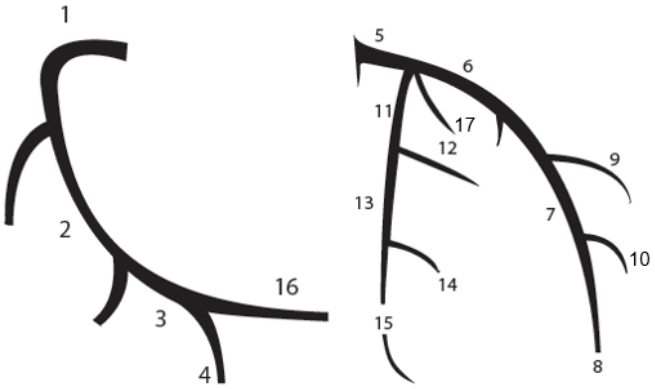
Automatic Labeling of X-Ray Images Based on Deep Learning

Version of September 17, 2018



Yin Yunchao

Automatic Labeling of X-Ray Images Based on Deep Learning

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE

by

Yin Yunchao
born in Tianjin, China

Computer Graphics and Visualization Group
Department of Intelligent Systems
Faculty EEMCS, Delft University of Technology
Delft, the Netherlands
www.ewi.tudelft.nl

Philips Healthcare
Gagel 1, Best
the Netherlands

Automatic Labeling of X-Ray Images Based on Deep Learning

Author: Yin Yunchao
Student id: 4717856
Email: yyc134039@gmail.com

Abstract

Coronary artery disease is the most common type of heart disease, which influences 110 million people's health and causes 8.9 million deaths in 2015. Physicians can visualize the lesion in coronary arteries by cardiac angiography (X-ray image) during diagnosis and treatment of coronary artery disease. The pathological findings in cardiac angiography are reported per segment or per artery of the coronary artery tree, therefore, it requires to annotate the name of each segment or artery in the coronary artery tree.

This thesis proposes a data-driven method as a first attempt at annotating cardiac angiography based on deep learning. The method aims at automatically regressing segment points between different segments on the coronary artery tree as the annotation of the cardiac angiography. The proposed data-driven cardiac angiography annotation methods can learn and generalize from manually annotated cardiac angiography examples, but its performance is limited by the number and quality of examples for learning.

Thesis Committee:

Chair:	Prof. Dr. A. Vilanova, Faculty EEMCS, TU Delft
University supervisor:	Prof. Dr. A. Vilanova, Faculty EEMCS, TU Delft
Company supervisor:	Dr. Javier Oliván Bescós & Dr. Roy Van Pelt, Philips Healthcare
Committee Member:	Prof. Dr. Jan Van Gemert, Faculty EEMCS, TU Delft

Preface

I would like to thank Dr. Javier Oliván Bescós and Dr. Roy Van Pelt from Philips Healthcare for the patient guidance and constructive feedback during this project. I would also like to thank Dr. A. Vilanova for her excellent suggestions and support in the research aspect of this thesis. Without their cooperation I would not have been able to complete this thesis.

Yin Yunchao
Delft, the Netherlands
September 17, 2018

Contents

List of figures

Chapter 1

Introduction

Coronary arteries are the vascular tree that supply blood, oxygen and nutrients to the heart. Unhealthy lifestyle including smoking, diabetes, lack of exercise, obesity, high blood cholesterol, poor diet, depression, and excessive alcohol may cause the accumulation of cholesterol-containing deposits (plaque) in coronary arteries. The plaque will narrow coronary arteries [Mehta et al., 2015][Charlson et al., 2013] that decreases blood flow, and leads to chest pain (angina), shortness of breath, or other coronary artery disease signs and symptoms. A complete blockage can cause a heart attack. The diseases caused by damage of coronary arteries are called Coronary artery disease, for instance myocardial infarction, and sudden cardiac death [Wong, 2014]. Coronary Artery Disease influences 110 million people's health and causes 8.9 million deaths in 2015, which is 15.9% of all-cause mortality [noa, 2016].

Percutaneous coronary intervention is the most common treatment for coronary artery disease. It dredges and enlarges the stenosis in the coronary artery to improve myocardial perfusion. Figure 1.1 shows an overall procedure of PCI. First, a catheter accesses coronary artery through the radial or femoral artery to deliver a deflated balloon and stent. When the balloon and stent reach the stenosis, they will be inflated and open up the narrowed artery.

X-ray cardiac angiographies are used to visualize the coronary arteries in both medical imaging diagnosis and PCI treatment. However, the diagnosis of coronary artery disease and PCI treatment are usually done by different physicians in clinical practice. Therefore, medical imaging diagnostic reports are generated to facilitate communication between physicians. According to the Cardiovascular Computed Tomography guideline for the interpretation and reporting of coronary computed tomographic angiography [Raff et al., 2009], the pathological findings are reported per segment or per artery as it shows in Figure 1.2. The coronary artery tree needs to be segmented and annotated first, so that the physicians can report plaque locations in the artery.

However, it is very time and effort consuming to segment and annotate the coronary artery tree only for making report. In order to improve workflow intelligence in PCI treatment and save time from making report, this project designs an algorithm to segment and annotate the coronary artery tree automatically.

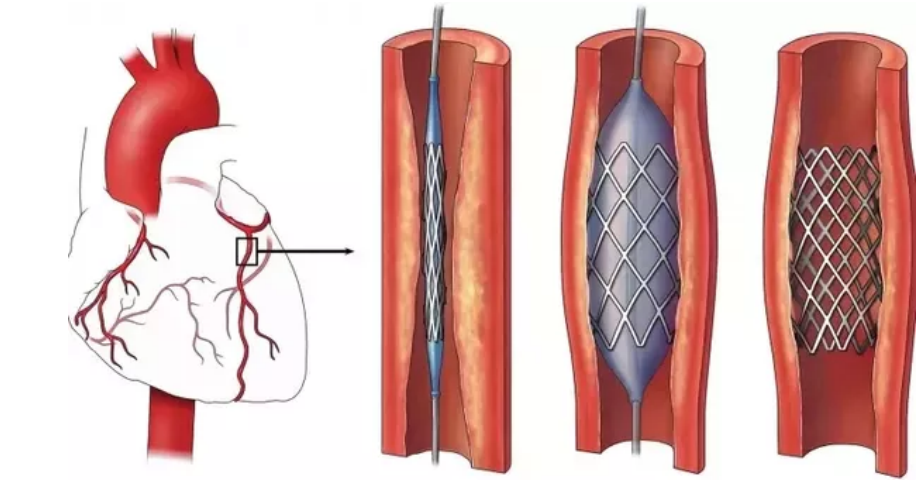


Figure 1.1: Overall procedure of PCI. The first image shows the location of stenosis in coronary artery, the balloon and stent are delivered by a catheter to the stenosis location as the second image, then the balloon is inflated to open up the stenosis.

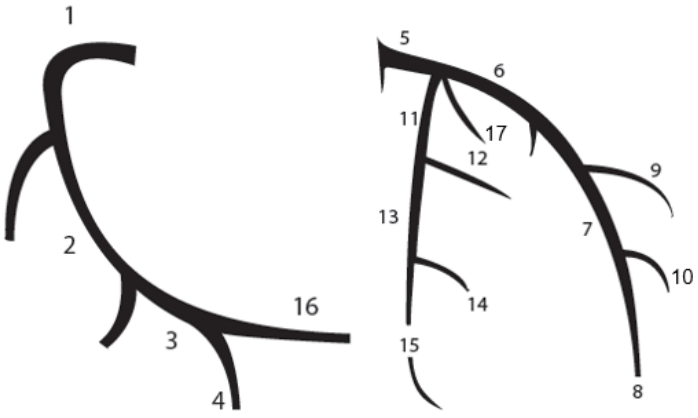
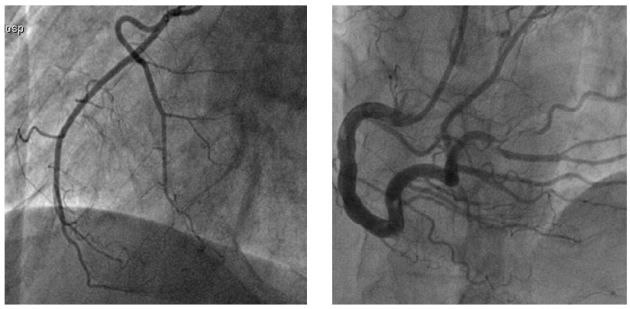


Figure 1.2: An example of segmented coronary artery tree in medical imaging diagnostic reports. On the left is right coronary artery, on the right is left coronary artery.



a) b)

Figure 1.3: The cardiac angiographies show right coronary artery trees from different patients, they have different anatomical structures.

1.1 Problem statement

Ideally, the coronary artery tree is divided into segments with a number of branches and bifurcations as Figure 1.2 [Yang et al., 2011]. However, coronary artery tree on cardiac angiography is difficult to identify in clinical practice for the following reasons:

1) The topological structures of coronary artery tree are different because of anatomical variations. Some patients have a lot of branches on the main arteries, while others do not. Angiographies in Figure 1.3 present two right coronary artery trees, but both look very different from the reference coronary topological model in Figure 1.2. Right coronary artery tree in the first cardiac angiography of Figure 1.3 does not bifurcate at the end of the artery, while the other one has a lot of branches at the same part. Moreover, anatomical variation is not the only reason for topology difference. During image acquisition, contrast agent is injected into the artery to improve contrast and visualize the coronary artery tree on cardiac angiography. The stenosis in arteries and heart motions will lead to the uneven distribution of contrast agent as Figure 1.4. It causes the topology difference including invisible side branches and discontinuous artery.

2) Image acquisition angle also influences coronary artery tree structure on cardiac angiographies. The right coronary artery angiographies in Figure 1.5 are acquired from different image acquisition angles, their structure looks different although it is the same.

3) The coronary artery tree in cardiac angiography is a projection of the 3D coronary artery tree. Ambiguities caused by projection also bring difficulties to cardiac angiography labeling, for instance the overlapping and crossing blood vessels in Figure 1.6. The bifurcations and side branches are used as important landmark for cardiac angiography annotation. However, the crossing blood vessels create bifurcations on the main artery, it can be confused with the bifurcations of side branch. Overlapping blood vessels are also common in angiography, it makes the overlapped blood vessel invisible on the angiography and brings ambiguities to the topological structure of coronary artery.

4) The noisy background of cardiac angiography increases the difficulties in coronary topological structure extraction. It is unavoidable to include medical instruments and catheters in the angiography during image acquisition because they are used in PCI treat-

Figure 1.4: These angiographies are influenced by the uneven distribution of contrast agent. a) shows the example of right coronary artery tree with invisible side branches on angiography; b) shows the example of the discontinuous artery on angiography; c) shows the cardiac angiography when the artery is totally occluded by plaque, the red arrow shows the plaque location. Contrast agent cannot be pumped into arteries behind the plaque, so half of the right coronary artery is not visible on the angiography.

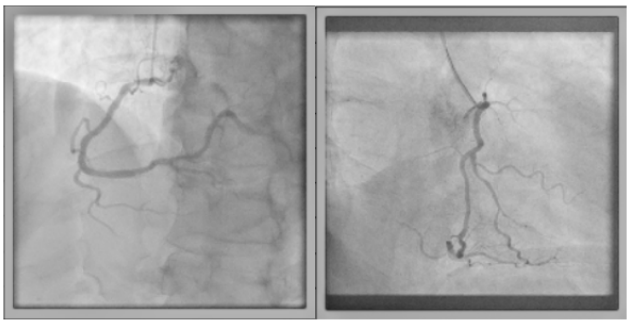


Figure 1.5: Two images show cardiac angiographies of right coronary artery acquired from different image acquisition angles, their topological structures are very different.

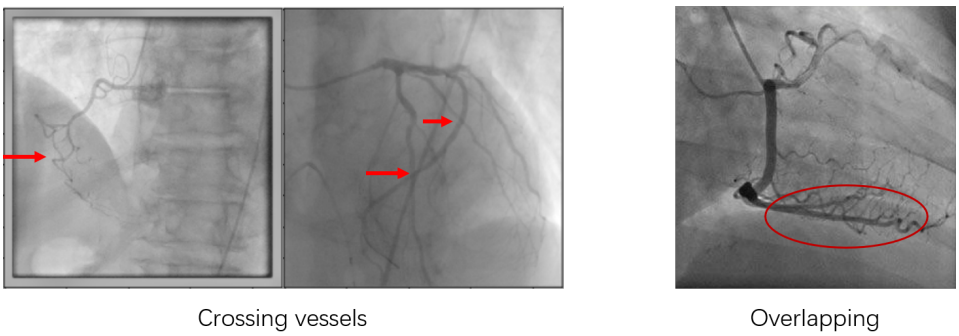
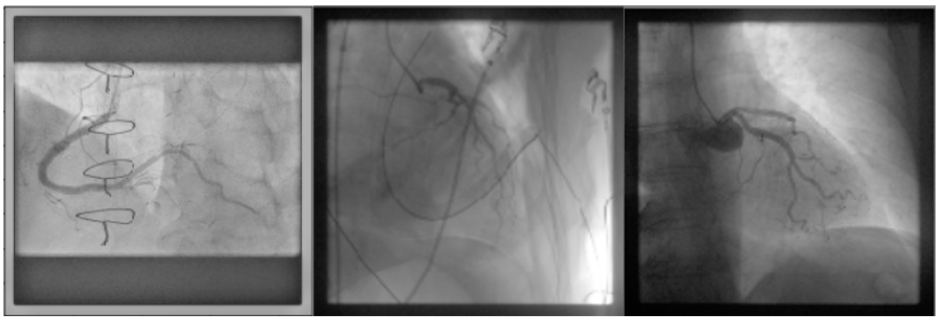


Figure 1.6: On the left, the images are crossing vessel examples in right coronary artery angiography and left coronary artery angiography respectively. On the right, the image shows an example of overlapping blood vessel.



a) Right coronary artery b) Left coronary artery c) Left coronary artery

Figure 1.7: Angiographies with noisy background.

ment. The tubular shape catheter on background of Figure 1.7 b) can be easily confused with blood vessels. Furthermore, the shadow of myocardium leads to the uneven greyscale on background as b) and c) in Figure 1.7, the gradient at the edge of myocardium shadow can also complicate the identification of the coronary artery tree.

In clinical practice, it is important that the cardiac angiography annotation method is robust and accurate under various coronary topological structure, ambiguities, and background noise.

Most previous methods for coronary artery tree labeling address it in a model-based approach. The topological structure of the coronary artery tree is extracted from cardiac angiographies first, and then matched with reference coronary artery tree topological model according to the similarity of their geometric features. The topological structure of coronary artery tree is segmented and labeled according to the reference model. However, the model-based approach is limited by coronary artery tree extraction accuracy and graph matching ability. The noisy background makes it hard to robustly and accurately extract the topological structure of the coronary artery tree. On the other hand, the definition of the model to match itself is also difficult due to multiple anatomical variations.

1.2 Assumption

In recent years, deep learning methods made great progress in a lot of image analysis tasks including image classification [Li et al., 2014][Shin et al., 2016][Krizhevsky et al., 2012][Szegedy et al., 2015], image segmentation [Long et al., 2015], object detection [Shin et al., 2016][Girshick, 2015] between others. Instead of predefined rules or models, deep learning methods learn the needed features for image analysis task from manually labeled data called training. If the training set is large enough and properly represents the variations in practice, the 'learn from examples' deep learning approach will be able to address the image analysis problem more accurately and robustly compared with traditional methods based on predefined rules or models.

In this thesis, we proposed the use of deep learning for cardiac angiography annotation

given the characteristics described above. Cardiac angiographies annotated by physicians are used as training set to train the deep learning model. The training set consists of cardiac angiographies acquired from different image acquisition angles, different coronary anatomical structures and cardiac angiographies with ambiguities. There is one assumption made to support this approach:

The cardiac angiography annotation model can be learned from annotated examples.

There are two contributions of this thesis:

1) This thesis proposes a method as a first attempt of annotating cardiac angiography based on deep learning. It shows that the cardiac angiography annotation method can be learnt from examples;

2) The method for cardiac angiography annotation proposed in this thesis is less dependent on manually designed feature extractors and coronary topological structure extraction compared with previous model-based approach.

Chapter 2

Clinical Background

In Chapter 1, we present that the topological structure of coronary artery tree is influenced by interpatient anatomical variations and image acquisition angles, we elaborate on the reason of anatomical variations and influence of image acquisition angles in section 2.1 and 2.3 respectively. In section 2.2, we introduce the coronary artery tree model for medical imaging diagnostic reports. In section 2.4, we introduce the available data for this thesis.

2.1 Coronary Artery Anatomy

The heart is a muscular organ, which pumps blood through blood vessels into the circulatory system. The coronary artery tree is part of the coronary circulation, which supplies blood to the cardiac muscle. The coronary artery tree is divided into left coronary artery tree and right coronary artery tree near the aorta valve. Figure 2.1 shows the branch names of the coronary artery tree. Right coronary artery tree runs through coronary sulcus and supplies blood to the right atrium, right ventricle, and the heart conduction. If right coronary artery tree continues supplying a portion of blood to posterior left ventricular wall via posterior descending artery, the patient is considered to have a right dominant circulation, which shown in 85% of the population. Left coronary artery tree supplies blood to the left part of the heart including the left atrium, left ventricle, and the interventricular septum. Left coronary artery tree has two major branches: the left circumflex artery and left anterior descending artery. If the posterior left ventricular wall is supplied by left coronary artery tree, the patient has a left dominant circulation, which represents 5% of the entire population. The other 10% population have a co-dominant circulation, which means the posterior left ventricular wall is supplied by both left and right coronary artery tree [Das, 2013][Betts, 2014].

2.2 17 segments model

The coronary artery tree model used for medical imaging diagnostic reports is shown in Figure 2.2, entitled '17 segments model'. It suits most coronary artery tree with right dominant circulation.

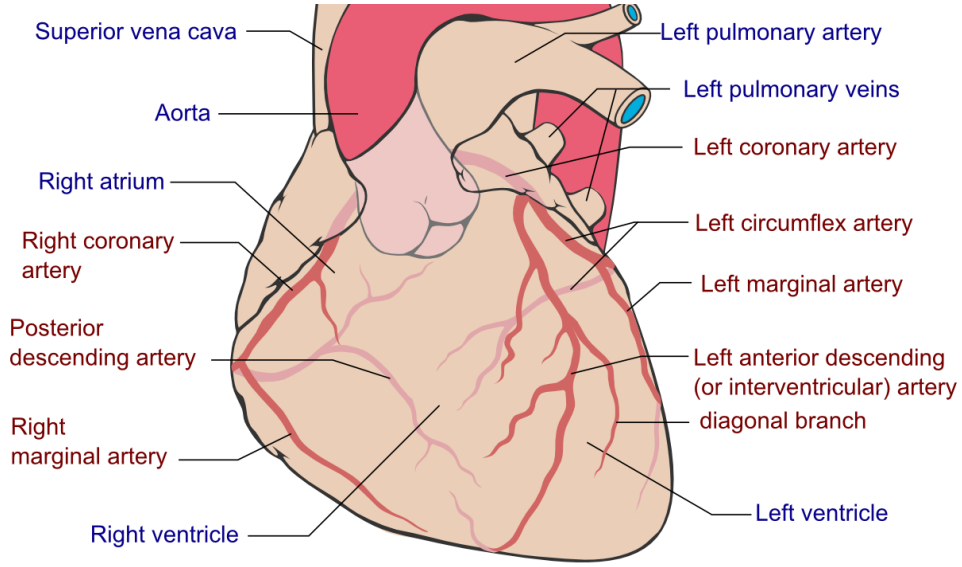


Figure 2.1: The anatomical terminology of coronary artery tree. The coronary arteries are labeled in red letters and other structures are labeled in blue letters.

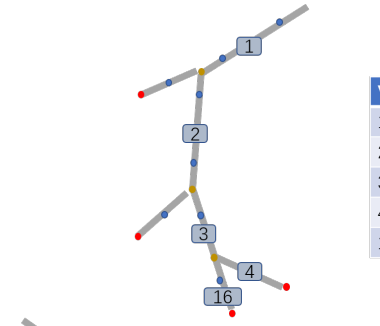
The main branch of right coronary artery is divided into proximal right coronary artery, mid right coronary artery, and distal right coronary artery. The right posterior descending artery and right posterolateral branch are used to supply blood to posterior left ventricular wall and they do not exist in coronary artery tree with left dominant circulation. The angiography on the left of Figure 2.3 shows an example of right coronary artery tree with right dominant circulation, which has the same structure as '17 segments model'. The artery bifurcates into right posterior descending artery and right posterolateral branch after distal right coronary artery. On the right side is an example of right coronary artery tree with left dominant circulation. It does not have right posterior descending artery and right posterolateral branch, because the posterior left ventricular wall is supplied by left coronary artery tree.

Left coronary artery tree has two main branches, namely left anterior descending artery and left circumflex artery. They are also divided into three segments: proximal, mid and distal. The side branches on left anterior descending artery are named as diagonal branches and the side branches on left circumflex artery are named as marginal branches. When the patient has a left dominant circulation, there will be more side branches on left circumflex artery to supply the posterior left ventricular wall.

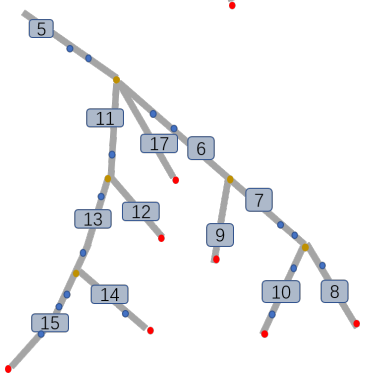
2.3 Projection angle

Except for the interpatient anatomical variations, different angiogram views also influence the shape and topological structure of coronary artery tree on the cardiac angiographies.

The angiographies used in this thesis are taken by 'Philips Mobile C-arm', which is



Vessel segment names in RCA
1. Proximal RCA
2. Mid RCA
3. Distal RCA
4. Posterior Descending Artery (RPDA)
16. Right posterolateral branch (RPLB)



Vessel segment names in LCA	
5 Left main	11 Proximal Ramus Circumflex Artery (PLCX)
6 Proximal Left Anterior Descending Artery (LAD)	12 Marginal branch (M1)
7 Mid LAD	13 Mid LCX
8 Distal LAD	14 Marginal branch (M2)
9 Diagonal branch 1 (D1)	15 Distal LCX
10 Diagonal branch 2 (D2)	17 Intermediate artery

Figure 2.2: The 17 segments model defined by American Heart Association

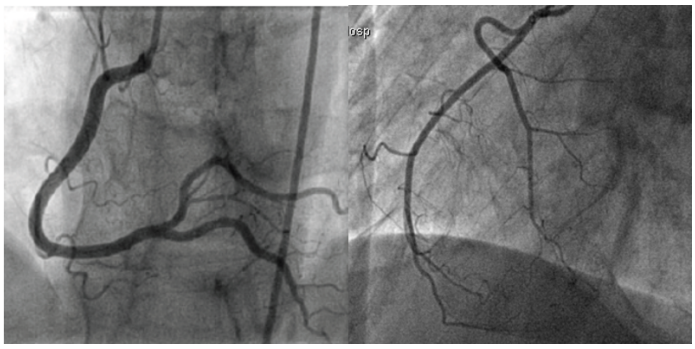


Figure 2.3: Right coronary artery tree with different circulation type: a) This is a typical right dominant right coronary artery tree, because the artery bifurcates to right posterior descending artery and right posterolateral branch after distal right coronary artery; b) This is a left dominant right coronary artery tree, it does not have right posterior descending artery and right posterolateral branch after distal right coronary artery segment.

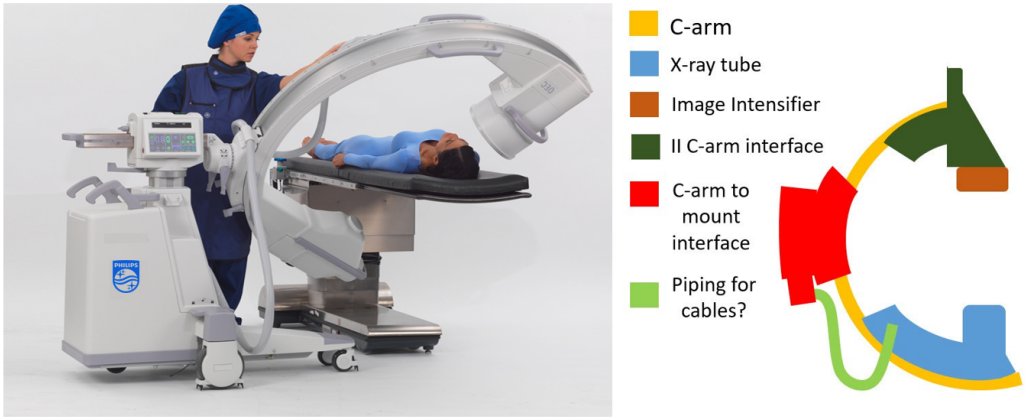


Figure 2.4: The image shows the name of each part in Mobile C-arm and how the cardiac angiography is taken by Mobile C-arm.

a medical imaging device based on X-ray technology. Figure 2.4 shows how the cardiac angiography is taken and name of each component on the Mobile C-arm. Patient lies between image intensifier and X-ray tube, the cardiac angiography is acquired as a projection of heart in the direction of the connection between image intensifier and X-ray tube. C-arm can rotate around the patient and acquire angiography from different projection angles.

Left anterior oblique and right anterior oblique are two commonly used cardiac angiography acquisition angles. Figure 2.5 shows same patient's angiographies acquired from left anterior oblique and right anterior oblique. The shape and topological structure of same patient's right coronary artery are different in two angiographies. In angiography taken from right anterior oblique, right posterior descending artery and right posterolateral branch are less obvious than they are in the angiography taken from left anterior oblique. It is challenging to annotate angiographies taken from different image acquisition angle correctly.

2.4 Data

In order to increase contrast and visualize the arteries during cardiac angiography acquisition, contrast agent is injected into coronary arteries. Therefore, the cardiac angiography is usually a sequence of images taken during contrast agent injection. Figure 2.6 selects several important frames in the sequence to present how the coronary artery tree changes in cardiac angiography during contrast agent injection.

The cardiac angiographies used in this thesis is provided two partner Dutch hospitals of Philips. Each cardiac angiography sequence consists of 20-100 frames. There are 3000 sequences from 300 patients in total, mixed by 1369 left coronary artery tree angiography sequences, 596 right coronary artery tree angiography sequences and 1035 angiography sequences without vascular tree.

The training set was annotated by the author of this thesis under supervision of clinical scientists. We select the best frames from 200 patients' right coronary arteries angiography

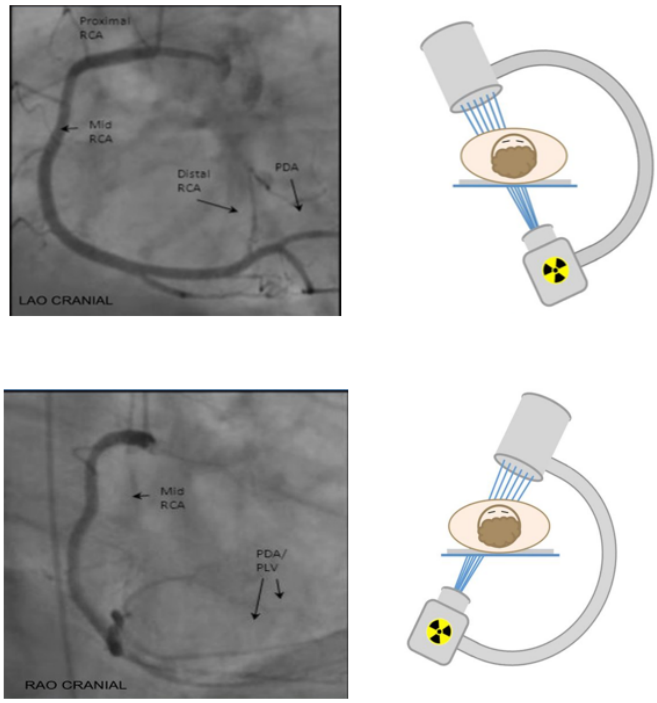


Figure 2.5: Right coronary artery angiography acquired form different angiogram views: a) on the top is left anterior oblique view and corresponding Right Coronary Artery angiography; b) on the bottom is right anterior oblique view and corresponding right coronary artery angiography.

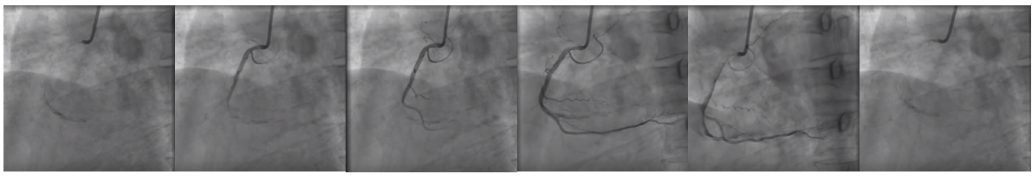


Figure 2.6: Some selected frames in the cardiac angiography sequence. The right coronary artery tree gradually becomes visible as the contrast agent injected into the artery, and then gradually disappear as the contrast agent is expelled from the right coronary artery.

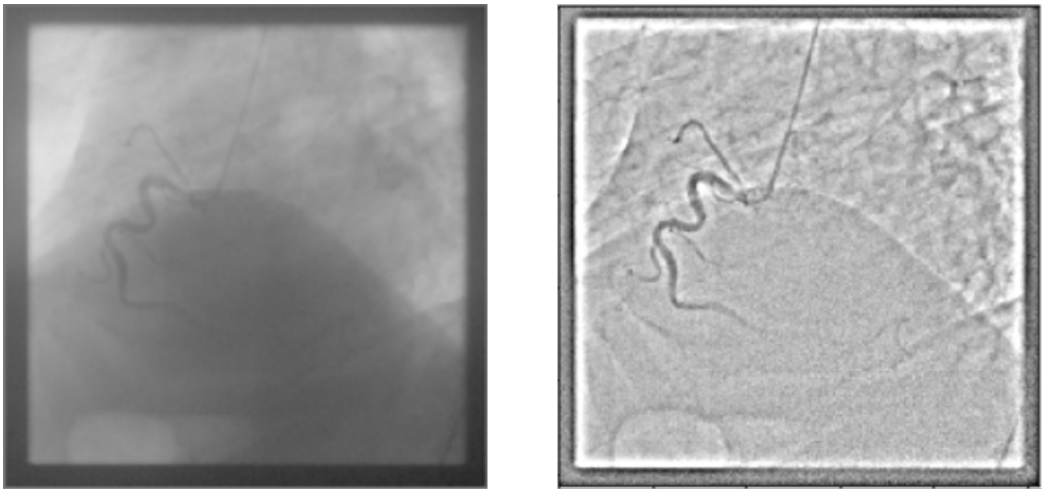


Figure 2.7: On the left is the original right coronary artery angiography; on the right is the preprocessed right coronary artery angiography.

sequences and annotate them based on 17 segments model for following experiment. 10.6% of the cardiac angiographies used to train the neural network have right coronary artery trees with left dominant circulation, while 89.6% of the data have coronary artery trees with right dominant circulation. Because of the time limitation and complexity of the left coronary artery tree, the initial focus was on annotation of the right coronary artery tree. Besides, the cardiac angiographies we used for training are preprocessed to boost the contras, an example is shown in Figure 2.7.

Chapter 3

Related work

According to our literature research, there is very limited research focusing on cardiac angiography annotation. Most of the proposed methods use a model-based approach, which has two main steps: 1) extracting the topological structure of coronary artery tree from cardiac angiography; 2) annotating the topological structure by segments. We present the literature on cardiac angiography annotation on these two main steps, section 3.1.1 reviews the coronary topological structure extraction methods, section 3.1.2 reviews the coronary artery tree annotation methods. In section 3.2, we review the human pose estimation methods based on deep learning, which inspired us to annotate cardiac angiography.

3.1 Cardiac angiography labeling

Previous methods extract coronary artery tree from cardiac angiographies first, then annotate it based on predefined reference model. Figure 3.1 shows two examples of reference model used in previous methods. Reference model is a predefined 2D topological structure of coronary artery tree, which encodes the prior anatomical knowledge as geometric features, including the hierarchy of coronary segments, the bifurcation locations, curvature, length proportion of different segments and so on. These geometric features can be used to distinguish segments on the coronary artery tree. Different parts on the topological structure are annotated as the segment on reference model that has the most similar geometric features with it.

The topological structure extraction methods are reviewed in section 3.1.1, and the coronary artery tree labeling methods are reviewed in section 3.1.2.

3.1.1 Coronary artery tree extraction

The centerline of the coronary artery tree is used as its topological structure. There are three kinds of points on coronary artery centerline as in Figure 3.2: simple points, bifurcation points, and end points.

Haris et al. [1999] and Metz et al. [2007] both track the centerline from a user defined start point. Haris et al. [1999] applied recursive sequential tracking method to follow the blood vessels after start point. Figure 3.3 shows how recursive sequential tracking method

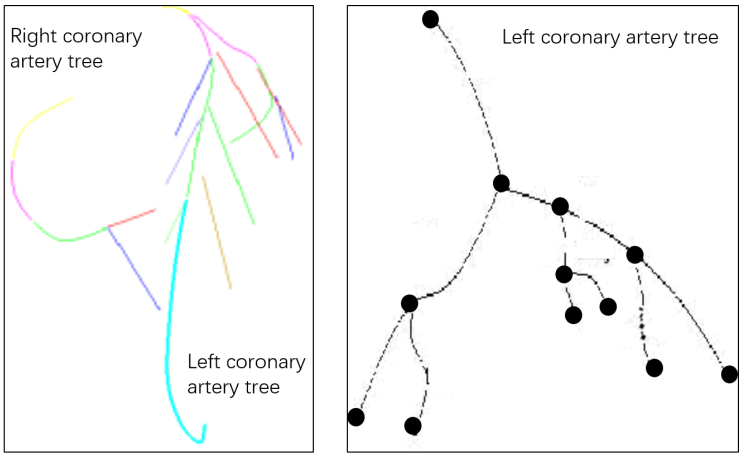
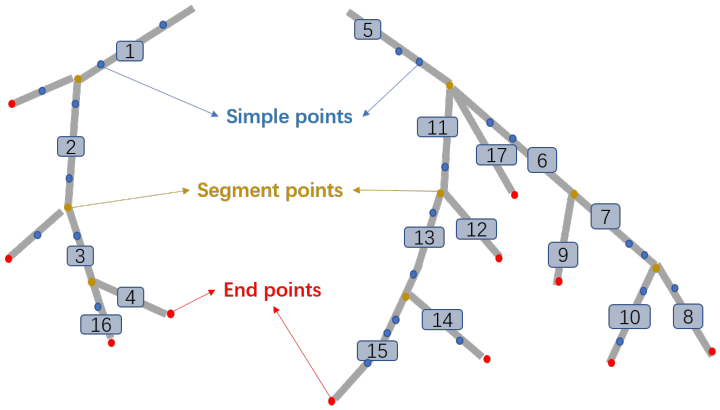


Figure 3.1: The images are reference models of coronary artery. a) shows the right and left coronary artery tree reference models used in Yang et al. [2011], b) shows the left coronary artery tree reference model used in Akinyemi et al. [2009].



Name	Function
Bifurcation point	The point where blood vessels bifurcate or cross with each other, in other words, the points link with more than one terminate point
End point	The point has no child points and locates at the end of the blood vessel
Simple point	The point links with only one terminate point, which is the majority of the blood vessel centerline

Figure 3.2: The graph on the top shows CAT centerline, yellow points are bifurcation points, red points are end points, blue points are simple points. The table below describes the difference between them.

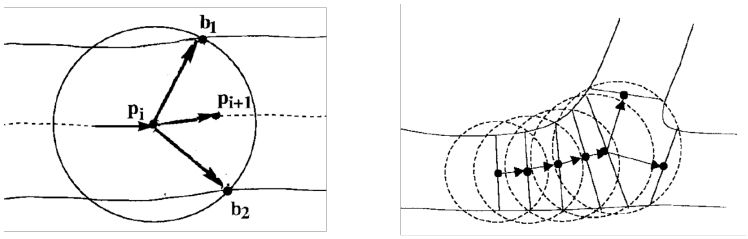


Figure 3.3: The images show the recursive sequential tracking method used in Haris et al. [1999]. a) shows the tracking method on simple points, b) shows the tracking method on bifurcation points.

works. A circular template with diameter slightly larger than the blood vessel radius is predefined and centered around the points on centerline. The grey level gradients on angiography within the circular template are analyzed to search for the optimal boundary of blood vessels. P_i is the current point on the centerline, and the optimal boundary intersects the circular template at b_1 and b_2, then the next point on blood vessel centerline P_{i+1} is given by the sum of vector $\overrightarrow{P_i b_1}$ and $\overrightarrow{P_i b_2}$function 3.1. The tracking starts from the user-defined start point on coronary artery centerline and operates recursively at artery bifurcation or cross areas.

$$P_{i+1} = P_i + \frac{\overrightarrow{P_i b_1} + \overrightarrow{P_i b_2}}{2} \tag{3.1}$$

Metz et al. [2007] also tracks the coronary arteries from a user-defined point on the centerline, but instead of the recursive sequential tracking method used in Haris et al. [1999], Metz et al. [2007] tracks the blood vessel by region growing method, which iteratively adds neighboring pixels within a predefined grayscale range to blood vessel region as in Figure 3.4. The newly added pixels are analyzed to see whether they are connected to each other or not. If the newly added pixels are connected as in Figure 3.4 a), then the newly added pixels are simple points, while if they are not connected as Figure 3.4 b), the bifurcation on coronary artery is detected.

The tracking based coronary artery extraction methods require a user-defined start point on coronary artery centerlines, for instance Haris et al. [1999] and Metz et al. [2007]. The required interaction will bring inconvenience in clinical practice.

In order to improve the automation of coronary artery centerline extraction, Liu et al. [2016] replaces the tracking methods by extracting centerline from a pixelwise segmented coronary artery tree. Figure 3.5 shows the algorithm scheme used in Liu et al. [2016]. First, the original cardiac angiography is preprocessed to enhance contrast between vessels and background. The enhanced angiography is processed by Hessian filter, which has a strong response on tubular structures including blood vessels. The Hessian filter outputs a binary map as shown in Figure 3.5 c), the detected tubular structures are labeled as 1 and other pixels are labeled as 0. The binary map is further refined by removing false positive responses on the background. Finally, the centerline of left coronary artery tree is extracted

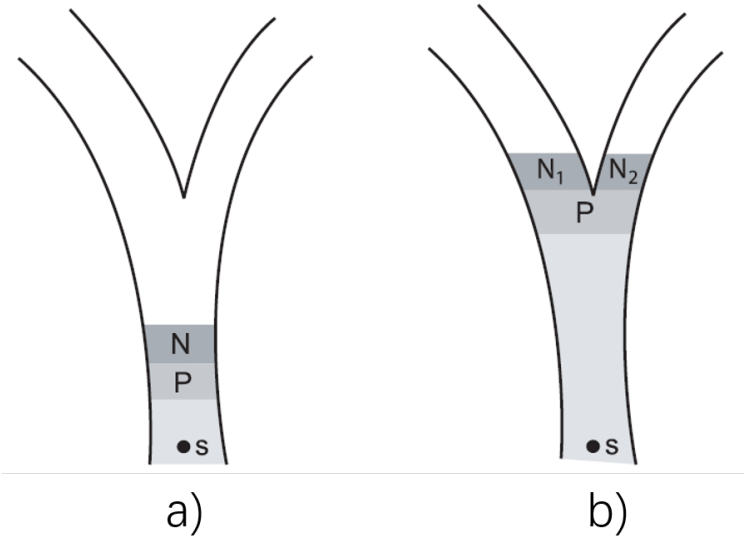


Figure 3.4: The images show the tracking method based on region growing in Metz et al. [2007]. 's' is the user defined start point on coronary artery centerline, 'N' and 'P' in a) are the newly added pixels in the last Two iterations. b) shows when there is a bifurcation points on the vessel, the newly added pixels 'N_1' and 'N_2' are not connected to each other.

from the refined binary map using fast marching algorithm [Bærentzen, 2001][Hassouna and Farag, 2007].

The centerline extraction method introduced in Yang et al. [2012] is similar to Liu et al. [2016], but improves the second step: generating binary map of coronary artery tree. They use an improved Hessian filter, which is less sensitive to the step-edge response cause by non-vessel structures including cardiac chamber.

Though Liu et al. [2016] and Yang et al. [2012] improved the automation of coronary artery centerline extraction by reducing interaction, but they can hardly extract coronary artery centerlines precisely when the contrast agent is not injected steadily as Figure 1.4 or background is noisy as Figure 1.7.

3.1.2 Coronary segments annotation

Extracting topological structure of coronary artery tree is the necessary precondition for annotation. The topological structure annotated by matching a predefined reference model, which encode the clinical prior knowledge as geometric features. Methods have been proposed for this step and will be presented next.

Ezquerra et al. [1998] generates the 2D reference model by projecting a 3D coronary artery tree model from the same angle as the test angiography. Figure 3.6 shows a left coronary artery angiography acquired from left anterior oblique view and the reference model acquired from corresponding angle. The geometric feature dissimilarities of topological

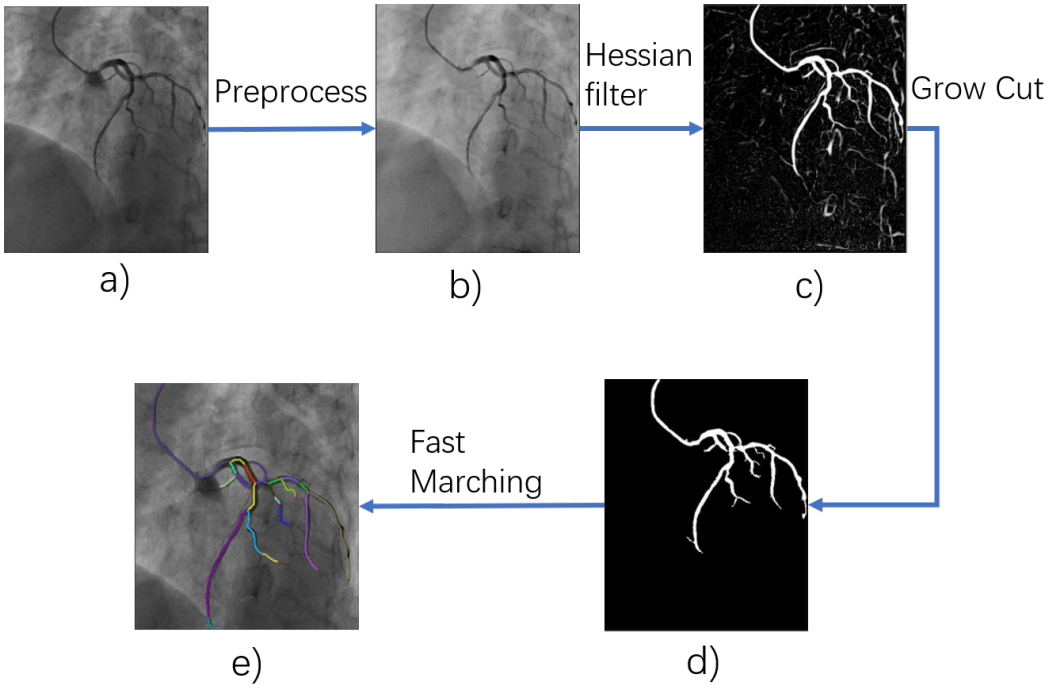


Figure 3.5: The flow chart of Liu et al. [2016]. a) the original LCA angiography, b) angiography enhanced by MSRCR, c) binary map produced by Hessian filter, candidate pixels on CAT are colored white, d) binary map refined by Grow Cut, false candidate pixels on the background are removed, e) the centerline of LCA extracted by fast marching algorithm

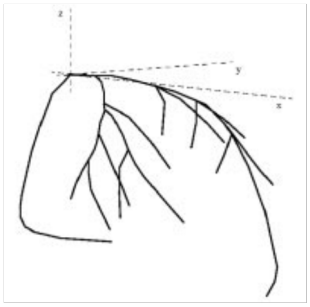
coronary structure in the angiography and reference model are defined as a cost function. The cost function is optimized iteratively to minimize the geometric feature difference, so that the most similar segment on topological structure and reference model of coronary artery tree are matched with each other. The topological structure is then segmented and labeled according to the reference model. They also use adjacent frames in the sequence of angiography to remove ambiguities, for instance, the crossing vessels can be removed when it does not exist on adjacent angiographies in the sequence. a) and c) in Figure 3.6 present the annotated left coronary artery angiography in Ezquerra et al. [1998] and the segment names.

Chalopin et al. [1998] also generates 2D reference models from 3D coronary artery tree. Ezquerra et al. [1998] generates the 2D reference according to the image acquisition angle during annotation, while citet731985 generates 98 reference models from difference projection angles beforehand (49 models for left coronary artery and 49 models for right coronary artery). The reference model that most closely resembles the coronary topological structure is chosen. The reference model and coronary topological structure are matched by similarity of curvature features on the artery, then each segment on topological structure is annotated with the names stored in curvature attributes of the reference graph.

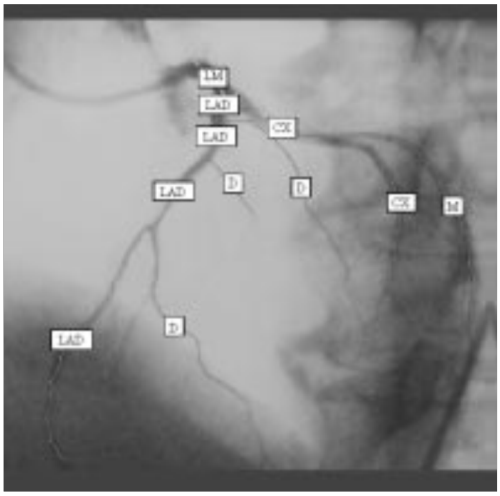
Recent studies tend to annotate coronary artery in a hierarchical manner, they first iden-

c) Vessel names abbreviations

Vessel	Abbreviation
Left Main	LM
Left Anterior Descending	LAD
Left Circumflex	CX
Septal	S
Diagonal	D
Marginal	M
Obtuse Marginal	OM
Median Ramus	MR



b) Reference model



a) Left coronary artery annotated in Ezquerra et al.[1998]

Figure 3.6: a) is annotated left coronary artery angiography acquired from left anterior oblique view in Ezquerra et al. [1998]; b) is the reference model projected from left anterior oblique view by the predefined 3D coronary artery tree model; c) is the name of each segment on left coronary artery tree.

tify the main arteries (right coronary artery, left anterior descending artery, and left circumflex artery) that every coronary artery tree has, then the side branches with more complex topological variations but less medical interest. Chalopin et al. [2001] first matches the main artery with reference model in each frame of angiography sequence. Dissimilarities of geometric features including blood vessel length, curvature and diameter are used as the criteria for graph matching. The annotated angiography with highest confidence is selected, and then it is used to modify the annotation on other angiographies in the sequence. Similar procedure is repeated to identify the side branches.

Yang et al. [2011] and Cao et al. [2017a] also annotate cardiac angiography in a hierarchical manner. They identified the CAT by point-set registration method [Myronenko and Song, 2010]. The main arteries in the test graph are normalized according to the registered point set and matched with reference model by rigid transformation. Then the difference of geometric features with reference model is minimize iteratively to annotate side branches.

These kind of hierarchical coronary annotation methods are more robust and accurate than Chalopin et al. [1998] and Ezquerra et al. [1998] which annotate the entire coronary artery tree by minimizing the difference of geometric features with reference model. Because the main arteries have less variations and easier to be annotated than side branches,

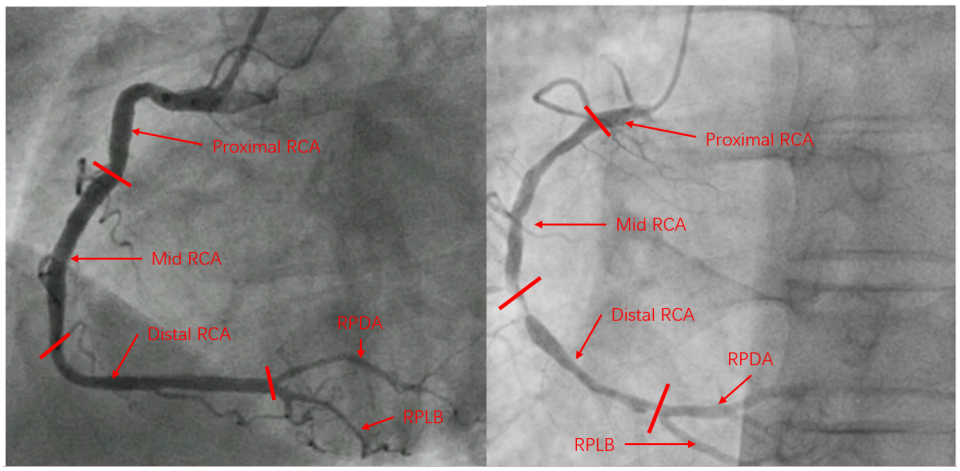


Figure 3.7: The images show right coronary artery(RCA) tree of different patients. The length of proximal RCA segments is different in two artery tree, the curvature of same segment in two artery tree is also different.

while based on the identified main branches, side branches will become easier to be correctly annotated than there is no information about main branches.

The graph matching approaches work well for the generic vasculature, but this approach will fail to correctly annotate the topological structure that is very different from the average. Therefore, Akinyemi et al. [2009] proposes a less strict model-based cardiac angiography annotation method. They generate a set of plausible candidate label for each segment by a pretrained multivariate Gaussian classifier, which uses geometric features including length proportion of each artery segment, the branch angle, average diameter, curvature of arteries as criteria. The candidate labels are slowly reduced by comparing the spatial properties, for instance the left anterior descending artery is always a child of left main. Finally, the labels are given to each vessel segment.

The annotation methods in Akinyemi et al. [2009] is less strict than graph matching, but predefining geometric features as classification criteria still limits algorithm's ability on coping with variations. For instance, the radius and proportion of each coronary segment are used as features for classifier in Akinyemi et al. [2009], but they are different according to patient's gender, age and dominant type. The branch angle and curvature of same coronary artery segment can also be different because of patient's anatomy and cardiac cycle phase. Figure 3.7 shows two right coronary artery trees with different length proportion and curvature at same segments.

On the other hand, the ability of transforming geometric features of coronary topological structure into a suitable representation or feature vectors also limits the model-based annotation approach. It is difficult to design a feature extractor transforming the structure of coronary artery tree into a suitable representation, which can be used for the graph matching cost function and multi-variate Gaussian classifiers.

The development of deep learning methods motivates us to address cardiac angiography

Figure 3.8: Result graph of human pose estimation task. Yellow lines: Head; Red lines: left arm; Blue lines: right arm; Green lines: right leg; Purple lines: left leg.

annotation by a data-driven approach. Deep learning methods discover the features needed for various image analysis tasks automatically from manually annotated examples (training set), instead of manually designing features extractors to cover all cases. The variations that are hard to be represented by reference models should be encoded in the training set for deep learning methods to learn, for instance different anatomical structure of coronary artery tree, various image acquisition angle, and complex background noise.

Though deep learning has not been used to address cardiac angiography annotation, it has made great progress in a lot of other image analysis tasks including human pose estimation, which shares a lot of common points with cardiac angiography annotation. Human pose estimation annotates names on different part of human body as it shows in Figure 3.8. The challenges in human pose estimation are also similar to cardiac annotation task: 1)human gesture graphs deform according to the limb movements, coronary topological structure also deforms according to different cardiac cycle phase; 2) the human gesture graphs are different in various image acquisition angle, the same problem also exists in cardiac angiography annotation; 3) human pose estimation is also a 3D to 2D problem, the human is projected on the image, therefore the limbs can overlap each other or be occluded by other stuff in the image, and the blood vessels can also cross and overlap each other; 4) the background for human gesture estimation changes a lot according to where images are taken, the background of angiography is also very variable. Because human pose estimation and cardiac angiography annotation has similar target and challenges, and corresponding method will be used in this thesis, we review the related studies of human pose estimation in next section.

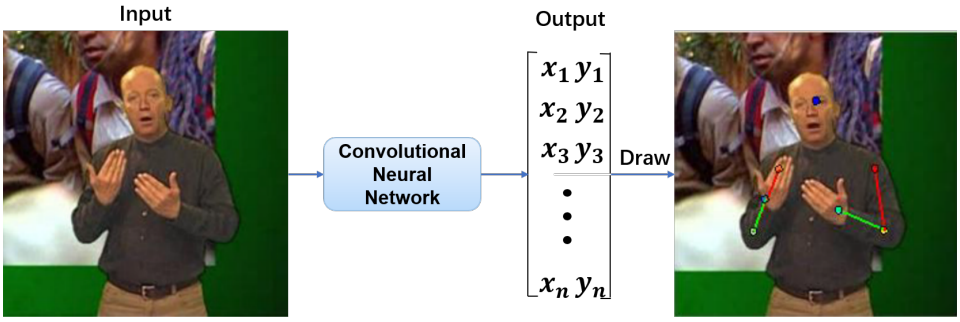


Figure 3.9: The overall scheme of joint coordinate regression method. The image is directly used as input of convolutional neural network, and the coordinates of n joints are outputted as a vector. Finally, the joints are drawn on the image and connected as a human skeleton.

3.2 Human pose estimation based on deep learning

Human pose is defined by the location of joints as Figure 3.8. Therefore, deep learning based human pose estimation methods including Wei et al. [2016], Pfister et al. [2015], Pfister et al. [2014], and Toshev and Szegedy [2014] first detect the joints location, then build human skeletons by connecting the joints in order.

Pfister et al. [2014] and Toshev and Szegedy [2014] regress the joint coordinates by a trained convolutional neural network. As it shows in Figure 3.9, convolutional neural network takes raw image as input and predicted the joint coordinates. The human skeleton is drawn by connecting the joint coordinates in order.

In another series of papers, [Wei et al., 2016], Pfister et al. [2015], Bulat and Tzimiropoulos [2016], and Payer et al. [2016] use convolutional neural network to regress heatmaps for each joint, the joint location is represented by a Gaussian blob on the corresponding heatmap as Figure 3.10. Convolutional neural network for heatmap regression takes original image as input and aims at outputting a series of heatmap with a Gaussian blob on it showing the joint location.

As mentioned in Pfister et al. [2015] and Tompson et al. [2014], joint coordinates regression is a highly non-linear and more difficult to learn mapping compared with heatmap regression. Heatmap regression has a multi-model output, which means multiple spatial locations can be predicted as the required joint (for instance, multiple gaussian blobs on the predicted heatmap). Such multi-model output gives more information about the image in every training iteration compared with a joint coordinate, so that learning becomes easier. Besides, the 'thinking process' of the heatmap regression is also easier to be understood and visualized. Figure 3.11 shows the predicted heatmap at different convolutional layers in Pfister et al. [2015]. The convolutional neural network first locks human as interesting area as it shows in a) and b) of Figure 3.11, false positive responses on predicted heatmap are gradually suppressed and less regions are predicted as target as shown in Figure 3.11 c), finally the joint locations are predicted as a Gaussian blob on corresponding heatmap.

The joint location regression used in pose estimation inspires us to use a similar method

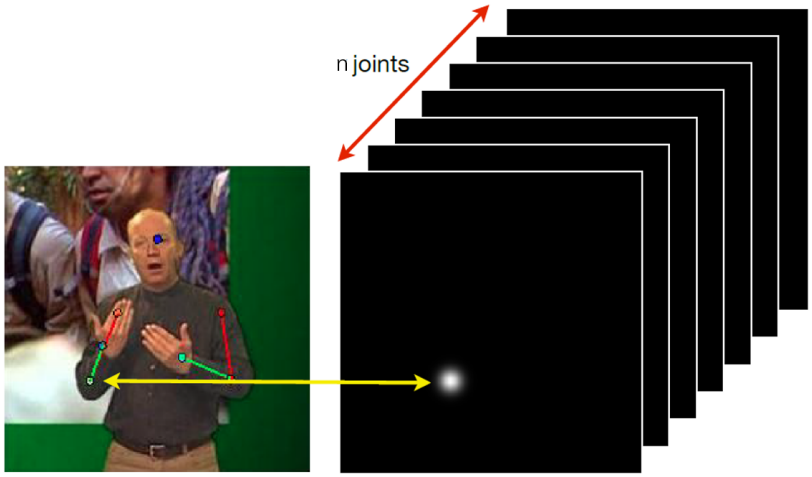


Figure 3.10: The images are the original image and corresponding joint heatmaps used as regressing target. There are n joints in image and each joint corresponds to one heatmap.

for annotating the cardiac angiography. Previous model-based cardiac angiography annotation methods first extract the topological structure of coronary artery tree in order to compare it with the reference model. In this thesis, we decide to annotate the cardiac angiography directly by regressing the segment point locating at the boundary of different segments. More details of the algorithm are introduced in next chapter.

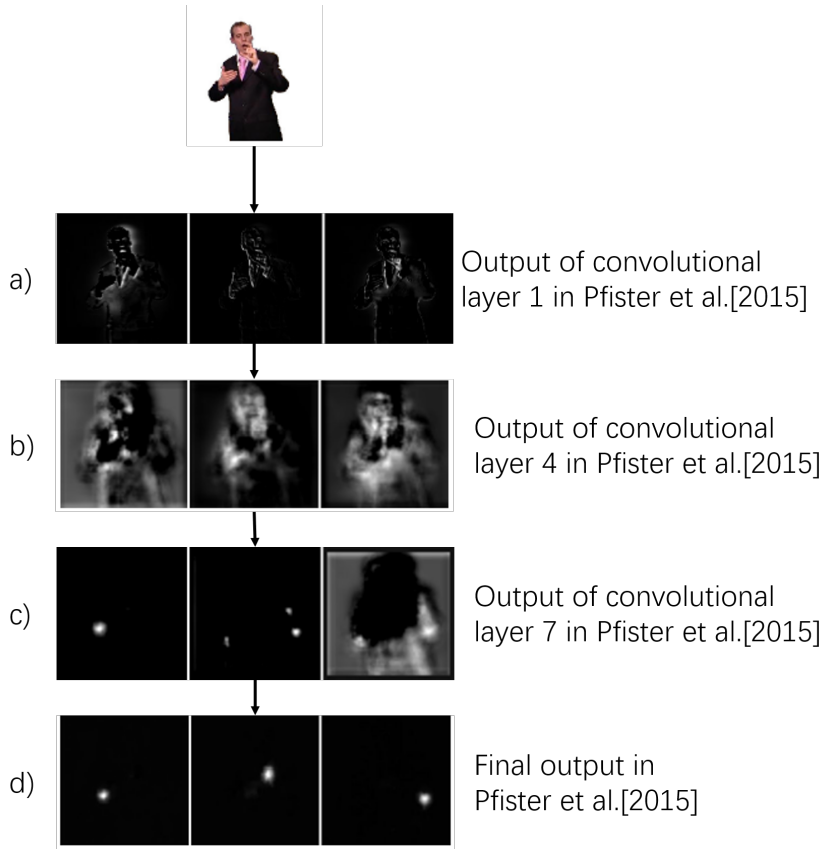


Figure 3.11: The images shows the predicted heatmap at different stages in the convolutional neural network used in Pfister et al. [2015]. In shallow layers of the network, area around human is locked as target. The false positives on the heatmap are then suppressed gradually as the network getting deeper. Finally, the predicted joint location is represented by a gaussian blob on heatmap.

Chapter 4

Cardiac angiography annotation methods

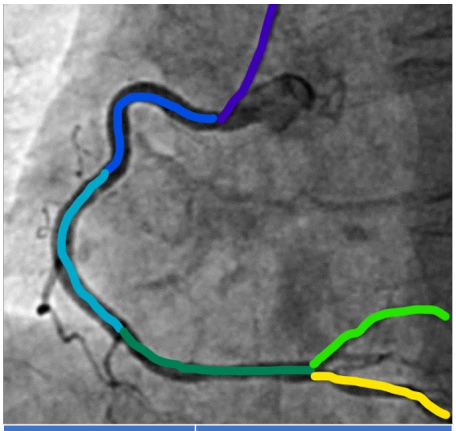
The model-based approach of cardiac angiography annotation works well for generic vasculature resembling the reference model. However, the model-based approach can hardly cope with anomalies, for instance the coronary artery has a special anatomy, or the side branches are different from reference models. The abilities of extracting coronary topological structure and transforming geometric features on coronary topological structure into a mathematical representation also limit the model-based methods. Therefore, this thesis uses deep learning methods which can automatically discover the features needed for annotation from training set, and the anomalous cardiac angiographies should be encoded in training set for deep learning method to learn.

Inspired by human pose estimation methods, this thesis proposes a method to annotate the coronary artery by regressing the segment points (see Figure 4.1). There are several benefits of segment point regression instead of labeling extracted coronary topological structure:

1) Segment point regression methods can be trained in an end-to-end fashion. It does not need to segment the coronary artery tree from angiography first, but directly regresses the segment point location on cardiac angiography. Therefore, the algorithm is no longer limited by the ability of centerline extraction.

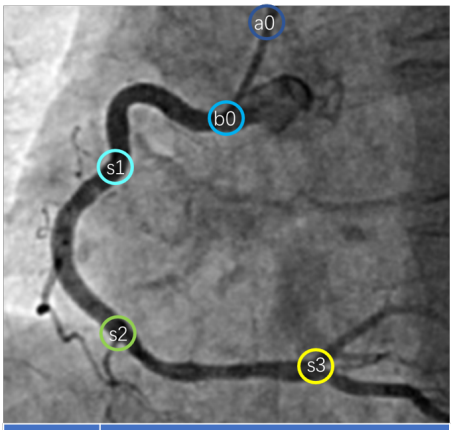
2) The training set for segment point regression is easy to get. As it shows in a) of Figure 4.1, we can use deep learning segment the coronary artery by pixel and assign label to each segment, in other words, we assign labels to each pixel to indicate which segment each pixel belongs to. However, deep learning method requires a lot of annotated data to get the neural network properly trained, getting all pixels labeled in hundreds of cardiac angiographies is extremely effort consuming. Besides, pixel-wise segmentation of the cardiac angiography is not necessary for coronary annotation, but the start and end locations of each segment is enough as shown in b) of Figure 4.1. Manually annotating the training set of cardiac angiographies as b) in Figure 4.1 is easier compared with a) and sufficient for our goal.

Given the benefits of segment point regression, we choose to annotate the cardiac angiography by segment points. The details of proposed method are introduced in this chapter. Section 4.1 presents the overall algorithm scheme. Modules in the algorithm scheme are



Color	Meaning
Violet	Catheter segment
Blue	Proximal RCA segment
Cerulean	Mid RCA segment
Fern green	Distal RCA segment
Bright green	RPDA segment
Yellow	RPLB segment

a)



Point	Meaning
a0	The start point of catheter
b0	The segment point between catheter and proximal RCA
s1	The segment point between proximal RCA and mid RCA
s2	The segment point between mid RCA and distal RCA
s3	The segment point between distal RCA, RPDA, and RPLB

b)

Figure 4.1: a) annotates the cardiac angiography based on the extracted right coronary artery; b) annotates the cardiac angiography based on segment points locating at the boundary of different segments.

elaborated in corresponding subsection.

4.1 Overall Scheme

As introduced in section 3.2, the multi-model output of heatmap regression makes learning easier, and the coarse-to-fine 'thinking process' of neural network is also easier to be visualized and understood compared with coordinate regression.

Such multi-model output of heatmap regression is very important for cardiac angiography annotation task. Because the local features around different segment points are very similar to each other as it shows in the second row of Figure 4.2. Such similar local features may make that the network confuses different segment points, which can be identified just by global information of the coronary artery tree. Multi-model output enables the network to first predict all possible locations on a coarse heatmap including the false positive responses caused by similar local features, then we can suppress false positive responses

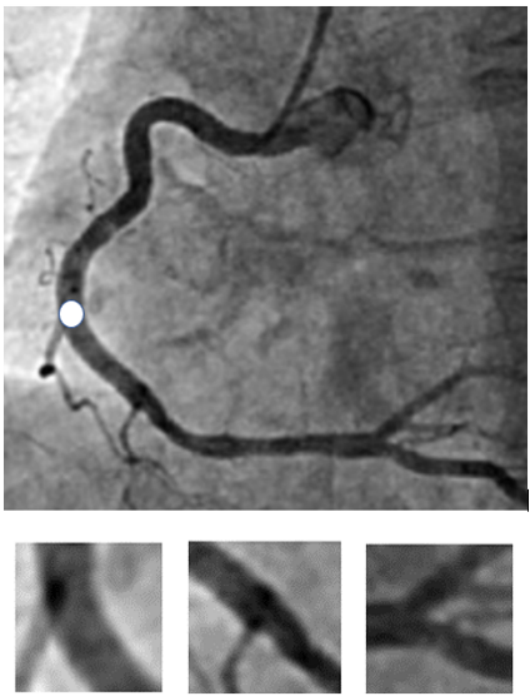


Figure 4.2: Above is RCA angiography with white point shows the vessel radius; the patches below are the local structures of different segment points on the angiography.

gradually. This coarse-to-fine procedure can prevent missing the correct segment point location and makes it easier to understand why the network fails to predict segment point location correctly.

On the other hand, segment coordinate cannot represent the boundary between different segments well. Because the boundary is not a precise coordinate but a small region according to the blood vessel radius as Figure 4.2. It is better to represent the segment point location as a Gaussian blob instead of an exact coordinate.

We detect the segment point location by heatmap regression. The overall scheme of proposed cardiac angiography annotation method is shown in Figure 4.3. In Module 1: angiography classification, we classify the angiographies according to the data character introduced in section 2.3. We are only going to work with right coronary artery angiography in the following experiments because of the time limitation and complexity of left coronary artery tree. Module 2: coarse heatmap regression takes the right coronary artery angiography as input and outputs a coarse heatmaps showing all possible segment point locations based on local features. Module 4: feature map extraction takes the original cardiac angiography as input and outputs a feature map. The feature maps from Module 4 are concatenated with the coarse heatmaps output from Module 2 as input of Module 3: False positive suppression. The false positive responses on coarse heatmaps are then gradually suppressed in Module 3 by learning global features. Finally, a series of heatmaps with a

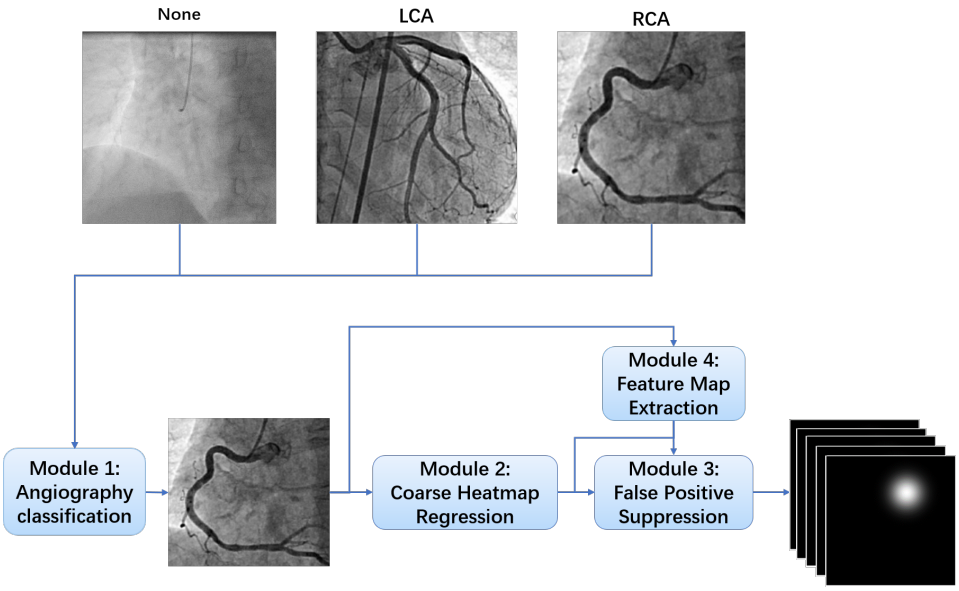


Figure 4.3: Overall scheme of right coronary artery annotation algorithm.

Gaussian blob showing the segment point location is produced by the algorithm scheme. More details of each module are introduced in following sections.

4.2 Module 1: Angiography classification

Convolutional Neural Network (CNN) is currently the state-of-art method for image classification. It performs very well in many image datasets, including MNIST, Cifar10, and Imagenet [Deng et al., 2009]. In medical image classification domain, CNN also achieves good results on interstitial lung disease image patches classification [Li et al., 2014] [Shin et al., 2016]. Therefore, we decide to use CNN to classify the angiographies.

As introduced in section 2.3, the data set we have is mixed by right coronary artery angiography, left coronary artery angiography, and angiography without vascular tree on it. The goal of angiography classification module is to train a CNN to classify the angiographies as 'RCA' for right coronary artery angiographies, 'LCA' for left coronary artery angiographies, and 'None' for background images without vascular tree. As it shows in Figure 4.4, our CNN design consists of three parts: convolution layer, max pooling layer, and fully connected layer.

The convolution layers extract the descriptive and discriminative features needed for classification task from training set, which can be presented mathematically as Function 4.1:

$$M_{x,y} = f(\sum_{i=x-n/2}^{x+n/2} \sum_{j=y-n/2}^{y+n/2} I_{i,j} * W_{i,j} + B_{x,y}) \tag{4.1}$$

$$f(x) = max(0,x) \tag{4.2}$$

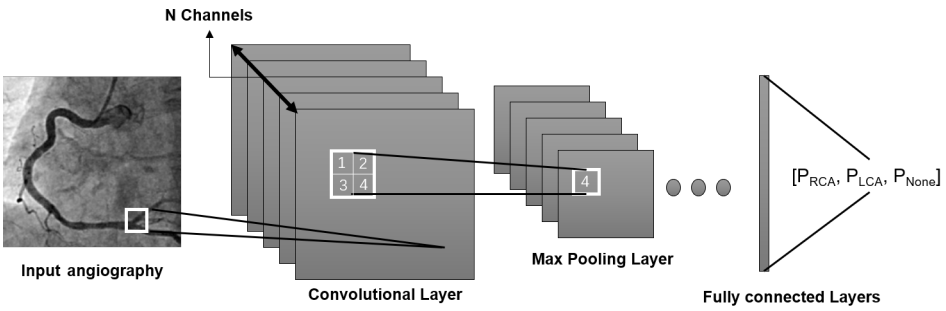


Figure 4.4: Structure of a Convolutional Neural Network

$I_{i,j}$ is the gray value at (i,j) on the angiography, W is a n*n weight matrix, which is used as the kernel parameters within a channel, and $B_{x,y}$ is a bias, both of them are updated during training. $M_{x,y}$ is the weighted sum of neighbor pixels around (x,y), which extracts the local features within a radius of n/2 pixels. The $M_{x,y}$ is then processed by Relu function, which activates only the positive values. Max pooling layer is another important component in the CNN, which downsamples the input and uses the maximum value from the neurons inside the sliding window at the prior layer as Figure 4.4. Pooling layer brings robustness on small pixel displacement and gradually reduce the spatial size of the image passing in the network, so that the amount of weight parameters and computation in the network are reduced and the risk of overfitting is controlled.

The convolutional layer and pooling layer are combined as one small module and repeated five times to generate a deeper architecture. Each module transforms the output of last layer into a higher and more abstract level representation. Thus, the CNN is generated as the complex function for cardiac angiography annotation [LeCun et al., 2015]. In the final layer of the network, two fully connected layers are applied, in which all the neurons are connected with neurons in next layer, so that the network can learn global features, instead of local features within the sliding window. The output of CNN is a 1*3 vector, which presents the probabilities of 'RCA', 'LCA', and 'None', and the cardiac angiography is classified as the class with highest probability.

The cross-entropy loss function in Equation 4.3 is used to calculate the difference between prediction and ground truth. 'x' represents the input angiography, p(x) is probability vector of angiography 'x' predicted by CNN, GT(x) is the ground truth vector of angiography 'x' corresponds to the probability vector p(x), the correct label in vector GT(x) is '1', which means angiography 'x' 100% belongs to this class. Loss(p,GT) is the cross entropy loss, which represents the difference between predicted probability vector p(x) and the ground truth label GT(x). The parameters in the network are updated iteratively by gradient decent optimization algorithm to minimize the cross-entropy loss.

$$Loss(p, GT) = -\sum_{i=0}^{2} GT_i(x) * log(p_i(x)) \tag{4.3}$$

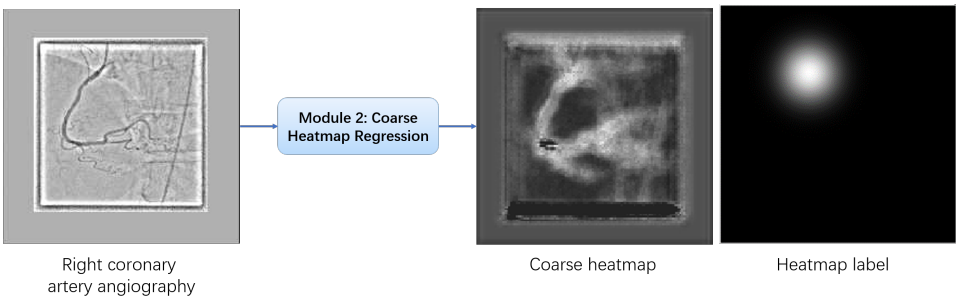


Figure 4.5: Input and output of module 2: coarse heatmap regression.

4.3 Module 2: Coarse heatmap regression

Module 2: coarse heatmap regression takes the right coronary artery angiography classified by Module 1 as input and produces a coarse heatmap for segment point as Figure 4.5. The good of this module is that all possible segment point locations including false positive responses are shown on the coarse heatmap, because it only makes use of local geometric features within a small patch to regress segment point location. Correct location of the segment point is less probably to be missed on such coarse heatmap. A coarse heatmap predicted by the network is shown in Figure 4.5, the coronary artery region is highlighted, which means the network locks it as interesting area, the location closes to the annotated segment point (see Heatmap label in Figure 4.5) is brighter, which means this location is more likely to be the segment point we are regressing.

In this section, we first introduced the generation of heatmap labels, which are used as training target of neural network, then the architecture of coarse heatmap regression neural network.

4.3.1 Generation of heatmap label for segment points

Heatmap label shows the segment point location by a Gaussian blob as Figure 4.6. The center of Gaussian blob locates at manually annotated segment point location, which is pixel with peak value on heatmap label. The value at each pixel decreases as it spatially getting further from the center of Gaussian blob, which means the probability of being the required segment point decreases.

The Gaussian blob is generated by function 4.4. P_{x_p,y_p} is the probability of being the required segment point at pixel P in the angiography, (x_p, y_p) is the coordinate of pixel P and (x_{GT}, y_{GT}) is the coordinate of manually annotated segment point location. The numerator of the right side of the function is the Euclidean distance between annotation and prediction, the denominator σ is the variance, which defines Gaussian blob's radius. $\frac{1}{2\pi\sigma^2}$ is used to normalized the value on the heatmap. The value at Gaussian center (annotated location) $P_{x_{GT},y_{GT}}$ is largest, which means (x_{GT}, y_{GT}) has largest probability of being the segment point.P_{x_p,y_p} decreases as the distance between P and annotated location increases, which

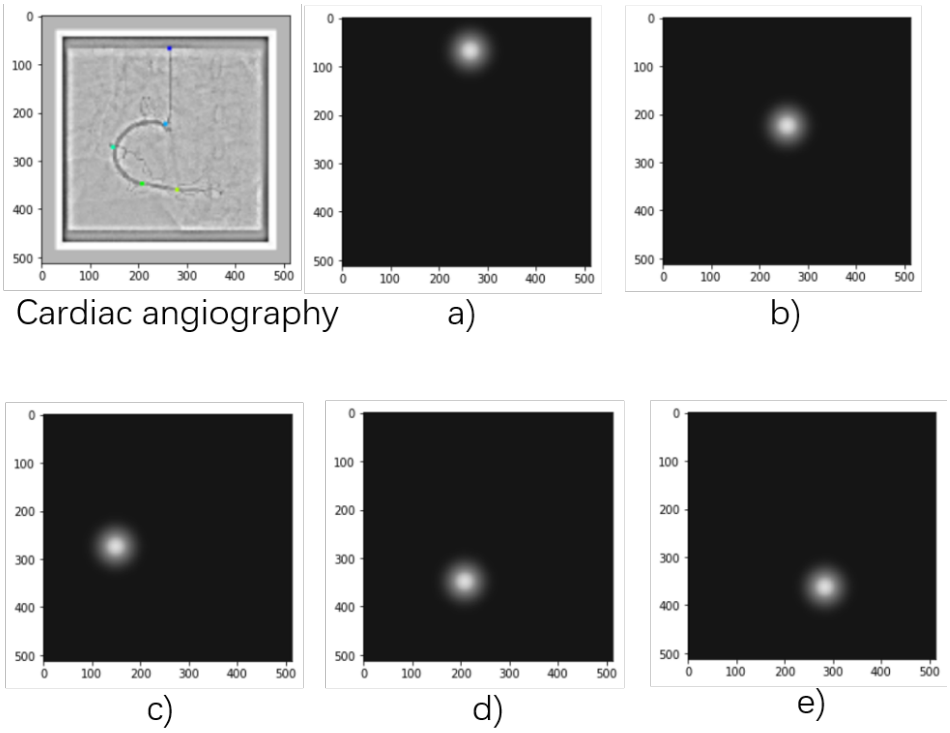


Figure 4.6: Example of a cardiac angiography and ideal heatmaps. a) heatmap of point 'a0'; b) heatmap of point 'b0'; c) heatmap of point 's1'; d) heatmap of point 's2'; e) heatmap of point 's3'

means probability of being segment point decreases as the pixel P spatially getting further to (x_{GT}, y_{GT}).

$$P_{x_p,y_p} = \frac{1}{2\pi\sigma^2} exp(-\frac{||(x_p,y_p)-(x_{GT},y_{GT})||_2^2}{\sigma^2}) \tag{4.4}$$

4.3.2 Coarse heatmap regression neural network

The network architecture for coarse heatmap regression is shown in Figure 4.7. There are two crucial design choices for this network:

1) Minimal kernel size for local feature extraction is used in convolutional layers.

In pose estimation methods, Pishchulin et al. [2016], Cao et al. [2017b], Bulat and Tzimiropoulos [2016], and Zhang et al. [2015] use the first five layers of VGG16 to regress joint heatmaps and it is proved to work better than other structures. VGG16 is a CNN originally used for nature image classification in The ImageNet Large Scale Visual Recognition Challenge (ILSVRC). It achieved 70.5% top-1 accuracy in 1000 classes classification.

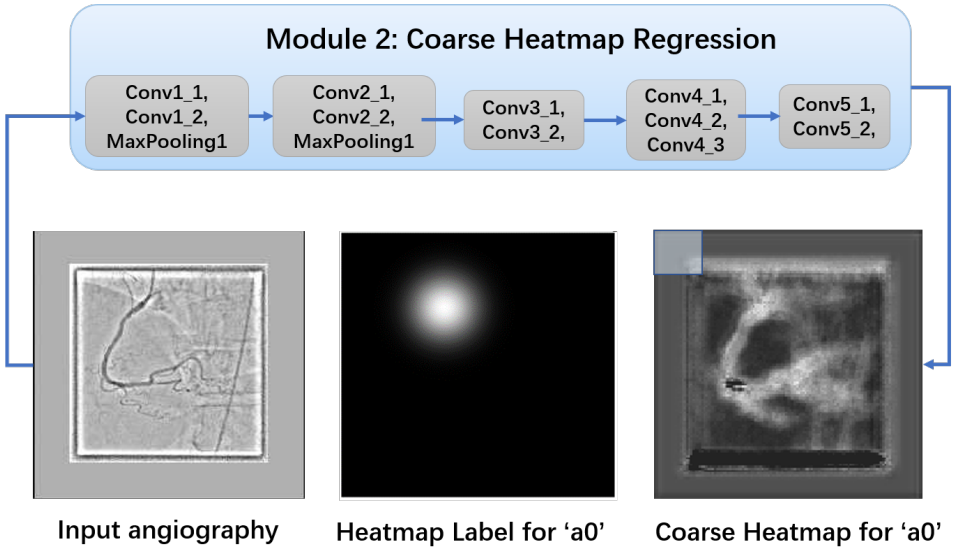


Figure 4.7: Structure of coarse heatmap regression network, the blue box is Module 2 and the grey boxes inside shows its inner structure. Input is the right coronary artery angiography and outputs are the coarse heatmap of segment point 'b0' predicted by different network structures after 10 epoch training. The transparent box on coarse heatmap indicates the patch size that the network extracted local feature from. The heatmap label shows the annotated location of segment point 'b0'.

An important reason for VGG16 to outperform other neural networks is using smallest kernel size for feature extraction on convolutional layer. Before VGG16, it was widely believed that the bigger kernel size in convolutional layer, the more information individual neuron can get. The convolution kernel size determines the region size on original input that each neuron is connected with, such region is called receptive field. As it shows in Figure 4.8 b), the convolutional layer with 5*5 kernel size enables one neuron to extract local feature from a patch of 5*5 pixels on input image, while the convolutional layer with 3*3 kernel size shown in Figure 4.8 a) only corresponds to a receptive field of 3*3 pixel. However, VGG 16 uses the smallest kernel for feature extraction (3*3 pixels) on convolutional layer but reaches a larger receptive filed by adding more layers. As it shows in Figure 4.8 c), the combination of two convolutional layers with 3*3 kernel can reach 5*5 receptive field as well.

The good of replacing 5*5 kernel by the combination of two 3*3 kernels is that less parameters are used. Two 3*3 kernels has 18 weights parameters, while a 5*5 kernel requires 25 weight parameters. Less parameters are easier to be updated during training. The deep structure also increases the fitting ability of network. Therefore, VGG16 has a better performance on both image classification and human pose estimation.

We also use the first five convolutional layers of VGG16 as network architecture in the network for coarse heatmap regression (see Figure 4.6), each convolutional layer is combined by two convolutional layers with 3*3 kernel.

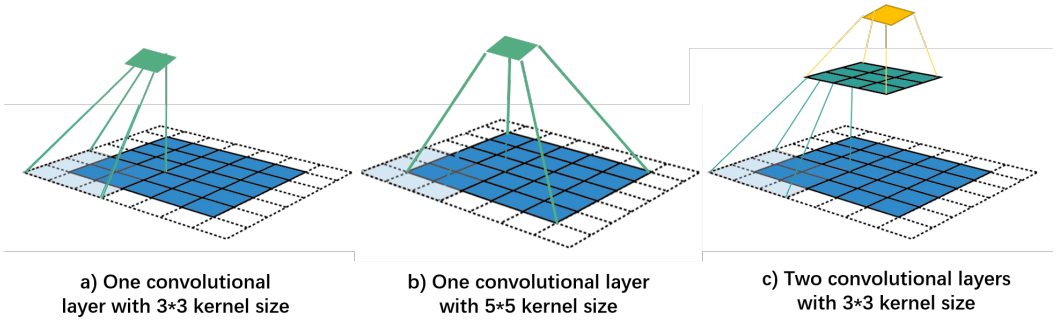


Figure 4.8: In a), the kernel used on convolutional layer converges 3*3 pixel area on the input image to one neuron in next layer, its receptive field is 3*3; in b), the kernel converges 5*5 pixel area on the input image to one neuron in next layer, its receptive field is 5*5; in c), two convolutional layers both have 3*3 kernel size. One neuron on yellow layer corresponds to 3*3 pixel area on green layer, 3*3 pixel area corresponds to 5*5 pixel area on the blue layer. Therefore, one neuron after two convolutional layer has a receptive of 5*5 pixels on the original input.

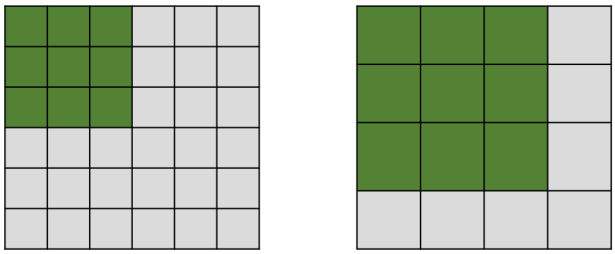


Figure 4.9: In a), the image has a resolution of 6*6, and 3*3 kernel converges 25% of the image into the receptive field; in b), the image has a resolution of 4*4, 56.25%(9/16) of the image is converged into the receptive field.

2) Another design choice is using minimal number of pooling layers in the network to keep the output resolution.

Except for the benefits of pooling layer introduced in section 4.2, pooling layer also accelerates converging larger image area into receptive field by down-sampling. As Figure 4.9 shows, the same kernel can converge larger proportion of the 4*4 image into receptive field compared with the 6*6 image. However, Payer et al. [2016] and Pfister et al. [2015] claim keeping output resolution is also very important for localization tasks, because it helps to remain detailed information on the image and the accuracy of predicted point location. In segment point regression task, we need to predict the segment point location as precise as possible. Therefore, we use two pooling layers in the network to balance the pros and cons. This choice will be further evaluated in Chapter 5.

4.4 False positive suppression

The coarse heatmap is predicted according to the local features within a limited receptive field as Figure 4.7. However, the similar local features around different segment points (see Figure 4.2) may confuse the network and lead to false positive responses on coarse heatmap. Therefore, we design Module 3: False positive suppression to learn global information in the angiography, for instance the hierarchy of coronary segments, the relative positions of segment points, distance between segment points and so on.

This module takes the coarse heatmap generated by previous network as input, and gradually removes the false positive responses by taking global features into consideration. Finally, this module aims at producing a heatmap that is the same as heatmap label. We first introduce the false positive suppression methods used by human pose estimation, in section 4.4.1 then our network architecture of false positive suppression module in 4.4.2.

4.4.1 False positive suppression in human pose estimation

False positive responses caused by similar local features exist in human pose estimation as well. Some human pose estimation methods use pre-defined energy functions, heuristic initialization of spatial probability priors, or dedicated network modules to suppress false positives [Pishchulin et al., 2016] [Tompson et al., 2015]. These methods select best location from the coarse heatmap according to pre-defined global features. However, it is hard to choose global features applicable to all kinds of cardiac angiographies and define them mathematically as an energy function.

Instead of predefined rules, the state-of-art pose estimation algorithm Wei et al. [2016] finds that neural network can learn implicit spatial dependencies between joints by enlarging receptive field. If the receptive field of the network covers entire input image, individual neuron at last layer will be indirectly connected with all pixels on the image, so that the neuron can organize global information on entire image. Learning global features by enlarging receptive field is not only simpler than manually defining global relationship between joints, but also achieves better result [Wei et al., 2016]. Therefore, we design a network in next section to let the receptive field cover entire right coronary artery tree, so that the network can automatically organize the global information on the vascular tree.

4.4.2 False positive suppression neural network

As we presented in section 4.3.2, both down-sampling the image by pooling layer and adding more convolutional layers can enlarge the receptive field. However, reducing image resolution will decrease the accuracy of predicted joint location. Therefore, we add more convolutional layers to enlarge the receptive field.

Though deep network structure is helpful to enlarge the receptive field, it will also bring the problem of vanishing gradient, which brings difficulties to back propagation and gradient descents. During training of a neural network, the parameters in the network are updated according to the partial derivative of the loss between training target and prediction with respect to current weight parameters in each training iteration. However, gradient used to

update parameters become vanishingly small as the network is getting deeper, and this will potentially stop network from further training. Relu activation function [Glorot et al., 2011] and batch normalization [Ioffe and Szegedy, 2015] are widely considered to be helpful on mitigate the impact of vanishing gradient, but the problem is not fully solved.

To further address vanishing gradient problem, we build multiple loss functions for intermediate supervision in 'Module 3: False positive suppression', the corresponding network structure is shown in Figure 4.10. The coarse heatmap from module 2 is used as input of three convolutional layers with 11*11 kernel, which enlarge the receptive field from 92*92 to 212*212 pixels (the original angiography has a resolution of 512*512 pixel). Then two convolutional layers with 1*1 kernel size are applied. The first convolutional layer with 1*1 kernel is used for cross-channel interaction and information integration [Szegedy et al., 2015]. The second one reduces output's dimensionality to the same as heatmap labels, so that we can build L2 loss function between its output and heatmap labels for intermediate supervision as equation 4.5. 'prediction' is the predicted heatmap and 'HeatmapLabel' is heatmap label used as training target in equation 4.5. The mean square error between them is calculated and summed in both dimensions first, then the mean square error of 5 heatmaps for different segment points are summed together.

The three convolutional layers with 11*11 kernel, two convolutional layers with 1*1 kernel and the L2 loss function built between output and heatmap label are combined as one submodule in Module 3. Such submodule is repeated three times (submodule 3.1, 3.2, 3.3) and the receptive field of the network achieves 452*452 at last layer as shows in Figure 4.10, entire right coronary artery on the angiography can be covered by the receptive field of the network.

$$Loss = \sum_{0}^{5} \sum_{i} \sum_{j} ||prediction(i,j) - HeatmapLabel(i,j)||^2 \tag{4.5}$$

From Figure 4.10, we can see the coarse heatmap outputted from module 2 remains the shape of right coronary artery tree to some extent and the start point of catheter 'a0' is correctly located on the heatmap. However, the predicted Gaussian blob on heatmap of Module 3.3 has a leftward offset from the annotated location on heatmap label, which initially appears on the output of Module 3.2. We assume this is because the information in original angiography is gradually diminished as the network getting deeper, which make it difficult for the network to predict correct location when mistakes have been made in previous layers. Therefore, we improved the network structure in next section.

4.5 Feature map extraction

As we presented in last section, it is hard to predict segment point location correctly if mistakes are made in previous layers. The information is diminished as the network getting deeper. Therefore, we design Module 4 to extract feature maps from the original angiography and periodically input the feature maps extracted by Module 4 to network in Module 3 to compensate information from the original angiography as it shows in Figure 4.11.

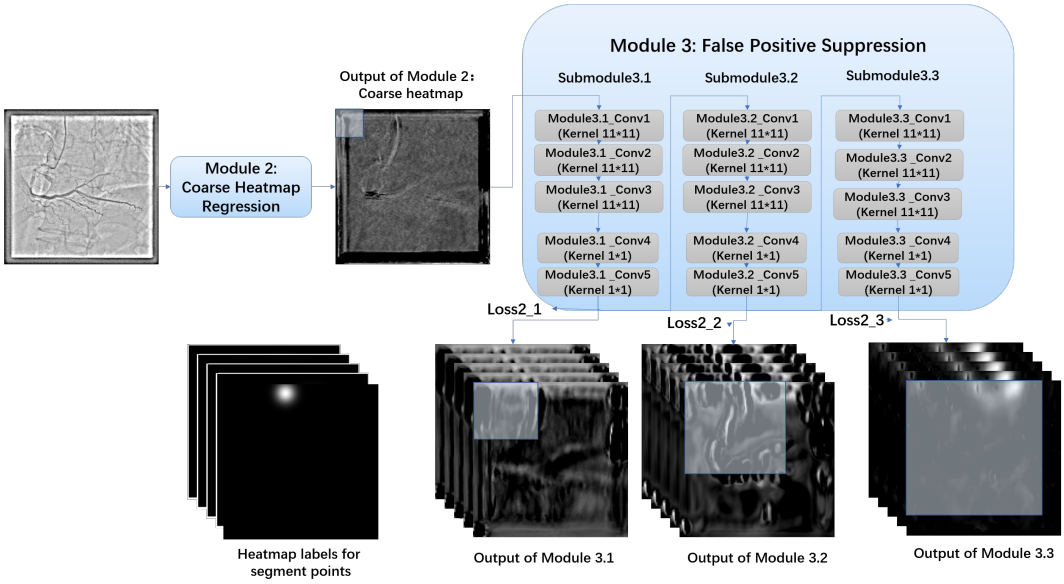


Figure 4.10: Network structure of false positives suppression.

The neural network for feature map extraction in Module 4 is the same as the neural network used for coarse heatmap regression shown in Figure 4.7, because of benefits of small kernel size and minimal pooling layers introduced in section 4.3. Figure 4.11 shows the overall network architecture for heatmap regression. The feature maps are concatenated by layer with the coarse heatmap and used as input for Module 3.1 (see Figure 4.12). Similarly, the feature maps are also concatenated with the output of Module 3.1 and 3.2 and used as input to next module. Module 4 has less layer compared with the place where the feature maps are inputted, therefore, the information from original angiography is less diminished compared with the output from previous submodule in Module 3. By combining them, we can compensate the information on original angiography to next submodule in Module 3.

The predicted heatmap of segment point 'a0' is shown as an example in Figure 4.11. The false positive responses are gradually removed correctly from the predicted heatmap as the network getting deeper, and the overall output is also less noisy compared with when Module 4 is not included in the algorithm in Figure 4.10.

As we described in this chapter, the algorithm scheme consists of 4 modules. Module 1 is used to classify cardiac angiographies into left coronary artery angiography, right angiography, and angiography without vascular tree. Module 2, Module 3, and Module 4 are used for segment point heatmap regression. Module 1 generates a label for each angiography, while segment point regression networks generate five heatmaps showing the corresponding predicted segment point location by a Gaussian blob on each heatmap, Figure 4.13 shows the predicted heatmaps of one angiography in training set. Given the different task and training target, we train Module 1 separately, and then train segment point regression

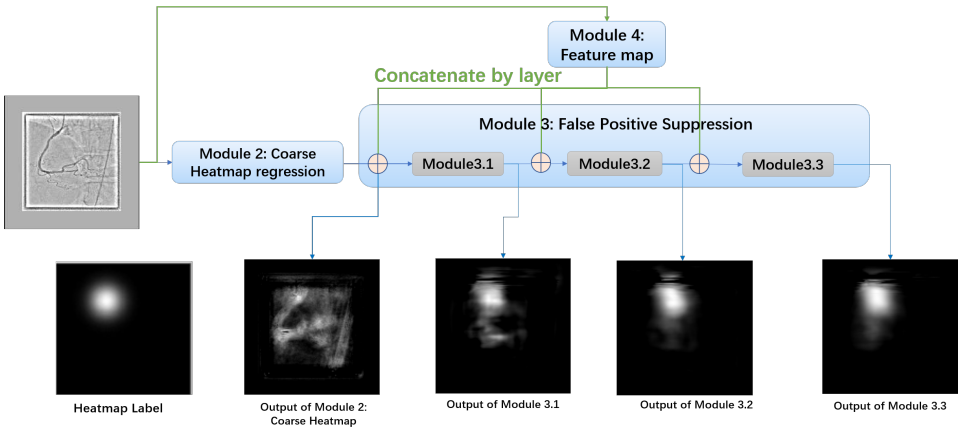


Figure 4.11: The image shows the overall network architecture for segment point location regression, and the location where feature maps extracted by Module 4 are inputted to the network. For the detailed inner structure of Module 2 and Module 4 please reference Figure 4.7, detailed inner structure of Module 3 including module 3.1, 3.2, 3.3 please reference Figure 4.10.

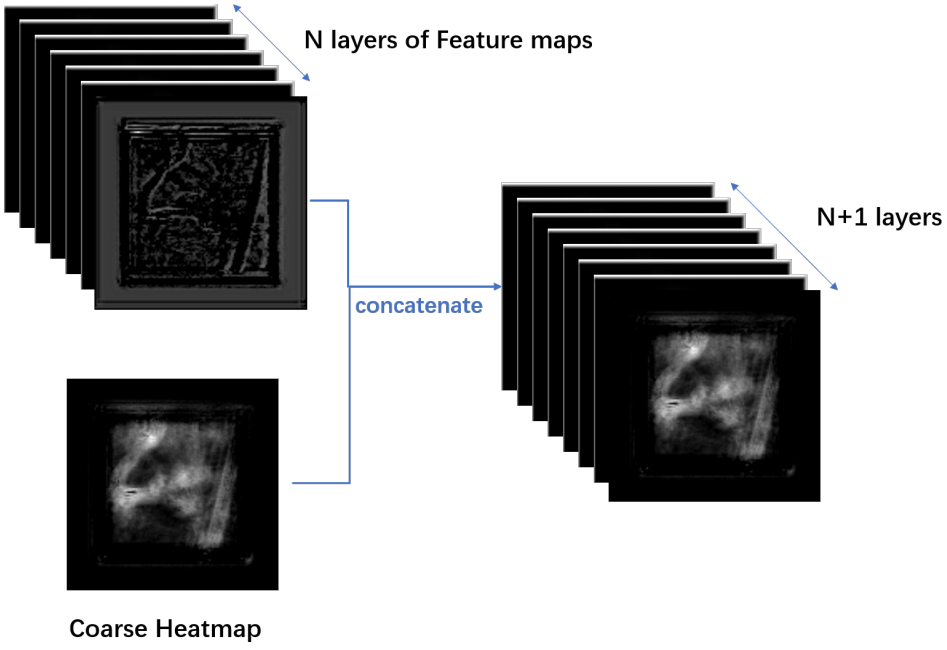


Figure 4.12: This image shows how the coarse heatmap is concatenate with feature maps.

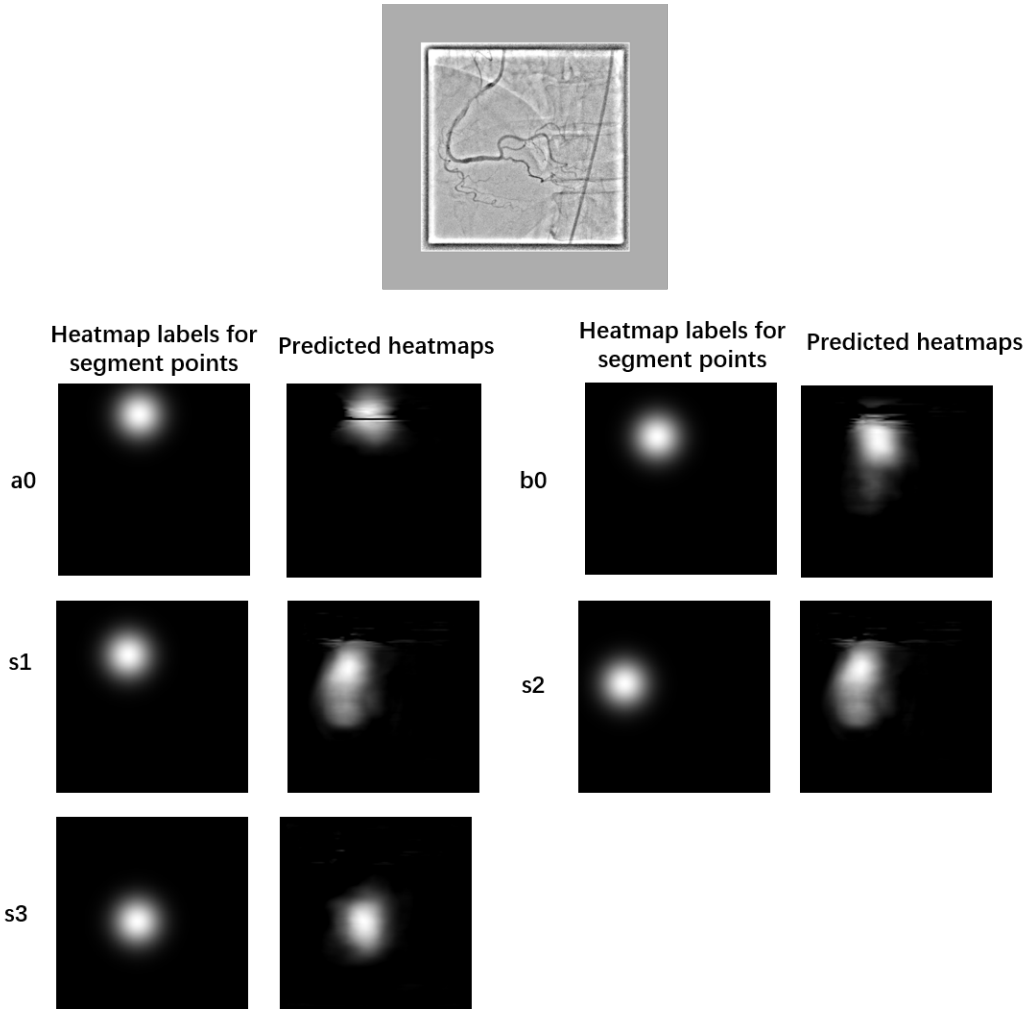


Figure 4.13: The angiography on the top is one example in training set. On the right are the segment point heatmaps generated by segment point regression networks, on the left are the corresponding heatmap labels.

networks in an end-to-end manner, which means each angiography in training set is passed through Module 2, Module 3, and Module 4 one by one each training iteration, the losses at different stages are summed together and minimized during training. The results of Module 1 and segment point heatmap regression networks are presented and evaluated in next Chapter.

Chapter 5

Evaluation

In chapter 4, we introduce the proposed algorithm scheme and visualize some predicted heatmap examples in training set at different stages of the network. In this chapter, we evaluate algorithm performance more comprehensively and standardizedly according to quantified accuracy metrics.

Module 1 for angiography classification and segment point heatmap regression networks consisting of Module 2, Module 3, and Module 4 are evaluated separately, because they are trained separately, and their output formats are different. Angiography classification module aims at classifying the angiography into correct class and outputs a label, while segment point heatmap regression networks aim at predicting segment point locations and output heatmaps representing the corresponding segment point location by Gaussian blob. In section 5.1, we evaluate the angiography classification network performance and present results. In section 5.2, we evaluate the segment point heatmap regression networks and present the annotated cardiac angiographies.

5.1 Evaluation of cardiac angiography classification

In this section, we introduced the accuracy metric used for angiography classification evaluation in section 5.1.1. In 5.1.2, we evaluate the performance of Module 1 network during training accoring to the accuracy metric introduced in 5.1.1 In 5.1.3, we evaluate the performance of trained module 1 network.

5.1.1 Accuracy metric for angiography classification network

The performance of Module 1 network is evaluated according to the metric called 'top-1 accuracy', which is one of the standard accuracy metrics in image classification task. As we introduced in section 4.2, the output of classification network is a vector representing the probability that the angiography belongs to 'RCA', 'LCA', and 'None'. Top-1 accuracy metric regards the angiography is correctly classified only when the probability of correct label ranks 1^{st} in the vector predicted by Module 1 network. Similarly, there are also top-n accuracy metric, which means the angiography is regarded as correctly classified when

the probability of the correct label ranks top n in the predicted vector. We have only three classes for angiography in total, therefore, we use top-1 accuracy metric.

Top-1 accuracy is calculated by the proportion of correctly classified angiographies in entire data set as Equation 5.1. 'acc' represents the accuracy of angiography classification task, $n_{correct}$ is the number of correctly classified angiographies, $N_{angiography}$ is the total number of angiographies. We evaluate the performance of Module 1 network by top-1 accuracy.

$$acc = \frac{n_{correct}}{N_{angiography}} \tag{5.1}$$

5.1.2 Evaluation of angiography classification network performance during training

The Module 1 network is trained iteratively by 4581 annotated cardiac angiographies called training set. In the process of training, we periodically pause the training, and test network by 693 cardiac angiographies other than training set, which is called validation set. When the misclassified angiographies in validation set increases continuously, it means the network has overfitted the training set and we should stop training in order to prevent overfitting.

The Module 1 network performance is evaluated according the top-1 accuracy metric. The performance during training is represented by the top-1 accuracy of both training set and validation set. Training set accuracy shows model's fitting ability on training set. Training set accuracy increases along with training iteration means the network is updating the parameters to fit the training set, in other words, the network is learning how to classify angiographies. Validation set provides an unbiased evaluation of network performance during training. The accuracy of training set represents network's ability to fit the training data, the accuracy of validation set represents network 's ability to generalize from the training set. Evaluating the network during training is very important because it helps us understand how well is the network trained. Achieving good performance on training set and validation set is also the precondition of trained network can classify other cardiac angiographies correctly.

We adjust the hyperparameter of Module 1 network including training epochs (one epoch means one pass of the full training set) to let the accuracies of both training set and validation set converge at a decent level. Graph in Figure 5.1 shows the network performance during training. The top-1 accuracy of training set represented by orange line increases steadily along with training iterations and achieves 90% when we stop training. It means the Module 1 network can fit training set well. The accuracy of validation set also increases along with training iterations and finally achieves 85%. It means generalization ability of Module 1 network also increases and achieves a reasonable level.

5.1.3 Evaluation of trained angiography classification network performance

The accuracy on training set and validation set converges at around 90% and 85% respectively, which means Module 1 network has already fitted both training and validation sets

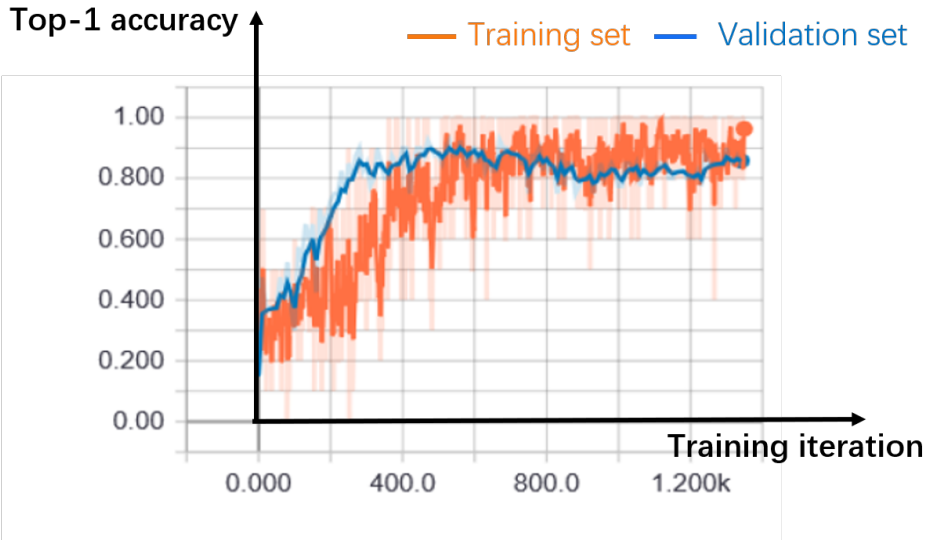


Figure 5.1: The graph shows the top-1 cardiac angiography classification accuracy of training set and validation set increasing along with training iterations.

well, then the network is saved as trained Module 1 network for cardiac angiography classification. We test the trained Module 1 network by 1348 cardiac angiographies other than training and validation set (test set) to evaluate the algorithm performance unbiasedly. The accuracy on test set achieves around 80%.

To understand the reason of those 20% misclassification better, we summarize some misclassified angiographies. Figure 5.2 shows two kinds of typical misclassified cardiac angiographies:

1) The Module 1 network misclassifies 'None' as 'LCA' or 'RCA' when there are tubular medical instruments on the background. Figure 5.2 a) shows an example of background angiography misclassified as 'LCA' by Module 1 network;

2) Module 1 network misclassified 'RCA' and 'LCA' angiographies as 'None', which means the network does not find any vascular tree on the angiography though there actually is. b) and c) in Figure 5.2 are 'LCA' angiography and 'RCA' angiography respectively, but both are misclassified as 'None'. We assume this is because the vascular structure too thin to be visualized on the noisy background.

5.2 Evaluation of cardiac angiography annotation network

In this section, we evaluate the performance of segment point heatmap regression networks. In section 5.2.1, we introduced the accuracy metric used for evaluation. In section 5.2.2, we evaluate the performance of segment point heatmap regression networks during training. In section 5.2.3, we give an unbiased evaluation to the trained segment point heatmap regression networks. In section 5.2.4, we evaluate the importance of each module in the

a) b) c)

Figure 5.2: Cardiac angiographies misclassified by the neural network according to top-1 accuracy metric, which means the predicted label with highest probability is different from annotated label.

segment point heatmap regression networks respectively. In section 5.2.5, we evaluate the design choice of using minimal number of pooling layers in Module 2 for coarse heatmap regression. In section 5.2.6, we evaluate the choice of three submodules in Module 3.

5.2.1 Accuracy metric for cardiac angiography annotation

In human pose estimation studies, heatmap regression is also used to predict the joint locations [Wei et al., 2016] [Pfister et al., 2015] [Bulat and Tzimiropoulos, 2016] [Payer et al., 2016]. They evaluate the performance of heatmap regression by first selecting the best joint location from predicted heatmaps, then calculating the Euclidean distance between predicted and annotated joint location.

In this thesis, we also evaluate the performance of segment point heatmap regression networks by the Euclidean distance between predicted and annotated segment point location, but we improve the method to extract segment point location from predicted heatmap. In this section , we first introduce the method to extract segment point location from predicted heatmap, then the accuracy metric for segment point heatmap regression networks performance.

Segment coordinate selection

Human pose estimation based on heatmap regression [Wei et al., 2016][Ge et al., 2018][Oberweger et al., 2015] selects the location with maximum value on predicted heatmap as predicted joint location, because the values on predicted heatmap represents the probability of being the required joint. This method is straight forward and easy to understand. However, the values of noise on predicted heatmap are possibly larger that the predicted segment point location as it shows in Figure 5.3 and 5.4.

In Figure 5.3, we can see that the Gaussian blob representing segment point ‘s3’ is generated on the predicted heatmap, but it is not the maximum on it. Noise caused by the gradient showing by red arrow on original angiography has larger value than Gaussian

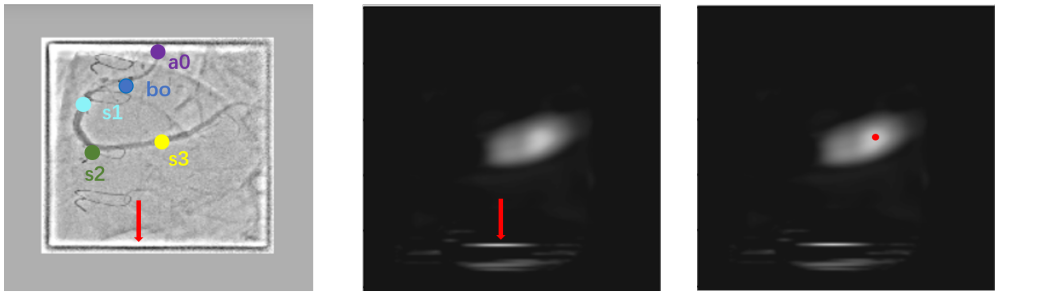


Figure 5.3: This image shows an example that maximum on predicted heatmap does not represent the predicted segment point location. a) is the input cardiac angiography, it has a strong gradient at the edge showing by the red arrow; b) is the predicted heatmap for segment point 's3', it responses to gradient on cardiac angiography and forms a strong noise at corresponding location, the red arrow shows the maximum on predicted heatmap, which is not the predicted segment point location representing by a Gaussian blob; on c), the red point is the segment point location chosen by our proposed method, it successfully find the predicted segment point location on heatmap representing by Gaussian blob.

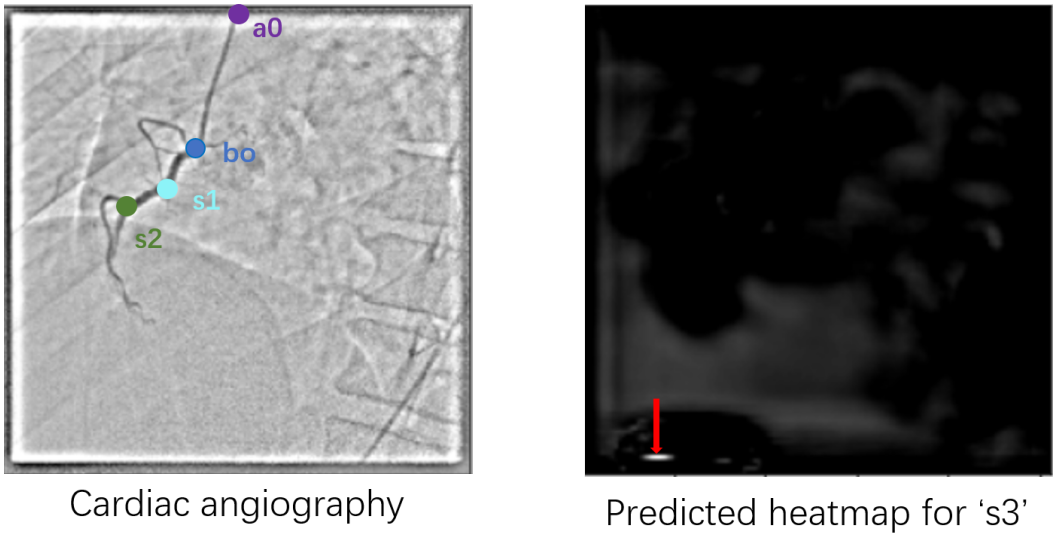


Figure 5.4: On the left, the angiography shows a right coronary artery without segment point 's3'; on the right, it is the predicted heatmap for 's3', there is no Gaussian blob on it, and the maximum on predicted heatmap is showing by the red arrow.

blob for 's3' on the predicted heatmap. In this case, simply selecting maximum will fail to correctly extract predicted segment point location, which is not the maximum on predicted heatmap.

The angiography in Figure 5.4 shows a right coronary artery tree with left dominant circulation, which does not have right posterior descending artery and right posterolateral branch, therefore, segment point 's3' does not exist in the angiography. In this case, the network should ideally predict 's3' heatmap as a map of 0, which is a black heatmap without Gaussian blob on it. However, the black heatmap is generally imperfect because it is unavoidable to have some noise on it as Figure 5.4. If we simply select the maximum from predicted heatmap, the noise will be predicted as segment point location by mistake, instead of return null for segment point location.

The segment point location is represented by a Gaussian blob on heatmap labels (training target), the predicted heatmaps are also supposed to show the predicted segment point location by Gaussian blob instead of impulse responses. Therefore, we decide to process the predicted heatmap by a Gaussian filter to suppress the noise as Figure 5.3 and 5.4. The Gaussian filter is generated using equation 5.2, which is same as equation 4.4 used to generate heatmap label (see section 4.2). (x_{center}, y_{center}) represents the center of Gaussian filter. The Gaussian filter is normalized by $\frac{1}{2\pi\sigma^2}$ to make the sum of elements in the Gaussian filter equals to 1, so that we can avoid amplifying or diminishing the value on the predicted heatmap. The predicted heatmap is then convolved with the Gaussian filter (the procedure of convolution is introduced in section 4.2), the filter weighted averages every pixel on the predicted heatmap with its neighbors, so that the Gaussian blob structures are remained on the processed heatmap, while the impulse noise is suppressed because the value on neighbor pixels of the impulse noise are low. Figure 5.5 shows the heatmap before and after processed by Gaussian filter, we can see that the noise caused by the gradient at the edge of the angiography is suppressed but the Gaussian blob is remained.

Though the noise can be suppressed by Gaussian filter, we still cannot return null as segment point location for black heatmap with noise. Because processed black heatmap is also not a map of '0'. Therefore, we use non-maximum suppression method [Neubeck and Gool, 2006] to first select a series of local maximums on the processed heatmap, then select segment point location by argmax from the local maximums. When there is a black heatmap, no local maximums can be detected by non-maximum suppression because values on it are all tiny, and thereby return null as segment point location for black heatmaps.

We use the method introduced above to select segment point location from the predicted heatmap in Figure 5.3 and 5.4, in which maximum on predicted heatmap does not represent the predicted segment point location. As it shows in Figure 5.3 c), the center of Gaussian blob is successfully selected as segment point location. The method also successfully returns null as segment point location for the black heatmap in Figure 5.4.

$$P_{x_p,y_p} = \frac{1}{2\pi\sigma^2} exp(-\frac{||(x_p,y_p)-(x_{center},y_{center})||_2^2}{\sigma^2}) \tag{5.2}$$

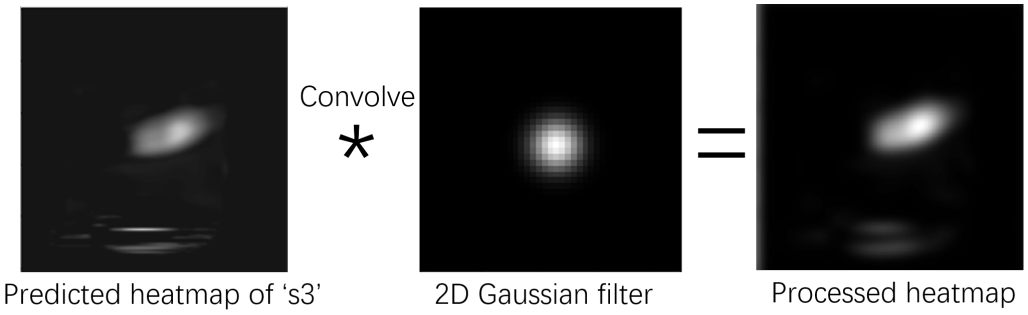


Figure 5.5: This image shows the effect on predicted heatmap after convolved with a 2D Gaussian kernel.

Accuracy of cardiac angiography annotation

The accuracy of predicted segment point heatmap is represented by the Euclidean distance between predicted and annotated segment point locations as Figure 5.6. The Euclidean distance shows how different is the predicted location and annotated location. The smaller Euclidean distance is, the more precise predicted segment point location is.

The Euclidean distance is calculated by equation 5.3. 'Edistance' is the Euclidean distance between annotated and predicted segment point location, (x_{GT}, y_{GT}) is the annotated coordinate of segment point, (x_s, y_s) is the segment point predicted by segment point heatmap regression networks.

$$Edistance = ||(x_{GT}, y_{GT}) - (x_s, y_s)|| \tag{5.3}$$

5.2.2 Evaluation of cardiac angiography annotation network performance during training

In this section we evaluate the performance of segment point heatmap regression networks during training. The performance is represented by the accuracy of training set (consisting of 145 angiographies from 100 different patients) and validation set (consisting of 10 angiographies from 3 different patients).

The full training set is passed through the network 10 times, in other words, the network is trained in 10 epochs. The accuracy of predicted segment point locations during 10 epochs training is shown in the graphs on the left side of Figure 5.7. The curve shows the Euclidean distance between predicted and annotated segment point location of the training set. The distance between predicted and annotated location (y-axis) decreases as the training iteration (x-axis) increases, which means the predicted segment points location gradually approaches annotated location and the accuracy on training set is increasing, segment point heatmap regression networks' ability to fit the training set is increasing.

On the right side of Figure 5.8 are the accuracy curve of validation set during training. The Euclidean distance between predited and annotated location on validation set also de-

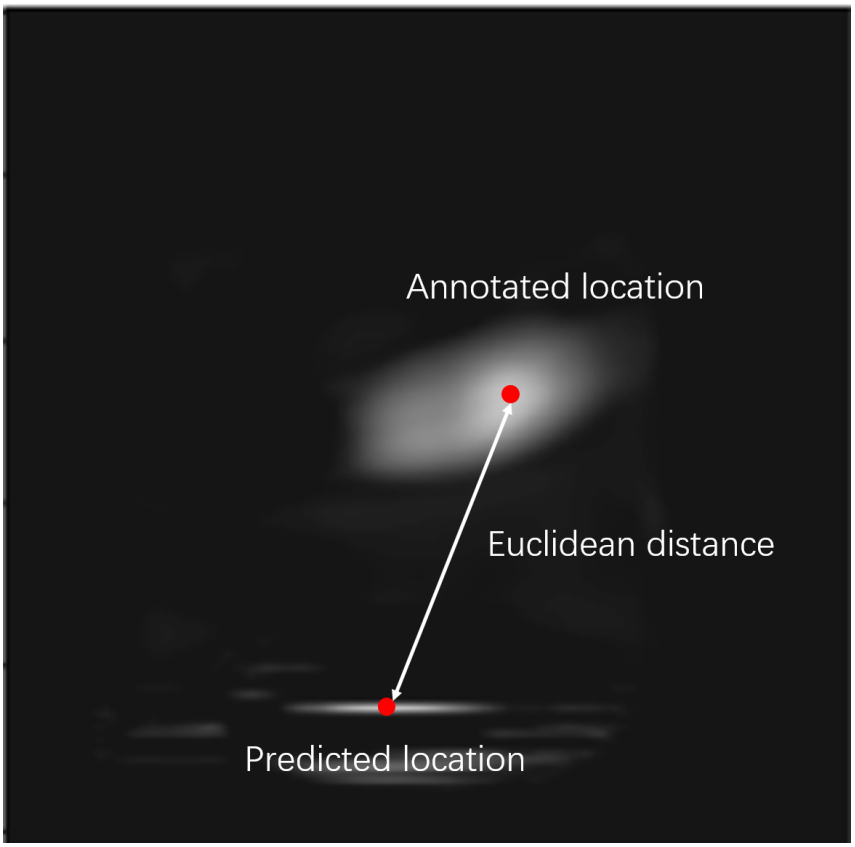


Figure 5.6: The white line linking predicted segment point location and annotated segment location shows the Euclidean distance, it is used as accuracy metric to evaluate the performance of predicted segment point location.

creases along with training iterations, which means the segment point heatmap regression networks' ability to generalize from training set is also increasing.

In figure 5.8, we put the predicted and annotated segment point locations on original angiography, then connect the predicted segment points in order to build the graph of right coronary artery. On the left side, the graphs are built according to the predicted segment point location. On the right side, the graphs are built according to the annotated location. The graph lines in 'violet', 'blue', 'aquamarine', and 'green' represent for 'catheter', 'proximal right coronary artery', 'mid right coronary artery', and 'distal coronary artery' respectively, the blood vessels after segment point 's3' are right posterior descending artery and right posterolateral branch.

The examples shown in Figure 5.8 include the difficulties of the angiographic annotations we introduced in Section 1.1: 1) angiographies a) and b) in Figure 5.8 are acquired from left anterior oblique and right anterior oblique respectively, which deforms the shape of coronary artery tree; 2) in c), the side branches around 's1' cross the main artery, which might be confused with bifurcations; 3) the side branches of right coronary artery trees in

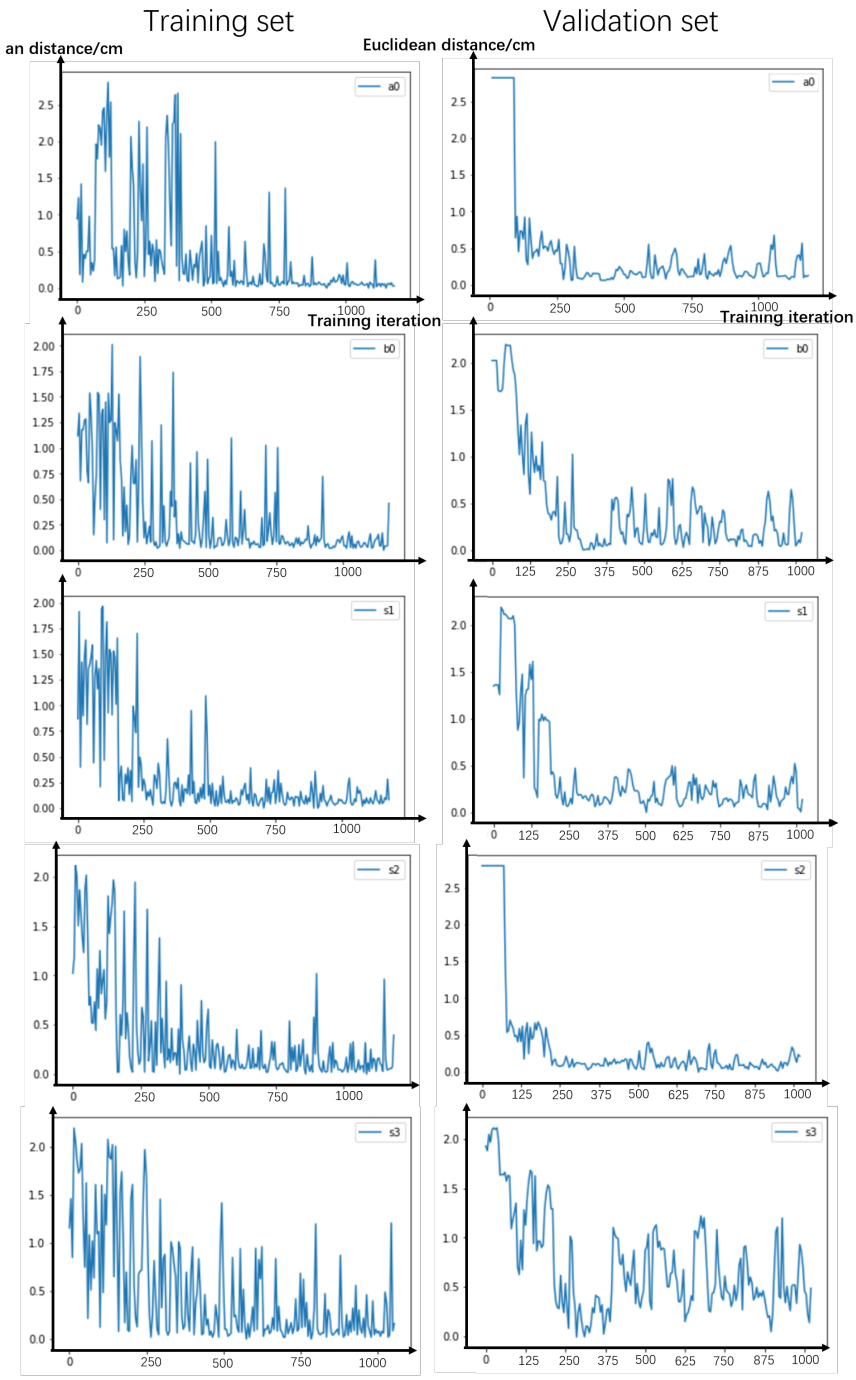


Figure 5.7: Accuracy curve shows the Euclidean distance between predicted segment point and manually annotated segment point. X-axis shows the training iterations, Y-axis shows the Euclidean distance in centimeters on the angiography.

da) are missing; 4) on the background of angiography c) and d), there are tubular structure catheters and strong gradient caused by the shadow of myocardium, which might also be confused with blood vessels. In these cases of training set, the segment point locations predicted by segment point heatmap regression networks are very close to the annotated locations. In Figure 5.8 b) and d), though there are offsets between predicted and annotated locations of segment point 's1', the offsets are along the blood vessel and within a reasonable range. The graphs build by predicted segment point locations can represent the structure of right coronary artery well and it is very similar to the graph built from annotated locations.

According to the accuracy curves of training set and validation set and the annotated examples shown in Figure 5.8, the network has learned how to annotate cardiac angiography in training set, then the networks are saved as trained segment point heatmap regression networks, which are further evaluated by test set in next section.

5.2.3 Evaluation of trained cardiac angiography annotation network

In order to give an unbiased evaluation on the segment point heatmap regression networks, we evaluate the trained network by a test set consisting of 46 cardiac angiographies, which are not included in training set and validation set.

The Euclidean distance between the predicted and annotated segment point locations in test set is shown by the box plot in Figure 5.9. y-axis represents the Euclidean distance in centimeters, and boxes from left to right represent for segment point 'a0', 'b0', 's1', 's2', 's3'respectively. We can see that predicted segment point locations are in general very close to the annotated location, but there are also some outliers, which means the proposed cardiac angiography annotation method cannot predict correct segment location in such case.

In order to understand the result better, we visualize and analyze some examples in test set annotated by trained segment point heatmap regression network. Figure 5.10 shows some correctly annotated cardiac angiographies in test set. In a), b), and c), there are strong gradient caused by the myocardium shadow and tubular shape catheter. In b) and d), some side branches are not visible on the angiography. In c), the catheter has very strange shape and curvature. The segment points are correctly annotated by trained network in these cases though the difficulties, which means that the trained network has the ability to generalize from training set.

Figure 5.11 shows right coronary artery with left dominant circulation. The segment point heatmap regression networks successfully predict that there is no 's3' on it by generating a black heatmap without Gaussian blob, but the precision of other segment points is influenced.

There are also cases that the network fails to correctly predict the segment location as it shows in Figure 5.12 and Figure 5.13. In Figure 5.12 a), we can see that segment point 's2' is influenced by the gradient caused by myocardium shadow, which crosses the main artery. The trained network confuses it with bifurcations on main artery. In b) and c) of Figure 5.12, segment point 'a0' are confused with other catheters start point on background, which are not connected with the vascular tree. However, angiographies with same problems in training set are correctly annotated (see Figure 5.8 a) and c)), part of angiographies with

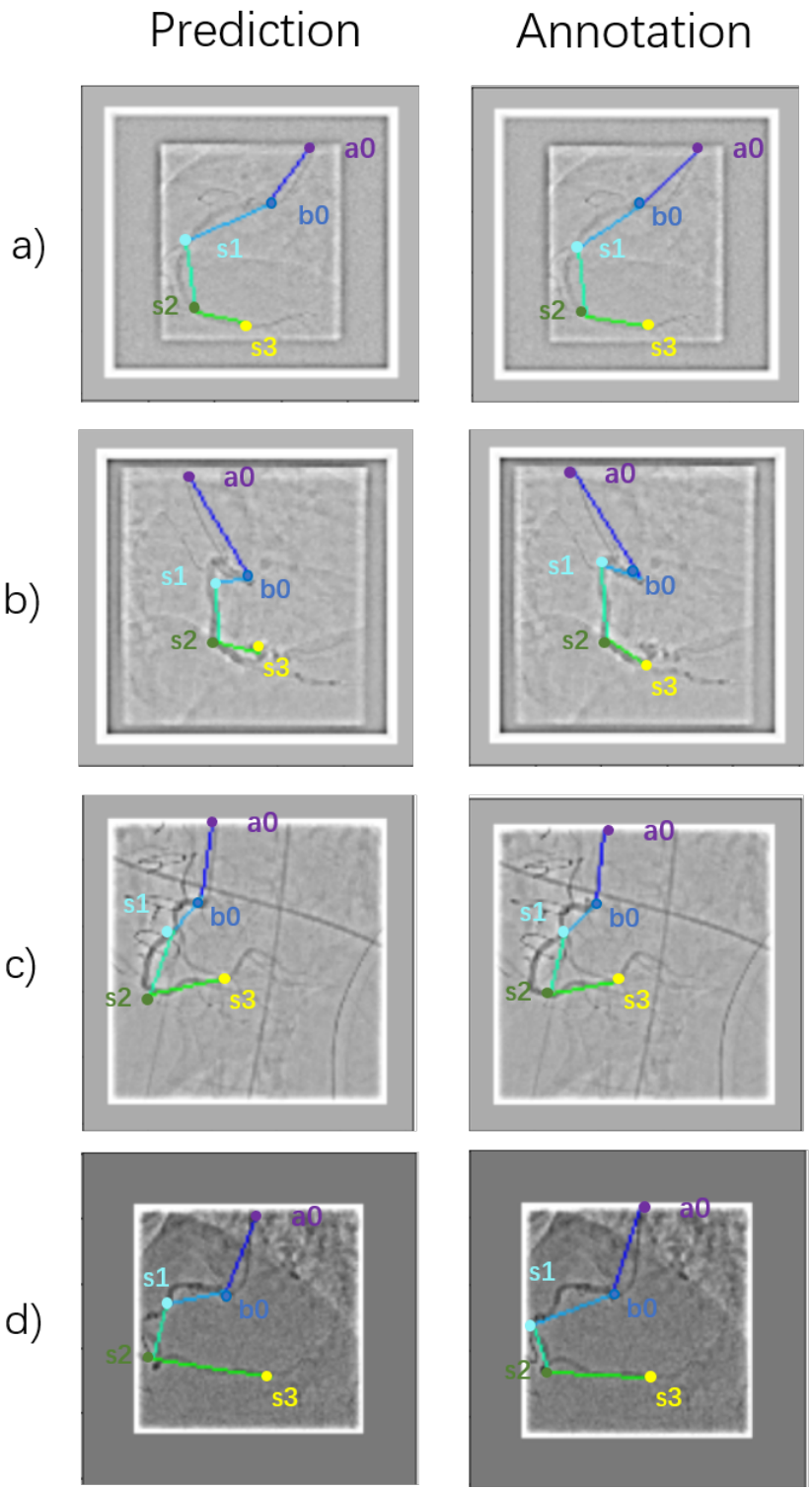


Figure 5.8: Some annotated right coronary artery angiographies from training set. The graphs on left side are drawn according to the predicted segment point locations, the graphs on right side are drawn according to annotated segment point locations.

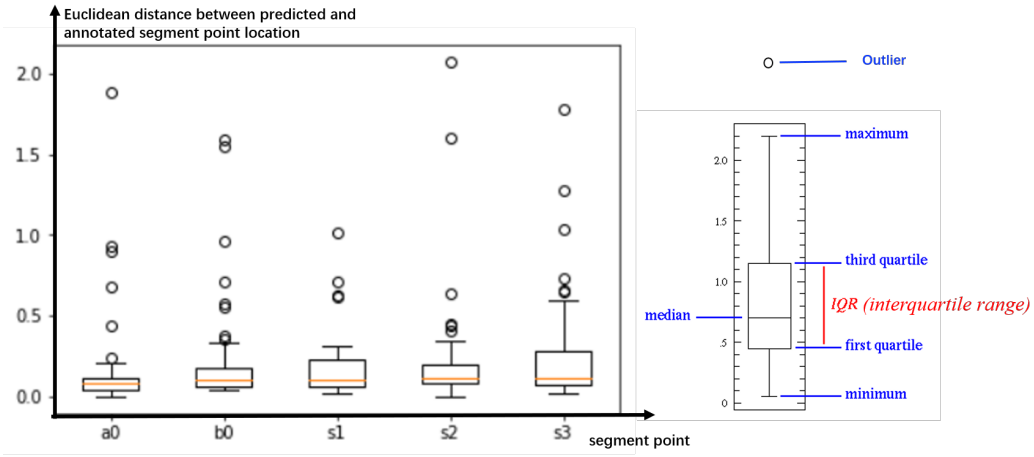


Figure 5.9: On the left is the box plot of Euclidean distance (centimeter) between predicted and annotated segment point locations. The boxes from left to right represent for segment point 'a0', 'b0', 's1', 's2', 's3'respectively. On the right is a box showing the meaning of different part of the box, the outliers are defied as 1.5 times IQR or more above the third quartile or 1.5 times IQR or more below the first quartile.

same problems in test set are also correctly annotated (see Figure 5.11a) and d)). We assume the unstable performance of trained network is because of the limited training set, in which there are merely 14 out of 145 cardiac angiographies has multiple catheters as b) and c) in Figure 5.12.

The errors in Figure 5.13 are more difficult to be understood compared with Figure 5.12. We assume such errors are because the gradient between blood vessels and background is too weak to distinguish them, but the assumption needs to be further justified by more test data.

5.2.4 Evaluation of different modules in segment point heatmap regression networks

In previous section, we evaluate the trained segment point heatmap regression networks consisting of Module 2, Module 3, and Module 4. In this section, we further evaluate the importance of each module in segment point heatmap regression networks by test set.

As we introduced in section 4.4, Module 3 enables the network to learn global features and suppresses the false positive responses on coarse heatmap. In order to evaluate the importance of Module 3, we train Module 2 by the same training set used to train the entire network, and then evaluate the trained Module 2 by same test set used to test the entire network. Module 2 predicts segment points only according to local features, by comparing the accuracy of Module 2 and entire network on the same test set, we can see whether the global features learned by Module 3 improve segment point heatmap regression networks performance. The accuracy of Module 2 on test set is shown by the box plot in Figure

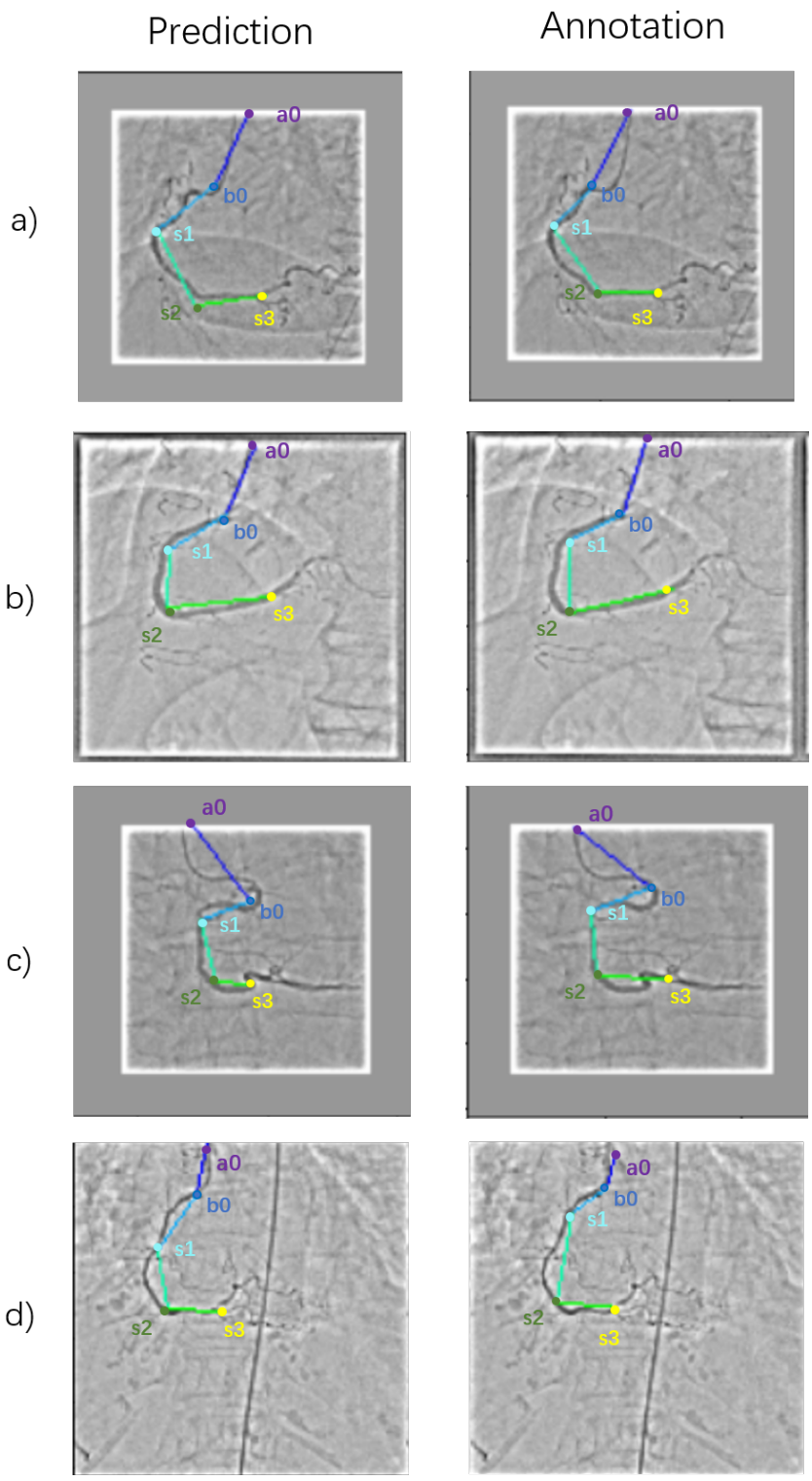


Figure 5.10: Annotated cardiac angiographies in test set. On the left side, the coronary artery graphs are built according to the predicted segment point locations. On the right side, the coronary artery graphs are built according to the annotated segment point locations.

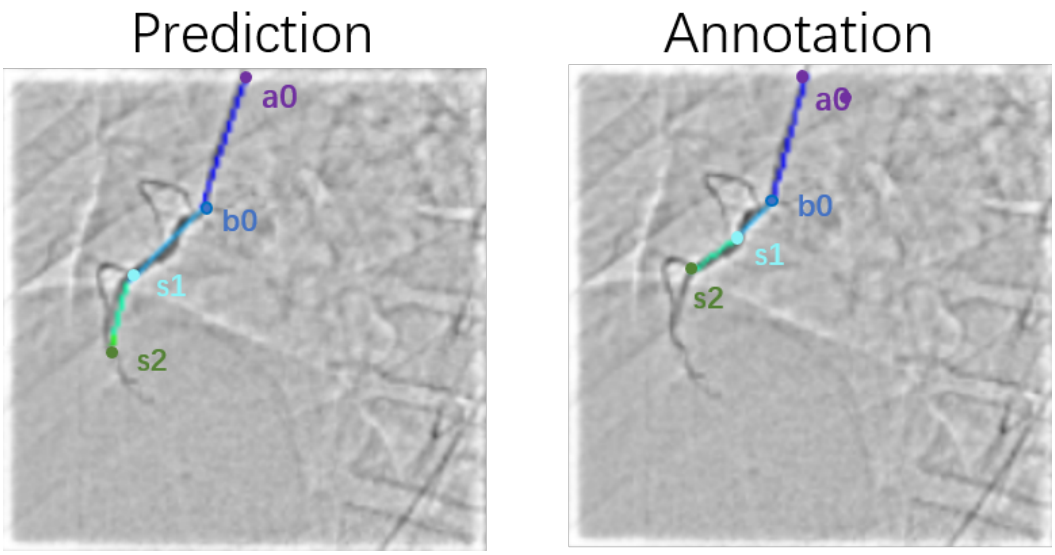


Figure 5.11: This cardiac angiography shows a right coronary artery with left dominant circulation, it does not have segment point 's3'.

5.14 b). For all segment points, Euclidean distance between the prediction of Module 2 and annotated segment point location is larger than Euclidean distance between prediction of entire network and annotated location shown in Figure 5.14 a). Therefore, we think learning global features by Module 3 can promote network performance.

As we introduced in section 4.5, the information on the original angiographies is diminished as the network getting deeper. Module 4 can help to remain information from original angiographies, so that the network can suppress the false positive responses correctly. In order to evaluate the importance of Module 4, we trained the network without Module 4, the network without Module 4 is shown in Figure 5.13 c). By comparing the accuracy of the network with and without Module 4 on the same test set, we can see whether Module 4 improves algorithm performance. According to the box plots in a) and c) of Figure 5.13, it is quite clear the that the performance of network without Module 4 is worse than the entire network. Module 4 can promote the performance of proposed cardiac angiography annotation method.

5.2.5 Evaluation of using minimal pooling layer in Module 2

In section 4.2, we make a design choice of using two pooling layers to remain the output resolution. We evaluate this design choice by comparing the performance of Module 2 network with two and three pooling layers, which are the most common choices in human pose estimation studies [Pfister et al., 2015][Wei et al., 2016]. The Module 2 network with two and three pooling layers are shown in Figure 5.15 and 5.16 respectively, their performance on test set is show below by a box plot. We can see that the predicted segment point location predicted by Module 2 with three pooling layers is less precise then the other

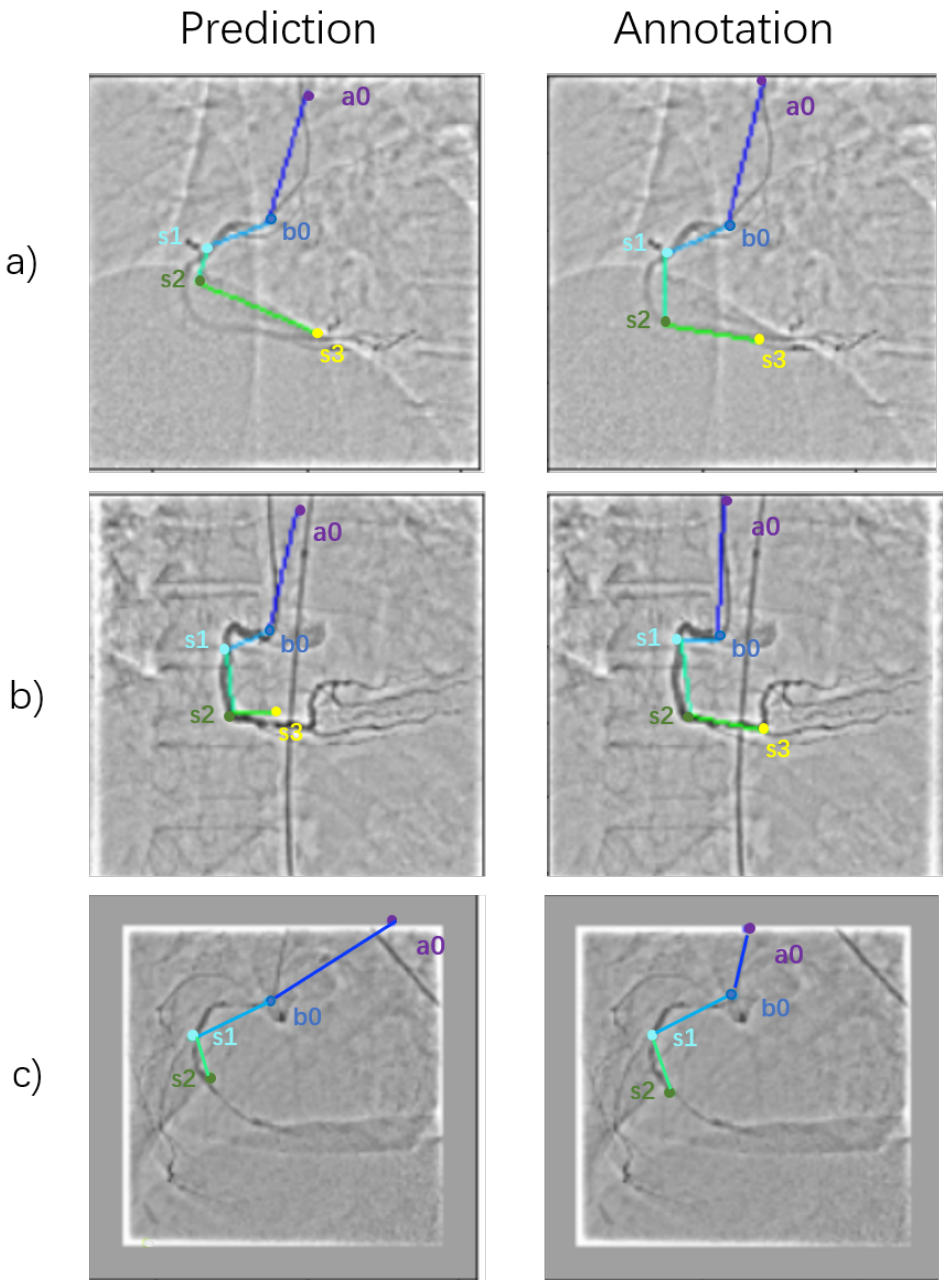


Figure 5.12: The angiographies are not correctly annotated by the network because of the noise on the background.

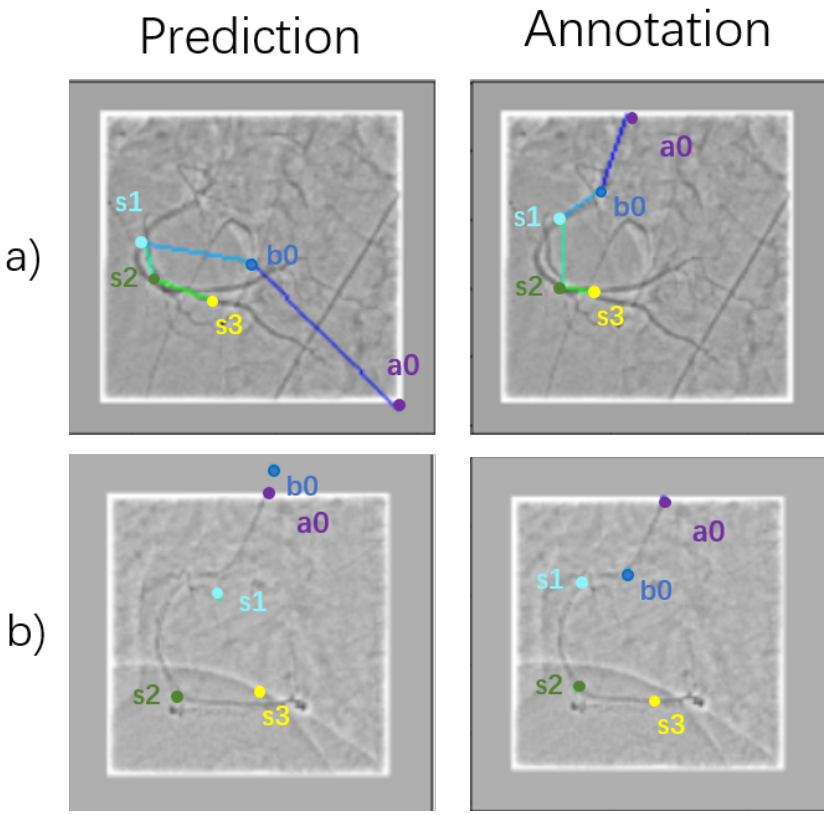


Figure 5.13: The cardiac angiographies are not correctly annotated, but the errors are more difficult to be understood.

one, which means the network with two pooling layers has better performance. Therefore, the Module 2 with two pooling layers is chosen to be used in the proposed segment point heatmap regression networks.

Evaluation of network structure in Module 3: False positive suppression

In Module 3: false positive suppression, we use three submodules to enlarge network's receptive field till it can cover entire right coronary artery tree in the cardiac angiography. We evaluate the choice of using three submodules by comparing the performance of Module 3 consisting of different number of submodules.

On the left side of Figure 5.17, it shows segment point heatmap regression networks with Module 3 consisting of two, three, and four submodules, the corresponding network performance on test set is shown on the right side. We can see that the segment point locations predicted by network with two submodules in Module 3 (first row of Figure 5.17) is less precise than the segment point locations predicted by network with three submodules in Module 3 (second row of Figure 5. 17). This is because the receptive field of network with two submodules in Module 3 is not large enough to cover entire coronary artery tree on the

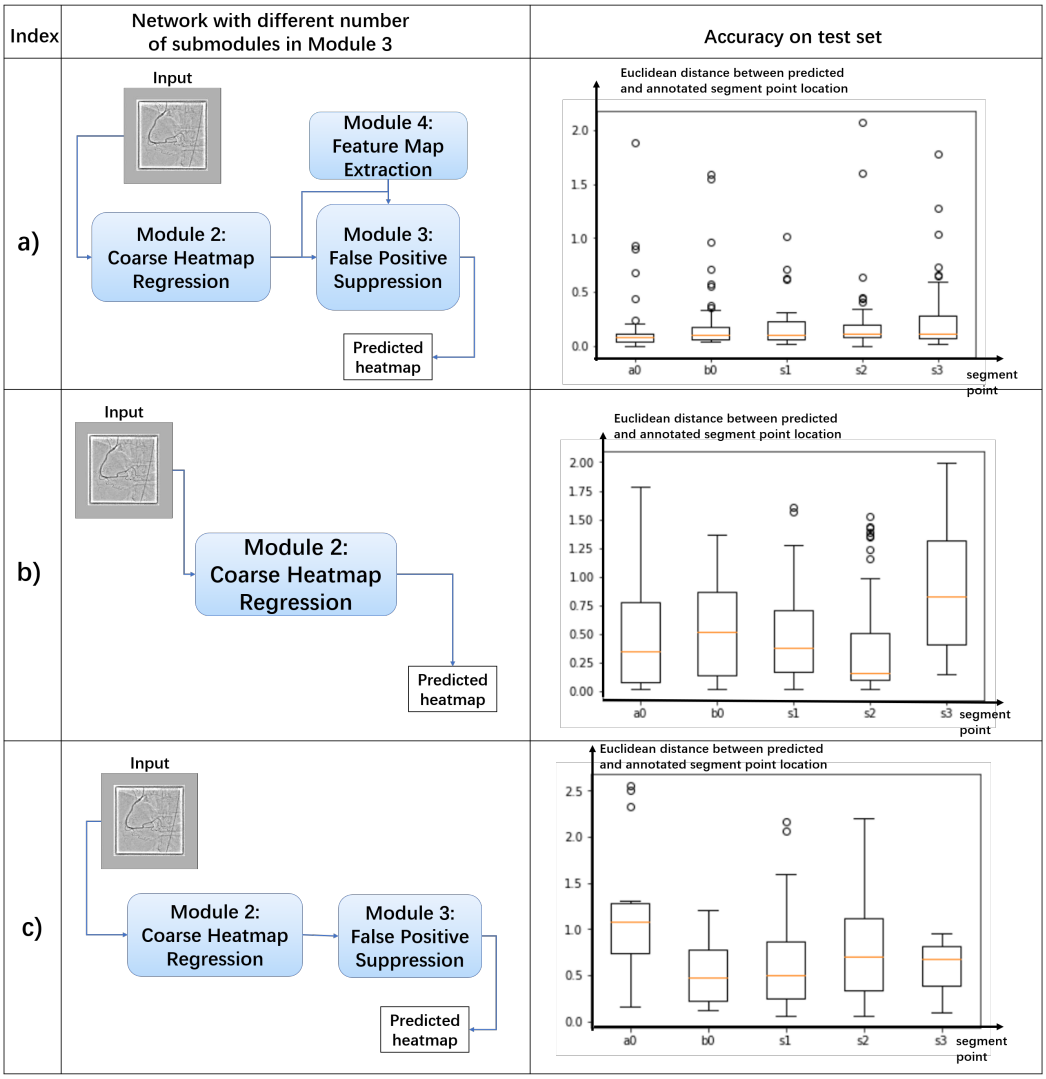


Figure 5.14: Left column of the table shows the network we use to evaluate the importance of each module in the segment point heatmap regression networks, right column shows the accuracy of corresponding network structure on test set.

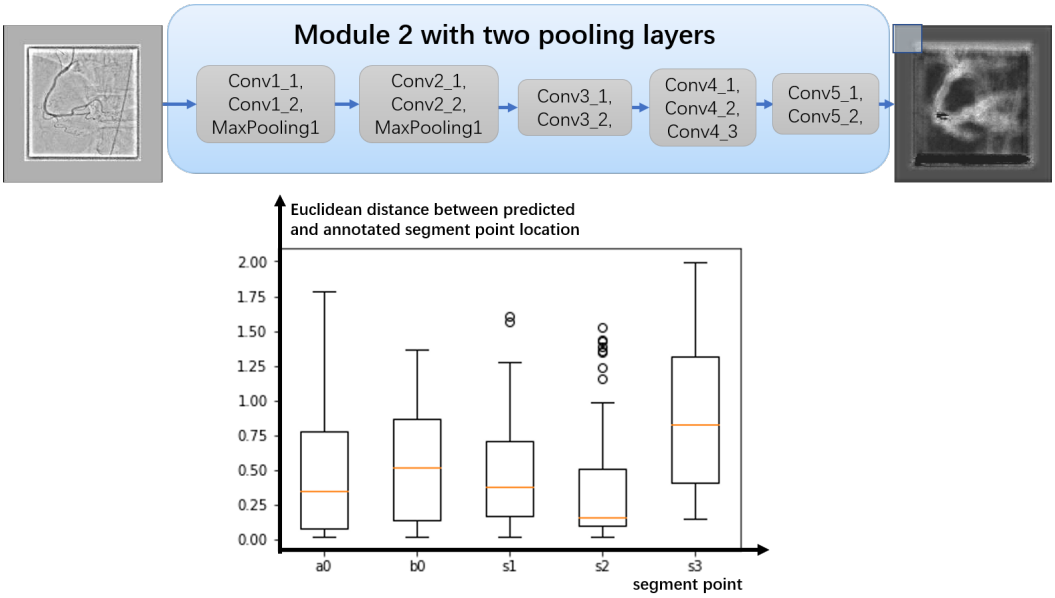


Figure 5.15: The graph and box plot represent for Module 2 network with two pooling layers and its performance on test set. The Box plot below shows the Euclidean distance between annotated and predicted segment point location, which is predicted by the Module 2 network with two pooling layers.

angiography, therefore, it is hard to organize global information from entire coronary artery tree. However, the accuracy hardly increases when we add more submodules after the third submodule in Module 3. This is because receptive field of the network has already covered the entire coronary artery when there are three submodules in Module 3, the network has already been able to organize global information form the coronary artery, therefore, it is not helpful to make the receptive field even larger.

In this chapter, we evaluate the performance of Module 1 network for cardiac angiography classification and segment point heatmap regression networks for cardiac angiography annotation. The performance of both networks increases during training, which shows that the network can learn how to classify and annotate the cardiac angiographies from examples. The performance of trained Module 1 network reaches a decent level, though it is influenced by the image quality in some cases. The performance of trained segment point heatmap regression networks is in general good according to the Euclidean distance between predicted and annotated segment point locations shown in Figure 5.9. However, our trained network is lack of robustness. There are very similar angiographies in test set, but only part of them are correctly annotated. We assume the size of training set limits the performance and generalization ability of segment point heatmap regression networks.

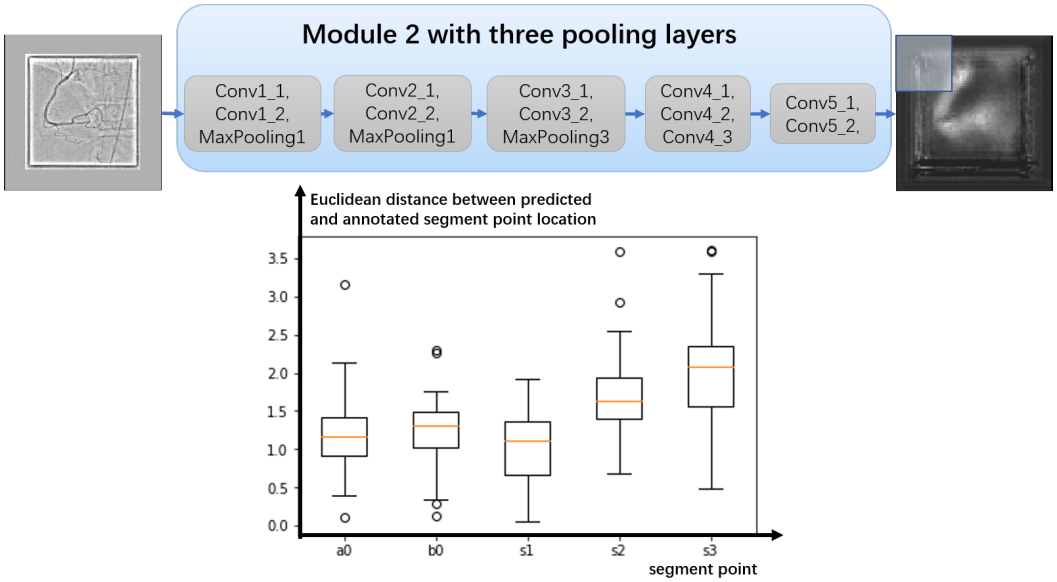


Figure 5.16: The graph and box plot represent for Module 2 network with three pooling layers and its performance on test set. The Box plot below shows the Euclidean distance between annotated and predicted segment point location, which is predicted by the Module 2 network with three pooling layers.

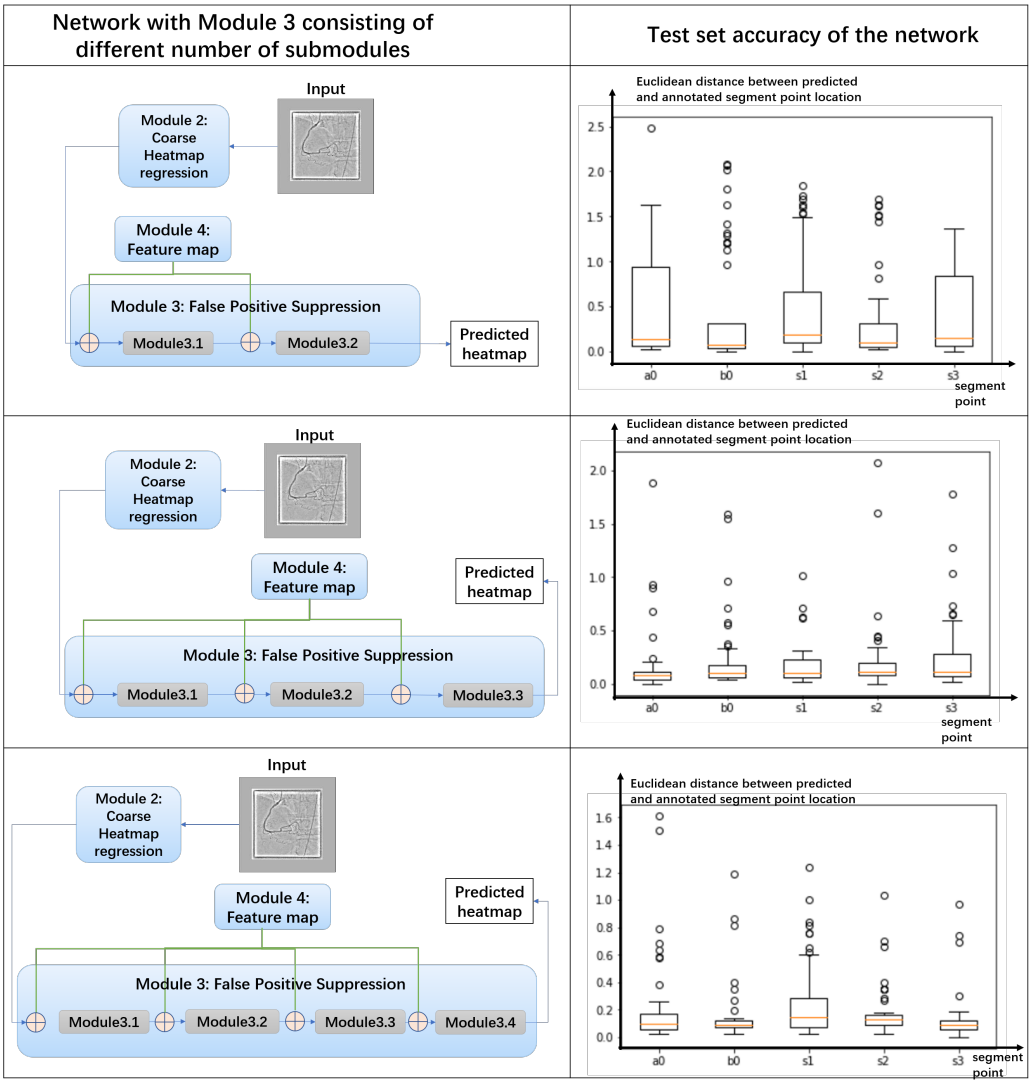


Figure 5.17: From left to right, the box plots represent for test set accuracy of Module 3 with two, three, and four submodules.

Chapter 6

Conclusions and future work

In this thesis, we propose a method as a first attempt for annotating cardiac angiography based on deep learning. This method differs from previous model-based cardiac angiography annotation methods, it is a data-driven approach, which learns cardiac angiography annotation from examples. In this chapter, we first conclude the proposed cardiac angiography annotation method and its contributions, then present the future work of this project.

6.1 Conclusions and contributions

Our proposed cardiac angiography annotation method annotates right coronary artery angiographies by regressing the segment points on the artery. From the experiments, we conclude the benefits of the proposed method:

1) Our proposed cardiac angiography annotation does not need manually designed feature extractors, it can learn and generalize from training set. It is hard to design feature extractors that cover all kinds of variations, but we can encode the variations in training set for the method to learn. In section 5.1.2 and 5.2.2, we can see that the ability of the proposed method to fit and generalize from training set is increasing, which means the cardiac angiography annotation method is learning from training set. The examples in Figure 5.8 show that the method can also learn how to cope with the difficulties in cardiac angiography annotation including various image acquisitions angles, different coronary topological structure, ambiguities caused by projection, and other tubular structure medical instruments on the background. According to the trained method performance on test set, we can see that the method is also able to generalize from training set and correctly annotate cardiac angiographies that are not included in training set but has similar difficulties with training set;

2) Our proposed method does not depend on coronary topological structure extraction, which is the precondition of previous model-based approach. Our method directly regresses segment point location on the cardiac angiography, therefore, the method is not limited to the precision of coronary topological structure extraction.

The contributions of this project can be summarized into two points:

1) This project proposed a method as the first attempts at annotating cardiac angiographies by means of deep learning. It shows that the cardiac angiography annotation method can be learned from the training set, and if the variations of cardiac angiography are represented well in the training set, the method also can generalize from them;

2) The proposed method for cardiac angiography annotation is less dependent on manually designed feature extractor and coronary topological structure extraction.

6.2 Future work

Due to the limitation on time and data, there are still opening questions and future work left for this project:

1) There are similar angiographies in the test set, but part of them are correctly annotated, the others are not. We assume the reason for lack of robustness of the proposed method is the limited size and quality of training set, but more cardiac angiographies are needed for testing before we come to a concrete conclusion;

2) The proposed method focuses on the annotation of right coronary artery angiography because of the time limitation and complexity of left coronary artery angiography. Applying the proposed method to left coronary artery angiography annotation is more challenging and complex, and it can further evaluate the ability of the proposed method to address cardiac angiography annotation;

3) Lack of annotated cardiac angiography is one of the important reasons limiting the proposed method. In future work, we can also develop semi-supervised deep learning methods for cardiac angiography annotation.

Bibliography

Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*, 388(10053):1545–1602, October 2016. ISSN 0140-6736. doi: 10.1016/S0140-6736(16)31678-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5055577/.

A. Akinyemi, S. Murphy, I. Poole, and C. Roberts. Automatic labelling of coronary arteries. In *2009 17th European Signal Processing Conference*, pages 1562–1566, Aug 2009.

J. A. Bærentzen. On the implementation of fast marching methods for 3D lattices, 2001. URL http://www2.imm.dtu.dk/pubdb/p.php?841.

J. Gordon(2013) Betts. Anatomy and physiology. pages 787–846. ISBN 1938168135., 11 edition, 2014.

Adrian Bulat and Georgios Tzimiropoulos. Human pose estimation via convolutional part heatmap regression. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 717–732, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46478-7.

Qing Cao, Alexander Broersen, Michiel A. de Graaf, Pieter H. Kitslaar, Guanyu Yang, Arthur J. Scholte, Boudewijn P. F. Lelieveldt, Johan H. C. Reiber, and Jouke Dijkstra. Automatic identification of coronary tree anatomy in coronary computed tomography angiography. *The International Journal of Cardiovascular Imaging*, 33(11):1809–1819, Nov 2017a. ISSN 1573-0743. doi: 10.1007/s10554-017-1169-0. URL https://doi.org/10.1007/s10554-017-1169-0.

Z. Cao, T. Simon, S. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1302–1310, July 2017b. doi: 10.1109/CVPR.2017.143.

C. Chalopin, I. E. Magnin, and G. Finet. Automatic labeling of the coronary tree using a three dimensional reference prior model. In *Computers in Cardiology 1998. Vol. 25 (Cat. No.98CH36292)*, pages 761–764, Sep 1998. doi: 10.1109/CIC.1998.731985.

C Chalopin, G Finet, and IE Magnin. Modeling the 3d coronary tree for labeling purposes. *Medical image analysis*, 5(4):301—315, December 2001. ISSN 1361-8415. doi: 10.1016/s1361-8415(01)00047-0. URL https://doi.org/10.1016/S1361-8415(01)00047-0.

Fiona J Charlson, Andrew E Moran, Greg Freedman, Rosana E Norman, Nicolas JC Stapelberg, Amanda J Baxter, Theo Vos, and Harvey A Whiteford. The contribution of major depression to the global burden of ischemic heart disease: a comparative risk assessment. *BMC Medicine*, 11:250–250, 2013. ISSN 1741-7015. doi: 10.1186/1741-7015-11-250. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4222499/.

Hirak Das. Variations in origin, course and termination of anterior interventricular artery -. *National Journal of Clinical Anatomy*, 2013.

J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. doi: 10.1109/CVPR.2009.5206848.

N. Ezquerra, S. Capell, L. Klein, and P. Duijves. Model-guided labeling of coronary structure. *IEEE Transactions on Medical Imaging*, 17(3):429–441, June 1998. ISSN 0278-0062. doi: 10.1109/42.712132.

L. Ge, H. Liang, J. Yuan, and D. Thalmann. Robust 3d hand pose estimation from single depth images using multi-view cnns. *IEEE Transactions on Image Processing*, 27(9):4422–4436, Sept 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2834824.

R. Girshick. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, Dec 2015. doi: 10.1109/ICCV.2015.169.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR. URL http://proceedings.mlr.press/v15/glorot11a.html.

K. Haris, S. N. Efstratiadis, N. Maglaveras, C. Pappas, J. Gourassas, and G. Louridas. Model-based morphological segmentation and labeling of coronary angiograms. *IEEE Transactions on Medical Imaging*, 18(10):1003–1015, Oct 1999. ISSN 0278-0062. doi: 10.1109/42.811312.

M. S. Hassouna and A. A. Farag. Multistencils fast marching methods: A highly accurate solution to the eikonal equation on cartesian domains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1563–1574, Sept 2007. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.1154.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015. URL http://arxiv.org/abs/1502.03167.

Lucian Mihai Itu, Saikiran Rapaka, Tiziano Passerini, Bogdan Georgescu, Chris Schwemmer, Max Schoebinger, Thomas G. Flohr, Puneet Sharma, and Dorin Comaniciu. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *Journal of applied physiology*, 121 1:42–52, 2016.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Yann LeCun, Y Bengio, and Geoffrey Hinton. Deep learning. 521:436–44, 05 2015.

Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen. Medical image classification with convolutional neural network. In *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, pages 844–848, Dec 2014. doi: 10.1109/ICARCV.2014.7064414.

X. Liu, F. Hou, H. Qin, and A. Hao. Robust optimization-based coronary artery labeling from x-ray angiograms. *IEEE Journal of Biomedical and Health Informatics*, 20(6): 1608–1620, Nov 2016. ISSN 2168-2194. doi: 10.1109/JBHI.2015.2485227.

J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, June 2015. doi: 10.1109/CVPR.2015.7298965.

Puja K. Mehta, Janet Wei, and Nanette K. Wenger. Ischemic heart disease in women: A focus on risk factors. *Trends in Cardiovascular Medicine*, 25(2):140 – 151, 2015. ISSN 1050-1738. doi: https://doi.org/10.1016/j.tcm.2014.10.005. URL http://www.sciencedirect.com/science/article/pii/S1050173814001807.

C. Metz, M. Schaap, A. V. Der Giessen, T. V. Walsum, and W. Niessen. Semi-automatic coronary artery centerline extraction in computed tomography angiography data. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 856–859, April 2007. doi: 10.1109/ISBI.2007.356987.

A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, Dec 2010. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.46.

A. Neubeck and L. Van Gool. Efficient non-maximum suppression. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 3, pages 850–855, Aug 2006. doi: 10.1109/ICPR.2006.479.

Markus Oberweger, Paul Wohlhart, and Vincent Lepetit. Hands deep in deep learning for hand pose estimation. *CoRR*, abs/1502.06807, 2015. URL `http://arxiv.org/abs/1502.06807`.

Christian Payer, Darko Štern, Horst Bischof, and Martin Urschler. Regressing heatmaps for multiple landmark localization using cnns. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 230–238, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46723-8.

T. Pfister, J. Charles, and A. Zisserman. Flowing convnets for human pose estimation in videos. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1913–1921, Dec 2015. doi: 10.1109/ICCV.2015.222.

Tomas Pfister, Karen Simonyan, James Charles, and Andrew Zisserman. Deep convolutional neural networks for efficient pose estimation in gesture videos, 11 2014.

L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele. Deepcut: Joint subset partition and labeling for multi person pose estimation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4929–4937, June 2016. doi: 10.1109/CVPR.2016.533.

Gilbert L. Raff, Aiden Abidov, Stephan Achenbach, Daniel S. Berman, Lawrence M. Boxt, Matthew J. Budoff, V. Cheng, Tony Defrance, Jeffrey C. Hellinger, and Ronald P. Karlsberg. Scct guidelines for the interpretation and reporting of coronary computed tomographic angiography. *Journal of cardiovascular computed tomography*, 3 2:122–36, 2009.

Z. Rahman, D. J. Jobson, and G. A. Woodell. Multi-scale retinex for color image enhancement. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 1003–1006 vol.3, Sep 1996. doi: 10.1109/ICIP.1996.560995.

H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, May 2016. ISSN 0278-0062. doi: 10.1109/TMI.2016.2528162.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, June 2015. doi: 10.1109/CVPR.2015.7298594.

J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler. Efficient object localization using convolutional networks. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 648–656, June 2015. doi: 10.1109/CVPR.2015.7298664.

Jonathan Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'14, pages 1799–1807, Cambridge, MA, USA, 2014. MIT Press. URL http://dl.acm.org/citation.cfm?id=2968826.2969027.

A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1653–1660, June 2014. doi: 10.1109/CVPR.2014.214.

S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4732, June 2016. doi: 10.1109/CVPR.2016.511.

Nathan D. Wong. Epidemiological studies of CHD and the evolution of preventive cardiology. *Nature Reviews Cardiology*, 11:276, 2014. URL http://dx.doi.org/10.1038/nrcardio.2014.26.

G. Yang, A. Broersen, R. Petr, P. Kitslaar, M. A. de Graaf, J. J. Bax, J. H. C. Reiber, and J. Dijkstra. Automatic coronary artery tree labeling in coronary computed tomographic angiography datasets. In *2011 Computing in Cardiology*, pages 109–112, Sept 2011.

Guanyu Yang, Pieter Kitslaar, Michel Frenay, Alexander Broersen, Mark J. Boogers, Jeroen J. Bax, Johan H. C. Reiber, and Jouke Dijkstra. Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography. *The International Journal of Cardiovascular Imaging*, 28(4):921–933, Apr 2012. ISSN 1573-0743. doi: 10.1007/s10554-011-9894-2. URL https://doi.org/10.1007/s10554-011-9894-2.

Ning Zhang, Evan Shelhamer, Yang Gao, and Trevor Darrell. Fine-grained pose prediction, normalization, and recognition. *CoRR*, abs/1511.07063, 2015. URL http://arxiv.org/abs/1511.07063.

Yefeng Zheng, Maciej Loziczonek, Bogdan Georgescu, S. Kevin Zhou, Fernando Vega-Higuera, and Dorin Comaniciu. Machine learning based vesselness measurement for coronary artery segmentation in cardiac ct volumes. 03 2011.

Yefeng Zheng, Huseyin Tek, and Gareth Funka-Lea. Robust and accurate coronary artery centerline extraction in cta by combining model-driven and data-driven approaches, 09 2013.