

CAS-Net: A Novel Coronary Artery Segmentation Neural Network

Rawaa HAMDI^{1,2}, Asma KERKENI^{1,3},
Mohamed Hedi BEDOUI¹ and Asma BEN ABDALLAH^{1,3}

1-Laboratory of Technology and Medical Imaging, Faculty of Medicine,
University of Monastir, Tunisia

2-Faculty of Sciences of Monastir, University of Monastir, Tunisia

3-Higher Institute of Computer Sciences and Mathematics,
University of Monastir, Tunisia

Abstract. In conventional X-ray coronary angiography, accurate coronary artery segmentation is a crucial and challenging step in the assessment of coronary artery disease. In this paper, we propose a new architecture (CAS-Net) for coronary artery segmentation. It is based on Residual U-Net and it includes both channel and spatial attention mechanism in the center part to generate hierarchical rich features of coronary arteries. Experiments are conducted on a private dataset of 150 images. The results show that CAS-Net outperforms the state-of-the-art methods achieving the highest accuracy of 96.91% and Dice of 82.70%.

1 Introduction

Coronary Artery Disease is one of the principal reasons of mortality worldwide [1]. Currently, X-Ray Coronary Angiography (XRCA) is the gold standard for the diagnosis and treatment of this disease since it reveals valuable morphological features of the coronary arteries such as length, diameter and tortuosity [2].

The analysis of these features by visual inspection of XRCA images is a laborious and time-consuming task and may lead to potential misinterpretations, due to noise, uneven illumination, complex vessel structure, stenosis and many vessel overlaps. Thus, automatic segmentation of vessel structures is a crucial tool to assist clinicians for a more accurate and fast analysis. So far, a large number of methods for blood vessel segmentation have been proposed such as filtering based, rules based, and machine learning based methods. However most of these methods are usually preceded by a tedious handcrafted feature engineering step. Furthermore, they heavily rely on empirical parameter adjustment.

Recently, deep learning methods have been proposed to handle the above issues by learning features directly from images in a fully supervised manner and have outperformed conventional methods with better performance, for different computer vision tasks. Particularly, in medical image, the U-Net architecture [3] is the baseline for most of the state-of-the-art semantic segmentation methods. This model allows for an end-to-end processing thanks to its encoder-decoder structure. An extended recent review of U-Net and its variants can be found in [4]. Despite the multitude of proposed architectures, only few works dealt

with vessel segmentation from XRCA images [5, 6, 7, 8]. Based on U-Net, Yang et al. [6] introduced a loss function based on the Generalize Dice Coefficient which penalizes false prediction in order to take into account the class imbalance problem during the main coronary vessels segmentation. Moreover, Jun et al. [7] extended the U-Net architecture by adding concatenate layers in such a way that features from all levels of the encoder are delivered for all of the blocks in the decoder path. In another work, in order to deal with the different vessel thickness, Li et al. [8] proposed to combine U-Net with a multi-scale feature fusion method named Feature Pyramid Networks. More recently, Mou et al. [9], proposed the Cs2-Net, a neural network for curvilinear structure segmentation. The model is based on two types of attention mechanism, spatial attention and channel attention to learn rich hierarchical representations of the structure of interest.

All the aforementioned segmentation methods are designed either for the main arteries in XRCA or to other tubular structures. In this paper and inspired by the Cs2-Net [9], we propose the CAS-Net architecture for XRCA segmentation. Our model takes advantages of deep residual networks, the shape of U-Net and the dilated convolution blocks. That leads to even more elegant architecture with better performance. Besides, since no standard public database is available for this task, we propose our private coronary artery database.

2 Proposed Method

The CAS-Net backbone is based on the encoder decoder framework with residual units and an attention block in the middle part (Fig.1.a). As we mentioned above, the attention module is inspired by [9]. We further add to the spatial attention block dilated convolution layers. The main advantage of using the dilated convolution is to obtain a larger receptive field without losing the spatial information of feature maps.

The encoder is a replication of residual units. The architecture of a residual unit is detailed in Fig.1.a where each unit includes: convolution layer, batch normalization layer, ReLU activation function and dropout layer. An additional 1×1 convolution layer is used to transform the input into the desired shape before the additional operation. Then, a maxpooling operation is applied to decrease the spatial dimensions.

Similarly, the decoder consists of a replication of residual units. Each unit is followed by an up-sampling of feature maps and a concatenation with those of the corresponding encoding path. Finally, a 1×1 convolution and a sigmoid function are used to recover the representations to a pixel-wise categorization.

In the middle path (Fig.1.c), we find the dual self-attention bloc composed of two types of attention mechanism: spatial attention block and channel attention block. The main objective of this block is to guide the model to highlight the most salient features and avoid useless features. The spatial and channel attention modules are detailed respectively in subsections 2.1 and 2.2.

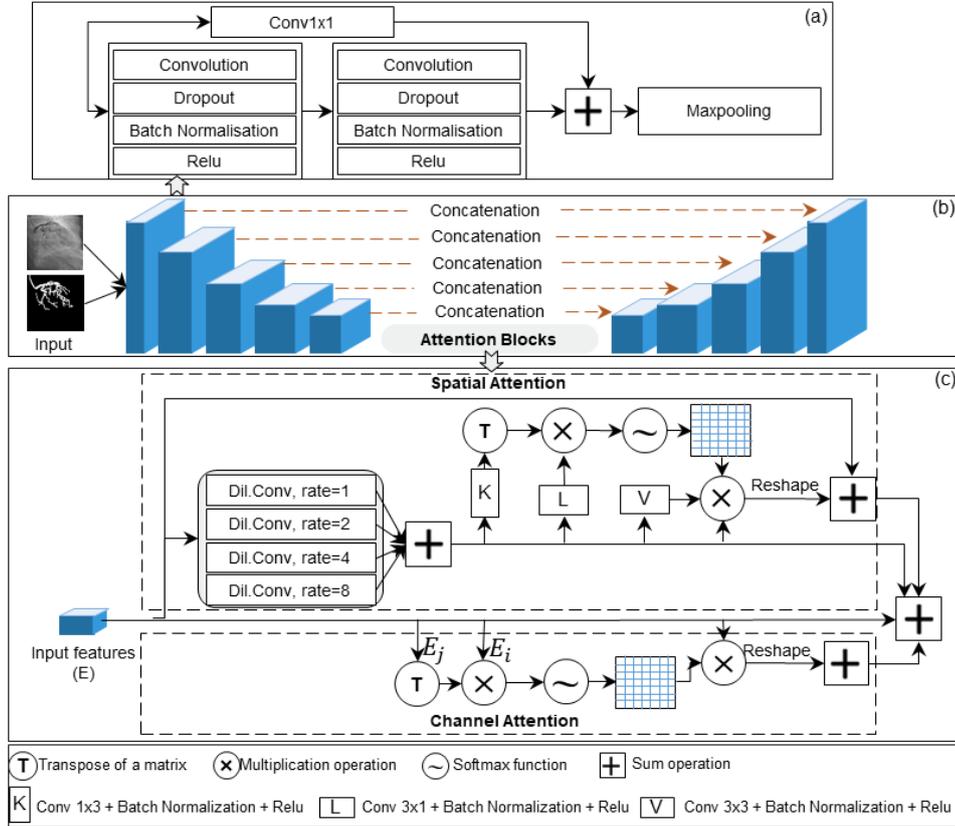


Fig. 1: Overview of the CAS-Net architecture. (a) A residual unit. (b) A high level Description of CAS-Net. (c) The Attention blocks

2.1 Spatial Attention Module

We adopt the spatial attention proposed in [9] and introduce a dilated convolution block to augment the receptive field of the feature maps in the spatial attention block and retain the detailed information. Four dilated convolution layers with skip connections are used with dilation rates 1, 2, 4 and 8, thus, the receptive fields are, respectively, 3×3 , 7×7 , 15×15 and 31×31 . The receptive field is usually calculated as follows:

$$E_{j+1} = [(2^{j+2} - 1) \times (2^{j+2} - 1)]$$

Where $j + 1$ is the dilatation rate. Considering $D \in \mathbb{R}^{C \times H \times W}$ the output of our proposed consecutive dilated convolution, and $E \in \mathbb{R}^{C \times H \times W}$ the input features, we feed D and E into two types of convolution layers (3×1) and (1×3) to produce two feature maps $K \in \mathbb{R}^{C \times H \times W}$ and $L \in \mathbb{R}^{C \times H \times W}$ respectively, where

C represents the number of features maps and $H \times W$ is the spatial size of the features maps. A multiplication between L and the transpose of k is done. Then, apply a softmax function on the result to obtain the intra-class association $I \in \mathbb{R}^{C \times H \times W}$, which represents the spatial attention map:

$$I_{(i,j)} = \frac{\exp(K_j^T \cdot L_i)}{\sum_{i'=1}^{H \times W} \exp(K_j^T \cdot L_{i'})}$$

$I_{(i,j)}$ is a correlation matrix between two points i and j . After applying the softmax function, the output is multiplied by a scale parameter and an element-wise sum operation with the input features E is performed yielding to the final output of this module.

2.2 Channel Attention Module

To efficiently derive the channel attention, we crush the spatial dimension of the input feature map by applying a softmax layer on the channel-wise similarity map between the input feature $E \in \mathbb{R}^{C \times H \times W}$ and its transpose $E^T \in \mathbb{R}^{C \times H \times W}$ as:

$$C_{(i,j)} = \frac{\exp(E_j^T \cdot E_i)}{\sum_{i'=1}^C \exp(E_j^T \cdot E_{i'})}$$

Where $C_{(i,j)}$ designates the correlation matrix of the i^{th} and the j^{th} channels. The idea behind is that two similar channels support each other whereas the different channels suppress each other. Then, a softmax function is applied to further discriminate vessel pixels from the background. We obtain a new feature maps named as E' . The final output of the channel attention module is defined as $E + E'$ over each pixel.

3 Experiments and Results

3.1 Experiments and evaluation

Since no public dataset is available for XRCA segmentation, we created our own private database by extending the dataset described in our previous work [10]. A total of 150 images obtained from 50 patients were selected by the clinicians. Each image has a size of 256*256 pixels and 256 gray levels with a resolution of 0.3x0.3 mm. The dataset is then augmented using random flipping, rotate and random crop. The model is implemented using pytorch framework and optimized by minimizing MSE loss and Dice coefficient loss through RMSprop algorithm. The 5-folds cross validation is adopted to divide the training and validation dataset. For each fold 80% were used for the training process and 20% for the validation process. The training is based on a mini batch size of 8 and the learning rate is set to 0.00001. Besides, early stopping was adopted to prevent overfitting.

In order to compare the proposed method with the existing methods, five performance metrics, namely: Accuracy (Acc), Sensitivity (Sen), Specificity (Sp),

Precision (Pre) and Dice metric are adopted. These metrics are the most commonly used evaluation metrics for quality assessment of vessel segmentation.

3.2 Results

To demonstrate the efficiency of CAS-Net, we test it on our dataset and compare it with three existing methods namely U-Net [3], Cs2-Net [9] and R2UNet [11]. For the qualitative evaluation, three example images from the testing process are illustrated in figure 2. The first column presents the original images, the second column presents the corresponding ground truth, the rest of the columns show the corresponding results of U-Net, R2U-Net, CS2-Net and CAS-Net, respectively. In general, the CAS-Net outputs are the nearest to the ground truth. Moreover, CAS-Net is less noise-sensitive.

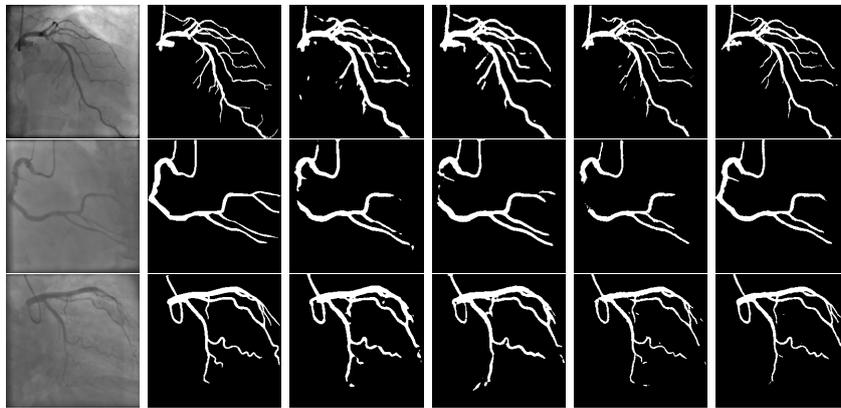


Fig. 2: Qualitative assessment of different methods performance. From left to right: original images, ground truth; and U-Net, R2-UNet, Cs2-Net, CAS-Net outputs.

Table1 presents the performance metrics values of the four methods. Our proposed architecture outperforms in all metrics values. It provides the higher precision (0.8718) yielding a higher Dice metric (2.2% better than Cs2-Net). However, U-Net provides the lowest values of different metrics: 0.9495 for accuracy, 0.7712 for sensitivity, 0.7182 for precision and 0.7437 for Dice. In the second place comes R2U-net which has achieved higher values but Cs2-Net comes in the first place with values close to those of CAS-Net.

4 Conclusion

In this paper, we proposed a new coronary artery segmentation architecture named CAS-Net. It adopts the Residual U-Net as backbone and includes both channel and spatial attention mechanism in the center part. We also built a real coronary artery dataset. The experimental results show better performance

Table 1: Experimental results of the proposed approach and comparison against U-Net, R2U-Net and Cs2-Net.

Methods	Accuracy	Sensitivity	Specificity	Precision	Dice
U-Net	0.9495	0.7712	0.9682	0.7182	0.7437
R2U-Net	0.9522	0.7691	0.9715	0.7389	0.7537
Cs2-Net	0.9645	0.7818	0.9836	0.8331	0.8066
CAS-Net	0.9691	0.7896	0.9879	0.8718	0.8286

compared to other state-of-the-art methods. In the future, we would like to extend this model for segmentation of other 2D and 3D curvilinear structures.

References

- [1] World Health Organization. Cardiovascular diseases (cvds). <http://www.who.int/mediacentre/factsheets/fs317/>, September 2012.
- [2] A. Banerjee, F. Galassi, E. Zacur, G. L. De Maria, R. P. Choudhury, and V. Grau. Point-cloud method for automated 3d coronary tree reconstruction from multiple non-simultaneous angiographic projections. *IEEE Transactions on Medical Imaging*, 39(4):1278–1290, 2020.
- [3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [4] Nahian Siddique, Paheding Sidike, Colin Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: theory and applications. *CoRR*, abs/2011.01118, 2020.
- [5] Ebrahim Nasr-Esfahani, Shadrokh Samavi, Nader Karimi, SM Reza Soroushmehr, Kevin Ward, Mohammad H Jafari, Banafsheh Felfeliyan, B Nallamotheu, and Kayvan Najarian. Vessel extraction in x-ray angiograms using deep learning. In *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 643–646. IEEE, 2016.
- [6] Su Yang, Jihoon Kweon, and Young-Hak Kim. Major vessel segmentation on x-ray coronary angiography using deep networks with a novel penalty loss function. In *International Conference on Medical Imaging with Deep Learning-Extended Abstract Track*, 2019.
- [7] Tae Joon Jun, Jihoon Kweon, Young-Hak Kim, and Daeyoung Kim. T-net: Nested encoder-decoder architecture for the main vessel segmentation in coronary angiography. *Neural Networks*, 128:216–233, 2020.
- [8] Rui-Qi Li, Gui-Bin Bian, Xiao-Hu Zhou, Xiaoliang Xie, Zhen-Liang Ni, and Zengguang Hou. Cau-net: A novel convolutional neural network for coronary artery segmentation in digital subtraction angiography. In *Neural Information Processing - 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23-27, 2020*, volume 12532 of *Lecture Notes in Computer Science*, pages 185–196. Springer, 2020.
- [9] Lei Mou, Yitian Zhao, Huazhu Fu, Yonghuai Liu, Jun Cheng, Yalin Zheng, Pan Su, Jianlong Yang, Li Chen, Alejandro F Frangi, et al. Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical image analysis*, 67:101874, 2021.
- [10] Asma Kerkeni, Asma Benabdallah, Antoine Manzanera, and Mohamed Hedi Bedoui. A coronary artery segmentation method based on multiscale analysis and region growing. *Computerized Medical Imaging and Graphics*, 48:49 – 61, 2016.
- [11] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.