Computerized Medical Imaging and Graphics 80 (2020) 101688

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Learning tree-structured representation for 3D coronary artery segmentation

Bin Kong [a,*], Xin Wang [b,*], Junjie Bai [b], Yi Lu [b], Feng Gao [b], Kunlin Cao [b], Jun Xia [c], Qi Song [b], Youbing Yin [b,*]

[a] Department of Computer Science, UNC Charlotte, Charlotte, NC, USA
[b] Research and Development Department, Shenzhen Keya Medical Technology, Co., Ltd., Guangdong, China
[c] Department of Radiology, The First Affiliated Hospital of Shenzhen University, Health Science Center, Shenzhen Second People's Hospital, Guangdong, China

ARTICLE INFO

Article history:
Received 5 August 2019
Received in revised form
13 November 2019
Accepted 6 December 2019

Keywords:
Coronary computed tomography angiography
Coronary artery segmentation
Vessel segmentation
Tree-structured segmentation
Tree-structured ConvGRU

ABSTRACT

Extensive research has been devoted to the segmentation of the coronary artery. However, owing to its complex anatomical structure, it is extremely challenging to automatically segment the coronary artery from 3D coronary computed tomography angiography (CCTA). Inspired by recent ideas to use tree-structured long short-term memory (LSTM) to model the underlying tree structures for NLP tasks, we propose a novel tree-structured convolutional gated recurrent unit (ConvGRU) model to learn the anatomical structure of the coronary artery. However, unlike tree-structured LSTM proposed for semantic relatedness as well as sentiment classification in natural language processing, our tree-structured ConvGRU model considers the local spatial correlations in the input data as the convolutions are used for input-to-state as well as state-to-state transitions, thus more suitable for image analysis. To conduct voxel-wise segmentation, a tree-structured segmentation framework is presented. It consists of a fully convolutional network (FCN) for multi-scale discriminative feature extraction and the final prediction, and a tree-structured ConvGRU layer for anatomical structure modeling. The proposed framework is extensively evaluated on four large-scale 3D CCTA dataset (the largest to the best of our knowledge), and experiments show that our method is more accurate as well as efficient, compared with other coronary artery segmentation approaches.

1. Introduction

As one of the leading worldwide health problems, cardiovascular disease can easily lead to the sudden death of the patients. Alarmingly, evidence shows that a great many seemingly healthy people suffer from this disease (Benjamin et al., 2018). Among all the cardiovascular diseases, most of them can be attributed to the coronary artery problems, which are caused primarily by the stenosis (i.e., narrowing) of either the left or right artery. Therefore, it is of vital importance to diagnose coronary artery disease, assess the risk, and plan treatment for the patients at the early stage (Norris et al., 1992). Currently, advanced cardiac imaging has been adopted as great tools to help physicians and surgeons for heart disease diagnosis as well as treatment planning. However, the image reviewing process regularly takes a considerable amount of time even for the experts, considering the massive size and the complexity of these images. Over the past two decades, coronary artery segmentation has drawn greater and greater attention because it not only greatly facilitates the reviewing process but also provides quantitative function analysis (Zhang, 2010). Unfortunately, the segmentation procedure still heavily relies on semi-automatic approaches, which are still time-consuming and error-prone. This is because fully-automatic approaches cannot produce sufficiently accurate results, as the coronary arteries exhibit extremely complex structures. Therefore, it is essential to accurately as well as efficiently segment coronary arteries.

Currently, it is standard procedure to evaluate coronary artery diseases with computed tomography angiography (CTA) as it provides high-resolution 3D imaging as non-invasiveness. The focus of this work is the accurate segmentation of the coronary artery in 3D coronary computed tomography angiography (CCTA) volumes, as illustrated in Fig. 1. Multiple reasons account for the difficulty of the coronary artery segmentation. First, the boundaries between the artery and background are often highly fuzzy, as is shown in Fig. 1(a). Second, the tubular structure of the coronary artery is extremely complex: the cross-section area changes gradu-

* Corresponding authors.
E-mail address: bkong@uncc.edu (B. Kong).

https://doi.org/10.1016/j.compmedimag.2019.101688
0895-6111/

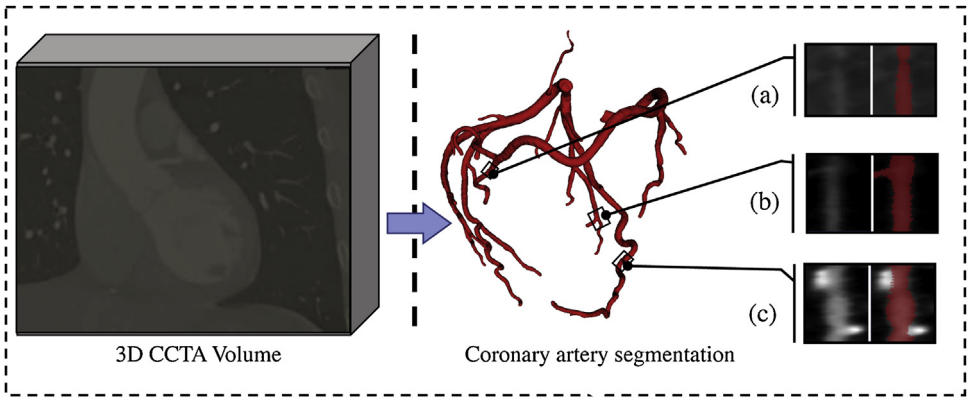

Fig. 1. From left to right: a 3D CCTA volume, the corresponding coronary artery segmentation, and three longitudinal views of the coronary artery. The coronary artery segmentation is denoted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ally along the artery and there exist a large number of bifurcations (see Fig. 1(b)). Third, the appearance and geometry of the coronary artery may vary considerably from one patient to another. Plus, the buildup of the plaque or calcification (extremely high-intensity regions in Fig. 1(c)) inside the coronary artery wall may further cause (Zhang, 2010) the variability from one patient to another. Finally, the image acquisition process may further introduce inherent image noise and artifacts (Boas and Fleischmann, 2012), making the segmentation even more challenging.

A substantial body of research has been devoted to the segmentation of the coronary arteries. Most of them (Lesage et al., 2016; Gao et al., 2019a) are based on the prior knowledge about the voxel intensity distributions, which suffer from multiple issues, e.g., holes and noisy contours. Additionally, they often fail to build a global tree structure as they only rely on local intensity information. To address this issue, geometry and topology prior have been employed to generate more anatomically reasonable segmentation result (Strandmark et al., 2013). Nevertheless, introducing these priors requires domain-specific expertise. Recently, deep learning has received growing attention over recent years in the medical imaging community (Su et al., 2017; Wu et al., 2019; Kong et al., 2018; Xia et al., 2019; Höfener et al., 2018; Mahapatra et al., 2019) and has been introduced to segment tree-like objects (Chen et al., 2018; Jin et al., 2017; Jiang et al., 2018). Compared with traditional methods for medical image analysis (Yan et al., 2015; Zhao et al., 2017; Gao et al., 2017; Xu et al., 2017), deep learning-based approaches achieve better performance and at the same time obviate hand-crafting features, as the hierarchical neural networks automatically learn the most discriminative features for the coronary artery purely from the training data. However, these methods either ignore the underlying anatomical structure in the coronary artery (Chen et al., 2018) or simply use traditional methods to post-process the segmentation results (Jin et al., 2017), which requires domain-specific knowledge and extensive tuning.

In this work, we explicitly model the anatomical structure of the coronary artery with a unified network. It consists of a fully convolutional network (FCN) model to extract discriminative features from CCTA dataset and a tree-structured ConvGRU layer to model the anatomical structure of the coronary arteries. We summarize the essential contributions as follows:

- A novel convolutional recurrent neural network (ConvRNN) layer, tree-structured convolutional gated recurrent unit (ConvGRU), is proposed to explicitly model the topological structure of the coronary artery.
- Accordingly, an end-to-end deep learning-based framework, consisting of a tree-structured ConvGRU layer and an FCN, is presented to accurately segment coronary arteries from 3D CCTA data. As far as we know, our approach is the first to incorporate tree-structured recurrent network into the coronary artery segmentation network for modeling the topological structure of the coronary arteries.
- Four large-scale CCTA datasets are employed to extensively evaluate the performance of the proposed framework. The results demonstrate that the proposed framework outperforms other baseline methods.

2. Related work

There is a large body of work on vessel (including coronary artery) segmentation. We divide them into different categories and briefly introduce them in this section. Then, previous RNN research related to this work is discussed.

2.1. Vessel segmentation

The simplest approach to segmenting vessel is to utilize the characteristics of the voxels, such as local geometry or super-intensity. For example, Pock et al. (2005) suggest using medial and adaptive thresholds to segment the liver portal vein. Featured with conceptual simplicity and computational efficiency, these approaches are often adopted in practice. However, they in general lead to issues such as leakage and holes in the segmentation and the parameters are often required to be tuned. A more refined scheme depends on iteratively delineating the vessels. In this section, we briefly categorize them into three groups. The methodologies in the first category require building a minimal cost path between the start and endpoints, which are pre-defined either manually or automatically (Wink et al., 2002; Li and Yezzi, 2007). Benefited from leveraging higher-level information, they tend to produce more anatomically reasonable results. However, accurately segmenting the coronary artery requires a well-designed cost function to control the iteration of the segmentation process. The basic strategy of the methods in the second category is to produce a pre-segmentation and then recover the missing structures and remove false-positive segmentation (Yang et al., 2012; Stefancik and Sonka, 2001). Nevertheless, they require a relatively accurate pre-segmentation to initialize the refining procedure. The methods in the third category approach vessel segmentation by tracking. They iteratively decide the next location and geometry (e.g., orientation and radius) of the vessel. This results in a significant reduction in the computational cost, as only a small portion of the image volume needs to be explored. Unfortunately, these techniques are also sensitive to noises, artifacts, and other local

perturbations, as they only rely on local information. We refer the readers to a more detailed review of the traditional vessel segmentation methods in Lesage et al. (2009).

All of the above approaches rely on hand-crafted features to segment vessels, which cannot fully capture the data's underlying features. As a consequence, generalizing to the unseen data is hard for the trained model. Deep learning-based techniques learn the most discriminative image features from training data. Deep learning-based vessel segmentation methods (Huang et al., 2018; Tetteh et al., 2019; Shen et al., 2019) have achieved better than ever segmentation results. Nevertheless, the trained deep learning models (more specifically, FCN) are still vulnerable to local disturbances owing to the extremely complicated anatomical structure and a large range of vessel sizes from the root to the terminal. To tackle this problem, Chen et al. (2018) suggests using multi-scale models. However, this only partly alleviates this issue. We are adopting a segmentation system that models the anatomical structure in an end-to-end manner. In this way, the anatomical information can be efficiently leveraged to guide the segmentation process.

2.2. Convolutional RNN models

Vessels with tubular structures and bifurcations gradually change geometry and elongation from proximal to the distal end. In this paper, we strive to use deep learning to model this special anatomical structure. Recurrent neural networks (RNNs) are great candidates for modeling long-term dependence (Kong et al., 2016, 2017). Until now, most of the past studies have used long short-term memory (LSTM) to deal with the notorious issue of vanishing or exploding gradients (Pascanu et al., 2013), which is a significant problem when training the vanilla RNN models. By incorporating several sophisticated gating functions, LSTM alleviates this issue. Nevertheless, the input-to-state and state-to-state changes are based on fully-connected layers in LSTM, which neglects local spatial correlations in input data. It is therefore not appropriate for the analysis of image sequences. The recently proposed convolutional LSTM (ConvLSTM) replaces the vector multiplication in LSTM with convolutional operations by preserving the spatial topology of the input while introducing sparsity and locality to the LSTM to reduce over-parameterization and overfitting. Unfortunately, vessels with highly branching and tubular structures are extremely complex, and ConvLSTM, which is originally designed for image sequence analysis, cannot deal with such complicated tree structures. While the tree-structured LSTM (Tai et al., 2015) is proposed for the analysis of tree-structured data (specifically, semantic relatedness as well as sentiment classification in natural language processing), the vector multiplication used in the tree-structured LSTM unit is not appropriate for image analysis. In contrast, our tree-structured ConvGRU design addresses both issues, i.e., a lack of consideration of complex tree structures and the local spatial correlation in the input data (Fig. 2).

3. Methodology

3.1. Preliminaries

We aim to utilize RNN models to extract anatomy related features. In this section, we review the preliminaries required to understand RNN models. Building on this, we further introduce our tree-structured ConvGRU model.

3.1.1. LSTM & GRU

Each unit in the LSTM model holds a memory cell, c_t, to maintain long-term memory. There exist three gates in an LSTM model, which are carefully designed non-linear functions: the input gate i_t, the forget gate f_t, and the output gate o_t. The information flow is controlled by these gates, which can be formally formulated as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1}), \tag{1}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1}), \tag{2}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1}), \tag{3}$$

$$m_t = \tanh(W_m x_t + U_m h_{t-1}), \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot m_t, \tag{5}$$

$$h_t = o_t \odot \tanh(c_t), \tag{6}$$

where sigmoid function is σ, the input vector at time step t is x_t. $\odot$ is the Hadamard product. W_i, U_i, W_f, U_f, W_o, U_o, W_m, and U_m denote the weight matrices,[1] which are shared at all time steps.

A new type of RNN model, gated recurrent unit (GRU) (Cho et al., 2014), is proposed recently, achieving comparable or even better performance in many sequential learning tasks. Unlike LSTM, it has no memory cell. Instead, two novel gates (i.e., an update gate u_t and a reset gate r_t) collaboratively decide how to update the hidden state h_t:

$$u_t = \sigma(W_z x_t + U_z h_{t-1}), \tag{7}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}), \tag{8}$$

$$\tilde{h}_t = \tanh(W x_t + r_t \odot U h_{t-1}), \tag{9}$$

$$h_t = (1 - u_t) \odot \tilde{h}_t + u_t \odot h_{t-1}, \tag{10}$$

where W_z, U_z, W_r, U_r, U, and W are the parameters to be learned.

3.1.2. ConvLSTM

The input-to-state as well as the state-to-state transitions are conducted by vector multiplications in the standard LSTM. It ignores the local spatial correlations in the input by vectorizing the input feature map. Therefore, it is not suitable for image sequence analysis. To address this issue, the vector multiplications are replaced by convolutions in ConvLSTM (Xingjian et al., 2015), to maintain the local correlations in the image sequence data. It defines a new mechanism to update the input-to-state as well as state-to-state transition:

$$i_t = \sigma\left(W_i * \mathcal{X}_t + U_i * \mathcal{H}_{t-1}\right), \tag{11}$$

$$f_t = \sigma\left(W_f * \mathcal{X}_t + U_f * \mathcal{H}_{t-1}\right), \tag{12}$$

$$o_t = \sigma\left(W_o * \mathcal{X}_t + U_o * \mathcal{H}_{t-1}\right), \tag{13}$$

$$\mathcal{M}_t = \tanh\left(W_m * \mathcal{X}_t + U_m * \mathcal{H}_{t-1}\right), \tag{14}$$

$$\mathcal{C}_t = f_t \odot \mathcal{C}_{t-1} + i_t \odot \mathcal{M}_t, \tag{15}$$

$$\mathcal{H}_t = o_t \odot \tanh(\mathcal{C}_t), \tag{16}$$

where $*$ indicates convolution, $\mathcal{X}_t$ is the current input image at time step t. The memory cell and hidden state are denoted by $\mathcal{C}_t$ and $\mathcal{H}_t$, respectively.

3.2. Tree-structured ConvGRU

Sequential ConvRNNs (Shi et al., 2017) cannot handle tree-structured data. For this reason, we propose a novel tree-structured ConvRNN network for extracting tree-structured anatomical information, in which the parent node selectively aggregates features from all its child nodes. Desirably, this tree-structured ConvRNN model is capable of automatically learning to emphasize important information in the data. For instance, it is desirable to emphasize the

[1] The bias terms in Eqs. (1–6) and other equations in this paper are ignored for simplicity.

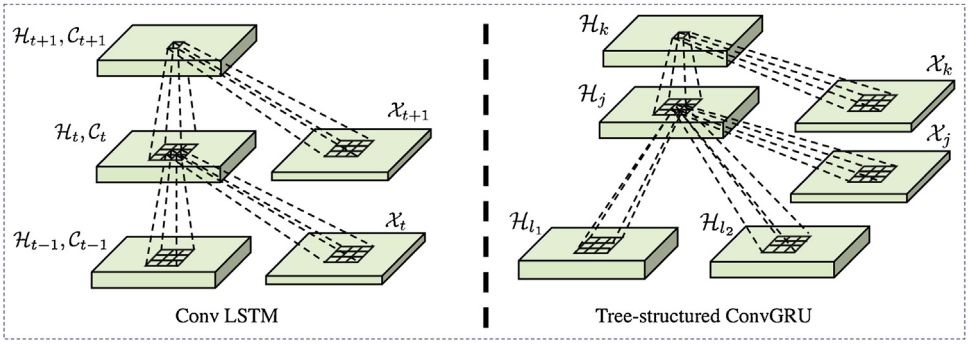

Fig. 2. From left to right: sequential ConvLSTM (Xingjian et al., 2015) and the proposed tree-structured ConvGRU. In ConvLSTM, the information, including the input $\mathcal{X}_t$, previous hidden state $\mathcal{H}_{t-1}$, and previous memory $\mathcal{C}_{t-1}$, is passed sequentially (from $t-1$ to t and then to $t+1$). As with tree-structured ConvGRU, there is no memory cell. The information is passed from all the children nodes to the parent node. For instance, node j in this figure incorporates the information (hidden state $\mathcal{H}_{l_1}$ and $\mathcal{H}_{l_2}$ from both its children l_1 and l_2 and the current input $\mathcal{X}_j$) to produce the current hidden state $\mathcal{H}_j$. Node k incorporates the information (hidden state $\mathcal{H}_j$ from its child j and its input $\mathcal{X}_k$) to produce the current hidden state $\mathcal{H}_k$. Note that although we only show one or two child nodes for the tree-structured ConvGRU model, it is capable of handling more than two child nodes.

geometry and direction of the main artery when there exists a much thinner artery merging with the main branch artery (e.g., Fig. 1(b)). In this work, we mainly focus on the extension of GRU, considering its lower computational requirement (Chung et al., 2014) than LSTM. Also, the experimental results demonstrate its superior performance than the LSTM extension on our datasets. Unlike LSTM, there is no memory cell or forget gate in GRU. Rather, for each node j in the tree, the memory cell is integrated into the hidden state $\mathcal{H}_j$ and the reset gate r_j controls the updating of the previous memory. As one unit may have multiple child nodes, we use a distinct rest gate r_{jk} for each child node to remove unimportant past information from each individual child node's memory. The whole procedure is detailed as follows:

$$\mathcal{H}'_j = \sum_{k \in \mathcal{N}_j} \mathcal{H}_k, \tag{17}$$

$$u_j = \sigma(W_z * \mathcal{X}_j + U_z * \mathcal{H}'_j), \tag{18}$$

$$r_{jk} = \sigma(W_r * \mathcal{X}_j + U_r * \mathcal{H}_k), \tag{19}$$

$$\tilde{\mathcal{H}}_j = \tanh\left(\sum_{k \in \mathcal{N}_j} r_{jk} \odot U * \mathcal{H}_k + W * \mathcal{X}_j\right), \tag{20}$$

$$\mathcal{H}_j = (1 - u_j) \odot \tilde{\mathcal{H}}_j + u_j \odot \mathcal{H}'_j, \tag{21}$$

where W_z, U_z, W_r, U_r, W, and U are the learnable parameters.

3.3. Artery centerline extraction

First, we extract the coronary artery centerline from the CCTA data, which captures the anatomical structure of the coronary artery. We use our earlier published approach (Guo et al., 2019b) for centerline extraction. It is a deep learning-based method, which is able to produce accurate (the error is within a single voxel) centerlines. The brief pipeline is summarized here. We refer the readers to Guo et al. (2019b) for more details.

- We pre-segmented coronary arteries with 3D U-Net (Çiçek et al., 2016). The anatomical structure is captured by the pre-segmentation. Nevertheless, there exists a lot of erroneous predictions (see Fig. 5 for more details). As the proposed tree-structured segmentation framework is comparatively resistant to imperfect segmentation, precise pre-segmentation is not needed in this work.
- We use our previously published deep learning based approach (Guo et al., 2019b) for centerline extraction. More specifically, the endpoints and distance map of the centerline are simultaneously predicted by a trained multi-task FCN network.
- The ultimate artery centerline is generated by minimal path algorithm. The generated centerline can be represented by a tree structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes (representing the centerline points) and adjacency matrix (representing connections among the centerline points) are denoted by $\mathcal{V}$ and $\mathcal{E}$, respectively.

3.4. Tree-structured segmentation network architecture

In this work, the coronary artery segmentation is formulated as a tree-structure segmentation problem, in which the training set is a collection of coronary artery trees and the predictions are also organized as a tree structure. The input tree in this work is produced as follows. For each node j in the artery tree $\mathcal{G}$, a cross-sectional view is cropped from the CCTA volume in the centerline's perpendicular direction. We further normalize this small patch with the aorta intensity and calcification threshold respectively to highlight both of these important regions. Finally, the normalized patches are concatenated with the original patch. The result is a tree-channel image $\mathbf{x}_j$ associated with node j. Formally, the goal is to learn a non-linear function, $(\mathcal{H}_1, \ldots, \mathcal{H}_J) = \sigma_W(\mathbf{x}_1, \ldots, \mathbf{x}_J)$, to map the tree-structured input to the tree-structured output, where J and W represent the number of nodes in the tree and the parameters to be learned.

Fig. 3 presents an overview of the proposed tree-structured segmentation framework. In our network, we model the structured information in a unified neural network, which can be trained end-to-end. It has three modules: an encoder, a tree-structured ConvGRU, and a decoder. The encoder ϕ extracts discriminative features from the input data, yielding a multi-scale representation $\mathcal{X}_j$ for each node j. The tree-structured ConvGRU ψ module models the anatomical structure of the coronary artery, generating a feature map $\mathcal{H}_j$, encoding the newly-extracted anatomically related features. Based on the feature map generated by the encoder and tree-structured ConvGRU, the decoder φ generates the final prediction $\mathcal{P}_j$.

3.4.1. Discriminative feature learning & tree-structured output generation

Fig. 4 illustrates the backbone network for feature extraction and final prediction. It's based on the U-Net (Ronneberger et al., 2015) architecture. The encoder ϕ and decoder φ divide the whole segmentation procedure into three separate stages: discriminative feature learning, anatomical structure modeling, and tree-structured output generation. During the discriminative feature learning stage, the image $\mathbf{x}_j$ associated with each node j is fed

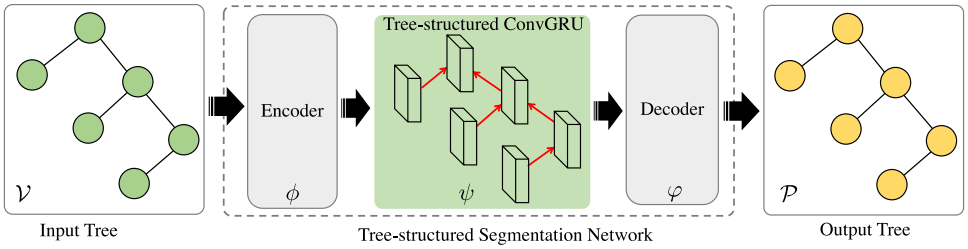

Fig. 3. An overview of the proposed tree-structured segmentation network. The input of the system is a input tree $\mathcal{V}$, i.e., images organized as a tree structure. The output $\mathcal{P}$ is also organized as a tree structure. The tree-structured segmentation network consists of two components: an FCN backbone with an encoder ϕ for discriminative feature learning and a decoder φ for prediction, and a tree-structured ConvGRU layer ψ for anatomical structure modeling. The FCN backbone and tree-structured ConvGRU layer are shared by all tree nodes. The detailed information is illustrated in Fig. 4.

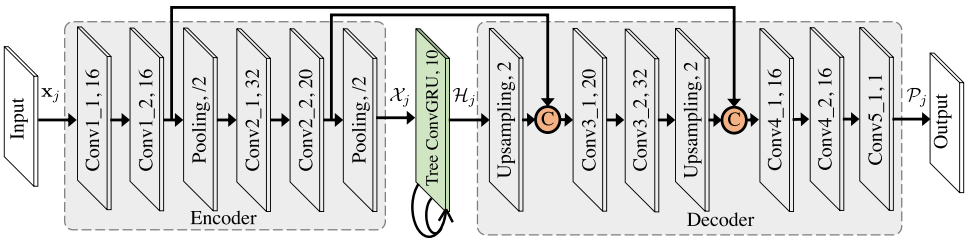

Fig. 4. Details of the proposed tree-structured segmentation network. Both the encoder and decoder consist of multiple convolutional layers (each is followed by a ReLU layer, which is ignored for simplicity). For the input image $\mathbf{x}_j$ associated with node j, it is passed into several convolutional layers and progressively downsampled by the pooling layers in the encoder, generating the feature map $\mathcal{X}_j$. The tree-structured ConvGRU layer takes input $\mathcal{X}_j$ and produces the hidden state $\mathcal{H}_j$. In the decoder, $\mathcal{H}_j$ from the tree-structured ConvGRU layer is progressively upsampled to the original dimension and at the same time incorporate the information passed from the encoder, yielding the final prediction $\mathcal{P}_j$.

into the encoder, which includes several 3×3 convolutional layers (each is followed by a ReLU layer). Two 2×2 layers are also used to downsample the feature map. The encoder is able to extract discriminative features from the input $\mathcal{X}_j = \phi(\mathbf{x}_j)$. After the anatomical structure modeling stage, a hidden state $\mathcal{H}_j$ is generated by the tree-structured ConvGRU layer (will be detailed in Section 3.4.2), the decoder progressively rescale the feature maps to the original dimension using deconvolution and at the same time incorporate the information passed from the encoder, yielding the final prediction $\mathcal{P}_j = \varphi(\mathcal{X}_j, \mathcal{H}_j)$ (see Eqs. (17–21) for more details). The details of the encoder and decoder are shown in Fig. 4.

3.4.2. Anatomical structure modeling

The introduction of the tree-structured ConvGRU ψ is motivated by the fact that there exists an inherent anatomical structure in the coronary artery tree. For instance, tubular artery gradually changes from the proximal to the distal end, with the elongation and radius changes smoothly from node to node. Using tree-structured ConvGRU in our system brings two benefits. First, by feeding the features extracted by the encoder to the tree-structured ConvGRU, the context information is propagated among the tree nodes. As a result, the final encoder makes prediction not solely by the features of one node but considering the topological changes along the coronary artery tree. Second, as mentioned in Section 3.2, there may exist multiple branches at each tree node. In these special locations, our system is capable of modeling these transitions. The tree-structured ConvGRU layer takes input $\mathcal{X}_j$ and produces the hidden state $\mathcal{H}_j = \psi(\mathcal{X}_j)$.

3.4.3. Loss function

The forward pass of the proposed tree-structured segmentation network for one input tree is illustrated in Algorithm 1. The proposed tree-structured segmentation forms a differentiable system, which can be trained end-to-end. Dice loss is applied node-wise and the final loss is the average dice loss (Milletari et al., 2016), as defined as follows:

$$\mathcal{L}(\mathcal{P}, \mathcal{G}) = \frac{1}{J} \sum_{j=1}^{J} \frac{2|\mathcal{P}_j \cap \mathcal{G}_j|}{|\mathcal{P}_j| + |\mathcal{G}_j|}, \tag{22}$$

where the output tree and all the ground truth segmentation are represented by $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_J)$ and $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_J)$, respectively. $\mathcal{P}_j$ and $\mathcal{G}_j$ are the prediction and ground truth for node j, respectively.

Algorithm 1. The forward pass of the proposed tree-structured segmentation network for one input tree

Input: $\mathcal{G}$ = input tree $(\mathcal{V}, \mathcal{E})$
Input: ϕ = encoder
Input: ψ = tree-structured ConvGRU layer
Input: φ = decoder
  $\mathcal{P} \longleftarrow \emptyset$
  for j in [1, *num_nodes*] **do**
    sample the input image $\mathbf{x}_j$ associated with node $j \in \mathcal{V}$
    extract features from $\mathbf{x}_j$ with $\mathcal{X}_j \longleftarrow \phi(\mathbf{x}_j)$
    generate the hidden state using $\mathcal{H}_j \longleftarrow \psi(\mathcal{X}_j)$
    produce the final prediction using $\mathcal{P}_j \longleftarrow \varphi(\mathcal{X}_j, \mathcal{H}_j)$
    $\mathcal{P}[j] \longleftarrow \mathcal{P}_j$
  end for
  return $\mathcal{P}$

4. Experiments

4.1. Dataset, evaluation metrics, and implementation details

We collected four large datasets (916 CT scans in total) from four hospitals. These collaborating hospitals are selected from different areas to represent the diversity of healthcare settings. 80%, 5%, and 15% scans were used for training, validation, and testing, respectively. The data splitting was carried out on the patient level. The ground truth was obtained by a semi-automatic approach. First, a vesselness based approach combined with the dynamic programming algorithm was used to obtain an initial entire coronary artery.

Table 1
Detatiled information of our datasets (CTA1, CTA2, CTA3, and CTA4). Apart from providing the number of training scans in each dataset, the average number of tree nodes and branches are also given.

Dataset	Number of scans	Number of nodes	Number of branches
CTA1	258	727	12.6
CTA2	273	806	11.1
CTA3	223	802	13.2
CTA4	162	694	12.9
Total	916	774	12.4

Table 2
Main comparison results. The proposed tree-structured segmentation network (TreeConvGRU) is compared with the recently proposed 3D densely-connected volumetric convnets (DenseVox) (Yu et al., 2017), sequential version of our tree-structured segmentation network (ConvGRU). All these methods are evaluated by the average dice loss.

Methods	DenseVox (Yu et al., 2017)	ConvGRU	TreeConvGRU
CTA1	0.8370	0.8399	0.8494
CTA2	0.8405	0.8427	0.8503
CTA3	0.8433	0.8436	0.8545
CTA4	0.8182	0.8237	0.8283
Total	0.8518	0.8614	0.8683

Two experienced image analysts then annotated the masks until their results passed the consistency check (dice score > 0.95 and average surface distance < 0.5 voxel). A more experienced analyst then checked both results and selected the better one as the ground truth. To the best of our knowledge, this dataset is largest available for evaluating coronary artery segmentation algorithms. These datasets are dubbed CTA1, CTA2, CTA3, and CTA4 in this work and include 258, 273, 223, 162 scans, respectively. The details of these datasets are shown in Table 1. To measure the performance of the segmentation methods, we use the average dice score of all the tree nodes. All the methods were trained and evaluated on a workstation equipped with a Tesla P40 GPU. To train the neural networks, the Adam optimizer (Kingma and Ba, 2015) was used. The initial learning rate, weight decay, and momentum are 0.001, 0.0005, and 0.9, respectively. Additionally, early-stopping was used to combat over-fitting.

4.2. Main results

First, the proposed approach is compared with a recently-introduced 3D object segmentation framework, 3D volumetric convnet (DenseVox) (Yu et al., 2017). For DenseVox, a $41 \times 41 \times 41$ subvolume around each tree node is fed into the DenseVox network. Unlike our approach, DenseVox does not consider long-range inter-node dependencies in the artery tree or the tree structure underlying in the artery tree. According to Table 2, the proposed tree-structured ConvGRU based segmentation framework (TreeConvGRU) consistently surpasses DenseVox (1.24%, 0.98%, 1.12% and 1.01%, and 1.01% on CTA1, CTA2, CTA3, and CTA4 respectively), indicating the essential role of modeling the long-range inter-node dependency and tree-structure in coronary artery segmentation.

Next, TreeConvGRU is compared with its sequential version, sequential ConvGRU (ConvGRU). Compared with TreeConvGRU, the tree structures are ignored by ConvGRU and the segmentation results are generated independently for each path in the tree. The results once again suggest the superiority of the proposed method over sequential models in modeling the inter-node dependency in tree structures: the average dice score of TreeConvGRU is better than ConvGRU by 0.95%, 0.76%, 1.09%, and 0.46% on CTA1, CTA2, CTA3, and CTA4, respectively. Additionally, to test the scalability of the proposed method, we evaluate the performance of the above methods on the aggregated dataset of CTA1, CTA2, CTA3, and CTA4, which are named Total. As is shown in Table 2, TreeConvGRU still consistently overperform ConvGRU and DenseVox (0.69% and 1.65%, respectively). We also provide some qualitative coronary artery segmentation results of our approach in Fig. 5. We compare the qualitative results of our network with a 3D U-Net based network citepshen2019coronary, i.e., the pre-segmentation of the coronary artery. As the segmentation is applied on every single voxel of the CCTA volume, the network is extremely sensitive to local perturbations. Therefore, the results suffer from a significant amount of false positives and false negatives. Even after post-processing (erosion, dilation, and connected component analysis), the false predictions on the coronary artery cannot be corrected. On the contrary, our network efficiently leverage the anatomical structure of coronary artery to guide its segmentation, generating a much more accurate segmentation result.

Table 3
Comparison of the segmentation accuracy around the bifurcation nodes (within 4 nodes' distance) and the average running time (seconds) on the testing set of the aggregated dataset (Total). The compared methods are: DenseVox (Yu et al., 2017), ConvGRU, TreeConvGRU.

Methods	DenseVox	ConvGRU	TreeConvGRU
Average dice	0.7806	0.8223	0.8537
Time (s)	58	26	**25**

Lastly, we also compare our method with another tree-structured extension of ConvRNN model, tree-structured ConvLSTM (TreeConvLSTM). This approach is different from our approach by substituting the GRU operations with LSTM. This change slightly decreases (0.14% in average on all datasets) the performance of our framework. This result matches the findings in Chung et al. (2014) regarding the comparison of non-convolution versions of LSTM and GRU. In this work, the information propagation is conducted from the root to the leaf nodes. It's possible to extend the proposed method to conduct the propagation in both directions with the technique in Zhang et al. (2015), i.e., from tree leaves to root as well as from root to leaves. However, the overall performance degraded by 0.04%. Here is one possible explanation for this performance degradation: the anatomical structure can be sufficiently modeled by the one-directional tree-structured ConvGRU, without needing to resort to more complex RNN models. In contrast, a more complex system may render the learning process even harder.

4.3. Comparisons on bifurcation nodes

Intuitively, it is much more challenging for the segmentation framework to generate good prediction at bifurcation nodes, compared with non-bifurcation nodes. This is because the dynamics around these nodes are much more complex. We conducted an extra experiment on Total to verify this hypothesis. In this experiment, we only evaluate the performance of the segmentation approaches on the nodes within 4 nodes' distance from bifurcation nodes. According to Table 3, our method consistently exceeds DenseVox and ConvGRU (7.31% and 3.14%, respectively). The results demonstrate the importance by introducing the true structure. DenseVox ignores the inter-node dependencies in the artery tree while ConvGRU only considers the dependencies along each vessel path. The proposed TreeConvGRU fully utilizes the tree structures, thus yielding the best performance at the bifurcation location. Additionally, we show the detailed local comparisons of our final networks with DenseVox in Fig. 6. According to these results, our tree-structured ConvGRU based methods are more robust to the local perturbations and the bifurcations.

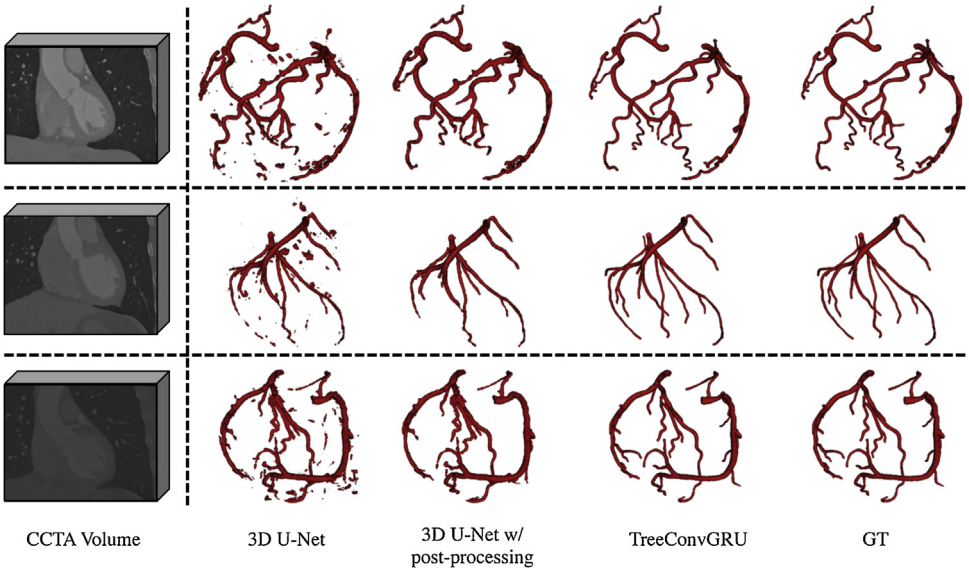

Fig. 5. Qualitative coronary artery segmentation result of 3D U-Net, 3D U-Net with post-processing, and the proposed method. From left right shows: the input 3D CCTA volumes, segmentation results of 3D U-Net based method (Shen et al., 2019), segmentation results of 3D U-Net with post-processing, segmentation results of the proposed tree-structured segmentation network, and the ground truth.

Fig. 6. Local segmentation result comparison. From left to right column, we show the comparison of some zoomed views segmentation results of DenseVox (Yu et al., 2017), TreeConvGRU, and the GT. Compared with DenseVox, our method generates our accurate results.

4.4. Comparisons of computational costs

The average running time (seconds) on the testing set of the aggregated dataset (Total) is shown in Table 3. The compared methods are: DenseVox (Yu et al., 2017), ConvGRU, TreeConvGRU. Regarding the computational cost, 58 s are required by DenseVox, as is shown in Table 3. This is the longest time among all the above methods. TreeConvGRU takes 25 s. It is almost 2.3 times more efficient than DenseVox. All in all, our methods generate more accurate coronary artery predictions while requiring less computational power.

5. Discussion

Extensive studies of coronary artery segmentation has been spurred by the arising concerns regarding cardiovascular diseases. However, owing to the complex nature of its anatomical structure, local image perturbations, and appearance or geometry variability, it is still challenging to apply fully automatic algorithms in clinical practices. To address these issues, we use the coronary artery tree to guide its segmentation. In this way, our network only needs to focus on the local artery segmentation. The reconstructed tree is a collection of nodes, with each of them highly dependent on others. Therefore, we propose tree-structured ConvGRU models to model the inter-node dependency. Accordingly, a tree-structured segmentation network is presented. Augmented with the tree-structured formulation to explicitly model the tree structure, our framework is able to achieve the state-of-the-art performance on four CCTA datasets, demonstrating the effectiveness of the proposed method in the segmentation of complex tree-structured objects.

There exist several limitations to this work. First, the two stages in our pipeline are isolated from each other and trained separately. It would be more desirable to design a unified network for coronary artery segmentation. In this way, the tree reconstruction network and the tree-structured segmentation network may be able to benefit one another by interacting and working collaboratively with each other during the end-to-end training process. Second, we rely on our previously published tree reconstruction approach (Guo et al., 2019b) in this work. It is able to generate relatively accurate artery trees for our datasets. However, it can be difficult to obtain accurate trees on some occasions, e.g., low-dose CT. There-

fore, quantitatively assessing its impact of tree reconstruction on the overall segmentation performance is also indispensable. Our future work will be devoted to addressing these issues.

In the current setting, we implicitly assume that the training and testing examples are drawn from the same distribution. However, in clinical practices, this may not be the case. For instance, the CCTA datasets from two different hospitals may be acquired by different devices. As a result, the appearance of these two datasets may vary significantly and the models trained on the first dataset may not generalize to the second one, resulting in their poor performance. The distribution shift between the training and testing data is formally referred to as the domain shift. It would be greatly desirable for one model trained on one dataset from one hospital to be generalizable to the datasets from other hospitals. In the future, we plan to explore using the domain adaptation techniques (Dou et al., 2018; Perone et al., 2019; Liu et al., 2019) to address this problem. Additionally, our formulation requires a pre-segmentation of the coronary artery, which is inefficient as it involves the dense evaluation of the whole 3D CCTA volume. Recently, with deep reinforcement learning (DRL), Zhang et al. (2018) significantly advanced the performance of vessel tracing. Without needing to densely evaluate the volume, DRL based vessel tracing approach is extremely efficient. We expect that employing DRL to trace the artery tree, instead of relying on pre-segmentation, is capable of further boosting the efficiency of our framework.

Finally, attention mechanism (Xu et al., 2015; Lu et al., 2016; Zhang et al., 2019) has been extensively used for avoiding the distractions of non-salient background regions. In this way, the performance can be potentially improved. In the future, we will extensively investigate its influence on our coronary artery segmentation model.

6. Conclusion

We present a unified framework with an FCN backbone and a tree-structured ConvGRU layer for coronary artery segmentation. Specifically, the FCN backbone extracts discriminative features from the CCTA images and produce the final tree-structured output. The tree-structured ConvGRU is responsible for considering the inter-node dependencies in the extracted features. Benefited from the careful design for the simultaneously discriminative spatial feature and inter-node dependency learning, the proposed framework achieved superior performance, compared with other methods. In the future, we plan to apply our approach to similar tree-structured object segmentation problems such as airway segmentation (Charbonnier et al., 2017).

Authors' contribution

Bin Kong and Xin Wang conceived of the presented idea.

Bin Kong, Xin Wang, Youbing Yin, and Junjie Bai developed the theoretical formalism.

Bin Kong, Xin Wang, Youbing Yin, Yi Lu, and Junjie Bai conceived and planned the experiments.

Bin Kong and Xin Wang wrote the manuscript with input from all authors.

Kunlin Cao, Youbing Yin and Qi Song designed and directed the project.

All authors discussed the results and commented on the manuscript.

Conflict of interest

None declared.

Acknowledgement

The work received supports from Shenzhen Municipal Government under the grant KQTD2016112809330877 and GJHZ20180926165402083.

References

Benjamin, E.J., Virani, S.S., Callaway, C.W., Chamberlain, A.M., Chang, A.R., Cheng, S., Chiuve, S.E., Cushman, M., Delling, F.N., Deo, R., et al., 2018. Heart disease and stroke statistics-2018 update: a report from the American Heart Association. Circulation 137 (12), 67–492.

Boas, F.E., Fleischmann, D., 2012. Ct artifacts: causes and reduction techniques. Imaging Med. 4 (2), 229–240.

Charbonnier, J.-P., et al., 2017. Improving airway segmentation in computed tomography using leak detection with convolutional networks. Med. Image Anal. 36, 52–60.

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. Proc. SPIE-Bio. Appl. Mol. Struct. Funct. Imag., Vol. 10578, 105782R.

Çiçek, Ö., et al., 2016. 3d u-net: learning dense volumetric segmentation from sparse annotation. Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent., 424–432.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. Proc. Annu. Meeting Assoc. Comput. Linguistics and Int. Joint Conf. Natural Lang. Proc., 1724–1734.

Chung, J., et al., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. Adv. Neural Inf. Process. Syst. Workshop.

Dou, Q., Ouyang, C., Chen, C., Chen, H., Heng, P.-A., 2018. Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. Proc. Int. Joint Conf. Artif. Intell., 691–697.

Gao, Z., Xiong, H., Liu, X., Zhang, H., Ghista, D., Wu, W., Li, S., 2017. Robust estimation of carotid artery wall motion using the elasticity-based state-space approach. Med. Image Anal. 37, 1–21.

Gao, Z., Liu, X., Qi, S., Wu, W., Hau, W.K., Zhang, H., 2017. Automatic segmentation of coronary tree in CT angiography images. Int. J. Adapt. Control Signal Process.

Guo, Z., Bai, J., Lu, Y., Wang, X., Cao, K., Song, Q., Sonka, M., Yin, Y., 2019. Deepcenterline: a multi-task fully convolutional network for centerline extraction. Inf. Process. Med. Imaging, 441–453.

Höfener, H., Homeyer, A., Weiss, N., Molin, J., Lundström, C.F., Hahn, H.K., 2018. Deep learning nuclei detection: a simple approach can deliver state-of-the-art results. Comput. Med. Imaging Graph. 70, 43–52.

Huang, W., Huang, L., Lin, Z., Huang, S., Chi, Y., Zhou, J., Zhang, J., Tan, R.-S., Zhong, L., 2018. Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images. Conf. Proc. IEEE Eng. Med. Biol. Soc., 608–611.

Jiang, Z., Zhang, H., Wang, Y., Ko, S.-B., 2018. Retinal blood vessel segmentation using fully convolutional network with transfer learning. Comput. Med. Imaging Graph. 68, 1–15.

Jin, D., Xu, Z., Harrison, A.P., George, K., Mollura, D.J., 2017. 3d convolutional neural networks with graph refinement for airway segmentation using incomplete data labels. Mach. Learn. Med. Imag. Workshop, 141–149.

Kingma, D.P., Ba, J., 2015. A method for stochastic optimization. Int. Conf. Learn. Represent.

Kong, B., Zhan, Y., Shin, M., Denny, T., Zhang, S., 2016. Recognizing end-diastole and end-systole frames via deep temporal regression network. Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent., 264–272.

Kong, B., Wang, X., Li, Z., Song, Q., Zhang, S., 2017. Cancer metastasis detection via spatially structured deep network. Proc. Int. Conf. Inf. Process. Med. Imaging, 236–248.

Kong, B., Sun, S., Wang, X., Song, Q., Zhang, S., 2018. Invasive cancer detection utilizing compressed convolutional neural network and transfer learning. In: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. Springer, Cham, pp. 156–164.

Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G., 2009. A review of 3d vessel lumen segmentation techniques: models, features and extraction schemes. Med. Image Anal. 13 (6), 819–845.

Lesage, D., Angelini, E.D., Funka-Lea, G., Bloch, I., 2016. Adaptive particle filtering for coronary artery segmentation from 3d CT angiograms. Comput. Vis. Image Underst. 151 (C), 29–46.

Li, H., Yezzi, A., 2007. Vessels as 4-d curves: global minimal 4-d paths to extract 3-d tubular surfaces and centerlines. IEEE Trans. Med. Imaging 26 (9), 1213–1223.

Liu, P., Kong, B., Li, Z., Zhang, S., Fang, R., 2019. CFEA: collaborative feature ensembling adaptation for domain adaptation in unsupervised optic disc and cup segmentation. Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.

Lu, J., Yang, J., Batra, D., Parikh, D., 2016. Hierarchical question-image co-attention for visual question answering. Adv. Neural Inf. Process. Syst., 289–297.

Mahapatra, D., Bozorgtabar, B., Garnavi, R., 2019. Image super-resolution using progressive generative adversarial networks for medical image analysis. Comput. Med. Imaging Graph. 71, 30–39.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. International Conference on 3D Vision, 565–571.
Norris, R.M., White, H.D., Cross, D.B., Wild, C.J., Whitlock, R.M., 1992. Prognosis after recovery from myocardial infarction: the relative importance of cardiac dilatation and coronary stenoses. Eur. Heart J. 13 (12), 1611–1618.
Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks. Proc. Int. Conf. Mach. Learn., 1310–1318.
Perone, C.S., Ballester, P., Barros, R.C., Cohen-Adad, J., 2019. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. NeuroImage.
Pock, T., Beichel, R., Bischof, H., 2005. Multiscale medialness for robust segmentation of 3d tubular structures. Image Anal., Vol. 2005, 481–490.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent., 234–241.
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.
Shi, X., et al., 2017. Deep learning for precipitation nowcasting: a benchmark and a new model. Adv. Neural Inf. Process. Syst., 5617–5627.
Stefancik, R.M., Sonka, M., 2001. Highly automated segmentation of arterial and venous trees from three-dimensional magnetic resonance angiography (MRA). Int. J. Cardiovasc. Imaging 17 (1), 37–47.
Strandmark, P., Ulén, J., Kahl, F., Grady, L., 2013. Shortest paths with curvature and torsion. Proc. IEEE Int. Conf. Comput. Vis., 2024–2031.
Su, S., Hu, Z., Lin, Q., Hau, W.K., Gao, Z., Zhang, H., 2017. An artificial neural network method for lumen and media-adventitia border detection in IVUS. Comput. Med. Imaging Graph. 57, 29–39.
Tai, K.S., Socher, R., Manning, C.D., 2015. Improved semantic representations from tree-structured long short-term memory networks. Proc. Annu. Meeting Assoc. Comput. Linguistics and Int. Joint Conf. Natural Lang. Proc., Vol. 1, 1556–1566.
Tetteh, G., Efremov, V., Forkert, N.D., Schneider, M., Kirschke, J., Weber, B., Zimmer, C., Piraud, M., Menze, B.H., 2019. Deepvesselnet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-d Angiographic Volumes. arXiv:1803.09340.
Wink, O., Frangi, A.F., Verdonck, B., Viergever, M.A., Niessen, W.J., 2002. 3d MRA coronary axis determination using a minimum cost path approach. Magn. Reson. Med. 47 (6), 1169–1175.
Wu, E., Kong, B., Wang, X., Bai, J., Lu, Y., Gao, F., Zhang, S., Cao, K., Song, Q., Lyu, S., et al., 2019. Residual attention based network for hand bone age assessment. Proc. IEEE Int. Symp. Biomed. Imaging, 1158–1161.
Xia, C., Li, X., Wang, X., Kong, B., Chen, Y., Yin, Y., Cao, K., Song, Q., Lyu, S., Wu, X., 2019. A multi-modal network for cardiomyopathy death risk prediction with CMR images and clinical information. Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.
Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst., 802–810.
Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y., 2015. Show, attend and tell: neural image caption generation with visual attention. International Conference on Machine Learning, 2048–2057.
Xu, L., Huang, X., Ma, J., Huang, J., Fan, Y., Li, H., Qiu, J., Zhang, H., Huang, W., 2017. Value of three-dimensional strain parameters for predicting left ventricular remodeling after ST-elevation myocardial infarction. Int. J. Cardiovasc. Imaging 33 (5), 663–673.
Yan, Z., Zhang, S., Tan, C., Qin, H., Belaroussi, B., Yu, H.J., Miller, C., Metaxas, D.N., 2015. Atlas-based liver segmentation and hepatic fat-fraction assessment for clinical trials. Comput. Med. Imaging Graph. 41, 80–92.
Yang, G., Kitslaar, P., Frenay, M., Broersen, A., Boogers, M.J., Bax, J.J., Reiber, J.H., Dijkstra, J., 2012. Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography. Int. J. Cardiovasc. Imaging 28 (4), 921–933.
Yu, L., et al., 2017. Automatic 3d cardiovascular MR segmentation with densely-connected volumetric convnets. Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent., 287–295.
Zhang, S., et al., 2015. Bidirectional long short-term memory networks for relation classification. Proc. Pacific Asia Conf. Lang. Inf. Comput., 73–78.
Zhang, P., Wang, F., Zheng, Y., 2018. Deep reinforcement learning for vessel centerline tracing in multi-modality 3d volumes. Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent., 755–763.
Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-Attention Generative Adversarial Networks. arXiv:1805.08318.
Zhang, D.P., 2010. Coronary Artery Segmentation and Motion Modelling. Imperial College London (Ph.D. thesis).
Zhao, S., Gao, Z., Zhang, H., Xie, Y., Luo, J., Ghista, D., Wei, Z., Bi, X., Xiong, H., Xu, C., et al., 2017. Robust segmentation of intima-media borders with different morphologies and dynamics during the cardiac cycle. IEEE J. Biomed. Health Inform. 22 (5), 1571–1582.