and validation dataset comprise 106 hard cases and 394 easy cases. Including more hard cases can better highlight the advantages of our method in challenging scenarios. Therefore, we focus on evaluating the performance of models on the hard set in the following subsections.

We adopt several metrics to quantify the results, including the centerline recall (cl-rec), precision (cl-pre), and overlap (cl-ov) defined in (Schaap et al., 2009), a weighted Dice (wDice) coefficient and the commonly used 95% HD for segmentation task. The centerline metrics can evaluate the precision of the extracted vessel tree, while wDice and 95%HD can measure the accuracy at the voxel level. The wDice is proposed because we noticed that the traditional Dice could be easily affected by subtle errors of thick vessel branches but insensitive to clear errors on thin branches. However, the thick vessels are easier to segment than the thin ones in practice. Therefore, we design the wDice so that thin vascular structures can get a higher weight than thick branches. In order to differentiate the weights of IDD loss, we show the weights of wDice in the right of Fig.4 for comparison. The definition of wDice is as follows:

$$wDice = \frac{2 * \sum_{k=1}^K p_k * y_k * w_k}{\sum_{k=1}^K p_k^2 * w_k + \sum_{k=1}^K y_k^2 * w_k}, \quad (8)$$

where $w_k \in W$ represents the weighting index of k -th voxel. We compute W in three steps. (1) We calculate the distance transform D of vessels as mentioned in Sec.3.3, and get the inverse map $D_{inv} = 1/D + 1$. (2) The centerline points are extracted (Sato et al., 2000) from the ground truth labels. (3) We find the

corresponding centerline point for each vascular voxel as the closest one. Each vascular voxel is assigned the same weight as the value of its associated centerline point in D_{inv} , and the background weights are set to one.

Besides, we also measure the percentage of veins which are mistaken as arteries by computing arterial centerline false positive rate: $a-cl-fpr = \frac{FP_v}{T_v}$. Here, the FP_v represents the number of ground-truth centerline points of veins that are misclassified into arteries (at least one of the predicted arterial centerline points found within a certain distance). T_v represents the total number of venous centerline points.

4.2. Contribution of IDD loss

We first investigate the contribution of the loss function in our proposed method to the coronary artery segmentation task. We employ a modified VNet (mVNet) backbone (Milletari *et al.*, 2016) with three downsampling operations and RSD module for comparison. Table 1 reports the results of mVNet supervised by several different loss functions, including traditional Dice loss (DL), Hausdorff distance loss (Karimi and Salcudean, 2019) (HDL), radial distance loss (Wang *et al.*, 2019a) (RDL), and the proposed IDD loss. HDL (Karimi and Salcudean, 2019) can protect targets' boundaries and RDL (Wang *et al.*, 2019a) can recover thin tubular structures. From Table 1, we can see that DL has the lowest a-cl-rec rate with 0.8269 and the highest a-cl-pre rate with 0.9102. This is because, on the one hand, the imbalance problem exists not only between vessels and background but also between thin branches and thick branches. Some thin structures will be missed by this region-based loss. On the other hand, since the extracted vessels are thicker branches, there are fewer prediction errors, resulting in a higher a-cl-pre rate. Compared to DL, HDL has improvements in a-cl-ov and a-wDice, which means it can segment more details. RDL and IDD loss focus on thin structures, placing higher weights on the small targets. The difference between these two loss functions is the weights on the boundaries. At a similar a-cl-rec level, IDD loss achieves about 4% improvements on a-cl-pre and 5% improvements on a-wDice compared to RDL. This is because that IDD sets larger weights near the vascular

edge, taking good care of thin branches with only a few voxels and preserving the boundaries of thick branches, simultaneously. IDD loss also achieves the lowest average 95% HD of arteries (a-hd95), demonstrating its ability to preserve similar shapes to the ground-truth labels and improve accuracy at the vascular boundary, aligning with the goals of its design.

Table 1. Comparison of several different loss functions on the hard set. They are achieved by the mVNet, supervised by the Dice loss (DL), Hausdorff distance loss (Karimi and Salcudean, 2019) (HDL), radial distance loss (Wang *et al.*, 2019a) (RDL) and IDD loss, respectively.

Methods	a-cl-rec	a-cl-pre	a-cl-ov	a-wDice	a-hd95 (mm)
DL	0.8269	0.9102	0.8602	0.8481	6.8885
	± 0.1100	± 0.0625	± 0.0779	± 0.0619	± 11.0489
HDL	0.8433	0.9078	0.8682	0.8547	6.0457
	± 0.1093	± 0.0655	± 0.0778	± 0.0569	± 10.1133
RDL	0.9088	0.7999	0.8465	0.8135	5.6290
	± 0.0723	± 0.0864	± 0.0641	± 0.0427	± 6.8446
IDD	0.9111	0.8306	0.8647	0.8619	5.1610
	± 0.0737	± 0.0844	± 0.0613	± 0.0420	± 7.2966

More qualitative results are illustrated in Fig. 5, which shows both the segmentation results and prediction errors. We can see that the segmentation results of mVNet with DL miss a lot of details and even break in the thick branches. mVNet with HDL or RDL is better in completeness but still misses some thin structures. Please note that mVNet with IDD can even yield more details than ground-truth labels on thin branches. However, these extra branches may lead to a decrease in a-cl-precision, which is consistent with the results in Table 1. Prediction errors mainly exist at the boundaries of coronary vessels. Slight differences will not affect the results, but if the branches are too thick, they will be prone to adhesion, like the prediction of case3 produced by mVNet with RDL. Overall, experimental results show that mVNet with IDD loss is suitable to improve the detection ability of coronary vessels. In the following subsections, mVNet denotes the method using IDD loss by default.

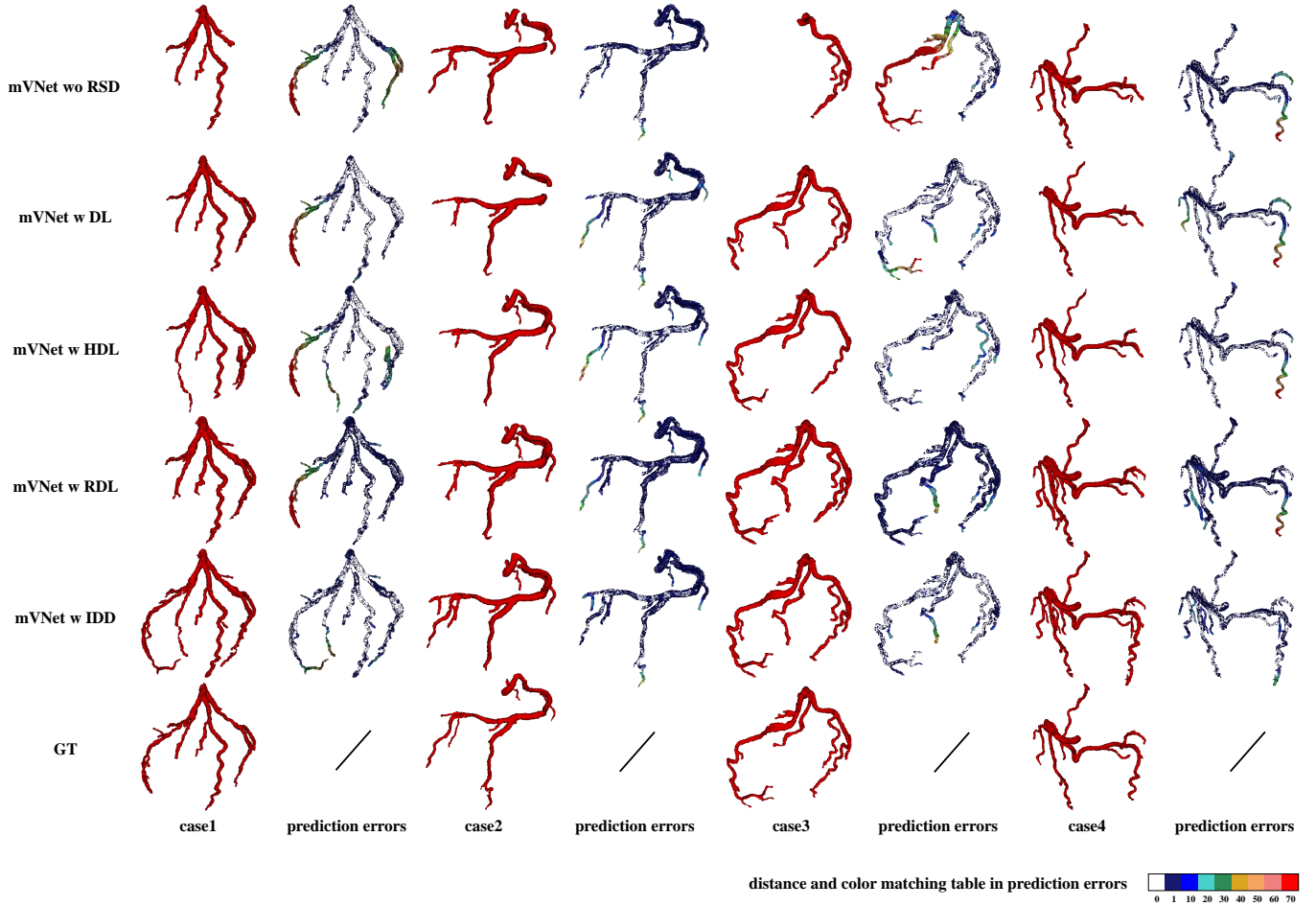


Fig. 5. Qualitative results of coronary artery segmentation. Each row is produced by a different method. Each two columns corresponds to a case, representing the segmentation result and its surface distance errors compared to the ground-truth (GT) label. The color map is plotted in the bottom. The red in the prediction error columns represent the surface errors of at least 70.

4.3. Different settings of AVDNet

To demonstrate the effectiveness of all components of AVDNet, we conduct multiple experiments with different settings. The baseline model is mVNet. Table 2 shows the results on hard set under different settings. In the first two rows of Table 2, we only segment the arteries and merge veins into the background label. These two rows show that the RSD module can help improve the a-cl-rec, a-cl-ov, and a-wDice significantly. From qualitative examples in the first row of Fig. 5, we observe that mVNet without RSD tends to undersegment arteries, especially the thin branches. Although mVNet without RSD has a lower probability of false positives, leading to higher a-cl-pre and lower a-cl-fpr, a more complete coronary artery tree is desired for subsequent plaque and stenosis diagnosis in practice. mVNet-av and IVRN-onedec directly recognize the voxel into

one of the three classes including background, artery and vein. The difference between mVNet-av and IVRN-onedec is just in the GCN module. The results of wDice and 95%HD show improvements with the GCN module. Comparison between the IVRN wo GCN and IVRN further attests to the efficacy of the GCN module. Compare to the methods for segmenting artery only, simultaneous artery and vein segmentation can reduce the a-cl-fpr while maintaining the a-cl-rec. In the IVRN, we use two decoders to split the multi-class segmentation task into two subtasks, and the results show that IVRN equipped with two decoders have significant improvements on the vein segmentation compared to the IVRN with one decoder (IVRN-onedec). This is because that the low response of coronary vein in CCTA images and large intersubject variance, feature learning on coronary veins is much more difficult than that of arteries. The coro-

Table 2. Effect of each component in the AVDNet. The results are generated on the hard set, by the mVNet without RSD module (mVNet wo RSD), mVNet only segmenting artery (mVNet-artery), mVNet segmenting artery and vein simultaneously (mVNet-av), IVRN segmenting artery and vein directly in one decoder (IVRN-onedec), IVRN without GCN module (IVRN wo GCN), IVRN supervised by the wDice loss (IVRN-wDice), IVRN, IVRN with the number of base feature is 16 (IVRN-base16), IVRN followed by a mVNet for refinement (IVRN+mVNet) and AVDNet, respectively.

Methods	a-cl-rec	a-cl-pre	a-cl-ov	a-cl-fpr	v-cl-ov	a-wDice	v-wDice	a-hd95 (mm)	v-hd95 (mm)
mVNet wo RSD	0.8398 ±0.1013	0.8633 ±0.0794	0.8455 ±0.0754	0.0836 ±0.0730	/	0.8374 ±0.0572	/	6.8593 ±8.6271	/
mVNet-artery	0.9111 ±0.0737	0.8306 ±0.0844	0.8647 ±0.0613	0.1264 ±0.0968	/	0.8619 ±0.0420	/	5.1610 ±7.2966	/
mVNet-av	0.9169 ±0.0738	0.8404 ±0.0793	0.8726 ±0.0567	0.0971 ±0.0803	0.7293 ±0.1072	0.8668 ±0.0377	0.8570 ±0.0464	5.1903 ±8.2441	11.7618 ±11.2895
IVRN-onedec	0.9181 ±0.0642	0.8414 ±0.0776	0.8748 ±0.0558	0.0885 ±0.0789	0.7408 ±0.1029	0.8687 ±0.0367	0.8656 ±0.0492	4.1543 ±6.0040	11.6246 ±12.0214
IVRN wo GCN	0.9031 ±0.0695	0.8274 ±0.0816	0.8597 ±0.0583	0.0857 ±0.0703	0.7653 ±0.0927	0.8611 ±0.0393	0.8724 ±0.0425	5.5022 ±7.1055	7.4888 ±8.4008
IVRN-wDice	0.8615 ±0.0737	0.8815 ±0.0800	0.8679 ±0.0593	0.0970 ±0.0960	0.7444 ±0.0962	0.8454 ±0.0442	0.8592 ±0.0402	5.2166 ±8.2176	10.2862 ±11.1169
IVRN	0.9056 ±0.0720	0.8524 ±0.0740	0.8747 ±0.0564	0.0637 ±0.0586	0.7866 ±0.0892	0.8698 ±0.0335	0.8795 ±0.0382	4.2516 ±6.1170	6.0932 ±6.9813
IVRN-base16	0.9093 ±0.0837	0.8455 ±0.0693	0.8722 ±0.0620	0.0725 ±0.0704	0.7836 ±0.0875	0.8674 ±0.0449	0.8813 ±0.0369	4.9014 ±8.3712	7.3040 ±9.4342
IVRN+mVNet	0.9094 ±0.0718	0.8414 ±0.0767	0.8701 ±0.0578	0.0596 ±0.0589	0.7874 ±0.0876	0.8552 ±0.0332	0.8738 ±0.0364	4.3501 ±6.5160	5.9239 ±6.7688
AVDNet	0.9256 ±0.0571	0.8448 ±0.0730	0.8807 ±0.0537	0.0520 ±0.0481	0.8037 ±0.0828	0.8753 ±0.0326	0.8844 ±0.0366	3.4045 ±5.8828	5.8186 ±7.2026

nary vessels segmentation branch of IVRN wo GCN treats coronary artery and vein as the same category, which is beneficial to improve the coronary vein detection rate. We also replace the IDD loss with wDice loss to supervise IVRN, thus exploring the impact of different weights on the results. The results show that the IVRN-wDice have poor results than the IVRN with IDD loss (IVRN) on both arteries and veins, especially for the wDice and hd95 of the veins, which decrease by 2.3% and increased by 68.8%, respectively.

The TVRN as the second learning stage contains two encoders MFB and IFB to refine the results of IVRN. In our method, the total number of parameters of IVRN (2935840) and TVRN (2905022) is 5840862. To demonstrate the improvement of TVRN comes from the design of TVRN itself, and not from the increase of network parameters, we increase the

number of initial feature maps of the IVRN to increase its total number of parameters to 5215948, whose results are shown in the row named "IVRN-base16" of Table.2. We can see that the results of IVRN-base16 do not show any improvements compared to the original IVRN. Another ablation study replaces the TVRN with a mVNet which has a similar structure as the MFB+decoder in TVRN, but only inputs the recognition masks with a single state to extract morphological features. Its results are shown in the row named "IVRN+mVNet" of Table.2. There are obvious differences between the results of IVRN+mVNet and AVDNet. The performances of IVRN+mVNet are even not better than those of the original IVRN. We believe that the key to TVRN's improvement lies on the contribution of MFB, which can learn to encode the topological consistency implicitly from the recognition masks in different states of IVRN.

Table 3. More testing results on the easy set. Most differences are not significant and we only compare mVNet-artery, mVNet-av, and AVDNet.

Methods	a-cl-rec	a-cl-pre	a-cl-ov	a-cl-fpr	v-cl-ov	a-wDice	v-wDice	a-hd95 (mm)	v-hd95 (mm)
mVNet-artery	0.9420 ±0.0519	0.8738 ±0.0615	0.9041 ±0.0393	0.0628 ±0.0665	/	0.8838 ±0.0237	/	2.4651 ±2.3985	/
mVNet-av	0.9455 ±0.0461	0.8761 ±0.0596	0.9076 ±0.0407	0.0510 ±0.0574	0.7445 ±0.1106	0.8852 ±0.0226	0.8522 ±0.0538	2.2597 ±2.6810	10.9508 ±13.3822
AVDNet	0.9470 ±0.0439	0.8763 ±0.0664	0.9081 ±0.0428	0.0314 ±0.0435	0.7998 ±0.0988	0.8876 ±0.0219	0.8732 ±0.0492	2.1022 ±2.8215	5.9922 ±9.5264

By using TVRN, AVDNet achieves the best results in a-cl-ov with 0.8807, v-cl-ov with 0.8037, a-cl-fpr with 0.0520, a-wDice with 0.8753, v-wDice with 0.8844, a-hd95 with 3.4045 and v-hd95 with 5.8186, whose improvements are up to 0.9%, 10.2%, 46.4%, 1.0%, 3.2%, 34.3%, and 50.5% compared to mVNet-av.

Furthermore, we utilize the Estrada Index(De La Peña *et al.*, 2007), a topological index of folding or 3D compactness for a graph, to measure the improvement on topology consistency by using TVRN. The Estrada Index was originally proposed to measure the compactness of molecular structures. It has been extended to all undirected graphs, including trees. In our study, we introduce a new metric for topology consistency by computing absolute difference between the Estrada Index of the segmentation result and the annotation, assuming that smaller difference corresponds to better topology consistency. Firstly, we extracted the skeletons of the arteries from the segmentation and constructed a tree based on these skeleton points. Next, we formed a graph using the edges of the tree and calculated the Estrada Index for this graph. As a result, each case in the testing dataset will be associated with an Estrada Index for a given method, collectively forming a group. Different methods have different groups of Estrada Index values. We then measured the case-by-case absolute difference between IVRN and labels, as well as between AVDNet and labels. The mean value and standard deviation were computed based on two sets of absolute differences (IVRN and labels vs AVDNet vs labels), resulting in 174.0388 ± 184.6441 and 152.6851 ± 177.4118 , respectively. Furthermore, out of the 100 hard cases, AVDNet exhibited a lower absolute difference in Estrada Index compared to IVRN

in 64 cases, indicating clear improvement by AVDNet in terms of topology consistency.

Some methods such as mVNet-artery, mVNet-av, and AVDNet are also tested on the easy set and their results are reported in Table. 3. CCTA images in the easy set have a strong intensity distinction between coronary artery and vein, and its tested results are easily satisfying. As we can see, the main differences between the methods are located in a-cl-fpr, v-cl-ov, v-wDice and v-hd95. There is a strong link between these four metrics. When more coronary veins can be correctly predicted, the values of v-cl-ov, v-wDice and v-hd95 increase, while the a-cl-fpr value decreases. By integrating the coronary vein segmentation into the coronary artery segmentation task, wrong segmentation of veins into arteries is alleviated. In Table. 3, compared with mVNet-artery, mVNet-av improves the a-cl-ov by over 8% and AVDNet improves by 50%. Different from the easy set, images in the hard set have weak intensity distinction between coronary artery and vein, making it difficult to distinguish coronary artery from the vein if the voxels cross each other in close vicinity. In Table. 2, the a-cl-fpr of mVNet-artery ups to 0.1264.

Due to the limited space and the importance of coronary artery, we just place parts of results of coronary vein in the main text. More comprehensive results including a-cl-rec, a-cl-pre, a-cl-ov, a-cl-fpr, v-cl-rec, v-cl-pre, v-cl-ov, v-cl-fpr, a-wDice, v-wDice, a-hd95 and v-hd95 can be found in Table.S2 and Table.S3 in the supplementary materials.

More qualitative results on coronary artery and vein segmentation are shown in Fig. 6. From the magnified parts of original images, we can see that all coronary arteries and veins have

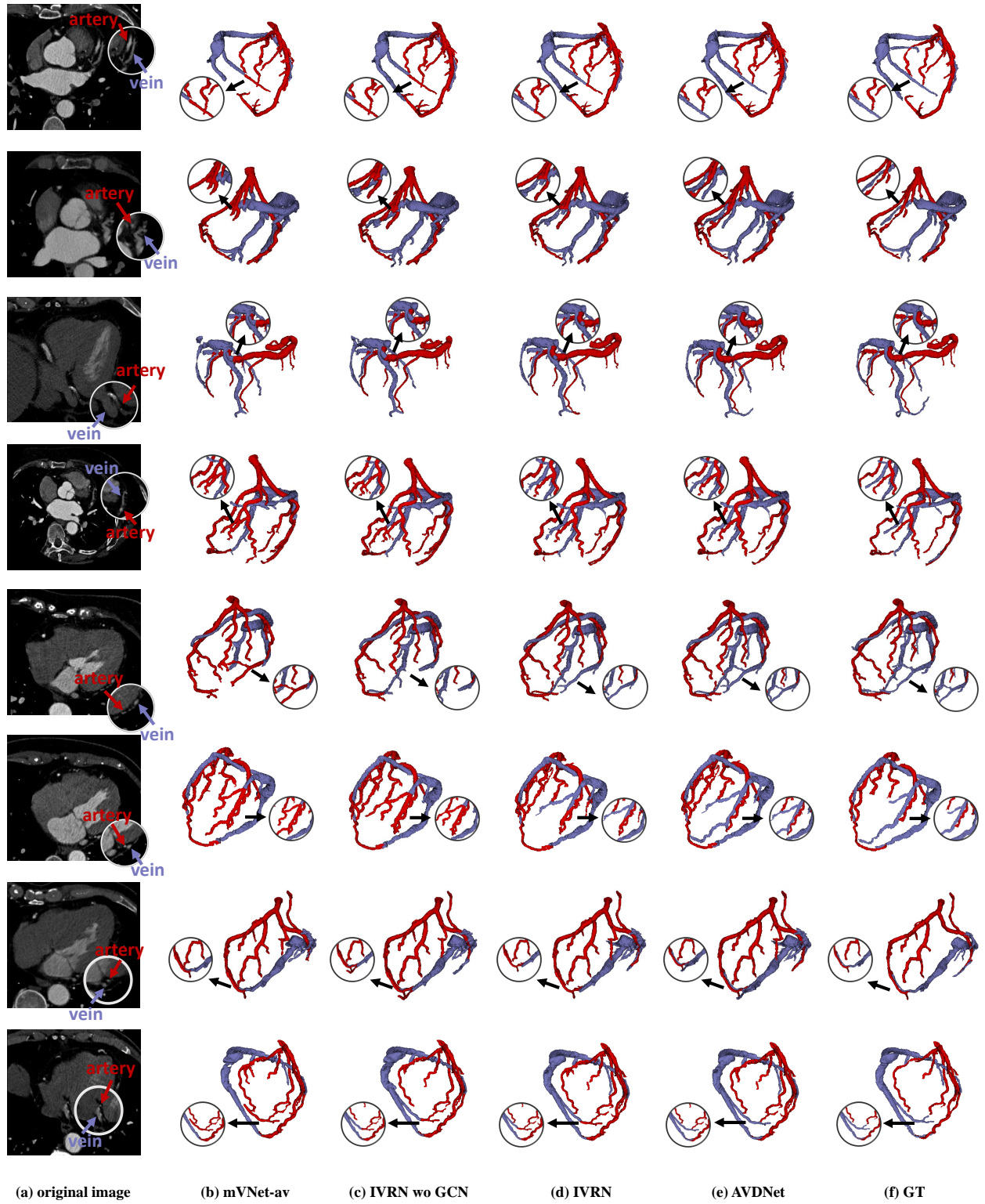


Fig. 6. Qualitative results of coronary artery and vein segmentation. Some parts of the original image are magnified for a better view of the arterial and venous vessels. The highlighted regions show the advantage of our method in disentangling arteries and veins when they are difficult to distinguish in the image. Red: artery; blue: vein.

similar intensity in the local location, which can easily lead to misclassification. We also zoom in on some parts of the pre-

dictions in Fig. 6, in order to compare the methods intuitively. Consistent with those quantitative results (such as a-cl-fpr and

Table 4. Comparisons with the state-of-the-art methods on the hard set, including Attention-FCN(Lei et al., 2020), Dense-UNet(Pan et al., 2021), and FFR-UNet(Song et al., 2022). For a fair comparison, they all use IDD loss during the training process.

Methods	a-cl-rec	a-cl-pre	a-cl-ov	a-cl-fpr	v-cl-ov	a-wDice	v-wDice	a-hd95 (mm)	v-hd95 (mm)
Attention-FCN	0.8096	0.8022	0.8007	0.0910	0.6213	0.8290	0.7947	10.7667	15.0600
with IDD	± 0.1159	± 0.0805	± 0.0797	± 0.0785	± 0.1151	± 0.0576	± 0.0631	± 12.4486	± 10.6353
Dense-UNet	0.8658	0.7827	0.8124	0.1102	0.6501	0.8469	0.8310	7.9494	14.3192
with IDD	± 0.1162	± 0.0886	± 0.0737	± 0.1031	± 0.1080	± 0.0570	± 0.0552	± 9.9011	± 11.3017
FFR-UNet	0.7658	0.6249	0.6703	0.2363	0.3343	0.7525	0.5064	26.7910	48.8348
with IDD	± 0.1403	± 0.1334	± 0.1050	± 0.1802	± 0.1855	± 0.0904	± 0.2003	± 33.5016	± 36.7690
AVDNet	0.9256	0.8448	0.8807	0.0520	0.8037	0.8753	0.8844	3.4045	5.8186
	± 0.0571	± 0.0730	± 0.0537	± 0.0481	± 0.0828	± 0.0326	± 0.0366	± 5.8828	± 7.2026

v-cl-ov) in Table. 2 and Table. 3, mVNet-av achieves the worst results and AVDNet has the best performance in Fig. 6. Compared with IVRN, TVRN takes topology consistency into account and further improves the ability to distinguish artery and vein in close vicinity. AVDNet combines IVRN and TVRN to produce the most robust and accurate predictions.

4.4. Comparisons to state-of-the-art methods

We compare our method with some latest coronary artery segmentation methods, including Attention-FCN(Lei et al., 2020), Dense-UNet(Pan et al., 2021) and FFR-UNet(Song et al., 2022). We re-implemented these methods accordingly. For a fair comparison, all methods were extended to segment the coronary artery and vein simultaneously, and IDD loss was employed for optimization. Attention-FCN(Lei et al., 2020) and Dense-UNet(Pan et al., 2021) are trained by $512 * 512 * 32$ and $512 * 512 * 16$ sliding windows along the z-axis, respectively. FFR-UNet(Song et al., 2022) filtered out the non-coronary slices and cropped the images into $64 * 64 * 64$ cubes as input.

The comparison results are shown in Table 4. In contrast to the training strategies of cropped patches, we feed a complete heart ROI to the AVDNet by downsampling the CCTA images. On the one hand, the downsampled input makes the network learn better global features and predict more continuous coronary arteries and veins. On the other hand, network convergence is accelerated. Note that all the metrics are computed in the original resolution. We can see that there are signifi-

cant gaps between these comparing methods and the AVDNet in Table 4. The AVDNet achieves the best performance on all metrics, especially maintaining a higher vascular detection rate and lower false positives.

4.5. Prediction results on ASOCA dataset

We further validate our method on the Automated Segmentation of Coronary Arteries (ASOCA) Challenge dataset. ASOCA provides a publicly available dataset of 40 CCTA images, comprising 20 healthy cases and 20 patients with confirmed coronary artery disease. Each case was annotated into two classes, coronary artery and background, by three annotators, which is not available for training the AVDNet. Therefore, we use the trained AVDNet model from our dataset to infer the ASOCA data. Ten prediction results are illustrated in Fig. 7. We compare the results of left coronary artery (LCA) and right coronary artery (RCA) between our predictions and the ASOCA annotations. We can see much more details of the arteries in the inference results from our method than in the ASOCA annotations, which leads to an average Dice score of 0.5304 and an average 95% HD of 20.8647 mm. In the last row of Fig. 7, we show the integrated prediction results for these examples, including left coronary artery, right coronary artery, and coronary vein. Our prediction results, including both arteries and veins, of 40 ASOCA data are available at <https://github.com/WennyJJ/Coronary-Artery-Vein-Segmentation> for follow-up research.

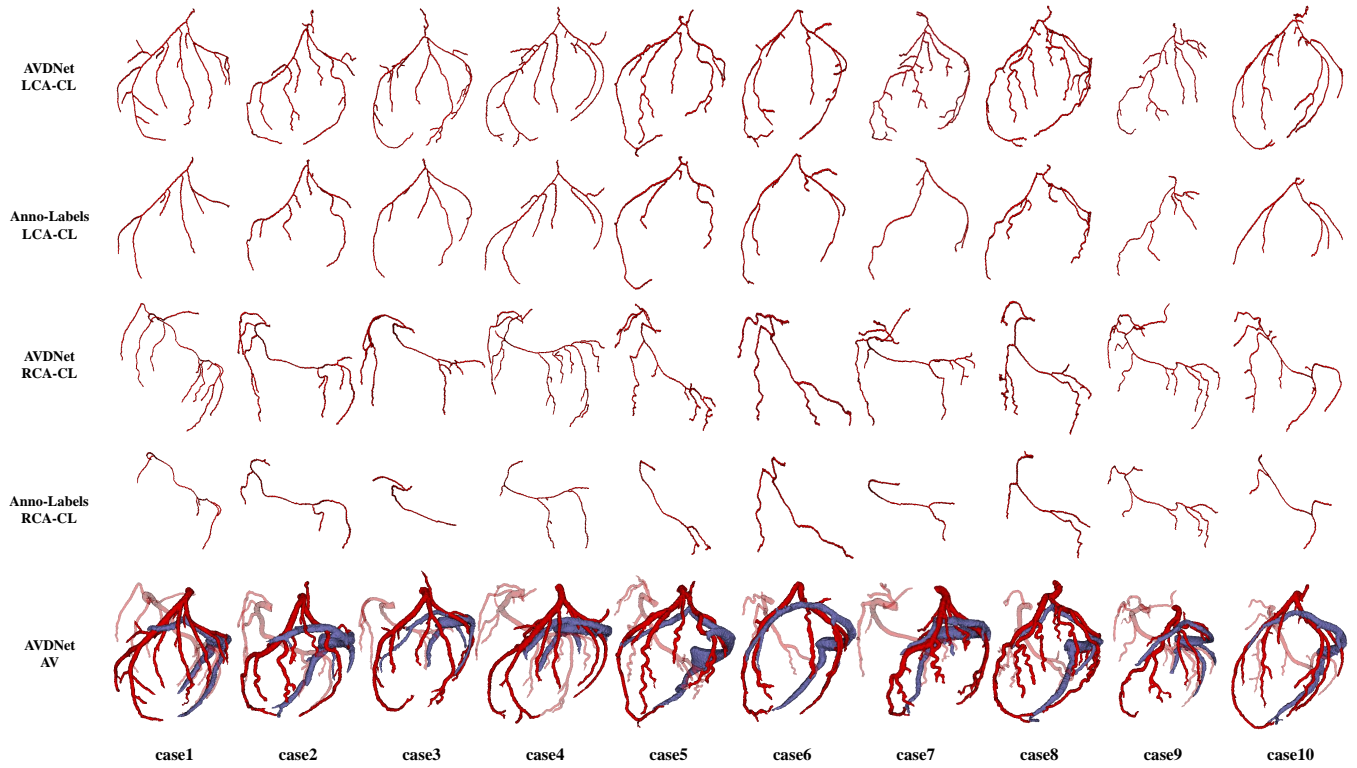


Fig. 7. Evaluation on the ASOCA dataset. We show the extracted centerlines of segmentation inferred by our AVDNet and of the ASOCA annotations (Anno-Labels). Notably, the centerline points are dilated for a better view. LCA-CL: centerlines of left coronary artery; RCA-CL: centerlines of right coronary artery. The last row shows 3D views of our prediction for artery and vein (AV). The RCA is displaced in transparency for clearer identification. Red: artery; blue: vein.

4.6. A real-world hard case study

To further elucidate the relationship between coronary artery segmentation and subsequent steps for diagnosis, we present a real-world hard case study with similar intensities between coronary arteries and veins (shown in Fig.8). The first row of Fig.8 displays segmentation results generated by the mVNet-artery, which directly segments coronary arterial branches from the image. Due to the similar intensities, many coronary veins are misidentified as coronary arteries (indicated by green-color arrows), such as the vessels between the RCA and LCX or those at the end of the OM. False positive vessels located between the RCA and LCX cause adhesion of the left and right coronary artery trees. When centerlines are extracted using the shortest path algorithm(Sato et al., 2000), only one vessel tree can be obtained and the LCX (shown in the first row, second column of Fig.8) and the LM (shown in the first row, forth column of Fig.8) miss their centerlines. The failure in centerline extraction can mislead the curved planner reconstruction (CPR). Since in

clinical practices, both plaques and stenosis are diagnosed and assessed on CPR images, an experienced doctor has to manually correct the segmentation and the trace of centerlines in this case for an accurate diagnosis, or the patient may be asked to retake the CT scan for a better quality image. On the contrary, the AVDNet produces significantly better segmentation results (shown in the second row of Fig.8) in such a complex scenario, which can benefit both the patient and the doctor.

5. Conclusion

In this paper, we integrated coronary vein segmentation into the artery segmentation task to address the issue of false positives in arterial prediction results. A cascaded network named AVDNet was proposed for coronary artery and vein segmentation, which consists of IVRN and TVRN. IVRN produced the initial segmentation results and TVRN revised the misclassification parts. Furthermore, IDD loss was designed to segment more thin vessel branches and protect vascular edges. Experi-

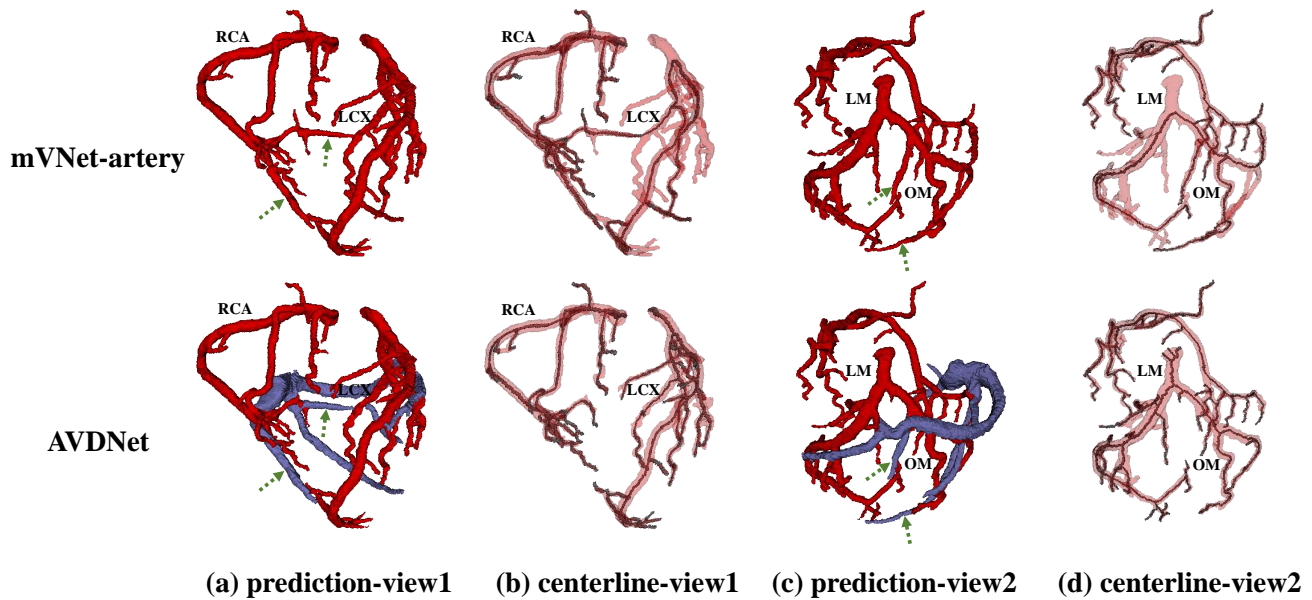


Fig. 8. A real-world hard case with similar intensities between coronary arteries and veins. The results are inferred by the mVNet-artery and the AVDNet. (a) prediction shown on a certain perspective (view1) (b) the extracted centerline from prediction shown on a same perspective with (a) (c) prediction shown on another perspective (view2) (b) the extracted centerline from prediction shown on a same perspective with (d). All centerline points are dilated for a better view. Red: artery; blue: vein; grey: centerline; green arrows: the misclassified parts by the mVNet-artery

ments were conducted on a multi-center dataset and methods were evaluated by different metrics. Quantitative and Qualitative results demonstrated that the proposed AVDNet has a strong ability to distinguish coronary artery and vein in CCTA images with whether strong discrimination of arterial and venous intensity values or weak ones.

Acknowledgments

This work was supported by the Beijing Nova Program (Z201100006820064, Z211100002121165), the National Key Research and Development Project of China (2020YFC2004800), and the research cultivation fund of Capital Medical University (pyz20003).

References

- Chang, Q., Yan, Z., Zhou, M., Liu, D., Sawalha, K., Ye, M., Zhangli, Q., Kanski, M., Aref, S.A., Axel, L., et al., 2022. Deeprecon: Joint 2d cardiac segmentation and 3d volume reconstruction via a structure-specific generative method. *arXiv preprint arXiv:2206.07163*.
- Chen, F., Li, Y., Tian, T., Cao, F., Liang, J., 2018. Automatic coronary artery lumen segmentation in computed tomography angiography using paired multi-scale 3d cnn, in: *Medical imaging 2018: Biomedical applications in molecular, structural, and functional imaging*, SPIE. pp. 652–658.
- Chen, Y.C., Lin, Y.C., Wang, C.P., Lee, C.Y., Lee, W.J., Wang, T.D., Chen, C.M., 2019. Coronary artery segmentation in cardiac ct angiography using 3d multi-channel u-net. *arXiv preprint arXiv:1907.12246*.
- Cui, H., Liu, X., Huang, N., 2019. Pulmonary vessel segmentation based on orthogonal fused u-net++ of chest ct images, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 293–300.
- Cui, X., Cao, Y., Liu, Z., Sui, X., Mi, J., Zhang, Y., Cui, L., Li, S., 2022. Trsa-net: Task relation spatial co-attention for joint segmentation, quantification and uncertainty estimation on paired 2d echocardiography. *IEEE Journal of Biomedical and Health Informatics*.
- Davies, R.H., Augusto, J.B., Bhuva, A., Xue, H., Treibel, T.A., Ye, Y., Hughes, R.K., Bai, W., Lau, C., Shiwani, H., et al., 2022. Precision measurement of cardiac structure and function in cardiovascular magnetic resonance using machine learning. *Journal of Cardiovascular Magnetic Resonance* 24, 1–11.
- De La Peña, J.A., Gutman, I., Rada, J., 2007. Estimating the estrada index. *Linear Algebra and its Applications* 427, 70–76.
- Duan, Q., Wang, G., Wang, R., Fu, C., Li, X., Gong, M., Liu, X., Xia, Q., Huang, X., Hu, Z., et al., 2020. Sensecare: A research platform for medical image informatics and interactive 3d visualization. *arXiv preprint arXiv:2004.07031*.
- Gu, L., Cai, X.C., 2021. Fusing 2d and 3d convolutional neural networks for the segmentation of aorta and coronary arteries from ct images. *Artificial Intelligence in Medicine* 121, 102189.
- Guo, Q., Song, H., Fan, J., Ai, D., Gao, Y., Yu, X., Yang, J., 2022. Portal vein and hepatic vein segmentation in multi-phase mr images using flow-guided change detection. *IEEE Transactions on Image Processing*.
- He, Y., Yang, G., Yang, J., Ge, R., Kong, Y., Zhu, X., Zhang, S., Shao, P., Shu, H., Dillenseger, J.L., et al., 2021. Meta grayscale adaptive network for 3d integrated renal structures segmentation. *Medical Image Analysis* 71, 102055.
- Huang, G., Sun, Y., Liu, Z., Sedra, D., Weinberger, K.Q., 2016. Deep networks with stochastic depth, in: *European conference on computer vision*, Springer. pp. 646–661.
- Hunter, I., Soraghan, J., McDonagh, T., 1995. Fully automatic left ventricular boundary extraction in echocardiographic images, in: *Computers in Cardiology 1995*, IEEE. pp. 741–744.
- Jimenez-Carretero, D., Bermejo-Peláez, D., Nardelli, P., Fraga, P., Fraile, E., Estépar, R.S.J., Ledesma-Carbayo, M.J., 2019. A graph-cut approach for pulmonary artery-vein segmentation in noncontrast ct images. *Medical image analysis* 52, 144–159.
- Karimi, D., Salcudean, S.E., 2019. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions*

- on medical imaging 39, 499–513.
- Keshwani, D., Kitamura, Y., Ihara, S., Iizuka, S., Simo-Serra, E., 2020. Topnet: topology preserving metric learning for vessel tree reconstruction and labelling, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 14–23.
- Kirbas, C., Quek, F., 2004. A review of vessel extraction techniques and algorithms. *ACM Computing Surveys (CSUR)* 36, 81–121.
- Kompatsiaris, I., Tzovaras, D., Koutkias, V., Strintzis, M.G., 2000. Deformable boundary detection of stents in angiographic images. *IEEE Transactions on Medical Imaging* 19, 652–662.
- Kong, B., Wang, X., Bai, J., Lu, Y., Gao, F., Cao, K., Xia, J., Song, Q., Yin, Y., 2020. Learning tree-structured representation for 3d coronary artery segmentation. *Computerized Medical Imaging and Graphics* 80, 101688.
- Laibacher, T., Weyde, T., Jalali, S., 2019. M2u-net: Effective and efficient retinal vessel segmentation for real-world applications, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0.
- Lei, Y., Guo, B., Fu, Y., Wang, T., Liu, T., Curran, W., Zhang, L., Yang, X., 2020. Automated coronary artery segmentation in coronary computed tomography angiography (ccta) using deep learning neural networks, in: *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, International Society for Optics and Photonics. p. 1131812.
- Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S., 2022. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Medical Image Analysis* 80, 102517.
- Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: *2016 Fourth International Conference on 3D Vision (3DV)*, IEEE. pp. 565–571.
- Nardelli, P., Jimenez-Cartero, D., Bermejo-Pelaez, D., Washko, G.R., Rahaghi, F.N., Ledesma-Carbayo, M.J., Estépar, R.S.J., 2018. Pulmonary artery–vein classification in ct images using deep learning. *IEEE transactions on medical imaging* 37, 2428–2440.
- Nekovei, R., Sun, Y., 1995. Back-propagation network and its configuration for blood vessel detection in angiograms. *IEEE transactions on neural networks* 6, 64–72.
- Niki, N., Kawata, Y., Satoh, H., Kumazaki, T., 1993. 3d imaging of blood vessels using x-ray rotational angiographic system, in: *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, IEEE. pp. 1873–1877.
- Pan, L.S., Li, C.W., Su, S.F., Tay, S.Y., Tran, Q.V., Chan, W.P., 2021. Coronary artery segmentation under class imbalance using a u-net based architecture on computed tomography angiography images. *Scientific Reports* 11, 1–7.
- Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large kernel matters—improve semantic segmentation by global convolutional network, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4353–4361.
- Qin, Y., Chen, M., Zheng, H., Gu, Y., Shen, M., Yang, J., Huang, X., Zhu, Y.M., Yang, G.Z., 2019. Airwaynet: a voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 212–220.
- Qin, Y., Zheng, H., Gu, Y., Huang, X., Yang, J., Wang, L., Yao, F., Zhu, Y.M., Yang, G.Z., 2021. Learning tubule-sensitive cnns for pulmonary airway and artery-vein segmentation in ct. *IEEE Transactions on Medical Imaging* 40, 1603–1617.
- Qu, Y., Li, X., Yan, Z., Zhao, L., Zhang, L., Liu, C., Xie, S., Li, K., Metaxas, D., Wu, W., et al., 2021. Surgical planning of pelvic tumor using multi-view cnn with relation-context representation learning. *Medical Image Analysis* 69, 101954.
- Sarwal, A., Dhawan, A.P., 1994. 3-d reconstruction of coronary arteries, in: *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE. pp. 504–505.
- Sato, M., Bitter, I., Bender, M.A., Kaufman, A.E., Nakajima, M., 2000. Teasar: Tree-structure extraction algorithm for accurate and robust skeletons, in: *Proceedings of the Eighth Pacific Conference on Computer Graphics and Applications*, IEEE. pp. 281–449.
- Schaap, M., Metz, C.T., van Walsum, T., van der Giessen, A.G., Weustink, A.C., Mollet, N.R., Bauer, C., Bogunović, H., Castro, C., Deng, X., et al., 2009. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical image analysis* 13, 701–714.
- Shen, Y., Fang, Z., Gao, Y., Xiong, N., Zhong, C., Tang, X., 2019. Coronary arteries segmentation based on 3d fcn with attention gate and level set function. *Ieee Access* 7, 42826–42835.
- Song, A., Xu, L., Wang, L., Yang, X., Xu, B., Wang, B., Yang, B., Greenwald, S., 2022. Automatic coronary artery segmentation of ccta images with an efficient feature-fusion-and-rectification 3d-unet. *IEEE Journal of Biomedical and Health Informatics* .
- Stansfield, S.A., 1986. Angy: A rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 188–199.
- Tolias, Y.A., Panas, S.M., 1998. A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Transactions on Medical Imaging* 17, 263–273.
- Wang, C., Hayashi, Y., Oda, M., Itoh, H., Kitasaka, T., Frangi, A.F., Mori, K., 2019a. Tubular structure segmentation using spatial fully connected network with radial distance loss for 3d medical images, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 348–356.
- Wang, K.N., Yang, X., Miao, J., Li, L., Yao, J., Zhou, P., Xue, W., Zhou, G.Q., Zhuang, X., Ni, D., 2022. Awsnet: An auto-weighted supervision attention network for myocardial scar and edema segmentation in multi-sequence cardiac magnetic resonance images. *Medical Image Analysis* 77, 102362.
- Wang, W., Wang, Y., Wu, Y., Lin, T., Li, S., Chen, B., 2019b. Quantification of full left ventricular metrics via deep regression learning with contour-guidance. *IEEE Access* 7, 47918–47928.
- Wang, W., Xia, Q., Hu, Z., Yan, Z., Li, Z., Wu, Y., Huang, N., Gao, Y., Metaxas, D., Zhang, S., 2021. Few-shot learning by a cascaded framework with shape-constrained pseudo label assessment for whole heart segmentation. *IEEE Transactions on Medical Imaging* 40, 2629–2641.
- Wang, Y., Wei, X., Liu, F., Chen, J., Zhou, Y., Shen, W., Fishman, E.K., Yuille, A.L., 2020. Deep distance transform for tubular structure segmentation in ct scans, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3833–3842.
- van der Weide, R., Bakker, C.J., Viergever, M.A., 2001. Localization of intravascular devices with paramagnetic markers in mr images. *IEEE Transactions on medical imaging* 20, 1061–1071.
- WHO, 2021. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
- Wolterink, J.M., Leiner, T., Išgum, I., 2019. Graph convolutional networks for coronary artery segmentation in cardiac ct angiography, in: *International Workshop on Graph Learning in Medical Imaging*, Springer. pp. 62–69.
- Wu, H., Wang, W., Zhong, J., Lei, B., Wen, Z., Qin, J., 2021. Scs-net: A scale and context sensitive network for retinal vessel segmentation. *Medical Image Analysis* 70, 102025.
- Xia, S., Zhu, H., Liu, X., Gong, M., Huang, X., Xu, L., Zhang, H., Guo, J., 2019. Vessel segmentation of x-ray coronary angiographic image sequence. *IEEE Transactions on Biomedical Engineering* 67, 1338–1348.
- Yun, J., Park, J., Yu, D., Yi, J., Lee, M., Park, H.J., Lee, J.G., Seo, J.B., Kim, N., 2019. Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. *Medical image analysis* 51, 13–20.
- Zhao, H., Li, H., Maurer-Stroh, S., Cheng, L., 2018. Synthesizing retinal and neuronal images with generative adversarial nets. *Medical image analysis* 49, 14–26.
- Zhuang, X., Xu, J., Luo, X., Chen, C., Ouyang, C., Rueckert, D., Campello, V.M., Lekadir, K., Vesal, S., RaviKumar, N., et al., 2022. Cardiac segmentation on late gadolinium enhancement mri: a benchmark study from multi-sequence cardiac mr segmentation challenge. *Medical Image Analysis* , 102528.