
**UNIVERSITÀ POLITECNICA DELLE MARCHE**

*Department of Information Engineering (DII)*

Master of Science in Biomedical Engineering

# Stenosis Detection in Coronarography Images via Deep Learning

Supervisor: Prof. Emanuele Frontoni

Co-Supervisor: Sara Moccia, PhD

Author:

Federica Turco

Academic Year 2019 - 2020

*Alla mia famiglia,*

*che da sempre mi sostiene e crede in me.*

# Ringraziamenti

A conclusione di questo elaborato, vorrei esprimere la mia gratitudine a tutte le persone senza le quali questo lavoro di tesi non esisterebbe nemmeno.

In primo luogo vorrei ringraziare il professor Frontoni, relatore di questa tesi, per avermi dato la possibilità di superare i miei limiti lavorando su questo progetto, in un ambito per me nuovo, dandomi l'opportunità di ampliare le conoscenze e competenze.

Un enorme ringraziamento va a Sara per la disponibilità e precisione dimostratemi durante tutto il periodo di stesura del lavoro di tesi. Colgo l'occasione per ringraziare di cuore Maria Chiara per la pazienza e la gentilezza che ha sempre mostrato nei miei confronti, sempre pronta a darmi le giuste indicazioni in ogni fase della realizzazione dell'elaborato.

Grazie alla mia famiglia per non avermi mai fatto mancare nulla ed avermi dato l'opportunità di studiare e intraprendere questa esperienza formativa, supportandomi e motivandomi. Grazie perché senza di voi non sarei mai arrivata fino in fondo a questo cammino.

Grazie a Mariachiara e Francesca per essere sempre al mio fianco e per avermi sempre ed incondizionatamente aiutata quando ne avevo bisogno.

# Abstract

Coronary artery disease (CAD) is the most common type of heart condition and a major cause of mortality worldwide. It involves the presence of stenosis, the constriction of at least one of the large coronary arteries responsible for supplying oxygenated blood to the heart muscle leading to the poor oxygen supply and subsequently resulting in an imbalance state between oxygen demand and supply.

In symptomatic patients, the assessment of CAD presence and its severity is critical for determining appropriate clinical management, and diagnosis is generally made or confirmed with some form of medical imaging. Due to its ability to assess coronary stenoses and its potential for immediate quality control of the performed revascularization, invasive coronary angiography (ICA) is currently regarded as the gold standard for coronary artery stenosis detection.

The introduction of a tool to automatically detect stenosis from images could provide a huge support to the clinicians, reducing bias and subjective interpretations while allowing to effectively measure the severity of the stenosis in a more reliable way.

In recent years, machine learning (ML) methods have demonstrated highly accurate and reliable performance across a variety of computer-vision related tasks, including image classification, object detection, and semantic segmentation. Consequently, deep learning (DL) may have the potential to replace or reduce manual burden in detection and quantification of coronary stenosis.

The CNN for object detection, pretrained on natural images, chosen to be performed

on the provided dataset to obtain stenosis detection from ICA images is the second version of the YOLO (You Only Look Once) family, the one that better performed on the dataset provided for this work.

# Contents

Chapter 1

# INTRODUCTION

## 1.1 Coronary Artery Disease

Coronary artery disease (CAD), among the group of disorders related to the heart and blood circulation refer to as cardiovascular disease, is the most common type of heart condition and a major cause of mortality worldwide [1]. It involves the presence of stenosis, the constriction of at least one of the large coronary arteries responsible for supplying oxygenated blood to the heart muscle leading to the poor oxygen supply and subsequently resulting in an imbalance state between oxygen demand and supply [2].

Two main arteries branch off the aorta namely Left Main Coronary Artery (LCA) and Right Coronary Artery (RCA) which supply blood to left and right parts of the heart respectively. These two main arteries then divide into a network of smaller coronary arteries which wrap themselves around the heart [3]. The stenosis often manifests in various positions along the artery, with those near major arterial junctions being most critical [4]. An example of stenosis can be seen in Figure 1.1.

The narrowing or blockage of these coronary arteries is due to the buildup of cholesterol and fatty deposits called plaque on the inner lining of the arterial wall [6]. The damaged arteries may become completely blocked, or become prone to clotting, so, over time, the heart has to work harder, possibly causing heart failure when the heart cannot pump efficiently or arrhythmia when the heart beats irregularly or too quickly. Indeed, if CAD occurs, the affected arteries can't supply enough oxygen-rich blood

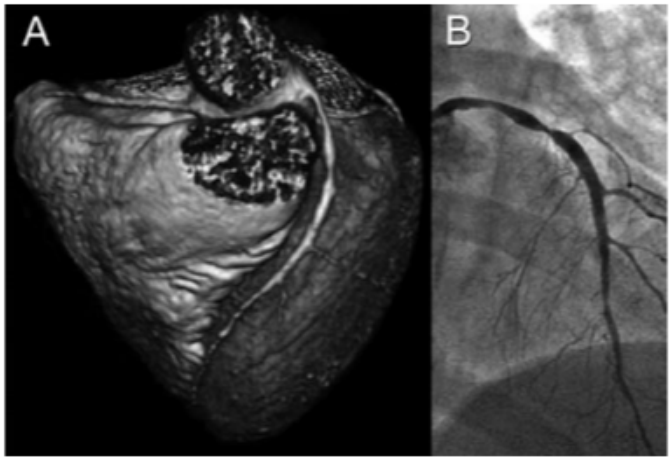


**Figure 1.1:** *Example of significant stenosis of the left anterior descending coronary artery from focal non calcific atherosclerosis with coronary computed tomographic angiography(A) and invasive coronary angiography (B). Figure adapted from [5].*

to the heart — especially when it's beating hard, such as during exercise. At first, the decreased blood flow may not cause any symptoms. As plaque continues to build up in the coronary arteries it can lead to a develop of the following CAD signs and symptoms:

- **Chest pain (angina)**. Pressure or tightness in the chest occur. This pain, called angina, usually occurs on the middle or left side of the chest. Angina is generally triggered by physical or emotional stress. The pain usually goes away within minutes after stopping the stressful activity. In some people, especially women, the pain may be brief or sharp and felt in the neck, arm or back.

- **Shortness of breath**. If the heart can't pump enough blood to meet body's needs, it may develop shortness of breath or extreme fatigue with activity.

- **Abnormal heart rhythm (arrhythmia)**. Inadequate blood supply to the heart or damage to heart tissue can interfere with the heart's electrical impulses, causing abnormal heart rhythms.

- **Heart attack**. A completely blocked coronary artery will cause a heart attack. The classic signs and symptoms of a heart attack include crushing pressure in

the chest and pain in the shoulder or arm, sometimes with shortness of breath and sweating.

- **Heart failure**. If some areas of the heart are chronically deprived of oxygen and nutrients because of reduced blood flow, or if the heart has been damaged by a heart attack, it may become too weak to pump enough blood to meet body's needs. This condition is known as heart failure.

The risk factors of CAD can be both environmental and genetic, and they can distinguish in modifiable, including smocking, obesity, hypertension and sedentary life style, or non-modifiable, for example age and gender (women have generally less typical signs and symptoms of heart attack than men). According to the data presented by World Health Organization (WHO), CAD is the most prevalent disease in the developing nations where it is a major threat and leading cause of mortality. The global burden of CAD in 2002 was 7.1 million which was predicted to be raised to 11.1 million by 2020 [2].

In symptomatic patients, the assessment of CAD presence and its severity is critical for determining appropriate clinical management, and diagnosis is generally made or confirmed with some form of medical imaging. Medical images take a significant part in patient diagnostics at different levels, including regular screening, diagnosis verification, preoperative planning, and follow-up. Accordingly, great efforts are invested into this field to improve the quality of images and facilitate accurate scan interpretation and avoid medical errors. At present, the CAD evaluation mainly depends on coronary computed tomographic angiography and invasive coronary angiography described in the following section [7].

CAD can be stabilized and treated through healthy living and the use of medications to control risk factors and symptoms. In some patients, the coronary arteries become more severely blocked and require a revascularization procedure. The two procedures to manage a blocked coronary artery are to implant a stent in the area of the blockage (angioplasty) as illustrated in Figure 1.2 or to entirely bypass the blocked segment of artery surgically (bypass surgery) [8].

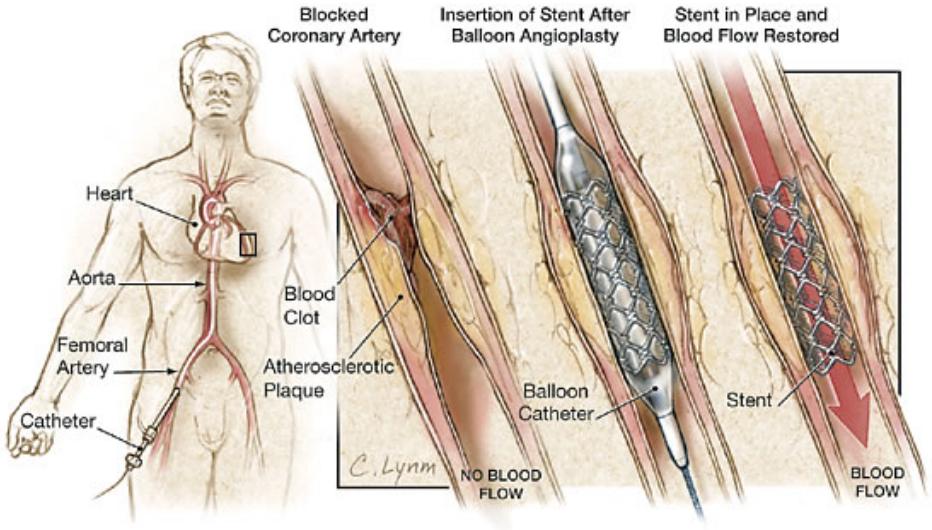


**Figure 1.2:** *Procedure of Percutaneous coronary intervention. Figure adapted from [9].*

**Percutaneous coronary intervention (PCI; angioplasty)** is a minimally invasive procedure in which a tube with an associated balloon is introduced via a peripheral artery (either the femoral artery in the groin or the radial artery at the wrist), avoiding the need for surgery and general anesthesia. The balloon is inflated in the area of the blocked artery to stretch it open. In most patients, a coronary stent is then placed. This spring-like looking device helps to keep the artery open and reduces the chance of recurrent narrowing. Approximately 15% to 20% of patients will develop renarrowing of the artery requiring a repeat angioplasty procedure within 6 to 12 months. Stents may be coated with medication that reduces the risk of renarrowing. To prevent clots developing in the stent, 2 medications that inhibit blood platelets are needed for up to 1 year after the procedure (usually aspirin plus an additional blood thinner).

**Coronary artery bypass graft (CABG) surgery** is a major surgical procedure requiring general anesthesia. In most patients, the procedure is performed after opening the chest through an incision with the breastbone. Veins taken from the leg and an artery taken from within the chest are used to bypass the coronary artery blockages. The bypass grafts have a high chance of remaining open in the first 5 to 8 years after the operation. However, by 10 years after the operation, about half of vein bypass grafts are either blocked or have developed a severe narrowing. In contrast, arterial bypass grafts are more likely to remain open. Although repeat bypass surgery is possible, many patients with a blocked vein graft can be treated medically and do not need

another operation.

Nevertheless, for decision making in the interventional catheterization laboratory with respect to revascularization, it is of paramount importance to determine whether a stenosis is functionally significant inducing reversible ischemia [10].

## 1.2 Medical images

The most used medical imaging techniques for the CAD assessment are the coronary computed tomographic angiography (CCTA) and invasive coronary angiography (ICA) [7].

### 1.2.1 Coronary computed tomographic angiography (CCTA)

A CCTA, shown in Figure 1.3, is a noninvasive imaging technique that combines advanced CT technology with an intravenous injection of contrast material to produce pictures images of the heart and its blood vessels. During CCTA, x-rays pass through the body and are picked up by detectors in the scanner, that produce cross-sectional images of the target on a computer screen.

In recent years, with the advent of improved hardware, CCTA has become feasible and practical having the potential to enable a comprehensive evaluation of coronary anatomy and myocardial perfusion in patients with known or suspected to have CAD in a noninvasive manner [12].

CCTA is increasingly used to assess CAD, providing high resolution three-dimensional images of the coronary arteries. Moreover, it can also provide additional information regarding the type of plaque (calcified, mixed or soft) [13].

Despite remarkable technical developments and even though it has the advantage of being noninvasive, CCTA acquisitions expose the patient to a higher dosage of radiation [3]. Other limitations are the calcification blooming artefacts, the limited spatial and temporal resolution, the unpredictability of hemodynamic significance of intermediate coronary lesions, the radiation exposure and the difficulties to acquire motion-free, high-quality images in patients with arrhythmias.

Coronary calcifications cause blooming artefacts of coronary calcific lesions which

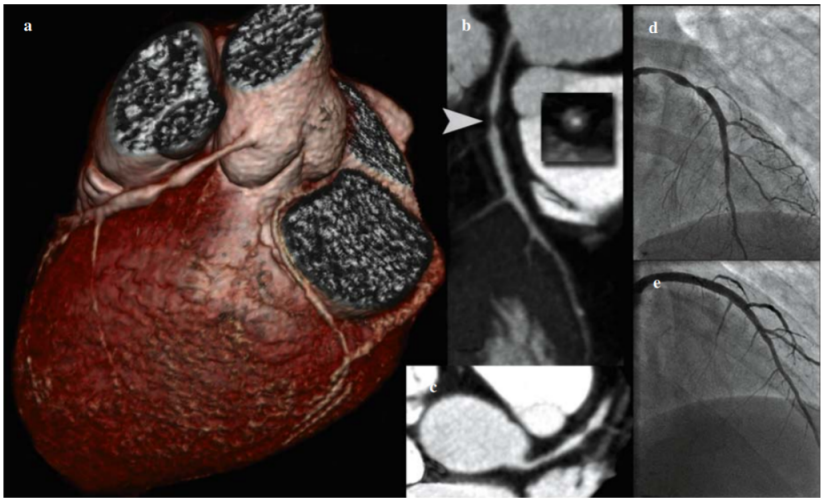


**Figure 1.3:** *Anterior descending coronary artery lesion with 64-layer CCTA. A patient with an anomaly of origin of the circumflex coronary artery (right coronary sinus and retro-aortic course) and significant stenosis on CT at the proximal tract of the anterior descendant (a; b, arrowhead). The lesion shows peculiar characteristics to the CT, i.e. absence of evident calcifications and presence of positive vascular remodeling (b, box; c). The conventional coronarography performed later (d) confirmed the presence and extent of the lesion that was treated by angioplasty and stenting (e). Figure adapted from [11].*

either obscure adequate evaluation of the underlying coronary lumen or induce an overestimation of the severity of a coronary obstruction. Both problems result in limitations in the diagnostic performance of CCTA and even the introduction of newly developed CT technology with improved spatial resolution or use of dual-energy CT may not fully reduce this problem [14].

### 1.2.2 Invasive coronary angiography (ICA)

ICA is a medical imaging technique that involves continuous X-ray (i.e. fluoroscopy) with simultaneous injection of radiopaque contrast into the coronary arteries [15]. During the procedure a long, thin, flexible tube called catheter is inserted into a blood vessel in the patient arm. Using X-ray images as a guide, the tip of the catheter is passed up to the heart and coronary arteries. A special type of dye called contrast medium is injected through the catheter and X-ray images, named angiograms, are taken. The contrast medium is visible on the angiograms, showing the blood vessels and the fluid travels through. In this way, any blood vessel narrowed or blocked can be more clearly

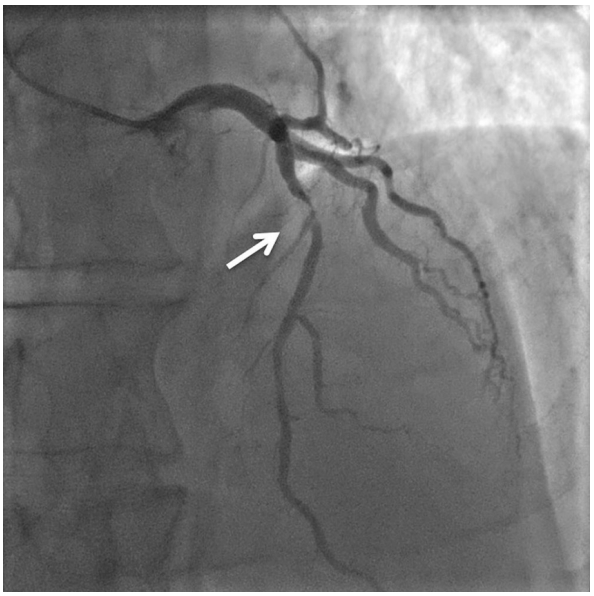

**Figure 1.4:** *Coronary angiography revealing a subtotal stenosis in the proximal left anterior descending artery. Figure adapted from [16].*

highlighted, as shown in Figure 1.4.

Due to its ability to assess coronary stenoses and its potential for immediate quality control of the performed revascularization, ICA is currently regarded as the gold standard for coronary artery stenosis detection. It can offer anatomical information of even very small vessels and enable cardiologists to observe dynamically from different projection angles. Cardiologists can then identify and locate each stenosis with a visual assessment [17].

The fundamental task required for the interpretation of coronary angiography is identification and quantification of the severity of stenosis within the coronary circulation [18].

Often, spurious dye and imaging artifacts can give a false appearance of stenosis in these sequences. To resolve such cases, clinicians watch the angiograms in cine loops exploiting the fact that areal stenosis will persist in each image frame through time. The surface representation of 2D Xray angiography sequences gives a much clearer indication of stenosis than conventional single frame analysis [4].

## 1.3 Disclosure

The actual imaging techniques used in clinical practice to diagnose CAD are still characterized by several limitations: they mainly rely on the experience of the clinician and his ability in the visual identification of the stenosis. Thus, the introduction of a tool to automatically detect stenosis from images could provide a huge support to the clinicians, reducing bias and subjective interpretations while allowing to effectively measure the severity of the stenosis more reliable.

In order to develop this diagnostic tool as a support in clinical practice, in this work the emerging Deep Learning (DL) techniques have been chosen to develop a system for stenosis detection from ICA images. This work was carried out in collaboration with the Cardiology Department of Ospedali Riuniti in Ancona: all the experiments and considerations done in this work were conducted on ICA images collected, annotated and provided in DICOM (Digital Imaging and Communications in Medicine) format with associated metadata by the doctors of this department.

Chapter 2

# STATE OF THE ART

Lesion detection in medical images has taken a significant diagnostic role in many aspects and the development of automatic techniques to perform detection increases strongly the efficiency of medical imaging interpretation tasks.

In this Chapter a literature review about the conventional methods for stenosis detection is presented. Starting from the explanation of ICA images visual inspection for stenosis detection, an overview of the computer-based methods for the detection is provided, including semi-automated and fully automated software reaching finally the application of DL techniques in the field of medical imaging. At last, the limitations in the state of art are highlighted and thus, the thesis objective is presented.

## 2.1 Conventional stenosis detection in clinical practice

### 2.1.1 Visual assessment

Much of the clinical care of patients with ischemic heart disease is based on research that has relied heavily on visual interpretation of the coronary angiogram [19].

Coronary angiography is performed to determine the presence and severity of coronary stenosis, thus guiding the treatment for patients with CAD. Physician visual assessment (PVA) of stenosis severity remains the standard method for guiding revas-

cularization [20].

However, several studies employing different techniques have seriously questioned the notion that visual interpretation of the coronary angiogram permits an accurate assessment of the physiologic significance of a coronary obstruction. Although there is no problem with assuming that normal caliber coronary vessels with minimal irregularities are not significantly obstructed and that nearly totally occluded vessels are significantly obstructed, assessing the physiologic significance of narrowings of intermediate severely is much more uncertain. Nonetheless, almost all patient management decisions and the vast majority of clinical studies rely on visual assessment of the coronary angiogram, using percent narrowing to define the severity of coronary disease [19].

Strictures are usually defined in terms of the percentage reduction from normal diameter; various values, ranging from 50% to 80% have been suggested to represent a haemodynamically significant stricture, but such precise percentages give the misleading impression that strictures can be assessed by eye with an accuracy of better than 5%̇The accuracy of angiography is critically dependent on the density of opacification of the artery, and the technique is least accurate in the middle range of narrowing, where there is a systematic underestimation of lesion severity.

There are two different methods to evaluate visually stricture anatomy in ICA images. The most common depends on assumptions about the geometry of the stricture and the vessels, which must be opacified to high density and displayed parallel to the intensifier plane, without overlap. This method needs to be capable of very accurate measurement because, in the critical range, small changes produce large alterations of flow. Another method is the densitometric measurements of coronary strictures and its accuracy is not restricted by the shape of the cross-section of the artery or its stricture. For a given concentration of contrast medium in an artery, the X-ray transmission through it will depend on the depth of vessel traversed, and can reflect the relative depth of normal and strictured channels without assumptions about geometry [21].

However, manual detection of stenosis is subjective and time-consuming, requiring rich clinical experience and expert knowledge. In addition, it is also challenging due to complex vessel structures, poor contrast between vessels and surrounding tissues,

nonuniform illumination, and overlap of background structures with inhomogeneous intensities [17].

The implication of variation in interpretation of angiograms is serious: if readings are erroneous, some patients will undergo revascularization procedures unnecessarily and others will be denied an essential treatment [22].

Consequently, even though PVA is considered a clinical standard, it has known limitations, such as significant intra- and inter-rater variability, as well as high positive prediction bias, implicitly leading to over-utilization of clinical services [20].

### 2.1.2 Computer-based image analysis

Quantitative coronary angiography (QCA) has been introduced exploiting automated or semi-automated edge detection to provide more precise quantitative estimates of bidimensional data obtained with ICA when compared with more subjective measures like PVA. QCA provides more precise quantitative data in comparison to visual assessment. In fact, the visual interpretation of the severity of a coronary stenosis is usually expressed in intervals of percentage of stenosis. Conversely, QCA produces a single specific measure for stenosis diameter, improving the accuracy and reproducibility of the severity assessment.

It is a technique directly based on contrast coronary angiography that obtains parameters that quantify objectively and with interval measures the significance of a coronary stenosis, and also helps to estimate the immediate- and long-term results of PCI [23].

QCA is based on the use of a specific and dedicated software that allows the determination of some specific measures of coronary lumen in an operator-independent way. Whereas QCA has the inherent drawback of focusing only on the contrast-filled lumen of the vessel, it continues to provide important insights for clinical practice and, most commonly, clinical research.

When used in clinical practice, QCA is performed on-line immediately after ICA providing objective and independent parameter for the assessment of stenosis severity helpful to decide the type of intervention and to choose the suitable devices and their sizes, such as the device length and diameter, and also maximum balloon inflation

pressure. Thus, it challenges qualitative estimation (i.e. PVA) of coronary stenoses [23]. On-line QCA is widely available in most modern angiographic services allowing the quantification of vessel size and lesion length. Usually, QCA software utilizes measurement calibration by comparing it with an object of known dimension [24].

When employed in clinical research, QCA is performed off-line after finishing ICA procedure or intervention, based on data storage and transfer, and conducted by angiographic core laboratory experts, providing a visual annotation of the diseased coronary arterial segments and the area surrounding each stenosis to determine the percent diameter stenosis. Thus, providing a purely numerical evaluation of coronary stenoses and interventional procedures.

Standard workflow for QCA consists of a multi-step analytical pipeline. As QCA is based on coronary angiography, the first step in order to perform a QCA analysis is to acquire high-quality ICA images focused on the target coronary artery segment of choice obtaining a two-dimensional luminogram of a three dimensional structure. Then, a single-frame images that best demonstrate the stenosis are selected by the analyst, knowing that the greater is the contrast between the radiopaque contrast-filled coronary artery and the radiotransparent background, the greater are the accuracy of QCA analysis and reliability of the algorithm for the automated detection and reconstruction of the lumen edges. In addition, also a second frame where the coronary catheter appears completely filled by contrast medium must be selected for calibration purposes. Subsequently, the two frames are analyzed by using dedicated softwares that works more or less in the same way: first a calibration is performed in which the catheter dimension is measured in order to be then recognized automatically in the digital images, after having acquired the digital image the automatic reconstruction of the hypothetically normal coronary lumen is made. This technique is essentially based on the calculation of a mean value of the diameters of the lumen in the segments of reference located upstream and downstream to the lesion [23].

QCA measurement process currently offers the most accurate and reproducible measurements of anatomical coronary stenosis severity and thus is considered the clinical gold standard for measuring coronary stenosis. However, despite the improvements over PVA, calculating stenosis severity by QCA still requires satisfactory image ac-

quisition and minimal user input to identify imaging frames for analysis, which may introduce variability as well [25].

Although QCA provides morphological information of coronary arteries with objective quantitative measures, considerable training is required to identify the target vessels and understand the tree structure of coronary arteries. Even thought the use of computer-aided tools, manual correction is necessary for accurate segmentation of coronary vessels [26].

Limitations related to QCA include dependency on ICA image quality and challenges in assessing complex lesions, such as those with thrombus or calcification [20]. In addition, QCA makes evaluations considering two dimensions, which is not suited for three-dimensional structures such as coronary vessels and atherosclerotic lesions. Therefore, it is necessary to develop systems for image acquisition and analysis that are automated and able to integrate more two-dimensional quantitative analyses in a single three-dimensional model, as the acquisition of three or more angles may improve three-dimensional reconstruction by averaging [23].

In order to improve the detection and quantification of coronary stenosis in clinical practice, a variety of semi-automated computerized algorithms have been proposed. For example, one introduces the use of Hessian based vessel enhancement filter applied to the angiographic frame to be segmented for enhancement of vessel structures, then image fusion using wavelet transform is used to make an image on the selected frame as a threshold for detecting the vessels [27]. Finally, alternative method aimed to providing a tool to accurately quantify vascular lesion and limit the variability of measurement due to subjective interpretation developing a coronary extraction and stenosis quantification method in ICA using a deformable spline algorithm and a string-matching technique [28].

However, all the above-mentioned methods are defined semi-automated software because they are computer-based image analysis tools developed to support identification procedure and required human interaction [18].

## 2.2 Advanced stenosis detection via Deep Learning technique

The automatic algorithms are mainly classified as semi-automated and fully automated software. Semi-automated tools, indicated in the previous section, speed up the analysis procedure, but typically they require the interaction of an expert across different stages to manually set several parameters. Instead, fully automated software employs Computer Vision and/or Machine Learning (ML) techniques, avoiding user dependency and thus providing standardized and reproducible measurements [29].

Low productivity, inter-operator variability, and reliance on specialized expertise constitute serious limitations. These clinical challenges introduced with the semi-automated tools motivate the development of improved software for rapidly, accurately and reliably analyzing coronary stenoses. In recent years, convolutional neural network (CNN) methods have demonstrated highly accurate and reliable performance across a variety of computer-vision related tasks, including image classification, object detection, and semantic segmentation [15].

The recent advances in ML, and especially in DL and deep CNNs have potential to replace or reduce manual burden in detection and quantification of coronary stenosis [30].

### 2.2.1 DL for stenosis detection

DL techniques have recently revolutionized many fields, from computer vision and natural image classification until reaching a profound influence on medical imaging. Because of the large number of cardiac images that are routinely acquired with a wide range of modalities, there has been a rise in publications applying DL in the cardiac domain [31].

Advances in high-performance computing and the increasing accessibility of ML algorithms capable of performing complex tasks have improved clinical interest in applying these techniques in research and clinical care [32].

CNN is an end-to-end method with a strong feature extraction ability, generating detection results for every single frame quickly and directly. However, due to vessel

motion and contrast agent flow, stenosis-like structures such as bent vessels or instantaneous contrast agent inhomogeneity appear in some of the frames, which might mislead the network to generate false positive detections [17].

The increasing huge amount of data in this field opens the realm of possibilities to apply DL methodologies to completely automatize the overall approach while discovering unseen discriminative patterns in stenoses.

In literature, among the most recent researches on stenosis detection performed on ICA, the works conducted by *Chao Cong et al.* [33], *Wei Wu et al.* [17] and *Benjamin et al.* [15] result to be particularly interesting in relation with the object of the thesis.

#### 2.2.1.1 Chao Cong et al.

In this study, an automated method for stenosis detection in ICA images is proposed by using a CNN based workflow for image-level stenosis classification, without the need for a priori vessel segmentation.

Firstly, image preparation is performed through an automatic detection of the ideal candidate frames and the redundant frames, defined as the ones in a coronarography video with best image quality, full contrast-agent penetration, clearly contrasted vessel borders, and anatomical significance of stenosis (if stenosis is present). Then, a recognition of the ideal candidate frames and the redundancy frames is made by training a CNN architecture from the Inception family (i.e. inception-v3) and a recurrent neural network (RNN) - that differs from CNN in its ability to process temporal information or data that comes in sequences - called directional long-short-term memory (LSTM), having as output a 0 for redundant frame and 1 for candidate frame. Frames that did not meet the selection criteria for candidate frames were manually removed from the augmented training dataset (for classification training). Finally, after stenosis classification training, a class of activation maps is employed to identify the discriminative regions. Figure 2.1 illustrates an abstraction of the algorithmic workflow.

The dataset consists on a sequence of frames of ICA of 194 patients saved in DICOM format with resolution of 512*512 or 1024*1024, 15 frames per second, 60-120 frames. The stenosis location and their severity were previously analyzed with QCA in order to categorized them into three clinically relevant groups based on stenosis severity -

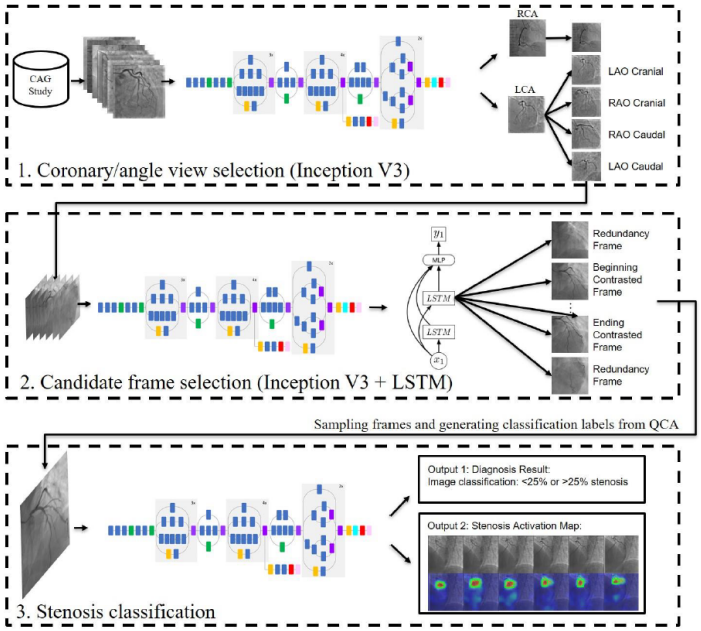


**Figure 2.1:** *Abstraction of the proposed end-to-end stenosis detection workflow. The input of our method is CAG dataset, and 4 main steps were described briefly in dash-line boxes from data preparation to stenosis detection. There are 3 types of outputs as Output 1: diagnosis result for an image-level stenosis classification; Output 2: stenosis activation map and Output 3: stenosis localization information. Figure adapted from [33].*

Category 0: <25 % stenosis, Category 1: 25 to 99 % stenosis, and Category 2: total occlusion. Thus, the main training setup is based on this three-category setup (named 3-CAT). However, also a binary setup is used with a distinction of two groups of stenosis severity - <25% vs. >25% - named 2-CAT. In addition, in 3-CAT total occlusion were separated from other stenosis, because total occlusions is seen to require a different and more urgent treatment strategy from stenosis lesions.

The Experimental results showed the potential of this method in stenosis detection in 2-CAT (<85% accuracy) and 3-CAT (<80% accuracy) in ICA images. However, training/validation overfitting was the most important issue to be handled [33].

A limitation of this study is that it was implemented at an image-level method, although it has demonstrated its ability in videos.

#### 2.2.1.2 Wei Wu et al.

Another DL approach to coronary artery stenosis detection was proposed by Wei Wu et al in 2020, which developed an object detection network with temporal constraints on the ICA sequence to achieve automatic coronary artery stenosis detection.

The flow diagram of this method, shown in Figure 2.2, involves three main parts. Firstly an automatic selection of the most contrast-filled frame from a coronarography sequence, considered the most appropriate for stenosis detection since they show complete coronary artery structures, is performed by using a CNN called U-Net. However, a single frame is not enough. Therefore, N frames before and after the most contrast-filled frame are selected. The selection of multiple frames can be regarded as a special data augmentation approach, which introduces the variance within the XCA sequence (e.g., motion and brightness variation) into the training set. Then, the Deconvolutional Single-Shot multibox Detector (DSSD) [34] was used to provide rough results for each selected frame. The backbone model of the applied DSSD is a CNN that requires a raw XCA frame as input and outputs the detected stenosis. The deconvolutional layers doubling the resolution of the high-level feature map with the learned deconvolutional layer and further combines feature maps from two different levels by elementwise summation. This step merges semantic information and location information, generating feature maps with richer contents.

Although the DSSD has high sensitivity, it is still influenced by a certain number of false positives. Due to this phenomenon, the potential temporal information of an ICA sequence was used to remove false positives by selecting the stenosis that most frequently appears in the sequence, thus filtering out remaining random false positives.

The dataset consists on ICA sequence data of 63 patient with resolution of 512*512 and a sequence length varying from 3 to 5 s at 14 frames per second.

Even though the proposed method has to face problem related to small dataset, it achieves sensitivity of 87.2% and positive predictive value of 79.5% [17].

#### 2.2.1.3 Benjamin et al.

In the Benjamin et al. study, three separate CNN models are proposed to solve the problems of localization, segmentation and classification, then integrated developing an

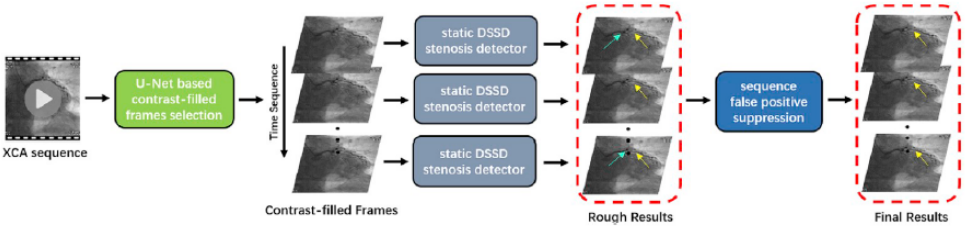


**Figure 2.2:** *Framework of the proposed method. The whole algorithm works as follows: first, the contrast-filled frames of an input XCA sequence are selected based on the U-Net segmentation results (shown in chronological order from top to bottom). Then, the DSSD provides rough results for each selected frame (yellow arrows for true positives and aqua arrows for false positives). Finally, the seq-fps module summarizes the rough results and removes false positives, generating the final results. Figure adapted from [17].*

end-to-end deep learning pipeline for characterizing stenosis in right coronary artery ICA images.

For the localization task, a CNN inspired YOLONet [35] is used to obtain a bounding box prediction around the stenosis. In particular, ICA images are given as input to the CNN generating a non-overlapping grid then used to obtain an image with a bounding box and stenosis coordinates overlaid. For the segmentation task, a CNN based on U-net is used to generate high-fidelity segmentation of the lesion of interest given a pre-localized ICA image input. For the classification task it is used a small CNN in order to to determine whether image inputs meet the clinical criterion for revascularization, i.e. $\geq$70% stenosis, provided pre-localized ICA images and segmentation annotation. Finally, a combination of the individual tasks is performed to form an end-to-end deep learning pipeline, classifying stenosis based exclusively on the frame ICA image input.

The dataset used is composed by single-frame image extracted from ICA of 1024 study participants. The data for each participant consists of the de-identification single-frame image extracted from ICA, the lesion segmentation annotations corresponding to each image, the coordinate location of stenosis, the QCA stenosis measurement and the normalized lesion silhouette annotation. In order to standardize data components, it was performed a preprocessing on the dataset consisting in the exclusion of the images containing more than one annotated stenosis lesion or no annotated stenosis lesion and

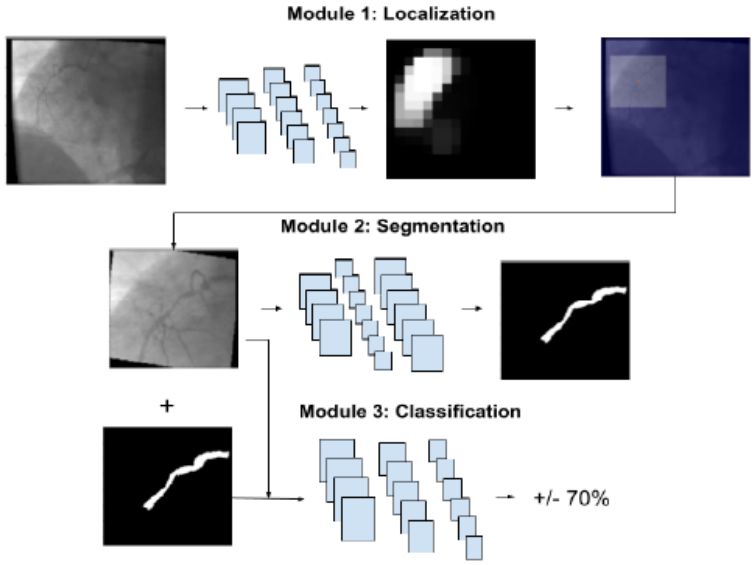


**Figure 2.3:** *Overview of Experimental Method. Module 1: Localize a bounding box around the stenosis lesion in the input image. Module 2: Segment the diseased lesion in the localized image. Module 3: Classify the combined image and segmentation input binarily via 70% stenosis threshold. Figure adapted from [15]*

the contrast standardization of angiography images and segmentation annotations, i.e. minimum and maximum pixel range were set to 0 and 1, respectively. Image sizes were standardized to 512*512 pixels. Data augmentation techniques are applied during the training phase of each CNN, including random shift, random rotation, random shear. An overview of the proposed method is shown in Figure 2.3.

This method, even if based on still-frame images, achieved a performance of 72.7% localization accuracy and relative to the current clinical standard for real-time RCA stenosis analysis, it showed a statistically significant improvement of 36.1% in false discovery rate at the 70% stenosis threshold. Thus, this proposed methodology obtained several benefits relevant in clinical setting, such as provide real-time inference and zero intra-rater variance, allowing for more reliable stenosis estimation.

## 2.3 Limitation in the state of the art

Considering that developing a tool to identify and quantify stenosis in a more accurate and rapid way should be an opportunity to actually improve clinical care, this thesis

aims to investigate the possibility to obtain stenosis detection on ICA images through the use of one of the most common object detection network.

However, although the use of DL has improved the analysis of biomedical images compared to manual inspection, these algorithms are not an universal remedy and also have limitations. The success of their application depends on the number of observations, number of features, selection and parameterization of features and algorithm chosen for the model [32].

The detection/classification tasks still present several challenges for fully automated strategy mainly due to the quality of the images, the variety in position of the stenoses and the difficulty to classify multi-level of damage. As a consequence of these limitations, these approaches employ hand-crafted features and the overall method is not at all fully automated and, for example, the experimental procedure and the features extraction stage is driven by the supervision of an expert operator [29].

Because DL requires very specific and precise labeling of the data for successful training, medical experts have to reexamine and annotate huge amounts of medical data to make them suitable for deep learning. As it is a time consuming and costly task, this has been one of the major hurdles for the application of deep learning in wider fields of medicine [36].

Although DL has shown incredible results across many fields, it remains a tricky technique to master. Given the same dataset, different researchers can obtain widely varying results. The main cause is that many decisions have to be made when training CNNs: how to pre-process the data; which network architecture to select; and how to optimize the coefficients of the network [31].

All the studies that obtain good performance when applying DL algorithms often differentiate themselves in aspects outside of the deep network, like data preprocessing or augmentation techniques. For example, by adding a stain normalization preprocessing step to improve generalization without changing the CNN, or by focusing on data augmentation strategies to make networks more robust, and they report that these strategies are essential to obtain good performance.

## 2.4 Aim of the thesis

Motivated by the opportunity to improve clinical care by offering a more accurate and rapid way to quantify coronary stenoses, this work aims to provide a tool based on object detection, one of the most important and challenging areas of computer vision, for the stenosis detection on images taken by ICA technique, due to the importance of the automatic detection of lesions in medical images to improve the clinical care.

An important part of the work is focused on the organization of the dataset containing the annotations made by clinician and images relative to the patients involved in the work, considered to be the main obstacle in the objective detection: a extensive job has been devoted to this task.

Then, a CNN for object detection, pretrained on natural images, has been chosen in accordance with the literature to be performed on the provided dataset to obtain stenosis detection from ICA images. The CNN model used for this purpose is the second version of the YOLO family, the one that better performed on the dataset provided for this work.

Chapter 3

# MATERIALS AND METHODS

In this chapter, a background of ML and DL and one of the fundamental task required when using them are reported respectively in Section 3.1 and 3.2. The YOLO architecture, a specific CNN based object detection model considered the most accurate real-time object detector, is presented in Section 3.2.1, while in Section 3.2.2 an explanation of the transfer learning strategy is described.

## 3.1 ML and DL overview

For a long time, engineers have been solving many problems using conventional programming, writing a set of commands that a computer needed to execute. Although it is very helpful and proven to be effective to execute repetitive tasks, it fails to implement tasks that are natural to humans. Tasks like vision, understanding natural language or learning are extremely difficult for a computer to master. For these kinds of problems, new sets of algorithms and techniques are needed. Techniques that will enable computers to mimic and replicate learning. These algorithms are called ML algorithms.

ML is the set of techniques, part of the artificial intelligence (AI), aimed at studying computer algorithms to perform unscheduled activities that improve automatically through experience. It is an evolving branch of computational algorithms that are designed to process input data to achieve a desired task without being literally pro-

grammed to produce a particular outcome and able to emulate human intelligence by learning from the surrounding environment.

The process of adaptation is called training, in which samples of input data are provided along with desired outcomes. The algorithm then optimally configures itself so that it can not only produce the desired outcome when presented with the training inputs, but can generalize to produce the desired outcome from new, previously unseen data [37]

Machine learning can be divided according to the nature of the data labeling into supervised, unsupervised, and semi-supervised as shown in Figure 3.1

- *Supervised learning* is used to estimate an unknown (input, output) mapping from known (input, output) samples, where the output is labeled (e.g., classification and regression). It allows the learner to deal with similarities and differences when the objects to be classified have many variable properties within their own classes but still have fundamental qualities that identify them. Most importantly, the successful learner should be able to recognize an apple or an orange that it has never seen before.

- In *unsupervised learning*, only input samples are given to the learning system rather than try to exactly program the kinematics a priori. This is unsupervised in the sense that the training doesn't associate a particular kinematic input configuration with a particular outcome. The algorithm finds its own way from the training input data.

- *Semi-supervised learning* is a combination of both supervised and unsupervised, where part of the data is labeled and other parts are unlabeled. In such a scenario, the labeled part can be used to aid the learning of the unlabeled portion. This kind of scenario lends itself to most processes in nature and more closely emulates how humans develop their skills.

Techniques based on ML have been applied successfully in diverse fields ranging from pattern recognition, computer vision, spacecraft engineering, finance, entertainment, and computational biology to biomedical and medical applications.

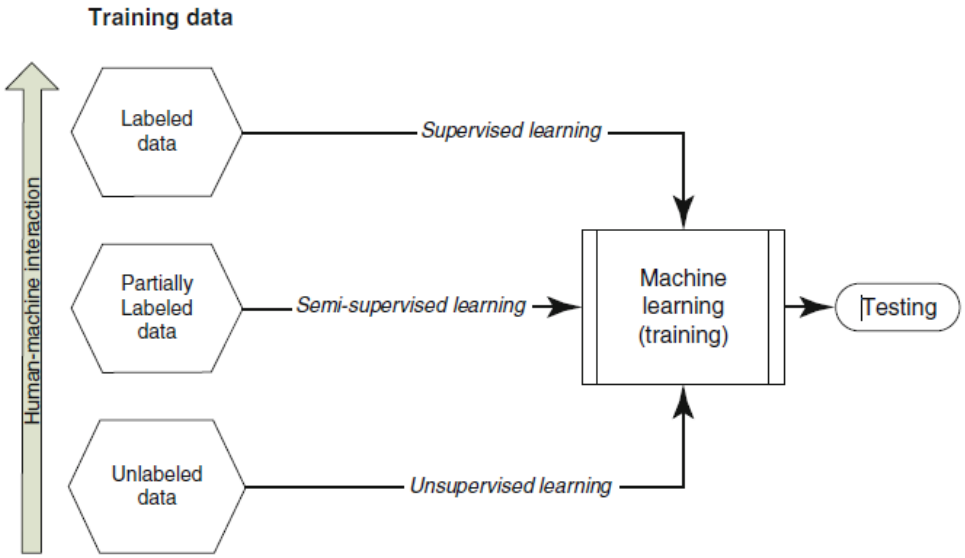


**Figure 3.1:** *Categories of machine learning algorithms according to training data nature. Figure adapted from [37]*

There are two particularly important advantages to a successful algorithm. First, it can substitute for laborious and repetitive human effort. Second, and more significantly, it can potentially learn more complicated and subtle patterns in the input data than the average human observer is able to do. Both of these advantages are very important to medical application [37].

A subset of ML is the DL that works at the same but with different skills; in fact, a DL model tries to simulate the human mind by correcting itself in case of errors. The term "Deep" comes from the fact that neural networks are equipped with multiple hidden levels of neurons, making them, in fact, deep and similar to structure of the human brain.

Various are the strong points of DL, such as its universality, robustness and generalization. Indeed, The DL approach is sometimes called universal learning because it can be applied to almost any application domain. Then, DL approaches do not require the precisely designed feature because optimal features are automatically learned for the task at handw obtaining a robustness to natural variations of the input data. In addition, the same DL approach can be used in different applications or with different data types and this is helpful where the problem does not have sufficient available data.

DL is about learning representations, i.e, learning intermediate concept or features which are important to capture dependencies from input variables to output variables in supervised learning, or between subsets of variables in unsupervised learning. Both supervised and unsupervised machine learning approaches are widely applied in medical image analysis [38].

## 3.2 Object detection

Although it is very helpful to know what class image belongs to, sometimes users also need to know the exact position of an object on a picture. The problem of localizing an object in an image is known as object detection.

Object detection is a fundamental problem in computer vision, and there have been many attempts to solve of this problem. The techniques used range from matching three-dimensional geometric models of object with images, to using two-dimensional view-based representations.

The detection of objects of interest or lesions in images is a key part of diagnosis and is one of the most labor-intensive for clinicians. Typically, the tasks consist of the localization and identification of small lesions in the full image space. There has been a long research tradition in computer-aided detection systems that are designed to automatically detect lesions, improving the detection accuracy or decreasing the reading time of human experts [39].

### 3.2.1 Model for object detection - YOLO

After years of development, the state-of-the-art object detection systems have been integrated with a large number of techniques. Most of the early object detection algorithms were built based on handcrafted features. Due to the lack of effective image representation at that time, people have no choice but to design sophisticated feature representations, and a variety of speed up skills to exhaust the usage of limited computing resources [40].

More recent approaches like Region-Based Convolutional Networks (R-CNNs) [41, 42, 43] use region proposal methods to first generate potential bounding boxes in an

image and then run a classifier on these proposed boxes, for this reason called two-stages detector. After classification, post-processing is used to refine the bounding boxes, eliminate duplicate detections, and rescore the boxes based on other objects in the scene. Those complex pipelines are slow and hard to optimize because each individual must be trained separately [35].

YOLO (You Only Look Once) is a state-of-the-art object detection algorithm that is incredibly fast and accurate. It is an end-to-end deep learning-based detection model that determines the bounding boxes of the objects present in the image and classifies them in a single pass. It does not involve any region proposal phase conversely to two-stage detectors.

The YOLO family of models are a series of end-to-end deep learning models designed for fast object detection, developed by Joseph Redmon et al. and first described in the 2015 paper titled "You Only Look Once: Unified, Real-Time Object Detection." [35].

There are three main variations of the approach: YOLOv1, YOLO9000 (also know as YOLOv2) [44] and YOLOv3 [45]. The first version proposed the general architecture, whereas the second version refined the design and made use of predefined anchor boxes to improve bounding box proposal, and version three further refined the model architecture and training process. Although the accuracy of the models is close but not good as R-CNNs, they are popular for object detection because of their detection speed, often demonstrated in real-time on video or camera feed input.

In general, YOLO network involves a single deep CNN that first splits the input image into a grid of S x S non-overlapping cells. For each cell directly predicts a bounding box and object classification. The result is a large number of candidates bounding boxes that are consolidated into a final prediction by a post-processing step.

As illustrated in Figure 3.2, for each cell, the network predicts B bounding boxes. Each of them contains five parameters (*x, y, w, h, sc*), where *sc* is the objectness confidence score of the box. Subsequently, the network calculates the probabilities of the classes for each cell [46].

The architecture used in this work is the YOLO v2 [44]. It improves the previous version in several ways, including some techniques that YOLO didn't use, such as Batch-Normalization and the use of anchor boxes to predict bounding boxes. Batch-

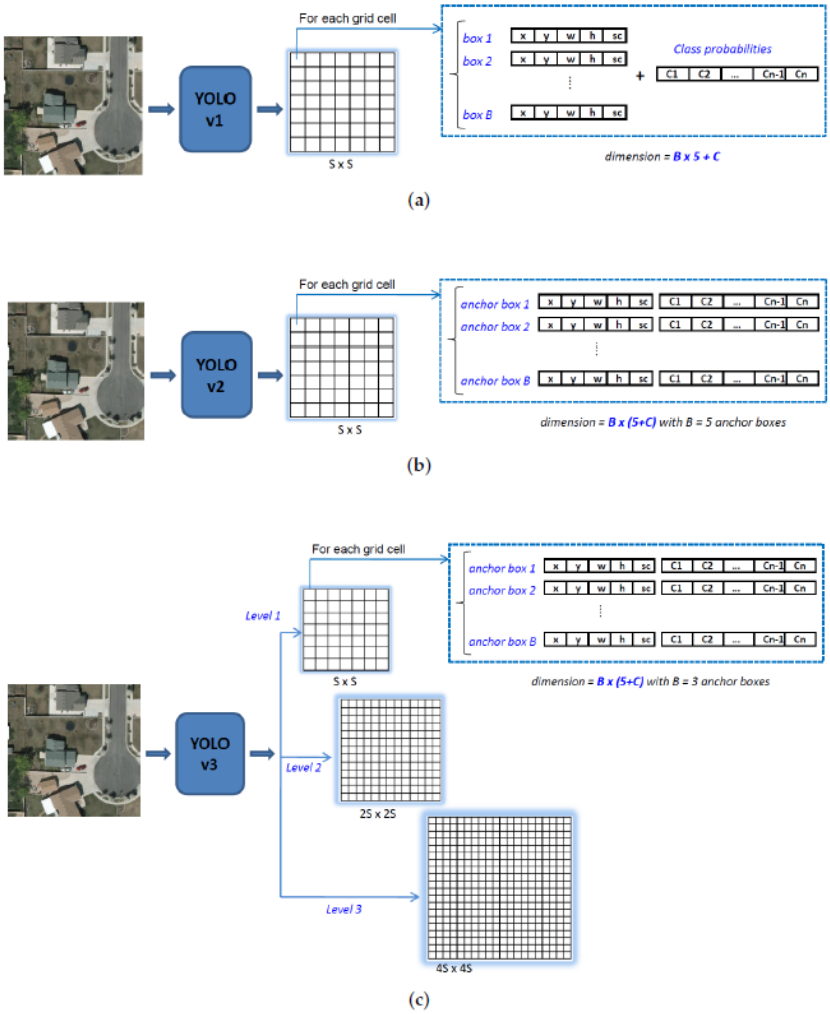


**Figure 3.2:** *Overview of the state-of-the-art You Only Look Once (YOLO) family for one-stage object detection. (a) In YOLOv1, the output is a tensor of dimension (S, S, B x 5 + C) with (S, S) the size of the grid, B the number of predicted boxes for each cell and C the number of classes. (b) In YOLOv2, the output is a tensor of dimension (S, S, B x (5 + C)). The difference is that the class probabilities are calculated for each anchor box. (c) In YOLOv3, the output consists of 3 tensors of dimension (S, S, B x (5 + C)), (2S, 2S, B x (5 + C)) and (4S, 4S, B x (5+C)) which correspond to the 3 detection levels (scales). Figure adapted from [46].*

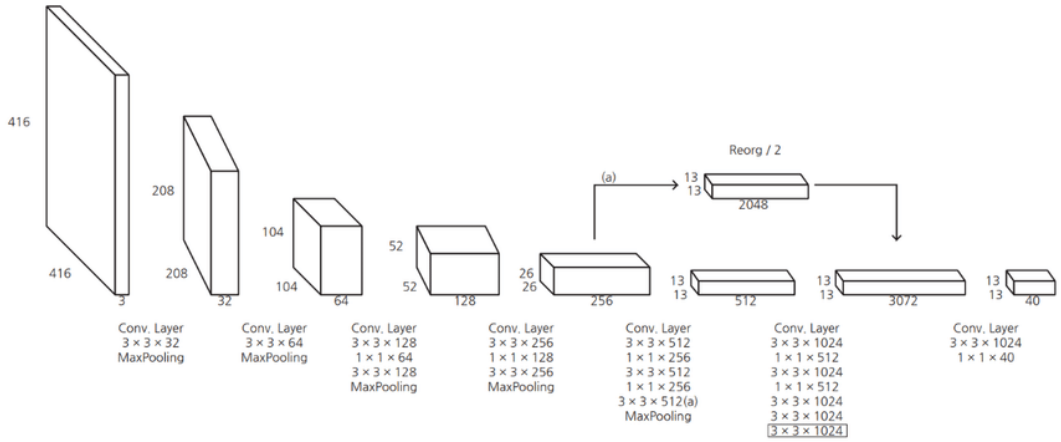


**Figure 3.3:** *The architecture of YOLOv2. Figure adapted from [47].*

Normalization is used to normalize the outputs of hidden layers, making the learning phase much faster. Anchor-Boxes is assumed by the shapes of the bounding boxes of the object that the network has to detect, making prediction much faster. The general YOLOv2 network structure is shown in Figure 3.3.

YOLO v2 uses anchor boxes to predict all the B boxes. The box is represented using different values. The prediction values:

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e_w^t \\ b_h &= p_h e_h^t \\ \sigma(t_0) &= Pr(object) \end{aligned} \tag{3.1}$$

Where Pr(object) is the probability that the box contains an object. The $\sigma()$ (sigmoid) function is used to constrain values between [0, 1]. The box center is expressed as an offset with respect to the top left corner of the grid cell $(c_x, c_y)$ and width and height are expressed relative to the anchor box dimensions ($p_w$ is the anchor box width, $p_h$ the height). The value $\sigma(t_0)$ gives an idea of how good is a box, it is the box confidence score. If this value is below a certain threshold (typically 0.3) the corresponding box is discarded. Figure 3.4 shows the values of a predicted box. The dashed box is the anchor box while the blue box is the prediction. The blue dot is the predicted box center, it is expressed as an offset with respect to the top left corner of the cell in which

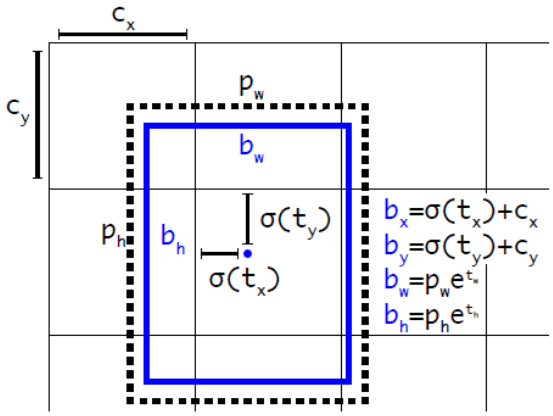


**Figure 3.4:** *Bounding boxes with dimension priors and location prediction. Figure adapted from [44].*

it is contained plus the coordinates of that corner.

YOLOv2 is state-of-the-art and faster than other detection systems across a variety of detection datasets. Furthermore, it can be run at a variety of image sizes to provide a smooth trade-off between speed and accuracy.

The employed pre-trained network and the relative weight are optimized on the created dataset. Thus, after loading the weights pre-trained on COCO dataset, the train was executed using the Adam optimizer with a learning rate set at $0.5 * 10^{-4}$ and performing 100 epochs.

### 3.2.2 Transfer learning

The field of DL has been widely and successfully used in many applications where patterns from past information (training data) can be extracted in order to predict future outcomes. Traditional ML is characterized by training data and testing data having the same input feature space and the same data distribution. When there is a difference in data distribution between the training data and test data, the results of a predictive learner can be degraded.

In certain scenarios, obtaining training data that matches the feature space and predicted data distribution characteristics of the test data can be difficult and expensive. Therefore, there is a need to create a high-performance learner for a target domain trained from a related source domain. This is the motivation for transfer learning. Ba-

sically, transfer learning is used to improve a learner from one domain by transferring information from a related domain [48].

In DL, transfer learning is a technique whereby a neural network model is first trained on a problem similar to the problem that is being solved. Transfer learning is essentially the use of pre-trained networks (typically on natural images) to try to work around the (perceived) requirement of large data sets for deep network training.

The general idea, illustrated in Figure 3.5, is to use the knowledge a model has learned from a task with a lot of available labeled training data in a new task that doesn't have much data. Instead of starting the learning process from scratch, we start with patterns learned from solving a related task.

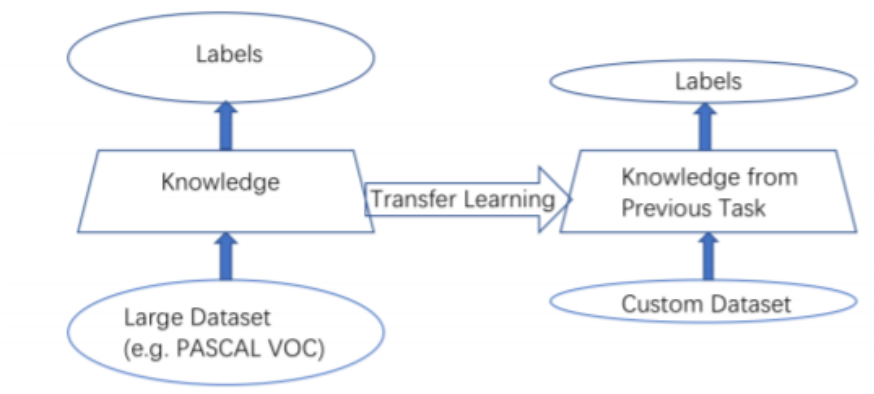


**Figure 3.5:** *Illustration of the transfer learning approach.*

Usually, a lot of data is needed to train a neural network from scratch but access to that data isn't always available, this is where transfer learning comes in handy. With transfer learning a solid machine learning model can be built with comparatively little training data because the model is already pre-trained.

There are two strategies when repurposing a pre-trained model by using transfer learning, which are fine-tuning the entire network and fine-tuning some layers with the others unchanged. Among them, the first strategy, fine-tuning the entire model refers to using the pre-trained network and modifying all the weights based on the custom dataset, leading to high computational cost and high demand for data amount. Consequently, this strategy is applied when we have a large dataset and different from the data in pre-trained data.

However, in the field of computer vision, the lower layers of the network are always similar to each other regardless the problem domain, whereas the higher layers of the network are updated to adapt to specific problems. So in the second strategy, only the higher layers are the parts that need to be optimized and fine-tuned during training, leaving the other lower layers frozen with the pre-trained weights. By reusing the parameters in the lower layers, the number of weights needs to be updated is largely reduced, which will rarely lead to overfitting. Both strategies are popular and have been widely applied.

Transfer learning has several benefits, but the main advantages are saving training time, better performance of neural networks (in most cases), and not needing a lot of data. With the technique, we can easily start a model from the pre-trained condition that have been trained with the other tasks.

Fine-tuning deep learning algorithms will help to improve the accuracy of a new neural network model by integrating data from an existing neural network and using it as an initialization point to make the training process time and resource-efficient.

In this work, the data is quite dissimilar to the one in ImageNet and is relatively large. Hence, the first strategy is applied to this task which is fine tuning the whole model with the initialization of the pre-trained model.

Chapter 4

# EXPERIMENTAL PROTOCOL

## 4.1 Dataset

The collection of the data used in this work has been provided by the Cardiology Department of Ancona's Ospedali Riuniti. The dataset was created from the collection of ICA images in DICOM format, the standard for the communication and management of medical imaging information and related data [49]. It consists on 1373 ICA images of 207 patients - with 5 patients having a double count due to the ICA acquisition made in different session. From the image acquisitions, a series of 512 $\times$ 512 pixel grayscale images with different gray levels (0–255) are generated.

Together with the images the dataset contains also annotations on the images made by visual inspection by clinician, listing the depicted stenosis and their positions in the image. The position information comes in the form of bounding boxes either squared or rectangular.

### 4.1.1 Data preparation

Training data consists of ICA images and annotation, in DICOM and *.csv* format respectively. In order to have format compatible with the CNN used, both files were converted into an unified format that the network can understand. In particular, ICA images were converted in *.jpeg* format while the annotation for each of the image into

*.xml* format. Pascal VOC (Visual Object classification) provides standardized image datasets for object detection.

After the conversion, the splitting of the dataset was performed. Data splitting technique is a crucial point: the dataset is divided into three part, to avoid overfitting and model selection bias, called training set, cross-validation set and testing set.

In particular, the training set consists in the sample of data used to fit the model, that is the actual subset of the dataset that is used to train the model (estimating the weights and biases). The model observes and learns from this data and optimize its parameters. Successively, the fitted model is used to predict the responses for the observations in a second dataset called the cross-validation set by minimizing the error. Finally, the testing set is used to provide an unbiased evaluation of a final model fit on the training dataset.It is only used once the model is completely trained using the training and validation sets.

In order to keep images associated with same patient in the same session (train, validation or testing phase) the dataset was organized obtaining a folder for each patient containing the annotations and associated images. Then, divided into train, validation and test folder performing a spitting by patient using a dataset split ratio of 80:10:10 for train,validation, and test sets respectively.

## 4.2 Training setting

By using the technique of transfer learning, the weights were randomly initialized in the head of the detector and also the rest of the network with pre-trained weights was initialized. After setting these as the starting point, the entire network was trained by updating all the weights to adapt to the custom dataset.

Network training aims to find the model parameters by minimizing a cost function ($J(y, \hat{y})$), where $\hat{y}$ denotes the output of the model (i.e. the prediction) and y the desired output (i.e. the label associated to the input x).

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[ \left( \sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left( \sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left( C_i - \hat{C}_i \right)^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left( C_i - \hat{C}_i \right)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

**Figure 4.1:** *Loss function used in YOLO v2. Figure adapted from [35].*

### 4.2.1 Loss

An important component for network training is the Loss Function, which defines the method with which the loss (i.e. the prediction error) is calculated. In other word, in the context of an optimization algorithm, the function used to evaluate a candidate solution (i.e. a set of weights) is referred to as the objective function. The goal is to maximize or minimize the objective function, meaning that it is searching for a candidate solution that has the highest or lowest score respectively. Typically, with neural networks, the aim is to minimize the error. As such, the objective function is often referred to as a cost function or a loss function and the value calculated by the loss function is referred to as simply "loss".

The Yolo v2 loss function is the one shown in the Figure 4.1.

Where:

- The 3 $\lambda$ terms are constants to take into account more one aspect of the loss function.
- The prediction of YOLO is a S x S x (B x 5 + C) vector. The 5 bounding box outputs of the box $j$ of the cell $i$ are coordinates of the center of the bounding

box $x_{ij}$, $y_{ij}$, height $h_{ij}$, width $w_{ij}$ and a confidence index $C_{ij}$.

- $1_i^{obj}$ denotes if object appears in cell $i$. It is equal to 1 when there is an object and 0 elsewhere
- $1_{ij}^{obj}$ denotes that the $j$th bounding box predictor in cell $i$ is "responsible" for that prediction. It is equal to 1 if there is an object in cell $i$ and confidence of the $j$th predictors of this cell is the highest among all the predictors of this cell.
- $1_{ij}^{noobj}$ is almost the same except it values 1 when there are NO objects in cell $i$

More general explanation of each term of the sum:

1. this term penalizes bad localization of center of cells
2. this term penalizes the bounding box with inaccurate height and width. The square root is present so that errors in small bounding boxes are more penalizing than errors in big bounding boxes.
3. this term tries to make the confidence score equal to the IOU between the object and the prediction when there is one object
4. Tries to make confidence score close to 0 when there are no object in the cell
5. This is a simple classification loss

### 4.2.2 Adaptive Moment Estimation (Adam) Optimization

Adam is an adaptive moment estimation of first-order and second-order moments. Since it adapts the learning rate $\eta$ to the parameters, performing larger updates for infrequent and smaller updates for frequent parameters [50], it is considered well suited for problems that are large in terms of data/parameters.

To be short, considering $g_{t,i}$ as the gradient of J, generic cost function, with respect to the parameter $\theta_i$ after $t$ batches. Using the Adam optimizer, every parameter after each batch is computed in this way:

$$\theta_{t+1,t} = \theta_{t,i} - \frac{\eta}{\sqrt{\hat{v}_t(g_{t,i}) + \epsilon}} \cdot \hat{m}_t(g_{t,i}) \tag{4.1}$$

where *epsilon* is an hyperparameter characterized by very small values and useful to prevent a null denominator, while $\hat{m}_t$ is the first moment (the mean) estimate of the gradient and $\hat{v}_t$ is the second moment estimate (uncentered variance) of $(g_{t,i})$.

The first and the second moment estimates are calculated in this way:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot (g_{t,i}) \tag{4.2}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot (g_{t,i})^2 \tag{4.3}$$

where the hyperparameters $\beta_1, \beta_2 \in [0, 1)$ control the exponential decay rates of moving average of the gradient $(m_t)$ and the squared gradient $(v_t)$ respectively.

The other very important parameter related to the gradient descent is the learning rate, $\eta$: if $\eta$ is too small leads to slow convergence, while if $\eta$ is too large can hinder convergence and cause the loss function to fluctuate around the minimum or even to diverge.

## 4.3 Metrics for the evaluation

When evaluating the performance of a network aimed at object detection, precise metrics are used, very different from those of the networks aimed at classification, because the variables are not Boolean, but pixels.

**Intersection over Union (IoU)**

Intersection over Union (Iou) can be considered as the basic metric for the evaluation of a network aimed at object detection; Iou measures the accuracy of the detector on a particular test set. It requires only two elements for each image to be evaluated:

- The *ground-thruth* box, that is the identifier box generated by the programmer at the moment in which the dataset is generated and represents the minimum size of the box enclosing the object.

- The *predicted* box, that is the model's ouput.

Mathematically, the parameter IoU is:

$$IoU = \frac{Area\,of\,Overlap}{Area\,of\,Union} \tag{4.4}$$

IoU is used mainly because it is impossible that the coordinates (x,y) of the predicted box are identical to those of the ground-truth box because of the multitude of parameters that are considered in the Object Detection.

**Average Precision (AP)**

Average precision (AP) is a popular metric in measuring the accuracy of object detectors. However, some concepts must be introduced before giving a definition.

- *Confidence* is the probability that a box contains the object;
- *True Positive* (TP) is an identification that meets three conditions: the confidence value greater than a certain threshold, the class corresponds to the correct one, the value of IoU is greater than a certain threshold. Consequently, *false positive* (FP) is an identification that violates one of the last two conditions
- *False Negative* (FN) is an identification that should be corrected but has a value of confidence inferior to the threshold
- *Precision* is given by the following equation:

$$precision = \frac{TP}{TP + FP} \tag{4.5}$$

- *Recall* is given by:

$$recall = \frac{TP}{TP + FN} \tag{4.6}$$

Setting multiple confidence threshold values results in different precision and recall pairs. Representing the two metrics on a Cartesian diagram we obtain the precision-recall curve. The average accuracy, which is based on this curve, is the average accuracy between all recall levels. To reduce the impact of the wave pattern on the AP value, the curve is first interpolated. Interpolated precision at a certain level r is defined as the highest precision found for each level $r' > r$:

$$p_{interp}(r) = max\ p(r') \quad with\, r' > r \tag{4.7}$$

Finally, the average precision can be defined as the area below the interpolated precision-recall curve which can be calculated with the following formula:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i)\, p_{interp}\,(r_{i+1}) \tag{4.8}$$

Where $r_1$, ... , $r_n$ are the recall levels where accuracy has been interpolated.

**Mean Average Precision (mAP)**

The AP covers a single class; as for all K classes present, we define the Mean Average Precision (mAP) as:

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{4.9}$$

## 4.4 Exclusion criteria

Since the major challenge in the state of the art is to have a good dataset, different session train of the model were made with different dataset used. The first attempt was made with all the ICA images, the other two with a reduced dataset obtain by performing some exclusion of the ICA images based on different criteria:

- Images with low contrast between coronary artery and background
- Images in which are present stenosis with a total occlusion
- Images with poor visibility of stenosis
- Images in which stenosis have unusual pattern and there is the presence of distraction such as pacemaker
- Images with wrong annotation, i.e. boxes placed in wrong position of the images or stenosis not completely included in it.

A representation of the rejected images is shown in Figure 4.2 and Figure 4.3. Based on this criteria, a manual elimination of this images was performed reducing the dataset in a heavily manner.

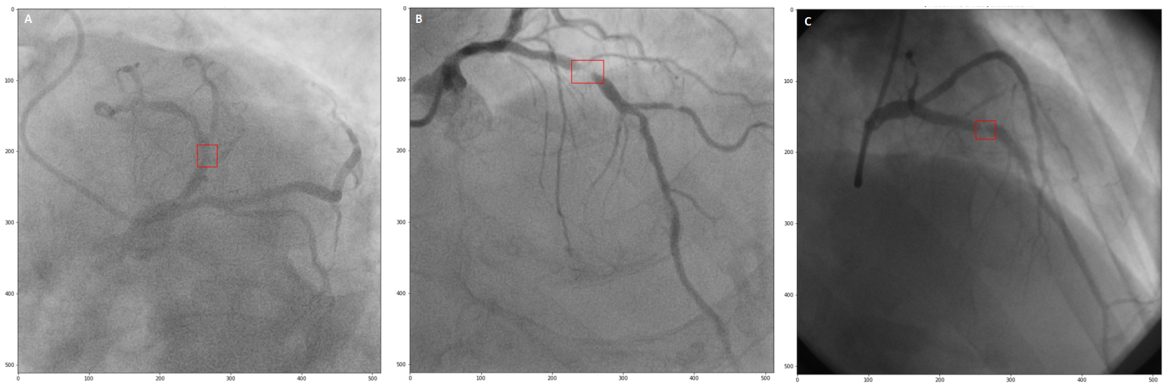


**Figure 4.2:** *Example of rejected images. Low contrast (A), total occlusion (B) and poor visibility of the stenosis (C).*

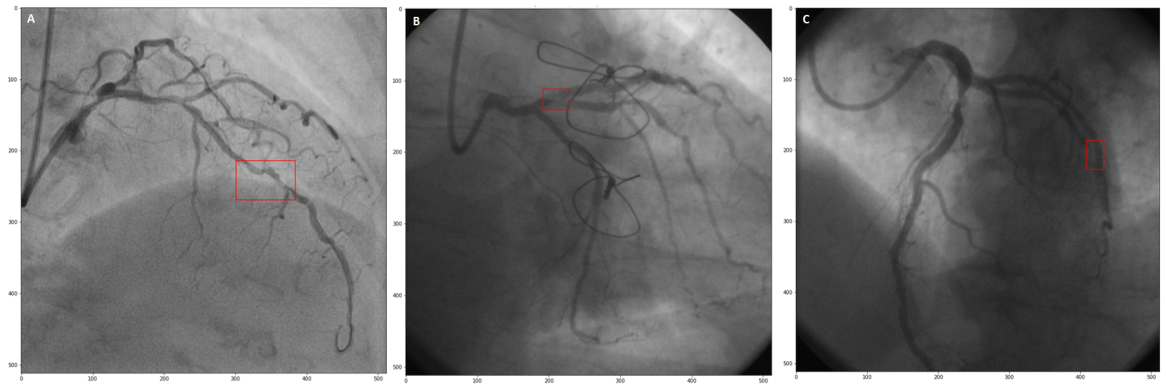


**Figure 4.3:** *Example of rejected images. Unusual pattern (A), distraction (B) and wrong annotation (C).*

## 4.5 Experiment

In Table 4.1 are shown the setting values used for the hyperparamethers used in all the three evaluation configuration. In summary, the first evaluation is made by using the entire dataset, the second one by performing a reduction of the original dataset based on some exclusion criteria shown in the following section. Finally, in the last evaluation it is performed a further reduction of the dataset and an offline augmentation of the data, in particular by operate resize and both horizontal and vertical flip of the images.

**Table 4.1:** *Hyperparameters settings used in all the trials.*

| Trial | Batch size | Momentum | Start learning rate | Optimazer | Epochs |
|---|---|---|---|---|---|
| 1 | 16 | 0.9 | $0.5 * 10^{-}4$ | Adam | 500 |
| 2 | 16 | 0.9 | $0.5 * 10^{-}4$ | Adam | 100 |
| 3 | 16 | 0.9 | $1 * 10^{-}4$ | Adam | 200 |

To increase the efficiency during the training procedure, a specific callback, named "ReduceLROnPlateau" has been applied. This function monitors the validation loss and perform a reduction of the learning rate by a factor equal to 0.1 if no improvement of the loss value is seen after 2 epochs.

Chapter 5

# RESULTS

This section summarizes the results based on experimental protocol described in Chapter 4. In particular, results of the statistical tests performed and the specific dataset used in each session will be shown in Section 5.1, while results relative to the prediction of the detection on the ICA images in Section 5.2.

## 5.1 Evaluation

To obtain an overall view of the results for the evaluation and understanding of the capability of the proposed object detection, the trends of the loss function are presented for each evaluation session. In addition, a table containing the dataset characteristics is reported. To summarize:

- **First trial**: complete dataset formed by 1373 ICA images of 207 patients;
- **Second trial**: reduced dataset formed by 740 ICA images of 154 patients;
- **Third trial**: further reduction of the dataset obtaining 164 ICA images of 54 patients. Then an augmentation of the train dataset was performed obtaining 378 ICA images.

### 5.1.1 First evaluation

**Table 5.1:** *Dataset design of the first trial.*

| | Number of patient | Number of images |
|---|---|---|
| Train | 141 | 814 |
| Validation | 25 | 240 |
| Test | 41 | 319 |

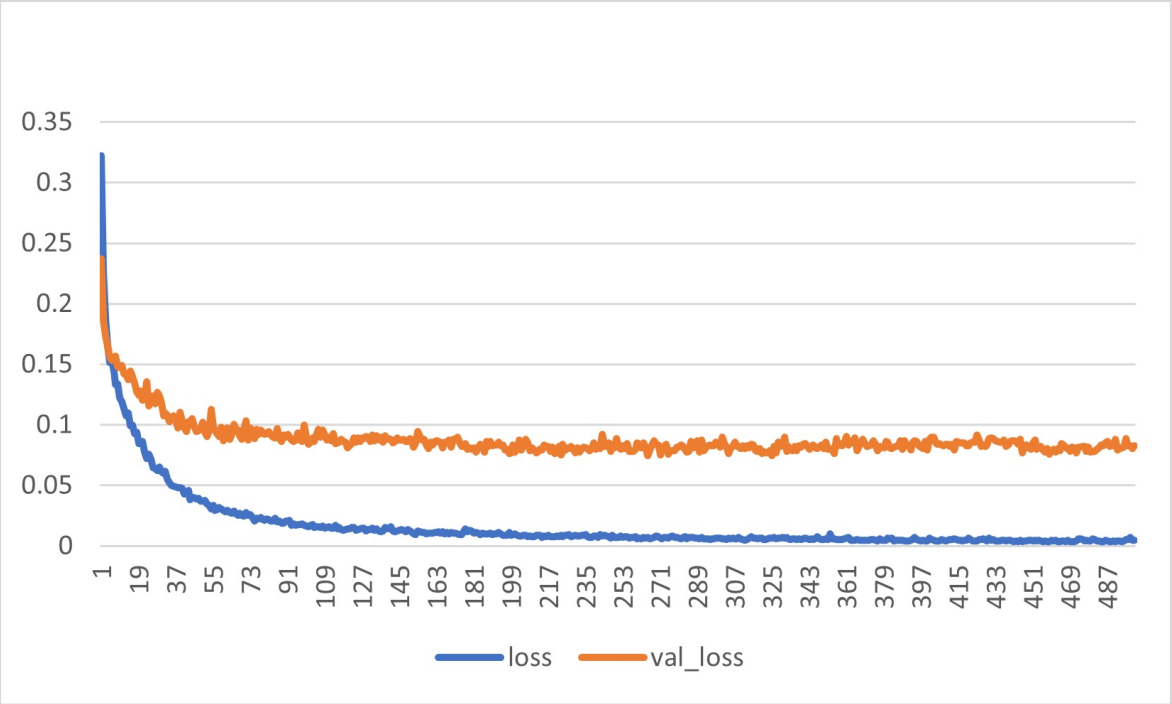


**Figure 5.1:** *Trend of the loss values for the first attempt (Complete dataset).*

### 5.1.2 Second evaluation

**Table 5.2:** *Dataset design of the second trial.*

| | Number of patient | Number of images |
|---|---|---|
| Train | 119 | 533 |
| Validation | 25 | 133 |
| Test | 10 | 74 |

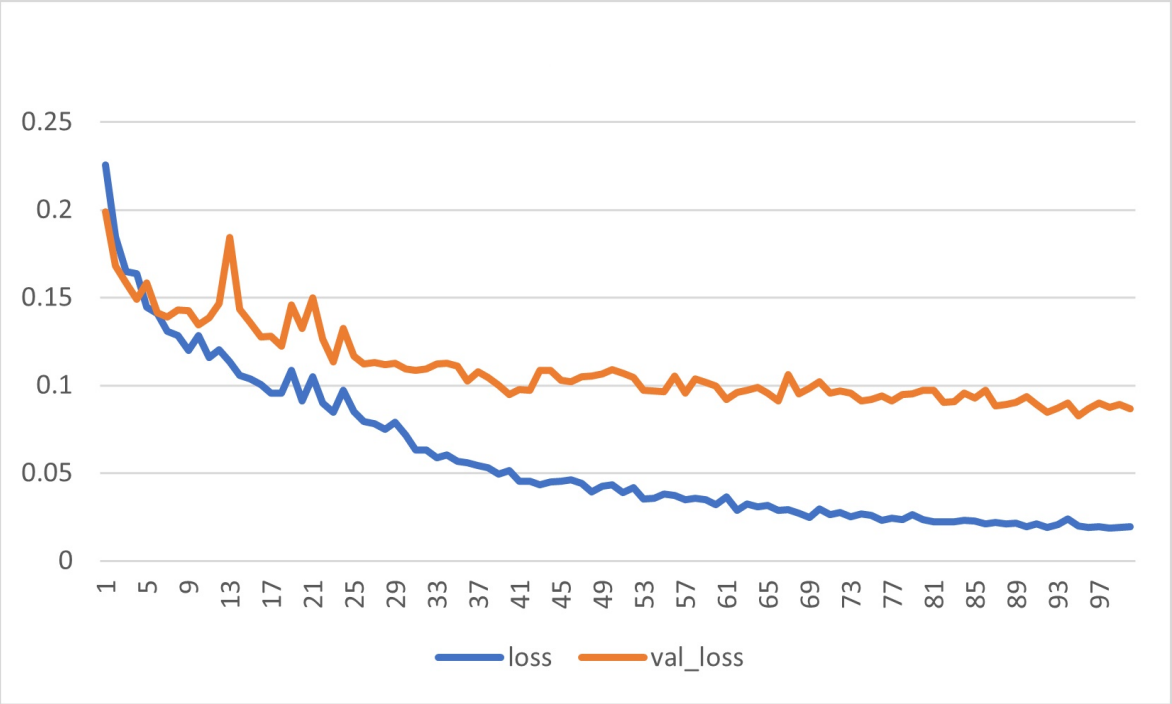


**Figure 5.2:** *Trend of the loss values for the second attempt (Reduced dataset).*

### 5.1.3 Third evaluation

**Table 5.3:** *Dataset design of the third trial.*

| | Number of patient | Number of ICA images |
|---|---|---|
| Train | 33 | 288 |
| Validation | 5 | 50 |
| Test | 16 | 40 |

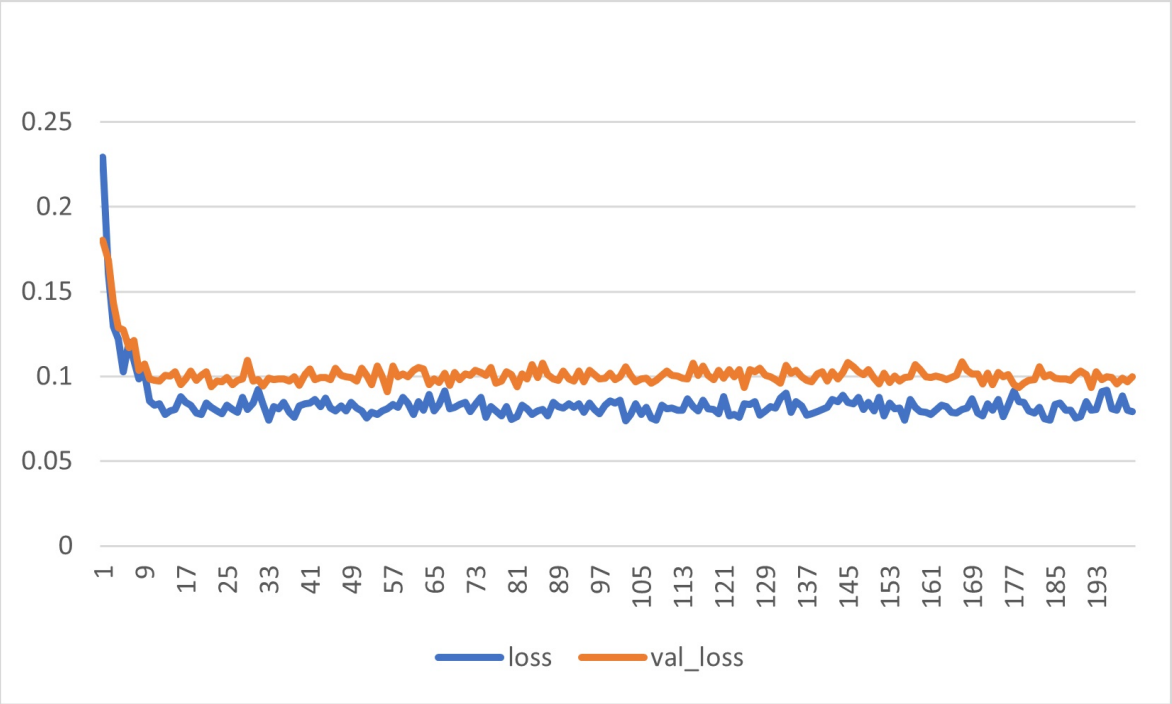


**Figure 5.3:** *Trend of the loss values for the third attempt (Reduced dataset plus augmentation).*

## 5.2 Prediction

In this section are shown all the prediction made after the training of the model on the dataset used in the three different evaluation session. The prediction of the last trial is not reported.

### 5.2.1 First evaluation

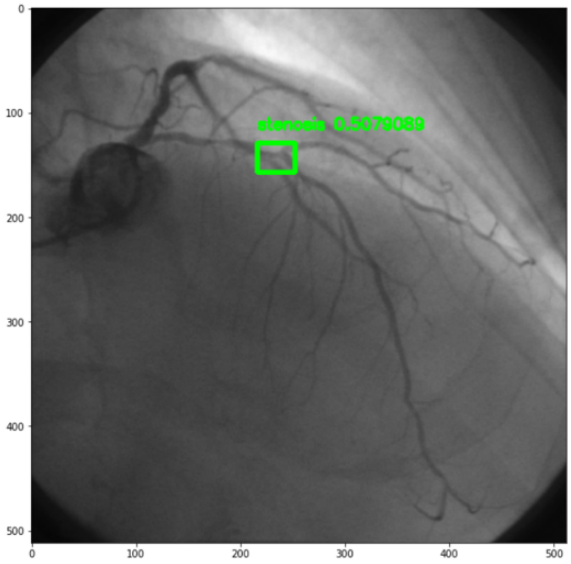


**Figure 5.4:** *Prediction first trial.*

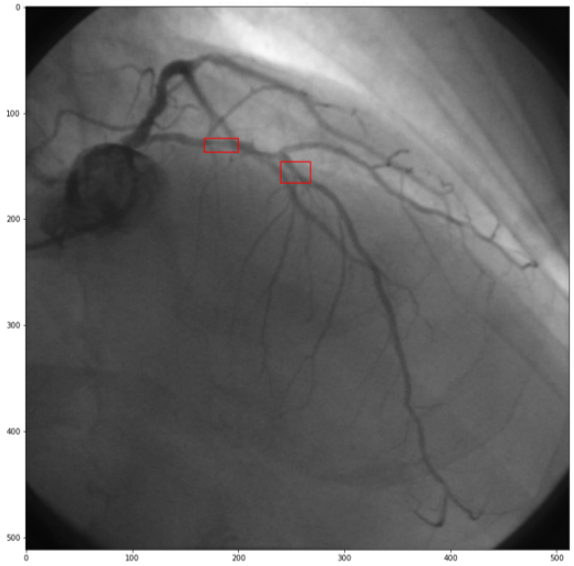


**Figure 5.5:** *Real annotation first trial.*

### 5.2.2 Second evaluation

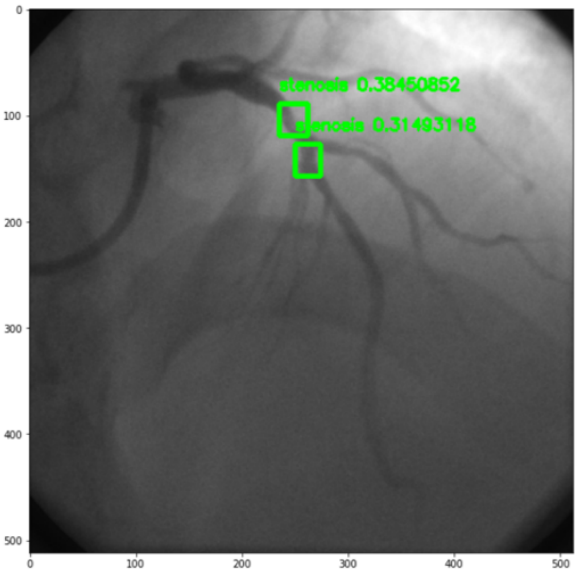


**Figure 5.6:** *Prediction second trial.*

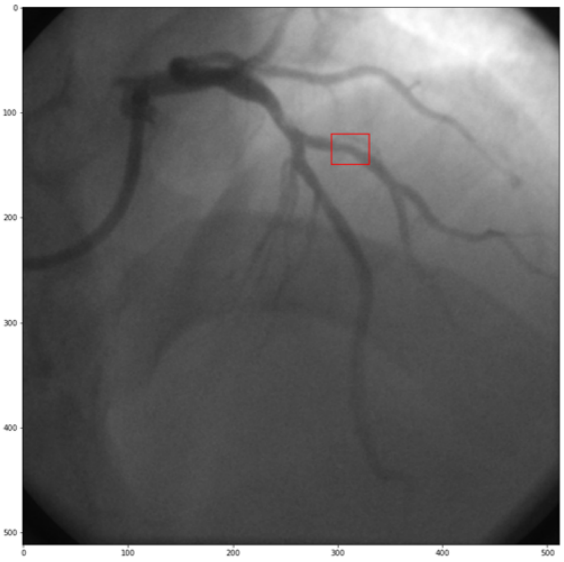

**Figure 5.7:** *Real annotation second trial.*

However, in few images the model was able to perform a right detection of the stenosis as shown in the Figure 5.8.

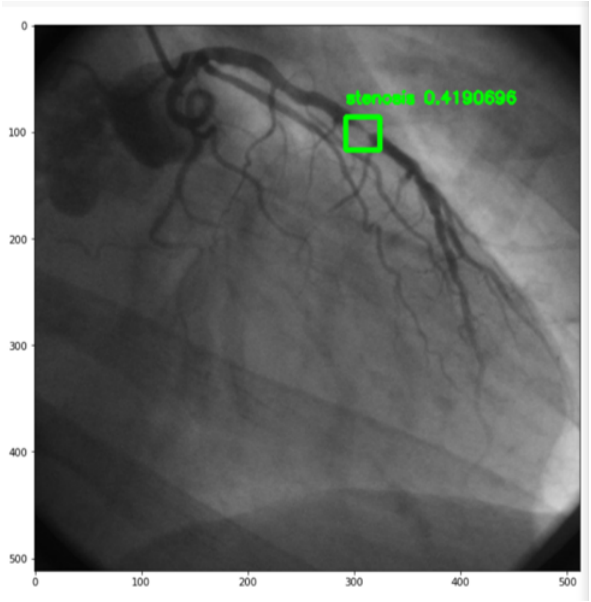


**Figure 5.8:** *Prediction second trial with the right detection.*

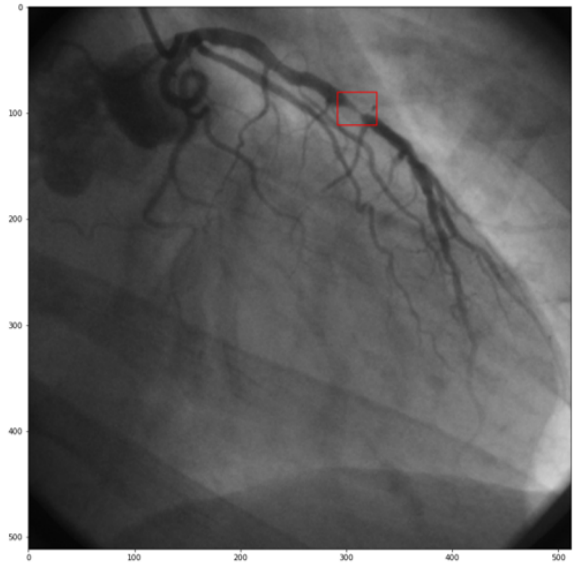


**Figure 5.9:** *Real annotation second trial with the right detection.*

Chapter 6

# DISCUSSION

In this Section results presented in Chapter 5 are discussed.

The variability of the resutlts obtained highlights the difficul task this thesis tries to face, due to the high variability of the target object to detect and the quality of the images. In fact, despite the different settings and the dataset selection in the different trials, the training of the network presented always an overfitting model, as shown by the trend of the loss in each trial, thus not generalizing correctly to unseen data. Moreover, the mAP calculated for each trial remains near the 1% thus showing a low accuracy in stenosis detection.

The bounding boxes, that define the position of the stenosis, are predicted in different location along the arteries in comparison with the ground truth, the manual annoteted position. The second trial results to be the only one to partially achieve correct predictions. Indeed, this can be this can be explained by the fact that using a more consistent dataset could be useful for the detection problem but, at the same time, the smaller dataset doesn't allow the model to predict well in all the images with the exception of image that shows clear analogies with the images used to the feed the network during the training session.

Moreover, the removal of frames from the dataset based on the criteria showed in the Chapter 4 retains the 45% of the original dataset in the second evaluation, and this greatly has affected the outcome of the training performed by using a pretrained network. For this precise reason, in the third trial, in which the greatest dataset

reduction has been applied, no prediction could be observed.

Several aspects of ICA, moreover, contribute to make this task challenging: ICA images contain significant image noise, at time including non-anatomical artifacts such defibrillator wires, and have limited image contrast. Furthermore, diseased lesions in ICA images represent a fraction of the overall image, adding a non-trivial search component to the image analysis problem.

Since, this network pretrained on natural images has been chosen for its ability to detect very well on other object detection tasks, the behaviour of the results suggests that the availability of a large and clean dataset has a fundamental role. In medical field, a very high number of images are acquired and stored everyday, but most of the time their quality is not sufficient to treat them as with natural images. Thus, even the most reliable detection network results to be ineffective on images with highly variability in patterns as targets and without considering an image enhancement through post-processing procedures.

Chapter 7

# CONCLUSIONS

With rapid development of DL, vascular extraction and stenosis detection technology-based AI is emerging as a fast and efficient diagnostic tool to detect CAD. However, all object detection applications in this particular medical field result to be effective only through a strong preprocessing of the images, enhancing the visibility of the target and quality of the images, thus reaching good results from a computational point of view, but unrealistically effective in clinical.

This thesis investigates the possibility for a common and reliable object detection network, like YOLOv2, to predict stenosis location based on raw images, as they are actually inspected by physicians with ICA.

Indeed, the largest part of DL methods presented in literature is applied on a single frame selected from an ICA frame-sequence to keep the image with the best resolution, and then, to obtain a contrast standardization of the ICA images, an intense preprocessing is performed before feeding them into the network.

In this work the YOLO v2 object detector model was trained on ICA images without performing a preprocessing altering their original characteristics and moving away from the real context in which the stenosis detection is significant.

Furthermore, the results show that dataset still remains the main problem when training a model on new data, especially in relation with particular medical contexts. Even though the storage and diffusion of digital biomedical images is rapidly increased, the detection task still presents several challenges for automated strategy mainly due

to the quality of the images and the variety of patterns that can characterize the same biological structure, like the stenosis in this case.

Despite all the advancements that DL technologies achieve everyday in many fields, the results of this work highlight that the medical images require a different processing, due to the meaning that they carry with them, and at least for this application we are still far to get a support tool that could be actually used in clinical practice. However, exploring this complex field with the latest technologies, experimenting different architectures over larger dataset and considering different frame-sequence or video strategies, can be a good starting point to obtain a support tool for physicians that could actually be deployed during ICA procedures.

# List of Abbreviations

CAD: Coronary artery disease

DICOM: Digital Imaging and Communication in Medicine

LCA: Left main coronary artery

RCA: Right coronary artery

PCI: Percutaneous coronary intervention

CABG: Coronary artery bypass graft

CCTA: Coronary computed tomographic angiography

ICA: Invasive coronary angiography

DL: Deep learning

PVA: Physician visual assessment

QCA: Quantitative coronary angiography

ML: Machine learning

CNN: Convolutional neural network

RNN: Recurrent neural network

LSTM: Long-short-term memory

DSSD: Deconvolutional Single-Shot multibox Detector

AI: Artificial intelligence

R-CNN: Region-Based Convolutional Network

YOLO: You only look once

VOC: Visual Object classification

ADAM: Adaptive Moment Estimation

IoU: Intersection over Union

AP: Average Precision

TN: True Negative

TP: True positive

FP: False positive

FN: False negative

mAP: Mean Average Precision

PPV: positive predictive value

# List of Figures

# List of Tables

# Bibliography


[1] A. B. I. Abubakar, T. Tillmann, "Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013:a systematic analysis for the global burden of disease study 2013," *Lancet 385 (9963)*, vol. 117, no. 171, 2015.

[2] S. Ashiq, K. Ashiq, S. Saleem, M. Shahid, H. Qayyum, and H. Shafi, "Prevalence and role of different risk factors with emphasis on genetics in development of pathophysiology of coronary artery disease (cad)," vol. 52, pp. 279–287, 01 2019.

[3] Z. Salahuddin, M. Lenga, and H. Nickisch, "Multi-resolution 3d convolutional neural networks for automatic coronary centerline extraction in cardiac ct angiography scans," 2020.

[4] C. B. Compas, T. Syeda-Mahmood, P. McNeillie, and D. Beymer, "Automatic detection of coronary stenosis in x-ray angiography through spatio-temporal tracking," *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 2014.

[5] E. Maffei, A. Palumbo, C. Martini, F. Notarangelo, C. Saccò, F. Ugo, D. Lina, A. Aldrovandi, C. Reverberi, C. Manca, and et al., "Predictive value of computed tomography coronary angiography for the evaluation of acute chest pain: single center preliminary experience," 2011.

[6] A. K. Malakar, D. Choudhury, B. Halder, P. Paul, A. Uddin, and S. Chakraborty, "A review on coronary artery disease, its risk factors, and therapeutics," *J Cell Physiol*, p. 1–12, Jan 2019.

[7] M. Chen, X. Wang, G. Hao, X. Cheng, C. Ma, N. Guo, S. Hu, Q. Tao, F. Yao, C. Hu, and et al., "Diagnostic performance of deep learning-based vascular extraction and stenosis detection technique for coronary artery disease," *The British Journal of Radiology*, vol. 93, no. 1113, p. 20191028, 2020.

[8] D. H. Fitchett, M. Gupta, M. E. Farkouh, and S. Verma, "Coronary artery revascularization in patients with diabetes mellitus," *Circulation*, vol. 130, no. 12, 2014.

[9] S. F. Afrin, M. H. Rahman, M. A. Millat, M. S. Quarashi, A. Begum, and M. Begum, "Minor myocardial injury: An early post intervention complication," *Delta Medical College Journal*, vol. 5, no. 2, p. 94–98, 2017.

[10] N. H. Pijls and J.-W. E. Sels, "Functional measurement of coronary stenosis," *Journal of the American College of Cardiology*, vol. 59, no. 12, p. 1045–1057, 2012.

[11] F. Cademartiri, E. Maffei, F. Notarangelo, F. Ugo, A. Palumbo, D. Lina, A. Aldrovandi, E. Solinas, C. Reverberi, A. Menozzi, L. Vignali, R. Malagò, M. Midiri, N. Mollet, G. Cervellin, and D. Ardissino, "64-slice computed tomography coronary angiography: Diagnostic accuracy in the real world," *La Radiologia medica*, vol. 113, pp. 163–80, 04 2008.

[12] M. Rief, M. Y. Chen, A. L. Vavere, B. Kendziora, J. M. Miller, W. P. Bandettini, C. Cox, R. T. George, J. Lima, M. Di Carli, and et al., "Coronary artery disease: Analysis of diagnostic performance of ct perfusion and mr perfusion imaging in comparison with quantitative coronary angiography and spect—multicenter prospective trial," *Radiology*, vol. 286, no. 2, p. 461–470, 2018.

[13] R. Shahzad, H. Kirişli, C. Metz, H. Tang, M. Schaap, L. van Vliet, W. Niessen, and T. van Walsum, "Automatic segmentation, detection and quantification of coronary artery stenoses on cta," *The International Journal of Cardiovascular Imaging*, vol. 29, no. 8, p. 1847–1859, 2013.

[14] A. C. Weustink and P. J. de Feyter, "The role of multi-slice computed tomography in stable angina management: a current perspective," *Netherlands Heart Journal*, vol. 19, no. 7-8, p. 336–343, 2011.

[15] B. Au, U. Shaham, S. Dhruva, G. Bouras, E. Cristea, A. Lansky, A. Coppi, F. Warner, S.-X. Li, and H. Krumholz, "Automated characterization of stenosis in invasive coronary

angiography images with convolutional neural networks," *ArXiv*, vol. abs/1807.10597, 2018.

[16] A. M. Adeoye, A. N. Adekunle, A. A. Adebiyi, A. Mullassari, S. Vijayakumar, and C. E. Nwafor, "A 45-year old man with recurrent syncope: an unusual presentation of coronary artery disease," *Pan African Medical Journal*, vol. 2, 2013.

[17] W. Wu, J. Zhang, H. Xie, Y. Zhao, S. Zhang, and L. Gu, "Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint," *Comput. Biol. Med.*, vol. 118, Mar. 2020.

[18] T. Wan, H. Feng, C. Tong, D. Li, and Z. Qin, "Automated identification and grading of coronary artery stenoses with x-ray angiography," *Computer Methods and Programs in Biomedicine*, vol. 167, p. 13–22, 2018.

[19] M. L. Marcus, D. J. Skorton, M. R. Johnson, S. M. Collins, D. G. Harrison, and R. E. Kerber, "Visual estimates of percent diameter coronary stenosis: "a battered gold standard"," *Journal of the American College of Cardiology*, vol. 11, no. 4, p. 882–885, 1988.

[20] H. Zhang, L. Mu, S. Hu, B. K. Nallamothu, A. J. Lansky, B. Xu, G. Bouras, D. J. Cohen, J. A. Spertus, F. A. Masoudi, and et al., "Comparison of physician visual assessment with quantitative coronary angiography in assessment of stenosis severity in china," *JAMA Internal Medicine*, vol. 178, no. 2, p. 239, 2018.

[21] M. Raphael and R. Donaldson, "A "significant" stenosis: Thirty years on," *The Lancet*, vol. 333, no. 8631, p. 207–209, 1989.

[22] L. L. Leape, R. E. Park, T. M. Bashore, J. Harrison, C. J. Davidson, and R. H. Brook, "Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures," *American Heart Journal*, vol. 139, no. 1, p. 106–113, 2000.

[23] P. Garrone, G. Biondi-Zoccai, I. Salvetti, N. Sina, I. Sheiban, P. R. Stella, and P. Agostoni, "Quantitative coronary angiography in the current era: Principles and applications," *Journal of Interventional Cardiology*, vol. 22, no. 6, p. 527–536, 2009.

[24] F. A. Pinton, B. d. Falcão, J. Mariani, L. J. Kajita, A. E. Filho, and P. A. Neto, "Accuracy and precision of online quantitative coronary angiography with automatic

calibration: a pilot study," *Revista Brasileira de Cardiologia Invasiva (English Edition)*, vol. 23, no. 1, p. 58–60, 2015.

[25] B. K. Nallamothu, J. A. Spertus, A. J. Lansky, D. J. Cohen, P. G. Jones, F. Kureshi, G. J. Dehmer, J. P. Drozda, M. N. Walsh, J. E. Brush, and et al., "Comparison of clinical interpretation with visual assessment and quantitative coronary angiography in patients undergoing percutaneous coronary intervention in contemporary practice," *Circulation*, vol. 127, no. 17, p. 1793–1800, 2013.

[26] S. Yang, J. Kweon, J.-H. Roh, J.-H. Lee, H. Kang, L.-J. Park, D. J. Kim, H. Yang, J. Hur, D.-Y. Kang, and et al., "Deep learning segmentation of major vessels in x-ray coronary angiography," *Scientific Reports*, vol. 9, no. 1, 2019.

[27] M. J. R. Fatemi, S. M. Mirhassani, and E. Ghasemi, "Article: Detection of narrowed coronary arteries in x-ray angiographic images using contour processing of segmented heart vessels based on hessian vesselness filter and wavelet based image fusion," *International Journal of Computer Applications*, vol. 36, pp. 27–33, December 2011. Full text available.

[28] J. Brieva, M. Galvez, and C. Toumoulin, "Coronary extraction and stenosis quantification in x-ray angiographic imaging," *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004.

[29] R. Rosati, L. Romeo, S. Silvestri, F. Marcheggiani, L. Tiano, and E. Frontoni, "Faster r-cnn approach for detection and quantification of dna damage in comet assay images," *Computers in Biology and Medicine*, vol. 123, p. 103912, 2020.

[30] W. Huang, L. Huang, Z. Lin, S. Huang, Y. Chi, J. Zhou, J. Zhang, R.-S. Tan, and L. Zhong, "Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018.

[31] G. Litjens, F. Ciompi, J. M. Wolterink, B. D. de Vos, T. Leiner, J. Teuwen, and I. Išgum, "State-of-the-art deep learning in cardiovascular image analysis," *JACC: Cardiovascular Imaging*, vol. 12, no. 8, p. 1549–1565, 2019.

[32] K. Shameer, K. W. Johnson, B. S. Glicksberg, J. T. Dudley, and P. P. Sengupta, "Machine learning in cardiovascular medicine: are we there yet?," *Heart*, vol. 104, no. 14, p. 1156–1164, 2018.

[33] C. Cong, Y. Kato, H. D. Vasconcellos, J. Lima, and B. Venkatesh, "Automated stenosis detection and classification in x-ray angiography using deep neural network," *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019.

[34] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "Dssd : Deconvolutional single shot detector," Jan 2017.

[35] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[36] H.-J. Jang and K.-O. Cho, "Applications of deep learning for the analysis of medical data," *Archives of Pharmacal Research*, vol. 42, no. 6, p. 492–504, 2019.

[37] I. El Naqa and M. J. Murphy, "What is machine learning?," *Machine Learning in Radiation Oncology*, p. 3–11, 2015.

[38] K. Raza and N. K. Singh, "A tour of unsupervised deep learning for medical image analysis," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 17, 2021.

[39] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, p. 60–88, 2017.

[40] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," 2019.

[41] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2014.

[42] R. Girshick, "Fast r-cnn," 2015.

[43] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, p. 1137–1149, 2017.

[44] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[45] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018.

[46] M.-T. Pham, L. Courtrai, C. Friguet, S. Lefèvre, and A. Baussard, "Yolo-fine: One-stage detector of small objects under various backgrounds in remote sensing images," *Remote Sensing*, vol. 12, no. 15, p. 2501, 2020.

[47] Seong, Song, Yoon, Kim, and Choi, "Determination of vehicle trajectory through optimization of vehicle bounding boxes using a convolutional neural network," *Sensors*, vol. 19, no. 19, p. 4263, 2019.

[48] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, 2016.

[49] NEMA, "Digital imaging and communications in medicine (dicom)," *The National Electrical Manufacturers Association (NEMA) Standards Publication PS 3.x–2008. Rosslyn, VA, 1992, 1993, 1994, 1995.*, 2008.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.