Università Politecnica Delle Marche

Department of Information Engineering

MASTER DEGREE IN BIOMEDICAL ENGINEERING

A Deep Learning Architecture for Stenosis Detection in Invasive Coronary Angiography Images

Supervisor:

Prof. Emanuele Frontoni

Candidate:

Eleonora Ceroni

Co-supervisors:

Prof. Lorenzo Scalise

Sara Moccia, PhD

Academic Year 2020 / 2021

C'è una forza motrice più forte del vapore, dell'elettricità

e dell'energia atomica: la volontà.

A. Einstein

Contents

Abstract

Coronary artery disease (CAD) is an atherosclerotic narrowing of the coronary artery lumen, that leads to angina or acute myocardial infarction. CAD is a leading cause of death and its prevalence has been increasing rapidly, especially in developing countries. It involves the presence of stenosis when the coronary arteries are narrowed or blocked induced by the atheromatous plaques building up inside, reducing oxygen-rich blood flow to the heart muscle and subsequently resulting in an imbalance state between oxygen demand and supply.

Invasive Coronary Angiography (ICA) is considered the reference gold standard imaging technique for the assessment of clinically significant CAD, which enables to reveal the initial CAD symptoms by the morphological features of the coronary arteries such as diameter, length, branching angle, and tortuosity. The challenging task for the interpretation of ICA images depends on complex vessel structure, image noise, poor contrast, and non-uniform illumination. In addition, the manual detection of stenosis is subjective and time-consuming, requiring rich clinical experience. Therefore, developing an ICA-based automatic detection algorithm could improve diagnostic efficiency and could provide huge support to clinicians, reducing bias and subjective interpretations.

Several approaches have been proposed in the literature for stenosis detection in ICA images, but these computer-based approaches owe their good results to strong pre-processing techniques. To reduce manual burden in the detection and quantification of coronary stenosis, deep-learning (DL) has been introduced.

Therefore, an end-to-end stenosis detection convolutional neural network (CNN) capable of automatically detect stenosis in ICA images is proposed. The

CNN for object detection chosen to be trained on the provided dataset is the Single-Shot Multibox Detector (SSD), since it could be adequate for the limited dataset and the number of parameters to train reduced.

The approach was validated on 5 models obtained through 5-fold cross-validation, to compensate for the limited dataset dimension and performance comparison of two different network versions, SSD300 and SSD7, is performed.

This approach has been shown to be quite good with mean values of Intersection over Union and Dice Similarity Coefficient (DSC) of 0.50 ± 0.06 and 0.64 ± 0.06 for SSD300 respectively and 0.30 ± 0.07 and 0.44 ± 0.08 for SSD7 respectively. Results of this work are good, despite they are slightly lower than values obtained in the literature, when performance metrics are comparable. It is necessary to highlight that the results depend on the annotations, subject to variability, since are drawn manually by operators. Moreover, it is possible to accept values with a certain margin of tolerance, since the aim is to identify coronary stenosis, which is correctly done by visual inspection of the stenosis prediction, except for some cases. Finally, this approach does not apply pre-processing on images, differently from reference works.

This work is a good starting point to improve the stenosis detection in ICA images but needs some improvements to be used directly in clinical, minimizing the risk of misinterpretation, and accelerating the decision-making regarding the proper treatment strategy.

Chapter 1

Introduction

1.1 Coronary Artery Disease

Coronary artery disease (CAD) is the most common type of cardiovascular disease and a major cause of mortality, resulting in an estimated 7.6 million deaths every year all over the world [1]. The main cause of CAD is atherosclerotic plaque accumulation in the epicardial arteries leading to a mismatch between myocardial oxygen supply and myocardial oxygen demand and commonly resulting in ischemia. It involves the presence of stenosis, which is a plaque of fat, cholesterol or other substances that can deposit and grow on the inner walls of a coronary artery. Thus, the lumen of the artery becomes narrower and less blood flows through the arteries, making it unable to supply sufficient blood or oxygen to the heart muscle, leading to permanent heart damage, such as heart attacks [2]. Therefore, it is important to identify and quickly treat the stenosis, before the heart is severely damaged.

Two main arteries branch off the aorta namely Left Main Coronary Artery (LCA) and Right Coronary Artery (RCA) which supply blood to the 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. An example of stenosis can be

seen in (Fig. 1.1).

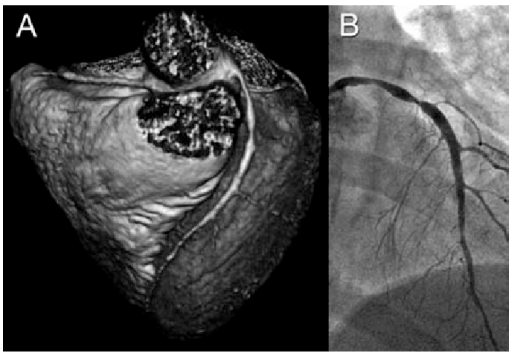


Figure 1.1: *Example of significant stenosis of the left anterior descending coronary artery from focal non calcific atherosclerosis with Computed Tomographic Coronary Angiography(A) and Invasive Coronary Angiography(B). Figure adapted from [4]*

Chest pain is the most likely symptom that occurs during physical and/or emotional stress, relieved promptly with rest or by taking nitroglycerin. This process can be modified by lifestyle adjustments, pharmacological therapies, and invasive interventions designed to achieve disease stabilization or regression [5]. 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) or to entirely bypass the blocked segment of artery surgically (bypass surgery) as illustrated in (Fig. 1.2) [6].

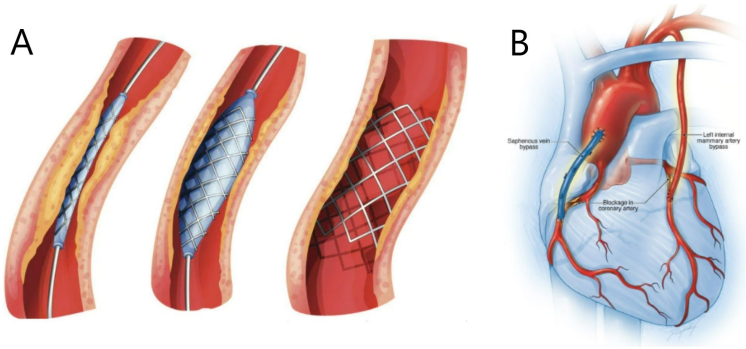


Figure 1.2: *Procedure of Percutaneous Coronary Intervention(A). Procedure of Coronary Artery Bypass Graft Surgery(B)*

Percutaneous coronary intervention (PCI; angioplasty) is a mini-

mally 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 repeated 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, two medications that inhibit blood platelets are needed for up to 1 year after the procedure (usually aspirin plus an additional blood thinner) [6].

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 through 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 [6].

However, it is of main importance to assess stenosis functional significance for decision making in the interventional catheterization laboratory with respect to revascularization [7].

1.2 Medical Images for CAD assessment

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 Computed Tomographic Coronary Angiography (CTCA) and Invasive Coronary Angiography (ICA) [8].

1.2.1 Computed Tomographic Coronary Angiography (CTCA)

CTCA, shown in (Fig. 1.3) is a non-invasive method that combines a CT scan and a iodine-based contrast medium to examine the arteries that supply blood to the heart. The images that are generated during the scan can be constructed to create a three-dimensional image, which may be viewed on a monitor or printed on a film.

The patient is connected to a cardiac monitor which shows the heart's electrical activity and an intravenous line is placed into an arm vein. During the scan, x-rays pass through the body and are picked up by special detectors in the scanner. During the procedure, contrast medium is introduced to the bloodstream to allow a clear definition of the blood vessels under examination by making them appear bright white. A beta blocker may also be administered to the patient through the same IV line, or given orally to help slow down the heart rate, while helping in improving image quality.

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) [9]. 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. Other limitations are the calcification blooming artefacts, the limited spatial and temporal resolution, the unpredictability of hemodynamic significance of intermediate coronary lesions and the difficulties to acquire motion-free, high-quality images in patients with arrhythmias [3].

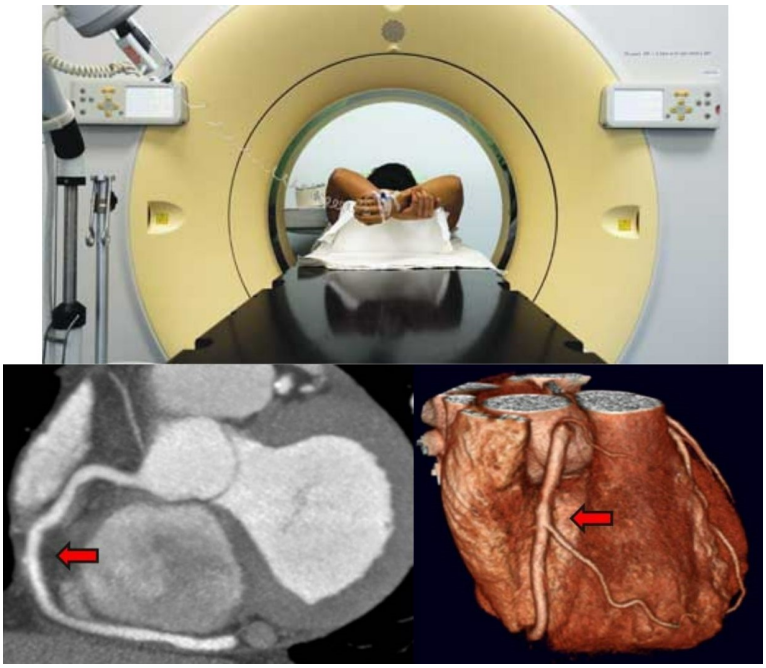


Figure 1.3: *In the upper panel the machinery of the CTCA. In the lower panel the CTCA image processed of the RCA.*

1.2.2 Invasive Coronary Angiography (ICA)

ICA, shown in (Fig. 1.4),i.e. the in vivo contrast study of the coronary artery tree and its lumen, is commonly used to investigate the anatomy of the coronary arteries and to assess the number, location, and severity of coronary stenosis. It is a medical imaging technique that involves continuous X-ray (i.e. fluoroscopy) with simultaneous injection of radiopaque contrast into the coronary arteries [10]. Despite novel imaging modalities, like CTCA, have been developed, ICA is currently regarded as the gold standard for coronary artery stenosis detection, according to the 2019 guidelines of the European Society of Cardiology [5]. With an injected contrast agent, ICA 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 [11]. During the procedure a long, thin, flexible catheter (a thin hollow tube with a diameter of 2–3 mm) is inserted into a blood vessel in the groin or 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 a contrast medium is injected through the catheter and X-ray images (angiograms) are taken. Both the right and left coronary arteries are injected multiple times after changing the position of the X-ray system to visualize the coronary tree from different perspectives. The sequential images that are produced over the 3–6 s acquisition time are called selective coronary angiograms, and the recorded images, if carefully and comprehensively gathered with skill, accurately reveal the extent and severity of all coronary arterial blockages. The contrast medium is visible on the angiograms, showing the blood vessels the fluid travels through. This clearly highlights any blood vessels that are narrowed or blocked. The procedure is usually carried out under local anaesthetic, so the patient is awake during the procedure, but the area where the catheter is inserted will be numbed. This reveals the structure of arteries on X-ray images allowing to visually detect stenosis and other visible abnormalities by a clinician. ICA is a fair and reliable diagnostic method, as well as monitoring the progress of revascularization during angioplasty.

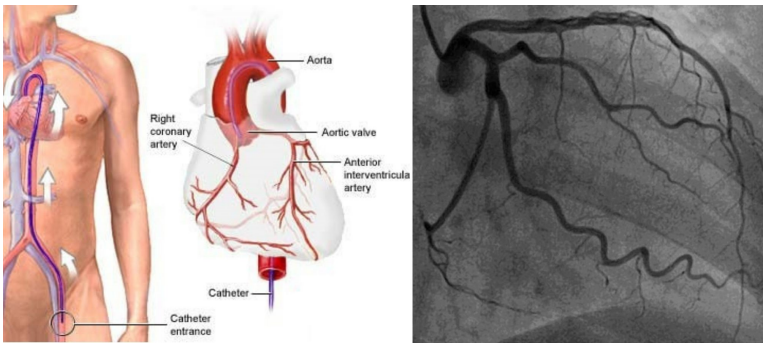


Figure 1.4: *On the left panel a representation of ICA procedure. On the right imaging of the left anterior descending dissection by ICA.*

1.2.3 Criticality in the Diagnosis

The fundamental tasks required for the interpretation of coronary angiography are the identification and the quantification of the severity of stenosis within the coronary circulation. Stenosis severity is typically determined in clinical practice by physician visual assessment. 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 a real stenosis will persist in each image frame through time. This approach, while considered as clinical standard, has known limitations, such as significant intra- and inter-rater variability, as well as high positive prediction bias. In addiction, the manual estimations of coronary artery stenosis are time consuming and the results from these readings are also dependent on subjective training. Thus, the introduction of a tool to automatically detect stenosis from images could provide a huge support to the clinicians, as it can not only eliminate the variability of intra- and inter-observers, but also provide a second opinion for cardiologists in enhancing the operational efficiency and productivity as well as improving diagnostic confidence.

1.3 Disclosure

The current imaging techniques employed to diagnose CAD is known to be significantly affected by intra- and inter-observer variation, due to the presence of other tissues, camera movements and the difficulties associated to the visual identification of the stenosis. To overcome all these limitations a support diagnostic tool in clinical practice is strongly suggested. In this work the emerging Deep Learning (DL) techniques have been chosen to develop a system for stenosis detection from ICA images automatically. Specifically, a Single Shot Multibox Detector has been developed, considering it to be suitable for the present clinical context. This work is carried out in collaboration with the Cardiology Department of Ospedali Riuniti in Ancona: all the experiments and considerations done in this work are 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

The process of interpreting complex coronary vasculature, image noise, low contrast vessels, and non-uniform illumination is time-consuming, thereby posing certain challenges to the operator. Automatic CAD detection and labeling may overcome these difficulties by supporting the decision-making process [5].

In this Chapter, a literature review about the conventional methods for stenosis detection in ICA images is presented, since the ICA is currently regarded as the gold standard for coronary artery stenosis detection. Moreover, this work tries to investigate the possibility of obtaining a CNN capable of automatically detect stenosis in ICA images. 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

Over the course of years, several methods have been proposed for stenosis detection: the existing methods based on ICA are relatively few compared to those

based on the CTCA, since the detection in ICA images is challenging, due to the low contrast between vessels and surrounding tissues as well as the complex overlap of background structures with inhomogeneous intensities [11].

2.1.1 Visual Assessment

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 revascularization [12]. 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. By definition, significant CAD is the narrowing of coronary artery lumen beyond the threshold of 50% diameter reduction in coronary angiography images, thus assessing the physiologic significance of narrowings of intermediate severity is uncertain. To examine the degree and severity of coronary artery stenosis in clinical practice, cardiology specialists select the key frames, and assess the stenosis manually by naked-eye visual inspection. The selected key frames are those in which the vessels are more visible, as the contrast agent passes through the arteries. However, the visual inspection of coronary artery stenosis is known to be significantly affected by intra- and inter-observer variation and time-consuming, requiring rich clinical experience and expert knowledge [11]. In addition, it is challenging due to camera movements, complex vessel structures, poor contrast between vessels and surrounding tissues, nonuniform illumination, and overlap of background structures with inhomogeneous intensities [13].

There are two different approaches to visually evaluate structure anatomy in ICA images [14]. The most common approach depends on assumptions about the geometry of structures and vessels, relying on the exact definition of vessel edges and requiring information on the luminal shape. Therefore, inaccuracies might occur in the evaluation of lesions with indistinct vessel edges or irregular luminal shapes. Although investigators have been working to make geometric

methods of angiographic analysis easier and faster, many geometric methods remain time consuming and difficult, and cannot be readily applied at the time of cardiac catheterization [15]. Percent stenosis may not accurately reflect lesion physiologic significance because the angiographically "normal" segment of the coronary artery may not be free of atherosclerotic disease.

The second and nongeometric approach 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 structure. It consists in measuring the x-ray attenuation through a contrast-filled coronary artery and from this information estimating the amount of contrast in the vessel, which is an index of luminal size. It can be used to estimate an absolute as well as relative (i.e., percent stenosis) index of coronary luminal size. In addition, densitometry is not dependent on the exact definition of vessel edges and it is theoretically free of assumptions as to shape of the lumen. Such characteristics may be important in analyzing lesions immediately after thrombolysis or angioplasty, when vessel borders are indistinct and the luminal shape is irregular [15].

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 [16]. Consequently, even though PVA is considered a clinical standard, it has known limitations, implicitly leading to over-utilization of clinical services [12].

2.1.2 Computer-based image analysis

Since the late 1980s, 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. 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 research and, in selected cases, clinical practice [17]. 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.

QCA 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. QCA is currently based on the use of a specific and dedicated software that allows determination of some specific measures of coronary lumen in an operator-independent way. 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. 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 stenosis 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. 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. The first step consists of calibration in which a catheter of a known diameter, often expressed in French, is measured. A central line is drawn by hand along the stretch of the catheter tip. The software automatically recognizes its margins by using specific algorithms that, by means of digital images, recognize the change from radiopaque pixels (black or dark grey)to radiotransparent pixels (light grey or white) according to a densitometric analysis. The software is in this way able to transform every pixel into a square with sides

characterized by known dimensions (mm). This parameter is then employed to measure the coronary region of interest [17].

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 acquisition and minimal user input to identify imaging frames for analysis, which may introduce variability as well [18]. Limitations related to QCA include dependency on ICA image quality and challenges in assessing complex lesions, such as those with thrombus or calcification [12]. 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 [17].

Semiautomatic methods require human interaction to locate a stenosis and only focus on assessing its severity. After vessel structures is extracted with different methods and the diameter of the target stenosis was manually measured, the severity is evaluated. In Brieva et al. [19] a semiautomatic segmentation technique based on a B-spline snake to extract the 2D coronary arteries is proposed. A string matching technique is then applied to compute the vessel diameter. A narrowing measure is then presented to evaluate the stenosis severity. deformable spline with string matching [19]. In Fatemi et al. [20] firstly Hessian based vessel enhancement filter (HBVF) is applied to the angiographic frame to be segmented for enhancement of vessel structures. Meanwhile, the filter is applied to other frames in the angiographic sequence. Afterward, angiographic frames in the sequence and their vesselness filtered versions are fused using 2-D wavelet transform to make an image which is used as a threshold for detecting the vessels. Next, for detecting the vascular structures, the fused image accompanying with a couple of thresholds is applied to the vesselness filtered frame

to be segmented. Finally, in order to find the narrowed coronary arteries, vessel contours processing is employed for the measurement of the vessel thickness and Hessian vesselness filter with wavelet-based image fusion [20].

All the above-mentioned methods are defined semi-automated software because they are computer-based image analysis tools developed to support identification procedure, ignoring the automatic detection task. These methods speed up the analysis procedure but require human interaction to locate a stenosis and only focus on assessing its severity.

2.2 Advanced stenosis detection via Deep Learning technique

Fully-automatic methods concentrate on achieving automatic stenosis detection, employing Computer Vision and/or Machine Learning (ML) techniques, avoiding user dependency and thus providing standardized and reproducible measurements [11, 21]. Even if the expert is capable of assessing stenosis in coronary vessels by visual inspection, this approach becomes very time-consuming in case of a huge amount of images. Moreover, the operator dependency results in the subjectivity of the stenosis quantification and thus in a high inter/intra-operator variability. In order to overcome these issues, computer-based image analysis tools are developed to support stenosis detection procedure. Despite a good performance, fully-automatic methods still have limitations: multiple pre-processing procedures, such as vessel enhancement, segmentation and skeleton extraction are performed in a single frame to detect stenosis. This is a time-consuming process, and the intrinsic complexities of ICA images prevent these substeps from obtaining ideal intermediate results. As a consequence, errors might accumulate, hampering final stenosis detection. To avoid the cumbersome pre-processing steps required for a single frame in traditional stenosis detection methods, CNNs have been introduced.

Recently, DL methods have been widely used in image analysis and has shown the capability to perform complex tasks like object detection, classification, and

semantic segmentation. The success is due to the ability of CNNs, a powerful DL method in the image processing field, to learn hierarchical representations of raw input data, without relying on hand-crafted features [22], that have shown promising results in medical imaging applications, having the potential to replace or reduce manual burden in detection and quantification of coronary stenosis [23]. Several studies have attempted the quantitative analysis of the degree of stenosis and segmentation of the coronary artery lumen using CNN. However, most of the previous approaches required intensive human interactions for data annotation, by giving lesion-specific information in each frame (i.e., pixel-level labeling of stenosis, or manual selection of key frame on each clip), which have limited the practical usage and application to large dataset [13].

In literature, among the most recent researches on stenosis detection performed on ICA, the works conducted by Cong et al. [24], Moon et al. [13], Wu et al. [11] and Danilov et al. [5] resulted to be particularly interesting in relation with the object of this thesis.

Cong et al. 2019

In this study an automated method for stenosis detection in ICA images is proposed, which makes use of a CNN based workflow for image-level stenosis classification, without the need for an a priori vessel segmentation. It depends solely on physician stenosis reports to achieve accurate classification/detection. By taking advantage of proposed redundancy training strategy and a class of activation maps, stenosis positioning and localization capability can be determined and visualized. Figure 2.1 illustrates an abstraction of the algorithmic workflow.

The study was performed on 194 patients from a multicenter dataset and the stenosis location and their severity were previously analyzed with QCA in order to categorize them into three clinically relevant groups based on stenosis severity. 77 patients (39.7% of the total) had a stenosis of $< 25\%$, indicated as Category 0; 97 patients had (50% of the total) had a stenosis of25%-99%, indicated as Category 1 and 20 (10.3%) were with 100% stenosis (occlusion), indicated as Category 2. All ICA studies were saved in the universal DICOM

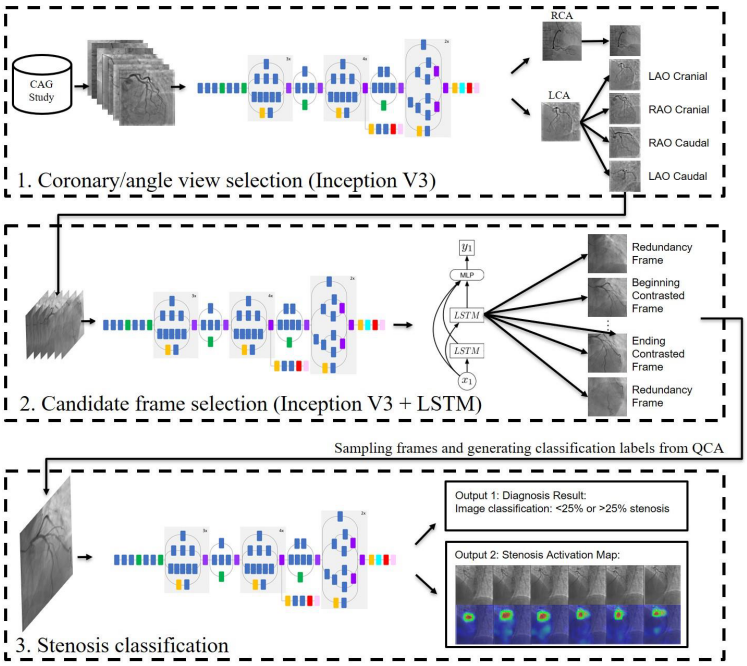


Figure 2.1: *Abstraction of the proposed end-to-end stenosis detection workflow. The input of the method is an ICA dataset, and the 4 main steps (from data preparation to stenosis detection) are described briefly in the dash-line boxes. The model returns 3 types of outputs: Output 1: diagnosis result for an image-level stenosis classification; Output 2: stenosis activation map and Output 3: stenosis localization information. Figure adapted from [24].*

format with a resolution of 512*512 or 1024*1024, 15 frames per second, 60-120 frames. The main training setup is based on the 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, 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. Firstly, LCA/RCA were classified initially by experts in a small subset and then leveraged by training a CNN classifier for automated classification of a view as either the LCA/RCA coronary artery. The image preparation is performed through an automatic detection of the ideal candidate frames and the redundant frames, defined as the one in a ICA video with best image quality, full contrast-agent penetration, clearly contrasted vessel borders, and anatomical significance of stenosis. It is trained an inception-v3 for recognizing the ideal candidate frames and the redundancy frames. To further improve the performance of candidate frame picking, they connect the fully-connection (FC) layer of inception to a pair of bi-directional long-short-term memory (LSTM) network, a recurrent neural network (RNN) that differs from CNN in its ability to process temporal information or data that comes in sequences. 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 with Inception-v3, a class of activation maps is employed to identify the discriminative regions. Results report $> 85\%$ accuracy in 2-CAT and $> 80\%$ accuracy in 3-CAT, showing the potential of this method in stenosis detection in ICA images. However, training/validation overfitting was the most important issue to be handled. A limitation of this study is that it was implemented at an image-level method, although it has demonstrated its ability in videos.

Moon et al. 2020

The proposed method consists of key frame detection, to train a DL model for classification of stenosis on each key frame, and visualization of the possible location of the stenosis. Firstly, they propose an algorithm that automatically

extracts key frames, essential for diagnosis, from 452 right coronary artery angiography movie clips. The DL algorithm is then trained with image-level annotations to classify the areas narrowed by over 50%. To make the model focus on the salient features, they apply a self-attention mechanism. The stenotic locations are visualized using the activated area of feature maps with gradient-weighted class activation mapping. Figure 2.2 illustrates an overview of the algorithmic workflow.

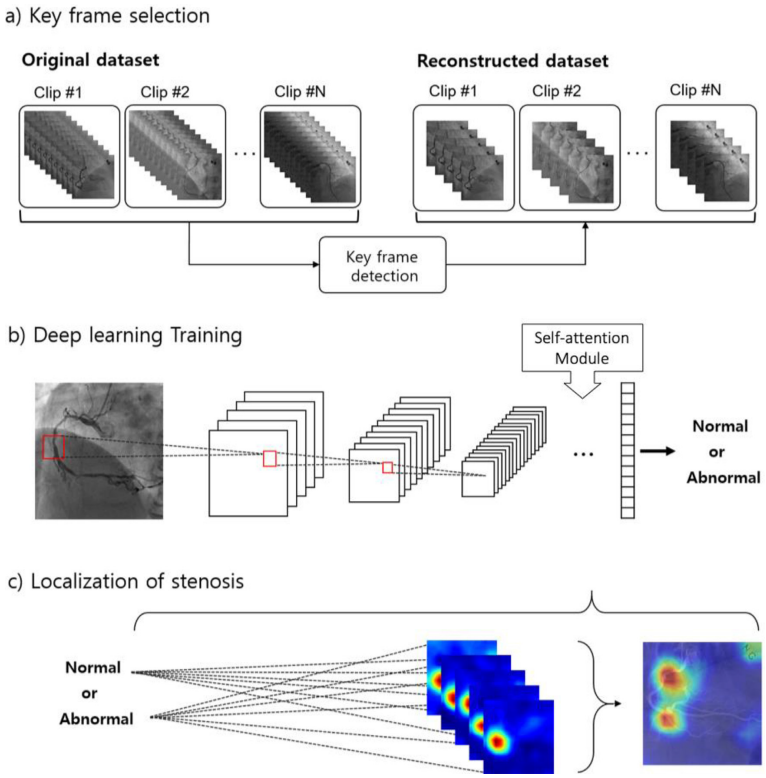


Figure 2.2: *Overview of the framework for the automated recognition of stenosis on ICA. Figure adapted from [13].*

DICOM-formatted coronary angiography data were retrieved anonymously from coronary angiography cases acquired in clinical practice. From the image acquisitions, a series of 512 × 512 pixel grayscale images with different gray levels were generated, and all cases were captured at the rate of 15 frames per second. The image dataset was ramdomly divided to conduct 5- fold cross validation. Firstly, they developed a fully automated key frame detection method that could significantly reduce the manual workload, and eliminate the unnec-

essary processing of a large number of video frames. After that, they utilized GoogleNet Inception-v3 as the base architecture of DL model. Each five key frames in a clip was fed into the CNN models to analyze the stenosis. Then, in order to maximize the attention of the baseline network (GoogleNet Inception-V3) to the characteristics of stenosis, they adopted the Convolutional Block Attention Module (CBAM) attention method. Since the stenosis can be observed in various different patterns and positions, they recalibrated the feature maps on both spatial and channel phases to make the model focus on the discriminative stenotic area, using Global Max Pooling (GMP) and Global Average Pooling (GAP). The training method achieved high frame-wise area-under-the-curve (AUC) of 0.971, frame-wise accuracy of 0.934, and clip-wise accuracy of 0.965 in the average values of cross-validation evaluations. Although there were different locations and types of stenosis with different shapes of blood vessels in the images, this approach showed promising performance. Furthermore, it showed high performance when tested on three external datasets with ensemble decision, which used different imaging protocols. The limitation of this study consists in the use of the right coronary arteries only in the angiogram image clips.

Wu et al. 2020

The authors propose a CNN-based method with a novel temporal constraint across ICA sequences. Specifically, they developed a deconvolutional single-shot multibox detector (DSSD) for candidate detection on contrast-filled X-ray frames selected by U-Net. Based on these static frames, the detector demonstrates high sensitivity for stenosis yet unacceptable false positives still exist. To solve this problem, they propose a customized seq-fps module that exploits the temporal consistency of consecutive frames to reduce the number of false positives. A flow diagram of the proposed framework is shown in Fig. 2.3. It includes three major parts: contrast-filled frames selection based on UNet, single frame stenosis detection based on DSSD and false positive suppression based on seq-fps.

The raw ICA sequence data used in this study were acquired from 63 patients,

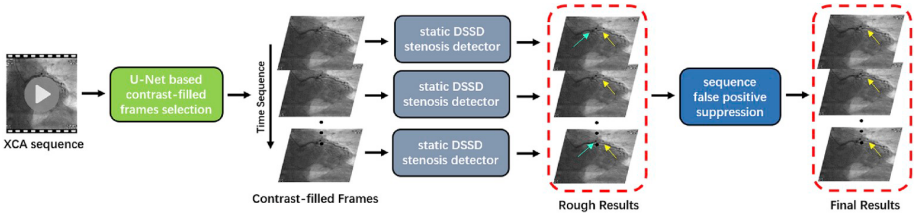


Figure 2.3: *Framework of the proposed method. The whole algorithm works as follows: first, the contrast-filled frames of an input ICA sequence are selected based on the U-Net segmentation results (shown in chronological order from top to bottom). Then, the DSSD provides rough results for each selected 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 [11].*

and sequence lengths vary from 3 to 5 at 14 frames per second. The resolution of each frame is 512 x 512 pixels. They obtained 148 ICA sequences in total and five-fold cross-validation experiments were conducted. Then, automatic identification of the N frames before and after the most contrast-filled frame from an ICA sequence is performed with U-net network. The DSSD network doubles 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 [25]. This step merges semantic information and location information, generating feature maps with richer contents that are beneficial for the detection task. The DSSD has high sensitivity; however, it is still influenced by a certain number of false positives. These false positives are generally stenosis-like structures generated by time-dependent contrast agent inhomogeneity and vessel motion. Based on a video object detection algorithm seq-nms (sequence-non maximum suppression), they design the temporal module called seq-fps. It is performed on the DSSD network results of consecutive contrast-filled frames selected by U-Net, which selects the stenosis that most frequently appears in an ICA sequence, thus filtering out remaining random false positives. The results show that the proposed method outperforms existing stenosis detection methods, achieving the highest sensitivity of 87.2% and positive predictive value of 79.5%.

Danilov et al. 2021

The very recent study of Danilov et al. is aimed at confirming the feasibility of real-time coronary artery stenosis detection using DL methods. To reach this goal they trained and tested eight promising detectors based on different neural network architectures (MobileNet, ResNet-50, ResNet-101, Inception ResNet, NASNet) using clinical angiography data of 100 patients. A total of 8325 grayscale images of 512 × 512 to 1000 × 1000 pixels were included for further study. Figure 2.4 presents the results of the comparative study of the neural networks. In addition to the absolute values of the metrics, the relative values are also reported. The metrics of SSD MobileNet V1 were used as a benchmark to compare with other models.

Model	Weights, mln		Training time, hours		Inference time, ms		F1-score		mAP@0.5	
	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.
SSD MobileNet V1	4.2	1.0×	16	1.0×	43	1.0×	0.72	1.0×	0.69	1.00×
SSD MobileNet V2	6.1	1.4×	20	1.3×	26	0.6×	0.80	1.10×	0.83	1.20×
SSD ResNet-50 V1	25.6	6.0×	47	3.0×	61	1.4×	0.73	1.01×	0.76	1.09×
Faster-RCNN ResNet-50 V1	25.6	6.0×	28	1.8×	98	2.3×	0.88	1.21×	0.92	1.33×
RFCN ResNet-101 V2	44.7	10.5×	55	3.6×	99	2.3×	0.96	1.32×	0.94	1.36×
Faster-RCNN ResNet-101 V2	44.7	10.5×	55	3.5×	118	2.7×	0.96	1.32×	0.94	1.35×
Faster-RCNN Inception ResNet V2	55.9	13.2×	93	6.0×	363	8.4×	0.94	1.30×	0.95	1.38×
Faster-RCNN NASNet	88.9	21.0×	147	9.5×	880	20.4×	0.82	1.13×	×	1.22×

Figure 2.4: *Comparative study of the selected models. Figure adapted from [5].*

Three neural networks have demonstrated superior results. The network based on Faster-RCNN Inception ResNet V2 is the most accurate and it achieved the mean Average Precision of 0.95, F1-score 0.96 and the slowest prediction rate of 3 fps on the validation subset. The relatively lightweight SSD MobileNet V2 network proved itself as the fastest one with a low mean Average Precision (mAP) of 0.83, F1-score of 0.80 and a mean prediction rate of 38 fps. The model based

on RFCN ResNet-101 V2 has demonstrated an optimal accuracy-to-speed ratio. Its mAP makes up 0.94, F1-score 0.96 while the prediction speed is 10 frames per seconds. The resultant performance-accuracy balance using the described neural networks has confirmed the feasibility of real-time CAD tracking supporting the decision-making process.

2.3 Limitation in the state of the art

Identification of significant coronary artery stenosis on ICA is an essential step in the exam interpretation and subsequent clinical planning [24]. Various automated or semi-automated stenosis detection methods have been proposed to aid the assessment. Despite these achievements, ICA images are characterized by the intrinsic complexities, due to complex vessel structures, poor contrast between vessels and surrounding tissues, nonuniform illumination, and overlap of background structures with inhomogeneous intensities [11]. Therefore, ICA is still indispensable for cardiologists in clinical practice. Despite a good performance, fully-automatic methods still have limitations: multiple preprocessing procedures, such as vessel enhancement, segmentation and skeleton extraction are performed in a single frame to detect stenosis. This is a time-consuming process, and the intrinsic complexities of ICA images prevent these substeps from obtaining ideal intermediate results [11]. As a consequence, errors might accumulate, hampering final stenosis detection. To avoid the cumbersome preprocessing steps for a single frame in traditional stenosis detection methods, CNNs are introduced to perform stenosis detection with state-of-the-art methods. 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 [26]. Moreover 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

[27].

DL has shown incredible results across many fields, but it still remains a tricky technique to master. Given the same dataset, different researchers can obtain widely varying results. All the studies that obtain good performance when applying DL algorithms often differentiate themselves in aspects outside of the deep network, like data pre-processing or augmentation techniques. For example, by adding a stain normalization pre-processing 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

CAD is the leading cause of death worldwide, however the process of interpreting complex coronary vasculature, image noise, low contrast vessels, and non-uniform illumination is time-consuming, thereby posing certain challenges to the operator. Real-time automatic CAD detection and labeling may overcome the above-mentioned difficulties by supporting the decision-making process.

Considering the limitations in the state of the art and the needs mentioned above, the work done in this thesis aims to attenuate these problems developing 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. The first part of the work is focused on the organization of the dataset, to allow a training through K-fold cross validation. A CNN for object detection, pretrained on natural images, has been chosen in accordance with the literature to be trained on the provided dataset to obtain stenosis detection from ICA images. The CNN model used for this purpose is the Single Shot Multibox Detector (SSD), in two different version, SSD300 and SSD7, in order to evaluate the best performing network between them. The choice is aimed by consideration on these architecures, that can be adequate for the limited dataset and the reduced number of parameters to train.

Finally, to the best of author's knowledge does not exist, in literature, an end-to-end stenosis detection network that does not use pre- and post- processing. This work tries to investigate the possibility to obtain a CNN capable of automatically detect stenosis in ICA images.

Chapter 3

Materials and Methods

In this chapter, a short overview on DL is reported in Section 3.1, describing the basics of Artificial Neural Network (ANN) and CNN, to give a background knowledge about the objective of this work. Hence, in Section 3.2 a deeper understanding of the object recognition task is presented, together with the related DL approaches. The SSD architecture, a specific CNN based object detection model considered one the most accurate real-time object detector, simple and fast, is presented in Section 3.2.1, while in Section 3.1.3 an explanation of the transfer learning strategy is described.

3.1 Deep Learning Overview

The trade-off between a precise and early diagnosis is a crucial aspect to allow optimal diagnostic, therapeutic and care work-up of several pathologies: in fact, wrong decision can yield in adverse health outcomes and psychological distress for the patients and in legal and financial impacts for healthcare system [28]

During the last decades, DL techniques faced a growing research interest thanks to their inherent capability of overcoming the drawbacks encountered with traditional Machine Learning (ML) algorithms, conventionally based on hand-crafted features for making decisions. Recent literature about DL techniques enlightened that the most used approach for medical image analysis are the CNNs, for several tasks, i.e. classification, detection and segmentation. The

application of DL algorithms to medical imaging faces many challenges, due to the inconsistency in the data (resolution, contrast, signal-to noise), the absence of image acquisition standards and the need for comprehensive medical image annotations. Moreover, the biggest challenge in this field is related to limited amount of data together with the complexities related to medical data sharing [29].

3.1.1 Artificial Neural Network (ANN)

ANNs are inspired by biological neural networks. An ANN, indeed, is composed of nodes, called artificial neurons, which model the neurons in the brain (Fig. 3.1). Physiologically speaking, each neuron takes inputs from its dendrites and inside the soma a weighted sum of inputs is performed. The signal arrives to the axon hillock. If the result is higher than the threshold limit, the signal propagates along the axon and it is run to the successive neurons through synapses.

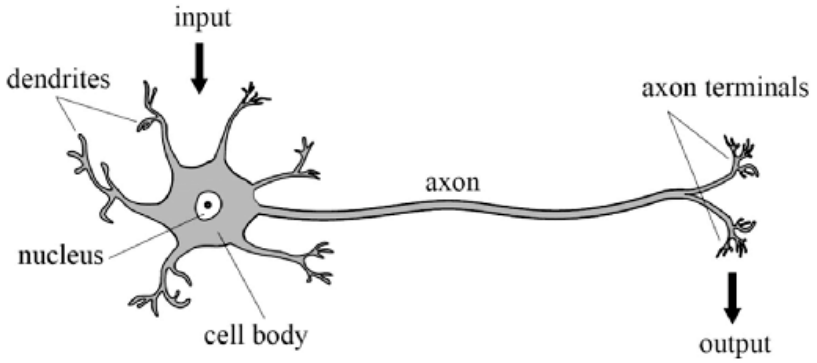


Figure 3.1: *Biological representation of neuron. Figure adapted from [30].*

In the same way, in the ANN, each connection between nodes transmits a signal that is a real number. Neurons and connections are adjusted during learning procedure by a weight. The nodes constitute layers, the first one is known as *input layers* while the last one as *output layer*. Among them there are one or more intermediate layers called *hidden layers*. Traditionally fully-connected ANNs compute a transformation of the input data by using weights and biases of the hidden layers in order to produce an output. To perform

classification, namely label each output, thresholds are introduced in the final layer. Each neuron of each hidden layer is connected to all the neurons of the previous layer but there are no connections among neurons belonging to the same layer. This structure allows to obtain an increasingly abstract representation of the input data. The simplest form of an ANN is the perceptron proposed by Frank Rosenblatt [31], represented in Fig. 3.2.

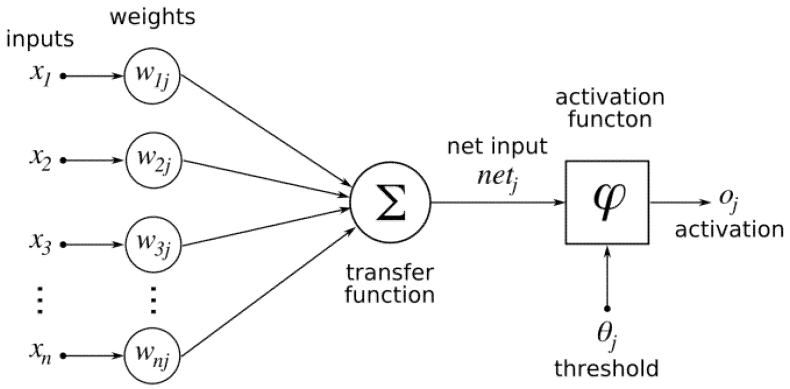


Figure 3.2: *Representation of the Rosenblatt perceptron. It receives M value inputs x_n and each of them is multiplied by the corresponding weight w_n. Subsequently all the products are added together and passed to the activation function. Figure adapted from [31].*

The perceptron mathematically represents the biological activity of the neuron. The input signals x_n communicate with the synapses w_{nj} and give $w_{nj} * x_n$ as product. The synaptic weights have different values because some inputs influence the output more than others. Moreover, they can be negative and therefore have an inhibitory influence. All the products are summed, if the result overcomes a threshold (bias b) the neuron fires. The weighted sum is modelled by an activation function g. The sigmoid function is usually applied and it is expressed as:

$$\sigma(x) = \frac{1}{(1 + e^x)} \tag{3.1}$$

The output can be expressed in the following way, in which N represents the number of inputs:

$$output = \hat{y} = g(\sum_{n=1}^{N} w_{nj} x_n + b) \tag{3.2}$$

The problem of the perceptron is that it is a linear classifier and thus it is never used individually. Its main application is to be used in more complex ANN, with multiple layers between the input and output layers. The CNN, described in the following section 3.1.2, represents a particular type of ANNs.

3.1.2 Convolutional Neural Networks (CNN)

CNNs, based on ANNs, are deep hierarchical neural models that roughly mimic the nature of mammalian visual cortex, and are the most promising architectures to recognize and localize objects within cluttered scenes[32]. Goodfellow et al. [33] defined CNNs as Artificial Neural Networks (ANNs) that perform convolution, in place of general matrix multiplication, in at least one of their layers [33]. The base of a CNN is the convolution which is a mathematical operation between two functions that produces a new function. The convolution theorem states that under certain conditions the Fourier transform of a convolution is a point-wise product of Fourier transforms. In other words, convolution in one domain (i.e., time domain) equals point-wise multiplication in the other domain (i.e., frequency domain). The convolution is defined as follow:

$$g(x) = f(x) * h(x) = \int_{-\infty}^{\infty} f(s)h(x-s)\,ds \tag{3.3}$$

where $f(x)$ and $h(x)$ are two functions and s a dummy variable. In case of image processing, $f(x)$ indicates the input, i.e. the image, $h(x)$ refers to the kernel, and $g(x)$ identifies the feature map. A CNN is composed of alternating convolutional layers, activation functions, pooling or subsampling layers and it ends with fully-connected layers. A generic representation of a CNN architecture is reported in Fig. 3.3:

The input is the whole image itself, this means that a depth is associated with the image and represents the number of channels. For example, an RGB image of $M \times N$ has 3 channels (one for R, B and G) so the full shape is actually $M \times N \times 3$. On the other hand, a scaled image of grays of the same size has only one value per pixel position, so its full shape is $M \times N \times 1$.

The purpose of a convolutional layer is to extract features from the input

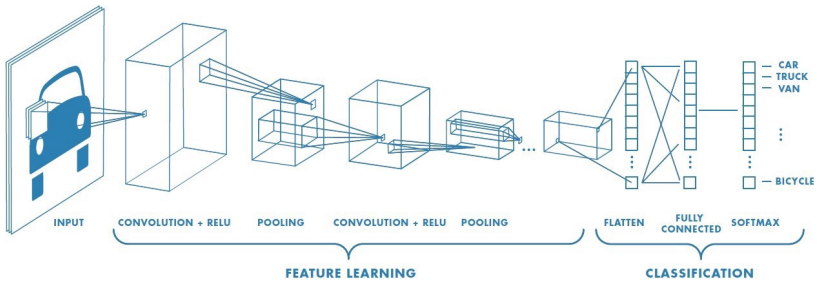


Figure 3.3: *Representation of a generic CNN architecture having convolutional, pooling and fully connected layers.*

layer. These layers are comprised of a series of filters or learnable kernels which aim at extracting local features from the input, and each kernel is used to calculate a feature map.The first convolutional layer extracts low-level meaningful features such as edges, corners, textures and lines. Next convolutional layers extract higher-level features, but the highest-level features are extracted in the last convolutional layer. Each kernel is a matrix, spatially smaller than the image, which convolves around the feature map or the input image, as represented in Fig. 3.4.

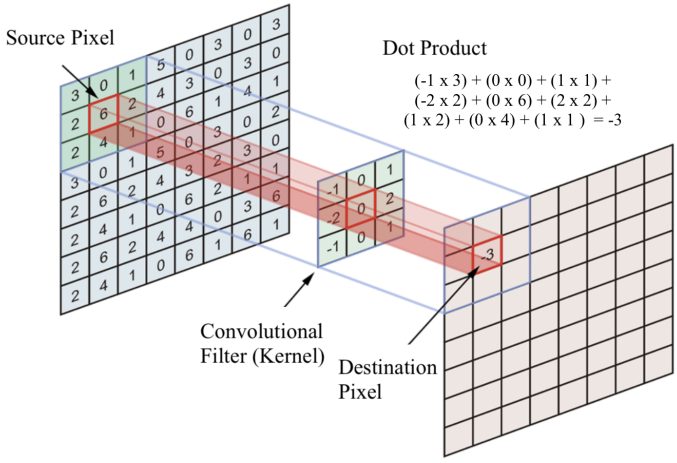


Figure 3.4: *Example of an image convolution. The 3x3 kernel convolves around the input image, computing at each step the dot product. The process is repeated for every pixel in the image. The source pixel is the anchor point at which the kernel is centered.*

The activation functions consist of non-linear layers that take the feature map generated by the the convolutional layer and create an activation map. The ac-

tivation function introduces the non-linearity into neural networks and it allows the learning of more complex features. There are several nonlinear activation functions such as $tanh(x)$, $sigmoid(x)$, and *Rectified Linear Unit (ReLU)*.

The pooling or subsampling layer reduces the resolution of the feature maps compressing features and the computational complexity of the network. The most common pooling is the max pooling, but even the average pooling is often implemented.The difference between them is how the output value is computed: max pooling returns the maximum value from the portion of the image covered by the window, while average pooling returns the average of all the values. The window is scrolled along the spatial dimension of the feature map by a fixed stride. An example of max pooling and average pooling is shown in Figure 3.5

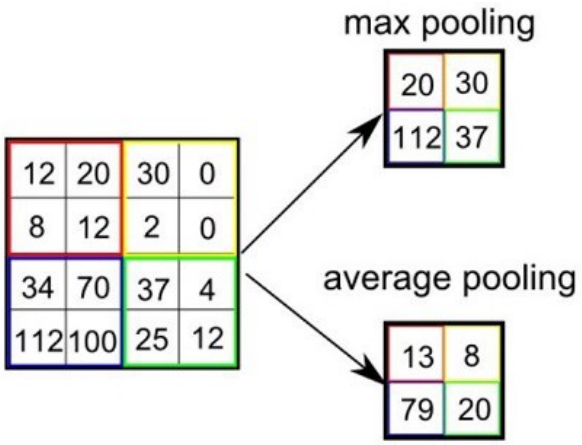


Figure 3.5: *Example of max pooling and average pooling operations. In this example a 4x4 image is downsampled to a 2x2 by taking the maximum value or the average value of each sub-region.*

A CNN ends with one or more fully connected layers that produce non-spatial output. The fully connected layers use the features arriving from all the previous layers for classifying the input image into the various classes. The number of neurons of the last fully connected layer equals the number of the classes.

A difference between CNNs and traditional ANNs is that the first ones are primarily used in the field of pattern recognition within images [34].

3.1.3 Transfer learning strategy

Data dependence is one of the most serious problem in DL. In particular, CNNs have a very strong dependence on massive training data, because they need a large amount of data to understand the latent patterns of images. Moreover, training a CNN from scratch requires extensive computational and memory resources. Thus, an alternative to training CNNs from scratch is transfer learning. Inspired by humans capability in applying learned knowledge to solve new problems faster and with better solutions, transfer learning aims to exploit and to transfer the knowledge or weights of a neural network trained on a large dataset for new tasks. Practically, it generalizes the knowledge (features, weights) of an existing solution to a new problem, leading to promising results also when the new task has significantly less data [35]. The process of transfer learning is illustrated in Fig. 3.6. One of the main aspect of CNNs is that the features extracted by the first layers are not specifically related to a dataset, but they are general and applicable to many dataset. Features computed by the last layer of a trained network must depend greatly on the chosen dataset and task. Thus, these last layer features are specific. If first layer features are general and last layer features are specific, then there must be a transition from general to specific somewhere in the network. Identifying the transition from general to specific can lead to saving of time and higher performances. Since general features can be applied to many dataset, avoiding their fine-tuning can save time; on the other hand, fine-tuning specific features layers can lead to higher performances [36].

3.2 Object detection

Object detection is a computer technology related to computer vision and image processing that deals with the process of finding and classifying objects in an image, thus it includes the problem of localization and the problem of classification of an object. Most of object detection models are designed for general image field, such as pedestrians detection and cars detection. Medical images have significantly different features from general images. These differences could result

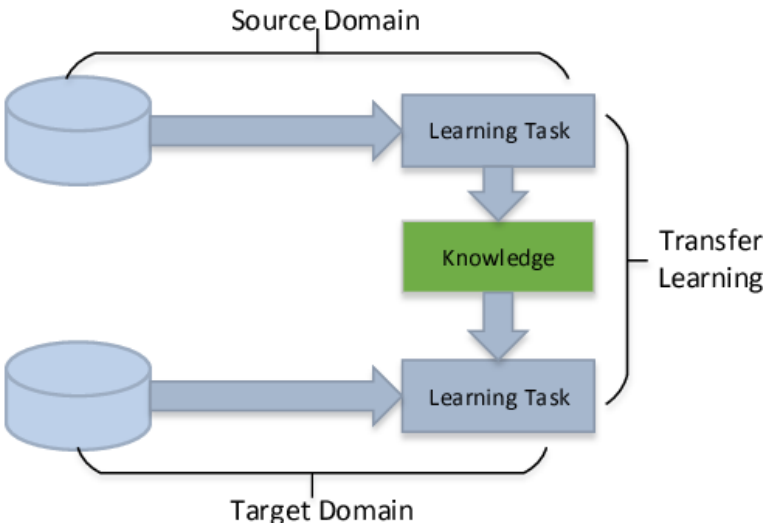


Figure 3.6: *Graphical representation of the transfer learning approach of previous works. The knowledge (features, weights) that a model has learned from a task (i.e., natural image classification) where a lot of labeled training data are available (source domain) is exploited and transferred to another task, such as medical image classification, with less data (target domain). Figure adapted from [35].*

in poor performance while models in general image field are directly applied to medical image field [37].

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 in localization and identification of small lesions in the full image space. Object detection in medical imaging aims to automatically detect lesions, improving the detection accuracy or decreasing the reading time of human experts [27].

3.2.1 Single Shot Multibox Detector (SSD)

One of the most popular object detection algorithm is the SSD, proposed by Liu et al. [38] in 2015. It is a one-stage detector, composed of a base network and additional convolutional layers. SSD takes only one shot to detect multiple objects present in an image using multibox. It is significantly faster than the other detection networks and it is high-accuracy object detection algorithm. Even with relatively low resolution images, high speed and accuracy pertain to

SSD, because of some specific characteristics: eliminates bounding box proposals like the ones used in Regional CCN's and includes a progressively decreasing convolutional filter for predicting object categories and offsets in bounding box locations. High detection accuracy in SSD is achieved by using multiple boxes or filters with different sizes, and aspect ratio for object detection. It also applies these filters to multiple feature maps from the later stages of a network. This helps perform detection at multiple scales. Multiple versions of SSD are developed, with different input images dimensions. Two of them are compared in this thesis, in order to evaluate which can be the most effective for stenosis detection and are presented in the Section 3.2.1 and 3.2.1.

SSD 300 architecture

The SSD approach is based on a feed-forward convolutional network that produces a fixed-size collection of bounding boxes and scores for the presence of object class instances in those boxes, followed by a non-maximum suppression step to produce the final detections. The early network layers are based on a standard architecture used for high quality image classification (truncated before any classification layers), called base network. Then, it is added an auxiliary structure to the network to produce detections [38]. SSD300 architecture is showed in Fig. 3.7.

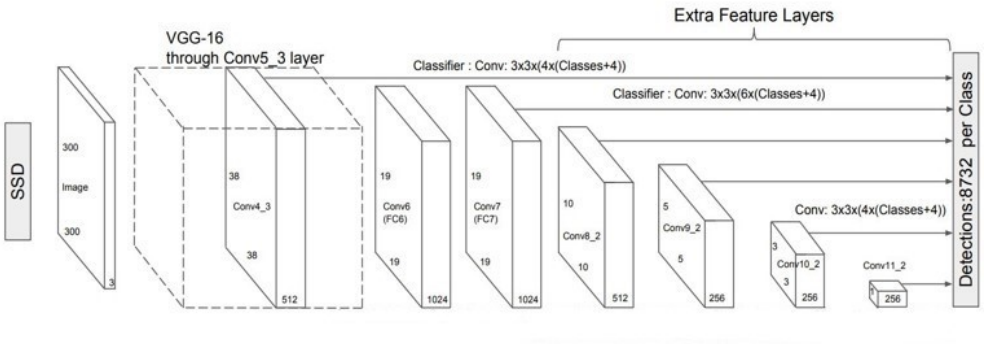


Figure 3.7: *A graphical representation of SSD300 architecture.*

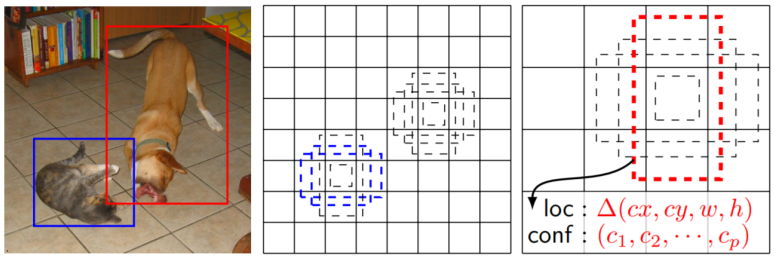
SSD has a base VGG-16 [39] network followed by multibox convolutional layers. The VGG-16 base network is a standard CNN architecture for high

quality image classification but without the final classification layers. VGG-16 is used for feature extraction. To the base VGG network, six additional convolutional layers are added for detection, to help with detection of objects at multiple scales. In three of those layers, it is made 6 predictions instead of 4. In total, SSD makes 8732 predictions per class. Each prediction is composed of:

- Bounding box with shape offset. Δcx, Δcy, h and w, representing the offsets from the center of the default box and its height and width.
- Confidences for all object categories or all the classes. Class 0 is reserved to indicate absence of the object.

The prediction of the bounding boxes and their confidence for different objects in the image is done by multiple feature maps of different sizes that represent multiple scales. Progressively decreasing convolutional layers decreases the feature map size and increases the depth. The deep layers cover larger receptive fields and construct more abstract representations. This is helpful in detecting larger objects. Initial convolutional layers cover smaller receptive fields and are helpful in detecting smaller objects present in the image [38].

To predict the bounding box, the default boxes are pre-selected manually to cover a wide spectrum of real-life objects. SSD also keeps the default boxes to a minimum (4 or 6) with one prediction per default box and instead of using global coordination for the box location, the boundary box predictions are relative to the default boundary boxes at each cell (Δcx, Δcy, Δw, Δh), i.e. the offsets (difference) to the default box at each cell for its center (cx, cy), the width and the height. During training time the default boxes are matched over aspect ratio, location and scale to the ground truth boxes. SSD predictions are classified as positive matches or negative matches and it only uses positive matches in calculating the localization cost. If the corresponding default boundary box has an Intersection over Union (IoU) - ratio between the area of overlap between the predicted bounding box and the ground-truth bounding box and the area of union - greater than 0.5 with the ground truth, the match is positive. Otherwise, it is negative [38].



(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

Figure 3.8: *In the image the network matches two default boxes. One with the cat and one with the dog. They are treated as positives bounding boxes and the rest are treated as negatives. Figure adapted from [38].*

SSD 7 architecture

The SSD7 architecure, showed in Fig. 3.9, is a smaller version of the original architecture. The detection approach is the same, but it consists of 7 convolutional feature layers and 4 convolutional predictor layers. Also in this case, the images in input have a shape of 300x300 pixels and it makes 3256 predictions.

In this work it is compared 2 different versions of the same SSD architecture, in order to investigate the behaviors both in the case of a more complex and less complex network.

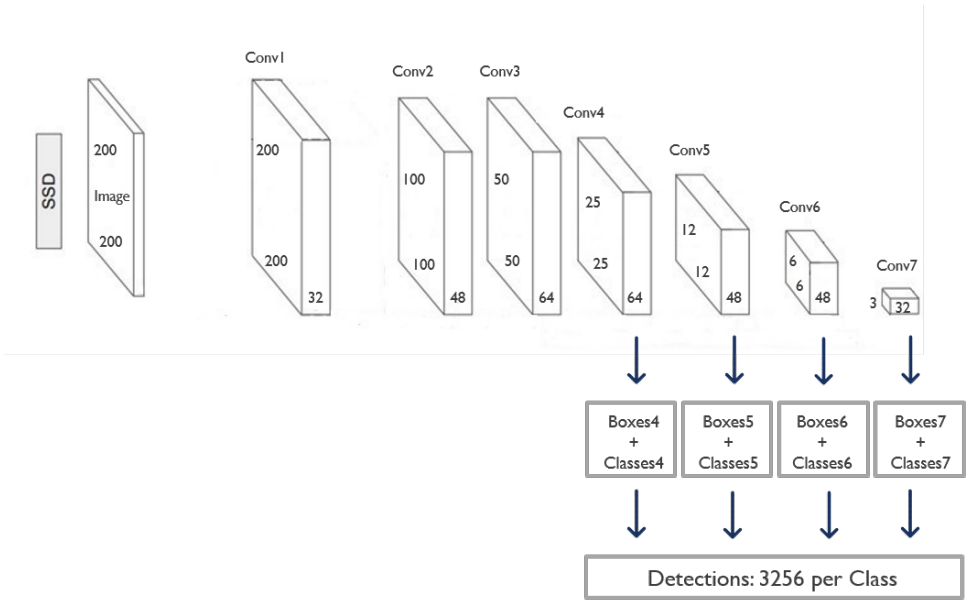


Figure 3.9: *A graphical representation of SSD7 architecture.*

Chapter 4

Experimental Protocol

4.1 Dataset

The raw ICA sequences used in this study are acquired from the Cardiology Department of Ancona's Ospedali Riuniti. They consist of 226 frames, extracted from the collection of 61 ICA images in DICOM format, which is a standard protocol for the management and transmission of medical images and related data [40]. 61 patients underwent ICA acquisition, of which 15 women and 46 men. From those acquisitions, a series of 512 $\times$ 512 pixels grayscale images with different gray levels (0–255) are generated. The acquired frames are manually labeled by the clinicians: the annotations on the images are made by visual inspection, 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

The dataset is composed of ICA images and annotations, in DICOM and .xml format respectively. The images are converted in .jpeg, while the annotations are converted in .csv format, in order to give the required inputs to the network. K-fold cross-validation is performed to split the dataset. Cross-validation is a resampling procedure used to evaluate ML models applied on limited data

samples. The procedure has a single parameter called K, in this study set equal to 5, which refers to the number of groups that a given data sample is to be split into. Each fold is composed of 12 patients, except for one fold that contains 13 patients. It is a popular method because of its simplicity and because it generally results in a less biased or less optimistic estimate of the model skill compared to other methods, such as a simple train/test split. The general procedure includes the following steps:

1. Shuffle the dataset randomly;
2. Split the dataset into k groups;
3. For each unique group:
    - Take the group as a hold out or test data set;
    - Take the remaining groups as a training data set;
    - Fit a model on the training set and evaluate it on the test set;
    - Retain the evaluation score and discard the model.
4. Summarize the skill of the model using the sample of model evaluation scores [41].

The schematic representation of the procedure is shown in Fig. 4.1 and the numerical subdivision of ICA images is reported in Table 4.1.

Table 4.1: *Dataset organization: 5-fold composition.*

K-fold	**Train**	**Validation**	**Test**
1-fold	134	46	46
2-fold	135	46	45
3-fold	139	45	42
4-fold	137	42	47
5-fold	133	47	46

The K-fold cross-validation approach is performed in testing phase. In particular, four datasets are used to train the network, and one of them is reserved

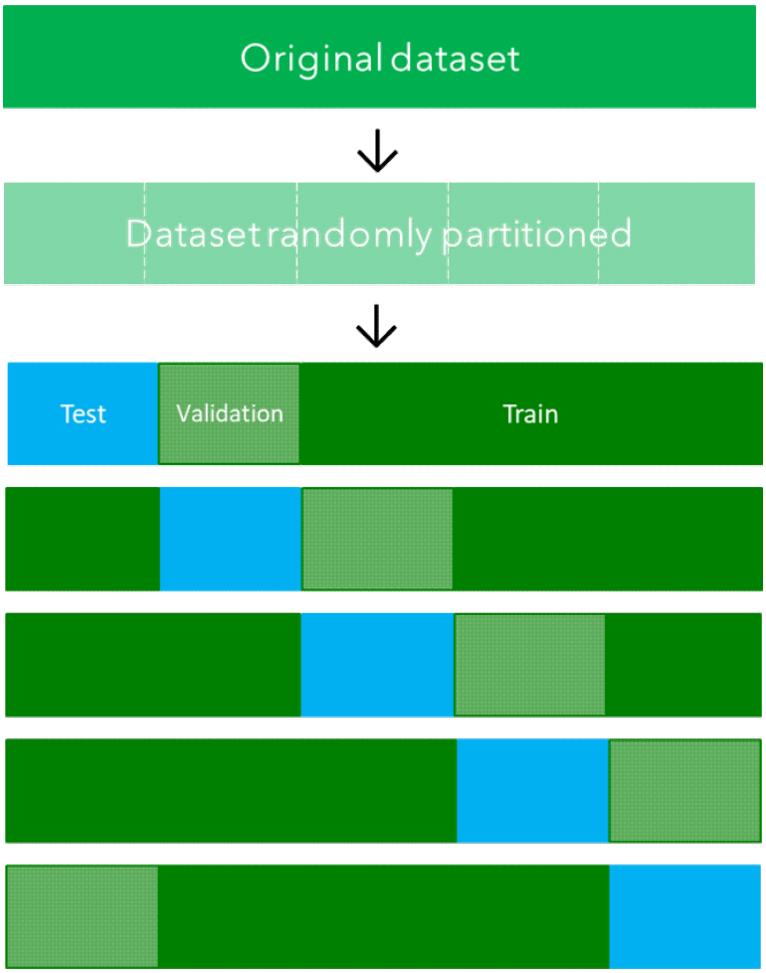


Figure 4.1: *A schematic illustration of K-fold cross-validation for K = 5. The original dataset (shown in dark green) is randomly partitioned into K disjoint sets (shown in light green). Then, for each subset, K–1 parts are used for training the model (shown in green), one of which is used for validation (shown in light green) and the remaining part is used for testing (shown in blue). This process is repeated K times for all possible choices of the test set, producing test errors. The final performance is reported by averaging the errors from each iteration.*

for validation. The last one is utilized for testing. The splitting is performed at the patient level. The training set contains the sample of data used to fit the model (weights and biases) and the model sees and learns from this data. The validation set contains the sample of data used to provide an unbiased evaluation of a model fit on the training dataset, while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. Hence the model occasionally sees this data, but never does it "learn" from this. Finally, the testing set concerns the sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. The test set provides the gold standard used to evaluate the model. It is only used once a model is completely trained (using the train and validation sets) [42].

4.2 Training setting

The overall algorithm is implemented with Keras, an open-source software library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library. Experiments are performed using Google Colaboratory: a free GPU cloud platform based on Jupyter notebook environment that supports 14858 MB of free GPU, 12.4 GB of free RAM, and a processor size of 900.6 MB.

Images are resized to 300 x 300 x 3 pixels or 200 x 200 x 3 pixels, for SSD300 and SSD7 respectively, in order to give as inputs to the network the correct image dimension. Before training, the images are pre-processed by removing the intensity mean. To train the SSD300, by using the technique of transfer learning, the pre-trained weights, related to the original SSD300 model trained on MS COCO [43], are initialized. While the SSD7 is trained from scratch, initializing the weights of the network to random values. After setting these as the starting points, the entire networks are 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 with the input x).

4.2.1 Data augmentation

Data augmentation is a technique that can be used to artificially expand the size of a training set by creating modified data from the existing one. It is a good practice to use data augmentation to prevent overfitting, or if the initial dataset is too small to train on, or to squeeze better performance from the model. In this work it is proposed an on-the-fly data augmentation that involves:

- Photometric transformations: brightness, contrast, saturation and hue.
- Geometric transformations: flip (horizontal and vertical), translation (both directions), scale (inward) and rotation (90°,180°,270°).

4.2.2 Loss function

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 words, 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 overall SSD objective loss function is a weighted sum of the localization loss (loc) and the confidence loss ($conf$):

$$L(x, c, l, g) = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \tag{4.1}$$

where N is the number of matched default boxes. If N = 0, wet set the loss to 0. The localization loss is a Smooth L1 loss between the predicted box (l) and the

ground truth box (g) parameters. Similar to Faster R-CNN, it is regressed to offsets for the center (c_x , c_y) of the default bounding box (d) and for its width (w) and height (h).

$$L_{loc}(x, l, g) = \sum_{i \epsilon Pos}^{N} \sum_{m \epsilon cx,cy,w,h} x_{ij}^{k} smooth_{L1}(l_i^m - \hat{g_j}^m) \tag{4.2}$$

$$\hat{g_j^{cx}} = \frac{g_j^{cx} - d_i^{cx}}{d_i^w} \tag{4.3}$$

$$\hat{g_j^{cy}} = \frac{g_j^{cy} - d_i^{cy}}{d_i^h} \tag{4.4}$$

$$\hat{g_j^w} = \log \frac{g_j^w}{d_i^w} \tag{4.5}$$

$$\hat{g_j^w} = \log \frac{g_j^j}{d_i^h} \tag{4.6}$$

The confidence loss is the softmax loss over multiple classes confidences (c). It is defined as:

$$L_{conf}(x, c) = -\sum_{i \epsilon Pos}^{N} x_{ij}^{P} log(\hat{c_i}^P) - \sum_{i \epsilon Neg} log(\hat{c_i}^0) \tag{4.7}$$

where

$$\hat{c_i}^P = \frac{exp(c_i)^P}{\sum exp(\hat{c_i}^P)} \tag{4.8}$$

The localization loss is the mismatch between the ground truth box and the predicted boundary box. SSD only penalizes predictions from positive matches. It is wanted the predictions from the positive matches to get closer to the ground truth. Negative matches can be ignored. The confidence loss is the loss of making a class prediction. For every positive match prediction, we penalize the loss according to the confidence score of the corresponding class. For negative match predictions, it is penalized the loss according to the confidence score of the class "0": class "0" classifies no object is detected.

4.2.3 Adaptive Moment Estimation (Adam) Optimization

Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iteratively based on training data. Adam's name is derived from adaptive moment estimation because uses estimates of the first and second moments of the gradient to perform updates, which can be seen as incorporating gradient descent with momentum (the first-order moment) and RMSProp algorithm (the second-order moment). Specifically, the algorithm calculates an exponential moving average of the gradient and the squared gradient, and the parameters β_1 and β_2 control the decay rates of these moving averages. Considering $g_{t,i}$ as the gradient of J, generic cost function, with respect to the parameter θ_i after t batches. Using the Adam optimizer, every parameter after each batch is computed in this way:

$$\theta_{t+i,t} = \theta_{t,i} - \frac{\eta}{\sqrt{\hat{v}_t(g_{t,i}) + \epsilon}} \cdot \hat{m}_t(g_{t,i}) \tag{4.9}$$

where ϵ 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:

$$m_t = \beta_1 \cdot m_{t,i} + (1 - \beta_1) \cdot (g_{t,i}) \tag{4.10}$$

$$v_t = \beta_2 \cdot v_{t,i} + (1 - \beta_2) \cdot (g_{t,i})^2 \tag{4.11}$$

where the hyperparameters β_1, β_2 ϵ [0, 1) control the exponential decay rates of the moving average of the gradient (m_t) and of the squared gradient (v_t) respectively. In this study β_1 and β_2 are set to 0.9 and 0.999 respectively. The mini-batch size is set to 16 as a balance between training speed and gradient convergence. The batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters. The

number of epochs, that define the number of times that the learning algorithm will work through the entire training dataset, is set to 100, with 10 steps per epochs. The performance of the mini-batch depends critically on how the learning rate (α) is tuned. α determines the step size of each iteration while moving toward a minimum of the loss function (cross-entropy). In setting the learning rate, there is a trade-off between the rate of convergence and overshooting. While the descent direction is usually determined from the gradient of the loss function, the learning rate determines how big a step is taken in that direction. A learning rate too hig can make the learning jump over the minima, while a learning rate too low might either take too long to converge or get stuck in an undesirable local minimum. Due to fine-tuning and to the assumption that the gradient of the cross-entropy is already in a good position, the learning rate is set to 10^{-3} in order to not move too far from the favorable area of where the gradient descent starts. To increase the efficiency during the training procedure, a specific callback, named "ReduceLROnPlateau" has been applied. This function monitors the validation loss and performs a reduction of the learning rate by a factor equal to 0.2 if no improvement of the loss value is seen after 8 epochs, until reaching a learning rate of 10^{-5}.

4.3 Evaluation Metrics

The performances of each tested CNN model are evaluated with respect to the manual annotation performed by the clinician, considered as the ground truth. A set of metrics commonly used to evaluate the performance of binary classification is employed. In the field of ML and in the problem of statistical classification, a confusion matrix, a specific table layout that allows visualization of the performance of an algorithm, is defined through these terms:

- True Positive (TP): number of predicted positives correctly classified: number of frames in which stenosis is correctly detected.
- True Negative (TN): number of predicted negative correctly classified as frames without stenosis.

- False Positive (FP): number of predicted positives incorrectly classified: frames in which the stenosis is mistakenly identified.
- False Negative (FN): number of predicted negatives incorrectly classified as frames without stenosis, although it presents stenotic areas.

Confidence is defined as the probability that a box contains the object.

The metrics used in this work for the evaluation of the proposed model are Intersection over Union (IoU), the Dice Similarity coefficient (DSC) and the Average Precision (AP), described in the following subsections [33].

4.3.1 Intersection over Union (IoU)

IoU is an evaluation metric used to measure the accuracy of an object detection algorithm on a particular dataset. In order to apply IoU to evaluate an object detector it is needed to clarify

- The ground-truth box, that is the hand-labeled bounding box from the testing set that specifies where, in the image, the object of interest is;
- The predicted box, that is the model's output.

The IoU is defined as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{4.12}$$

in which the *Area of Overlap* is the overlap between the predicted bounding box and the ground-truth bounding box. Instead, the *Area of Union* is the area encompassed by both the predicted bounding box and the ground-truth bounding box. Its range is from 0 to 1, with 1 signifying the greatest similarity between predicted and truth.

4.3.2 Dice similarity coefficient (DSC)

The DSC (also known as the F1 score) is one of the evaluation metrics most frequently used in medical image segmentation. It is a statistic used to gauge the similarity of two samples. The equation is as follows:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4.13}$$

It can be defined as the ratio between the double of the *Area of Overlap* and the *Total Number of Pixels* in both images. Its range is from 0 to 1, with 1 signifying the greatest similarity between predicted and truth, like IoU [36].

4.3.3 Average Precision (AP)

Average Precision (AP) is a metric mostly used in object detection, measuring the detection accuracy. Before, it is necessary to define the Precision and the Recall:

- *Precision* measures how accurate is the predictions. i.e. the percentage of correct predictions.

$$precision = \frac{TP}{TP + FP} \tag{4.14}$$

- Recall measures how good the network finds all the positives.

$$recall = \frac{TP}{TP + FN} \tag{4.15}$$

Generally, the AP represents the area under the precision-recall curve. Precision and recall are always between 0 and 1. Therefore, AP falls within 0 and 1 also, but it is often expressed as percentage. Precision-recall curves have a distinctive saw-tooth shape: if the $(k+1)^{th}$ terms retrieved is nonrelevant then recall is the same as for the top k terms, but precision drops. If it is relevant, then both precision and recall increase, and the curve jags up and to the right [33]. It is often useful to remove these jiggles and the standard way to do this is with an interpolated precision: the interpolated precision p_{interp} at a certain recall level r is defined as the highest precision found for any recall level $r' \geq r$:

$$p_{interp}(r) = \max_{r' \geq r} p(r')$$

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

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) p_{interp}(r_{i+1}) \tag{4.16}$$

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

mean Average Precision (mAP)

The AP calculates the accuracy for a single class [36]. When are present K classes, the mean Average Precision (mAP) is defined as:

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{4.17}$$

Chapter 5

Results

This chapter presents the results obtained following the experimental protocol with both SSD300 and SSD7 architectures. The results are reported in terms of performance metrics, statistical tests performed and prediction of the stenosis detection on the ICA images.

5.0.1 SSD300 evaluation

For each fold, a table containing the dataset characteristics, the trends of the loss function, the numerical values of performance metrics and the *precision–recall* curves are reported. Hence, it is possible to evaluate and understand the capability of the proposed networks in terms of stenosis detection in ICA images. In these trials, the network predicts a bounding box, therefore detects stenosis, for each image.

1-fold

Table 5.1: *Dataset organization: 1-fold composition.*

	Train	**Validation**	**Test**
Number of images	134	46	46
Number of patients	37	12	12

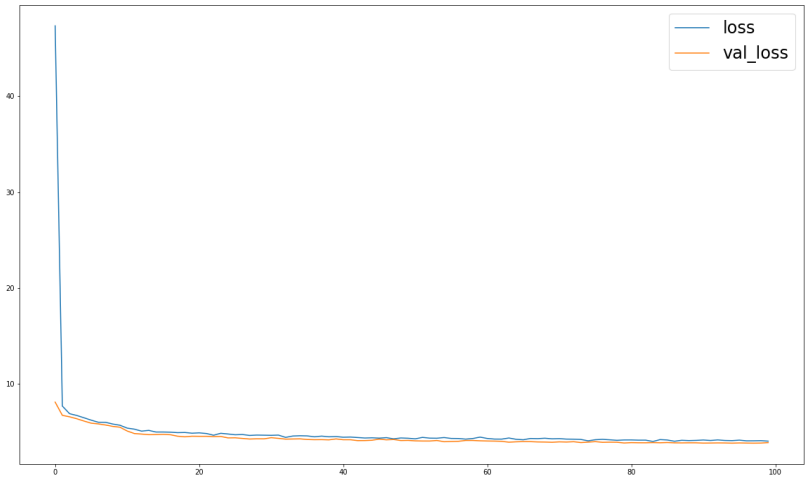


Figure 5.1: *Trend of the loss values for the 1-fold.*

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.45.

Table 5.2: *Performance metrics values for the 1-fold.*

	Mean value	**Standard deviation**
IoU	0.59	0.18
DSC	0.73	0.16

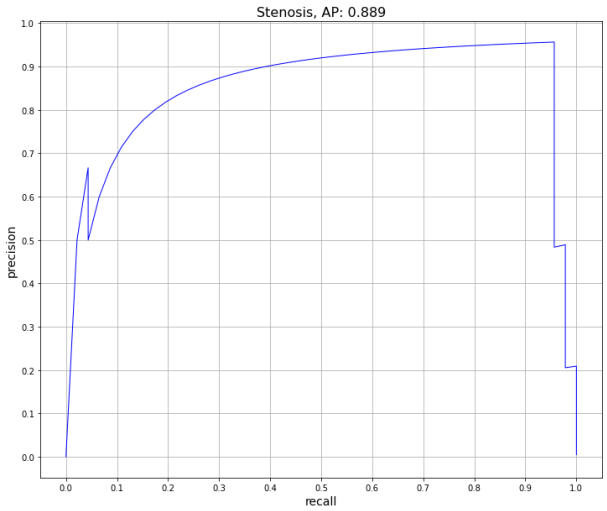


Figure 5.2: *Precision-Recall curve for the 1-fold. The mAP is 0.89.*

2-fold

Table 5.3: *Dataset organization: 2-fold composition.*

	Train	**Validation**	**Test**
Number of images	135	46	45
Number of patients	36	12	13

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.44.

Table 5.4: *Performance metrics values for the 2-fold.*

	Mean value	**Standard deviation**
IoU	0.46	0.14
DSC	0.61	0.13

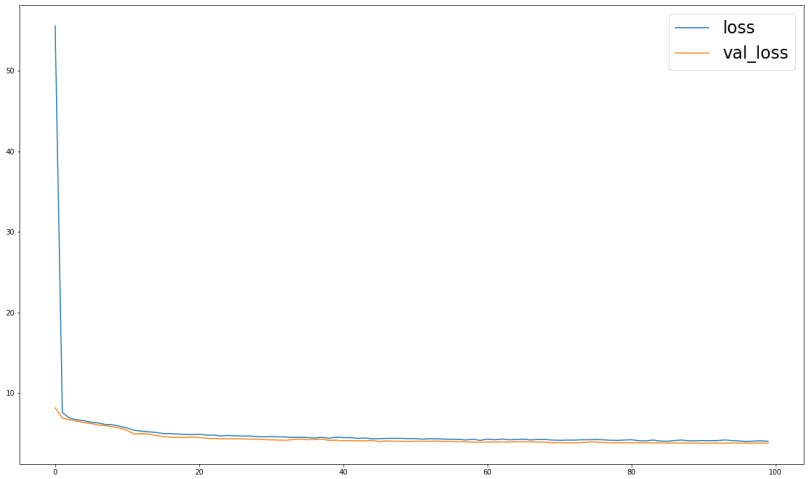


Figure 5.3: *Trend of the loss values for the 2-fold.*

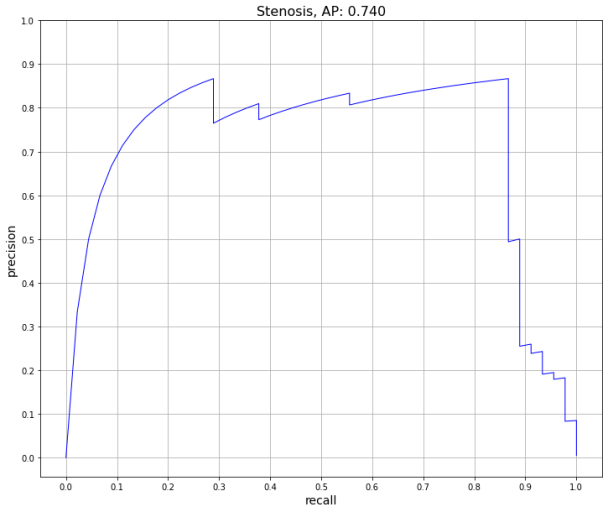


Figure 5.4: *Precision-Recall curve for the 2-fold. The mAP is 0.74.*

3-fold

Table 5.5: *Dataset organization: 3-fold composition.*

	Train	Validation	Test
Number of images	139	45	42
Number of patients	36	13	12

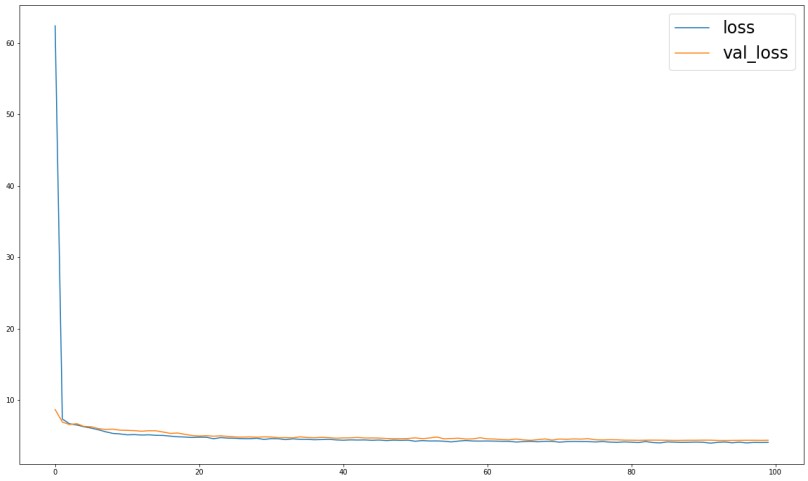


Figure 5.5: *Trend of the loss values for the 3-fold.*

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.49.

Table 5.6: *Performance metrics values for the 3-fold.*

	Mean value	Standard deviation
IoU	0.46	0.21
DSC	0.60	0.22

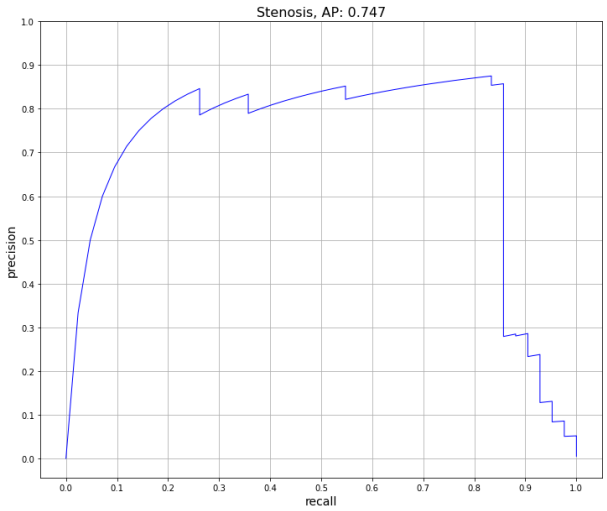


Figure 5.6: *Precision-Recall curve for the 3-fold. The mAP is 0.75.*

4-fold

Table 5.7: *Dataset organization: 4-fold composition.*

	Train	**Validation**	**Test**
Number of images	137	42	47
Number of patients	37	12	12

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.47.

Table 5.8: *Performance metrics values for the 4-fold.*

	Mean value	**Standard deviation**
IoU	0.45	0.23
DSC	0.59	0.24

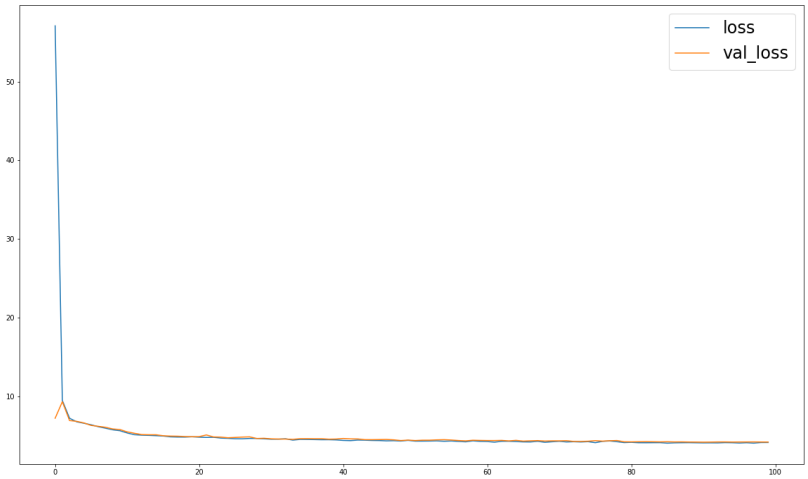


Figure 5.7: *Trend of the loss values for the 4-fold.*

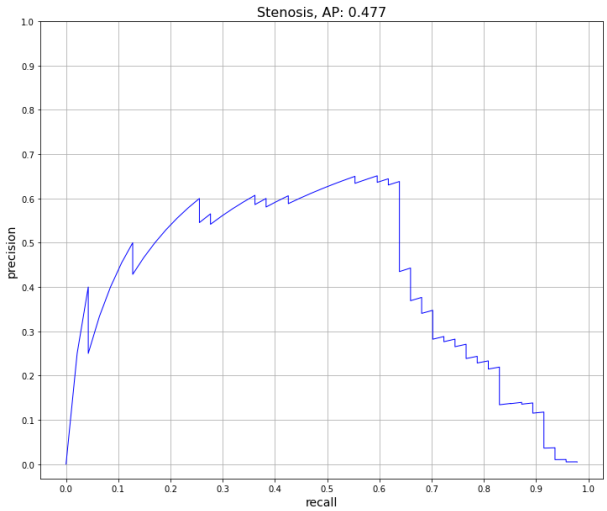


Figure 5.8: *Precision-Recall curve for the 4-fold. The mAP is 0.48.*

5-fold

Table 5.9: *Dataset organization: 5-fold composition.*

	Train	**Validation**	**Test**
Number of images	133	47	46
Number of patients	37	12	12

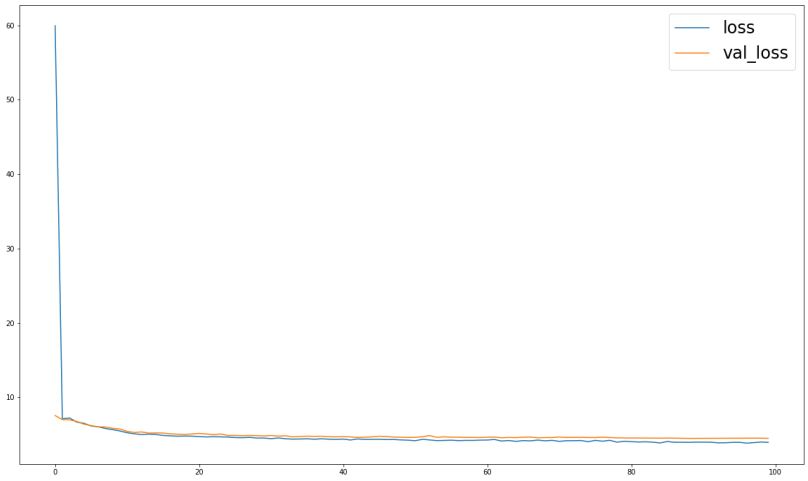


Figure 5.9: *Trend of the loss values for the 5-fold.*

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.59.

Table 5.10: *Performance metrics values for the 5-fold.*

	Mean value	**Standard deviation**
IoU	0.54	0.12
DSC	0.69	0.09

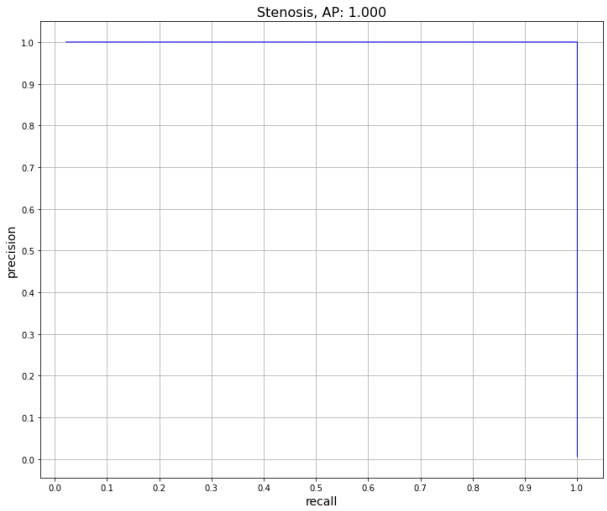


Figure 5.10: *Precision-Recall curve for the 5-fold. The mAP is 1.*

5.0.2 SSD7 evaluation

For each fold, a table containing the dataset characteristics, the trends of the loss function and the numerical values of performance metrics are reported, hence it is possible to evaluate and understand the capability of the proposed networks in terms of stenosis detection in ICA images.

1-fold

The dataset considered is the same described in Table 5.1 In this trial the network makes a prediction only in 9 images out of 46 images of test set.

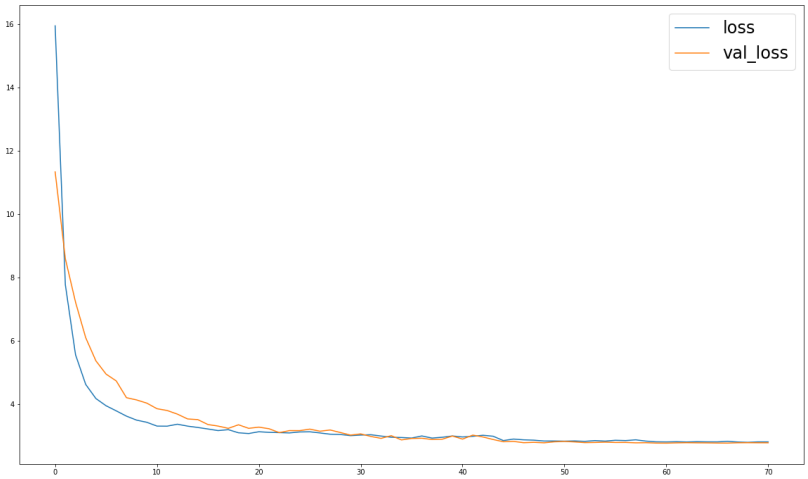


Figure 5.11: *Trend of the loss values for the 1-fold.*

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.30.

Table 5.11: *Performance metrics values for the 1-fold.*

	Mean value	**Standard deviation**
IoU	0.26	0.14
DSC	0.39	0.18

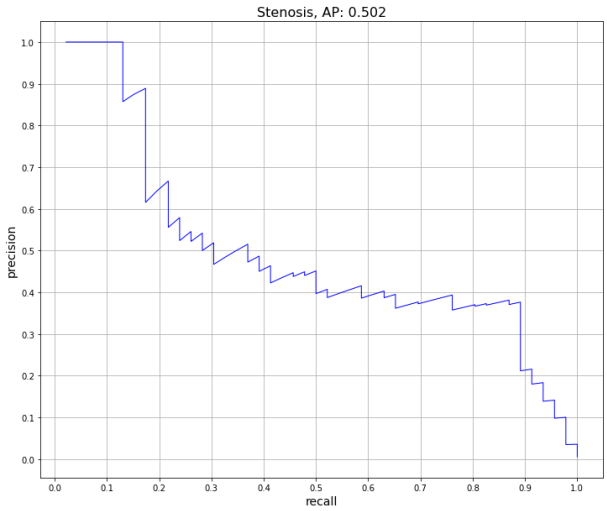


Figure 5.12: *Precision-Recall curve for the 1-fold. The mAP is 0.50.*

2-fold

The dataset considered is the same described in Table 5.3 In this trial the network makes a prediction only in 23 images out of 45 images of test set.

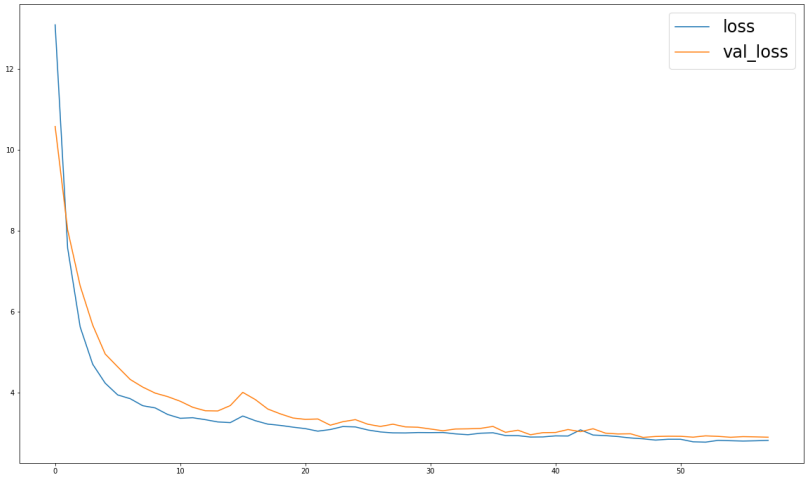


Figure 5.13: *Trend of the loss values for the 2-fold.*

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.30.

Table 5.12: *Performance metrics values for the 2-fold.*

	Mean value	Standard deviation
IoU	0.26	0.12
DSC	0.39	0.15

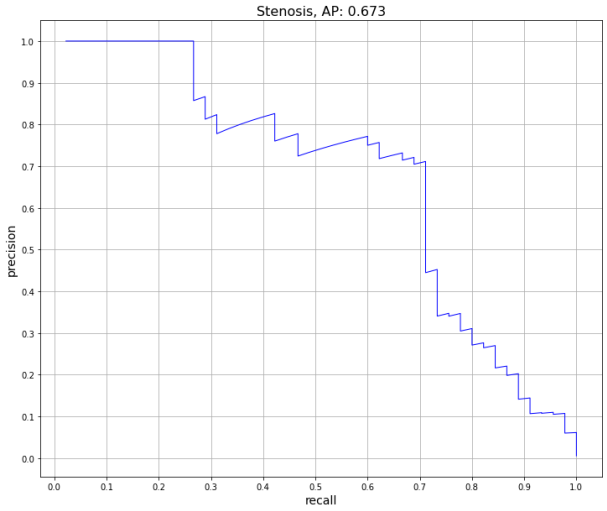


Figure 5.14: *Precision-Recall curve for the 2-fold. The mAP is 0.67.*

3-fold

The dataset considered is the same described in Table 5.5 In this trial the network makes a prediction only in 19 images out of 42 images of test set.

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.30.

Table 5.13: *Performance metrics values for the 3-fold.*

	Mean value	Standard deviation
IoU	0.34	0.15
DSC	0.49	0.17

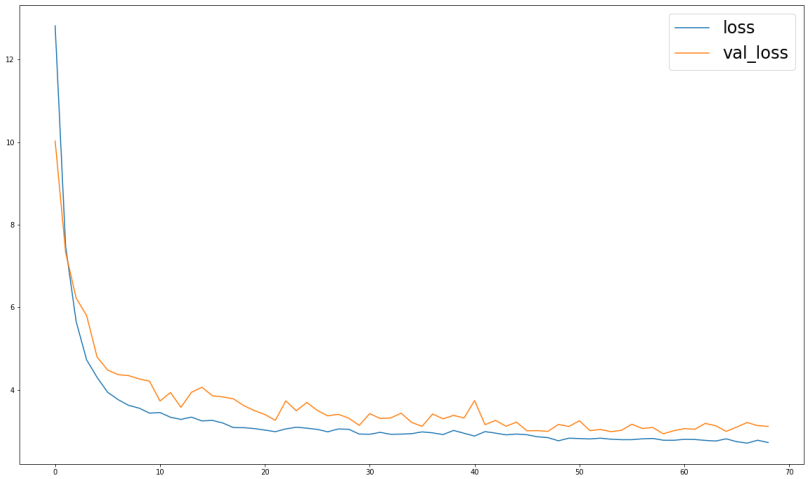


Figure 5.15: *Trend of the loss values for the 3-fold.*

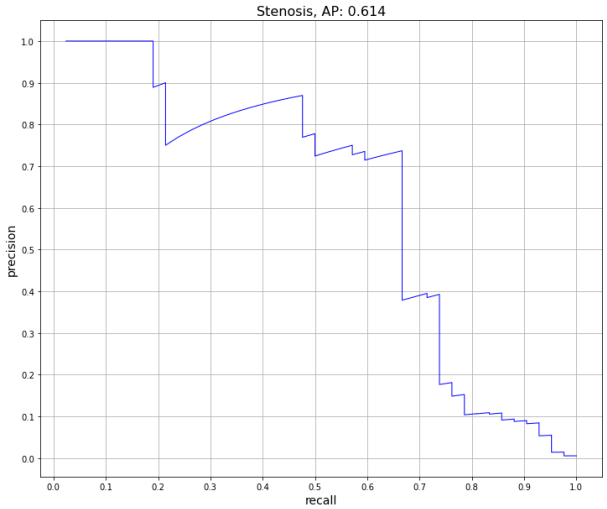


Figure 5.16: *Precision-Recall curve for the 3-fold. The mAP is 0.61.*

4-fold

The dataset considered is the same described in Table 5.7 In this trial the network makes a prediction in 42 images out of 47 images of test set.

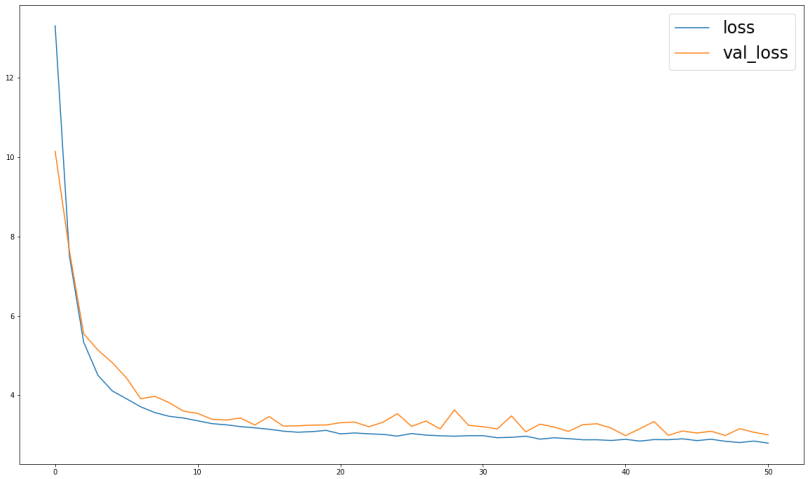


Figure 5.17: *Trend of the loss values for the 4-fold.*

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.30.

Table 5.14: *Performance metrics values for the 4-fold.*

	Mean value	**Standard deviation**
IoU	0.24	0.15
DSC	0.36	0.20

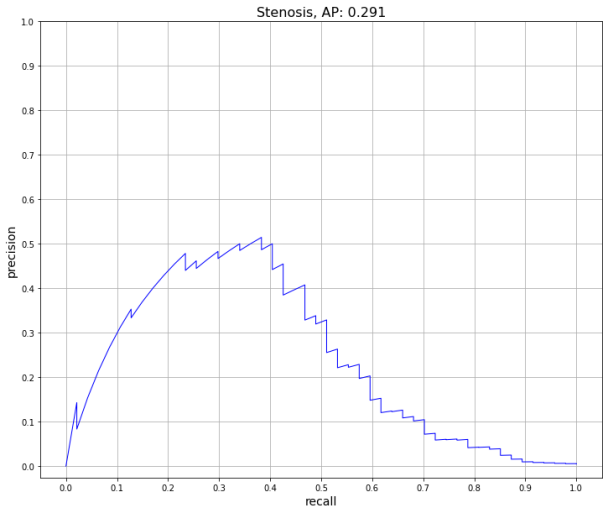


Figure 5.18: *Precision-Recall curve for the 4-fold. The mAP is 0.29.*

5-fold

The dataset considered is the same described in Table 5.9 In this trial the network makes a prediction only in 14 images out of 46 images of test set.

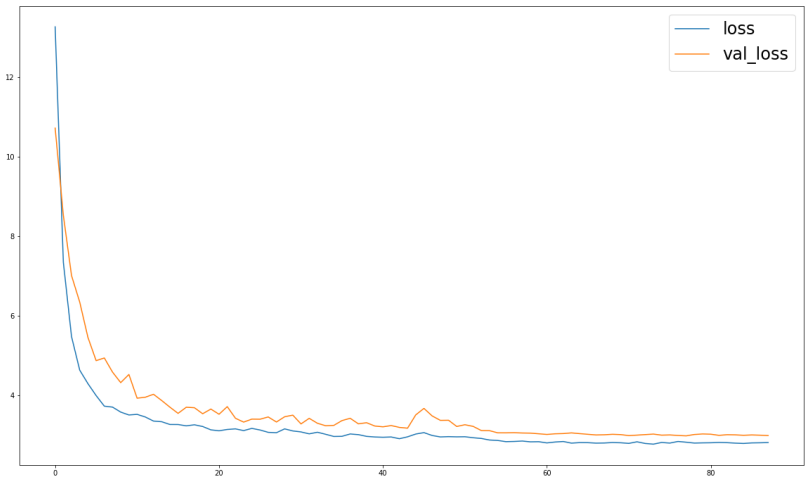


Figure 5.19: *Trend of the loss values for the 5-fold.*

The performance metrics results on the test set images are obtained setting the confidence threshold at 0.30.

Table 5.15: *Performance metrics values for the 5-fold.*

	Mean value	**Standard deviation**
IoU	0.41	0.20
DSC	0.56	0.19

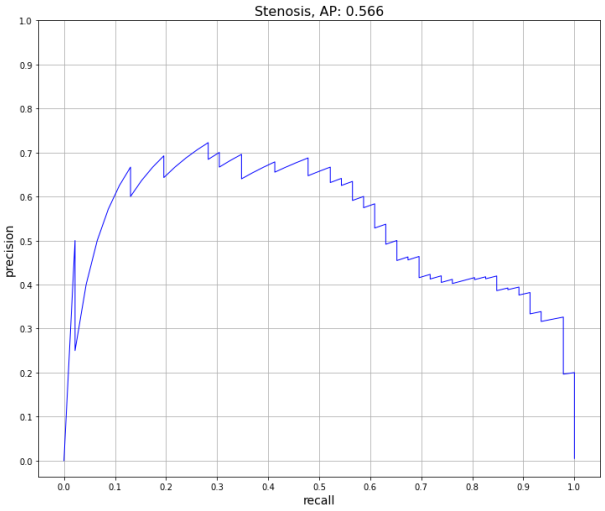


Figure 5.20: *Precision-Recall curve for the 5-fold. The mAP is 0.57.*

5.0.3 SSD300 prediction

In the Figures 5.21, 5.22 are reported the correct and the wrong predictions of stenosis in ICA images of test dataset.

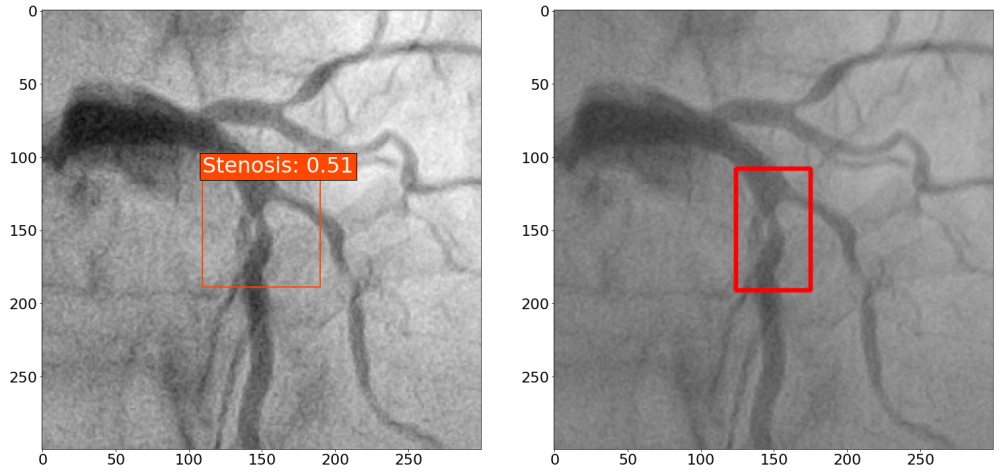


Figure 5.21: *A correct bounding box prediction, containing the stenosis, from 1-fold (on the left panel) and the corresponding bounding box annotated by the clinician (on the right panel).*

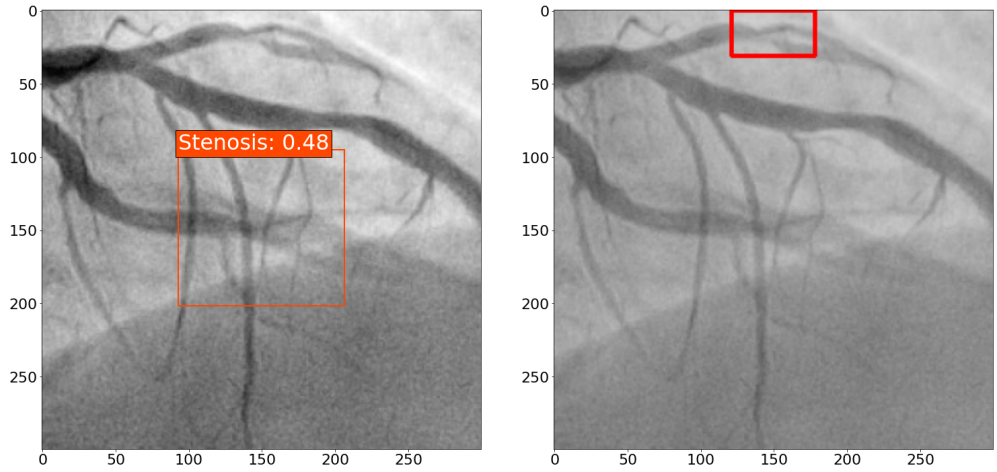


Figure 5.22: *A wrong bounding box prediction, without the stenosis, from 4-fold (on the left panel) and the corresponding bounding box annotated by the clinician (on the right panel).*

5.0.4 SSD7 prediction

In the figures 5.23, 5.24 are reported the correct and the wrong predictions of stenosis in ICA images of test dataset. In some cases multiple predictions for an image, lead to an inaccurate detection, since there is only one stenosis.

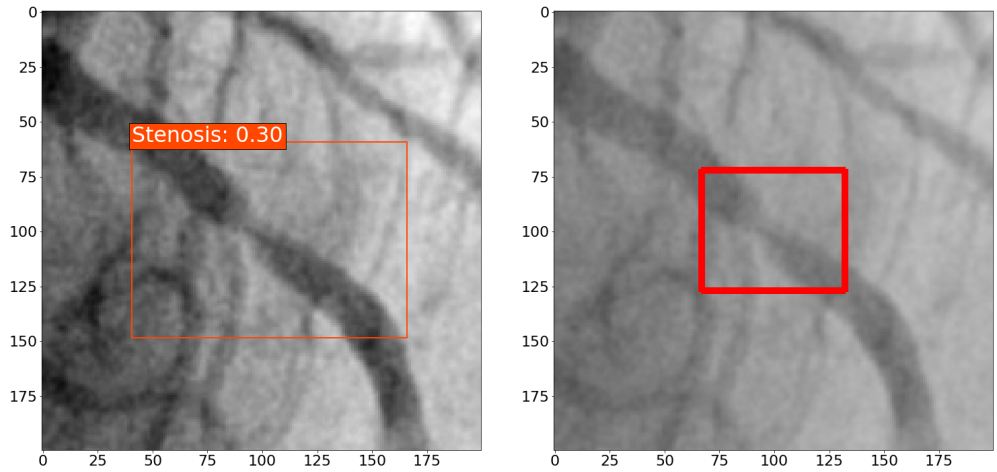


Figure 5.23: *A correct bounding box prediction, containing the stenosis, from 2-fold (on the left panel) and the corresponding bounding box annotated by the clinician (on the right panel).*

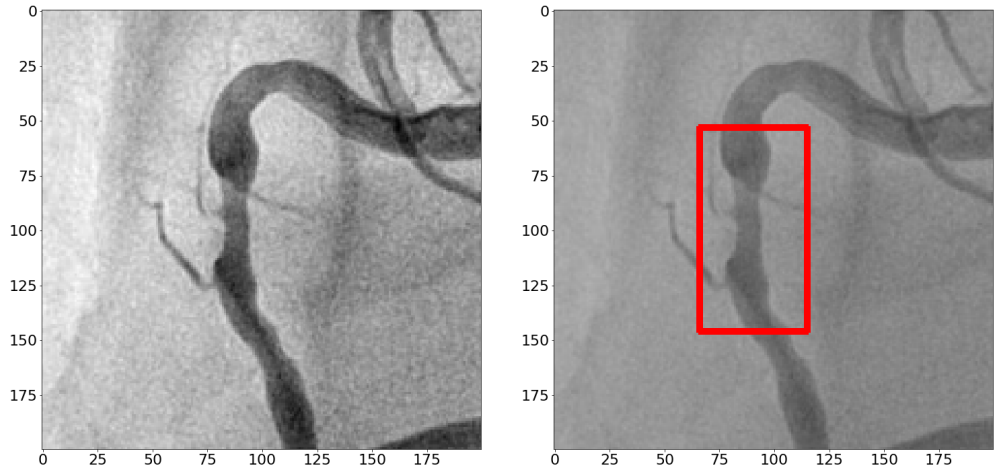


Figure 5.24: *A missing bounding box prediction, from 3-fold (on the left panel) and the corresponding bounding box annotated by the clinician containing the stenosis (on the right panel).*

5.0.5 Comparison between SSD300 and SSD7

In order to highlight similarities and differences between the SSD300 and SSD7 architectures, a statistical analysis is performed. Both IoU and DSC values obtained on the test set prediction for each fold are reported as boxplots in Fig. 5.25, graphically depicting numerical data through their quartiles. Moreover, the boxplots permit to point out the variability outside the upper and lower quartiles through whiskers. The outliers are plotted as individual points. The spacings between the different parts of the box indicate the degree of dispersion (spread) and skewness in the data. In addition they allow to estimate median and mean values. In the following Figures 5.25, 5.26, 5.27, 5.28, 5.29 the green triangle indicates the mean value, the blue line the median and the circle the outliers.

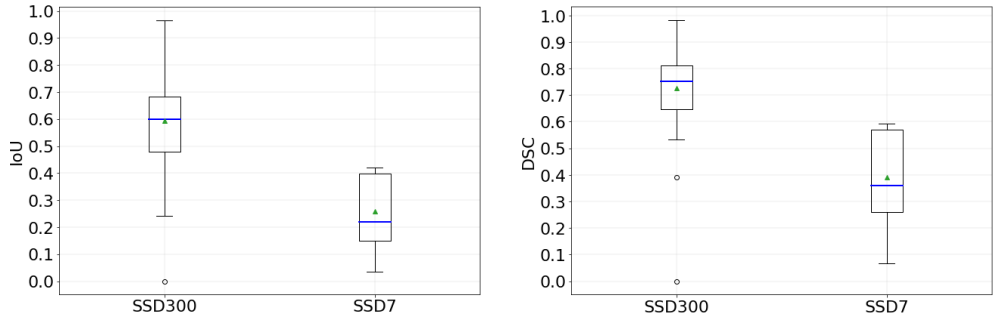


Figure 5.25: *Boxplots of the IoU (left panel) and DSC (right panel) for SSD300 (on the left) and SSD7 (on the right): results from the 1-fold.*

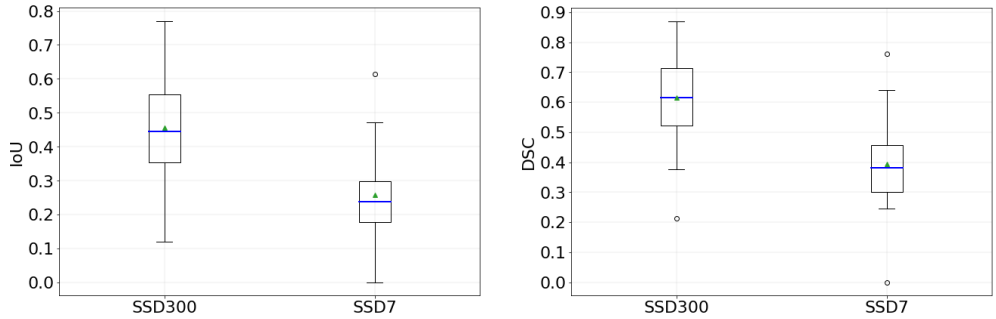


Figure 5.26: *Boxplots of the IoU (left panel) and DSC (right panel) for SSD300 (on the left) and SSD7 (on the right): results from the 2-fold.*

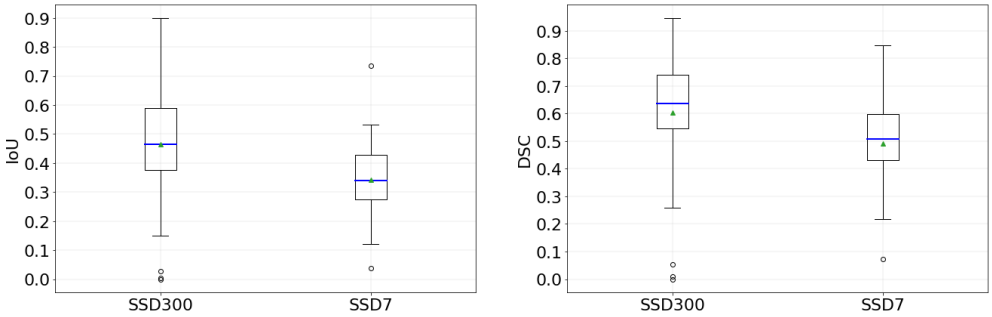


Figure 5.27: *Boxplots of the IoU (left panel) and DSC (right panel) for SSD300 (on the left) and SSD7 (on the right): results from the 3-fold.*

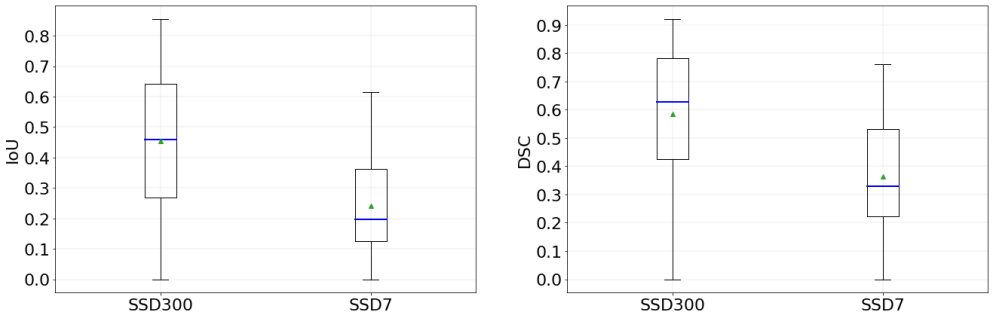


Figure 5.28: *Boxplots of the IoU (left panel) and DSC (right panel) for SSD300 (on the left) and SSD7 (on the right): results from the 4-fold.*

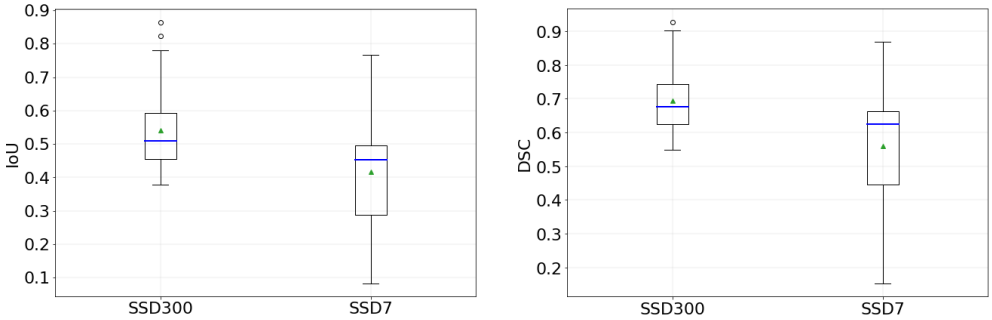


Figure 5.29: *Boxplots of the IoU (left panel) and DSC (right panel) for SSD300 (on the left) and SSD7 (on the right): results from the 5-fold.*

The goal of cross-validation is to test the model's ability to predict new data that was not used in estimating it, in order to flag problems like overfitting or selection bias and to give an insight on how the model will generalize to an independent dataset. Therefore, the mean value and their respective standard deviation of performances over the 5-fold both for SSD300 and SSD7 are reported in the following Table 5.16:

Table 5.16: *Performance metrics values for both networks SSD300 and SSD7 in all the 5-fold. The values are reported as mean value ±standarddeviation.*

	SSD300	**SSD7**
IoU	0.50 ± 0.06	0.30 ± 0.07
DSC	0.64 ± 0.06	0.44 ± 0.08
mAP	0.77 ± 0.20	0.53 ± 0.15

Chapter 6

Discussion

In this Chapter, the results presented in Chapter 5 are discussed.

In this work, an end-to-end stenosis detection CNN, without post- and pre-processing on images, capable of detect stenosis in ICA images automatically is proposed. Current literature comprises a small number of studies on stenosis detection through DL techniques, published in the very last few years; thus, there is plenty to explore on this subject. None of the state of art works has been capable to obtain good performances without including some image pre-processing, such as enhancing the visibility of the target and the quality of the images. Moreover, all the studies considered as references for this work base their effectiveness on the detection of the best video frames, on which then they perform the stenosis detection [24, 13, 11].

Therefore, to overcome the limitations met in literature and to investigate deeply this very recent topic, the SSD is chosen. In addition, this architecture can be suitable for the limited dataset and reduced number of training parameters.

The proposed approach results to be quite satisfactory especially in relation to the SSD300, keeping into account the preliminary stage in which the research is in the automatic detection of stenosis in ICA images. However, even though the SSD300 shows better performances in terms of IoU, DSC, and mAP compared to the SSD7. It should be further investigated in future studies ICA images with stenosis in a different location from the center, since the images considered

to train and evaluate these networks present the stenosis in most of the cases in the middle of the image. On the other hand, the SSD7 has lower performance than the SSD300 in all the 5-fold, but predicts bounding boxes even in areas other than the center of the image.

Furthermore, the values can be attributed to the lesser complexity of the architecture, unable to identify the ICA images features. The mean value and the standard deviation over the 5-fold are 0.50 ± 0.06 for IoU and 0.64 ± 0.06 for DSC in the case of SSD300; the value for SSD7 are 0.30 ± 0.07 and 0.44 ± 0.08 for IoU and DSC respectively, thus it is supposed that the network can generalize the problem in all the ICA images. It is noted that, in fold 4 the network under-performs in detecting the stenosis, in fact, the mAP, both for SSD300 and SSD7, is much lower than the average based on all the folds, affecting the overall fold's metrics. While, the best mAP is obtained in fold 5, evaluated with SSD300, which is equal to 1: the good predictions are recognized by the network itself, indicating also the greatest prediction confidence score of 0.59. Thanks to K-fold cross-validation, the results obtained promise good network behavior on data never seen before, in addition, it is possible to verify the data uniformity, since all the folds behave similarly, as demonstrated by the small standard deviation values. Results are slightly lower than values obtained in the literature and it is not always possible to make a numerical comparison, since the metric performances utilized in the reference studies are different from those used in this approach. It is necessary to highlight that the results depend on the annotations, affected by intra-operator variability, since they are drawn manually by three clinicians. Moreover, it is possible to accept values with a certain margin of tolerance, since the aim is to identify coronary stenosis, which is correctly done by visual inspection of the stenosis prediction, except for some cases. The variability of the results obtained highlights the difficult task this thesis tries to face, due to the high variability of the target object to detect and the quality of the images. Moreover, several aspects of ICA contribute to make this task challenging: ICA images contain significant image noise and have limited image contrast. In the medical field, a very high number of images are

acquired and stored every day, but most of the time their quality is not sufficient to treat them as with natural images. Nevertheless, this network, pre-trained on natural images, has been chosen for its ability to detect very well on other object detection tasks, and the results suggest that it can be effective also in medical images, but the availability of a large dataset could promise better performances.

Chapter 7

Conclusion

In the literature, DL approaches, in particular CNNs, were revealed to be very suitable to implement intelligent systems for supporting decisions in the clinical field. Among various methods for diagnosing CAD, ICA is the gold standard method. It helps to evaluate the severity and the extent of stenosis, as well as monitoring the progress of revascularization during angioplasty. However, detecting stenosis in these images is very challenging, since they have intrinsic complexities due to complex vessel structures, poor contrast between vessels and surrounding tissues, nonuniform illumination, and overlap of background structures with inhomogeneous intensities. Moreover, manual detection of stenosis is subjective and time-consuming, requiring rich clinical experience. Therefore, developing an ICA-based automatic detection algorithm can improve diagnostic efficiency and confidence. Driven by these limitations, this thesis proposes an end-to-end approach, through a CNN, to automatically detect stenosis in ICA images.

Results of this work are good, despite the fact that they are slightly lower than values obtained in the literature. State-of-the-art is very limited and dates back to the last few years. However, the studies considered perform a strong preprocessing on the raw images to be really effective. Thus, despite their better performances, the proposed approaches are difficult to apply in clinical practice. By visual inspection, SSD provides correct and accurate detection of stenosis in

ICA images, even though in some images the predictions are erroneous, such as multiple detections or in wrong positions. Although K-fold cross-validation is used to make up for the reduced number of images as much as possible, future works will widen the dataset, which will improve the performance.

In conclusion, this work is a good starting point to obtain automatic stenosis detection in ICA images but it needs further improvements to be used directly in clinical practice, minimizing the risk of misinterpretation and accelerating the decision-making regarding the proper treatment strategy.

List of Figures

List of Tables

Bibliography

[1] S. Jiangping, Z. Zhe, W. Wei, S. Yunhu, H. Jie, W. Hongyue, Z. Hong, and H. Shengshou, "Assessment of coronary artery stenosis by coronary angiography: A head-to-head comparison with pathological coronary artery anatomy," *Circulation: Cardiovascular Interventions*, vol. 6, no. 3, 2013.

[2] E. J. Benjamin, P. Muntner, A. Alonso, M. S. Bittencourt, and et al., "Heart disease and stroke statistics—2019 update: A report from the american heart association," *Circulation*, vol. 139, no. 10, 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," *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 91–95, 2021.

[4] 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," *Acta Biomedica*, vol. 81, no. 3, pp. 157–164, 2010.

[5] V. V. Danilov, K. Y. Klyshnikov, O. M. Gerget, A. G. Kutikhin, V. I. Ganyukov, A. F. Frangi, E. A. Ovcharenko, and S. R. Walani, "Real-time coronary artery stenosis detection based on modern neural networks," *Scientific Reports*, vol. 11, no. 7582, 2021.

[6] 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.

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

[8] M. Chen, X. Wang, G. Hao, X. Cheng, C. Ma, N. Guo, S. Hu, Q. Tao, F. Yao, and C. Hu, "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, 2020.

[9] 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, pp. 1847–1859, 2013.

[10] 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*, 2018.

[11] 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," *Computers in Biology and Medicine*, vol. 118, 2020.

[12] H. Zhang, L. Mu, S. Hu, B. K. Nallamothu, A. J. Lansky, B. Xu, G. Bouras, 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, pp. 239–247, 2018.

[13] J. H. Moon, D. Y. Lee, W. C. Cha, M. J. Chung, K.-S. Lee, B. H. Cho, and J. H. Choi, "Automatic stenosis recognition from coronary angiography

using convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 198, 2021.

[14] M. J. Raphael and R. M. Donaldson, "A "significant" stenosis: thirty years on," *The Lancet*, vol. 333, no. 8631, pp. 207–209, 1989.

[15] M. R. Johnson, D. D. Mcpherson, S. R. Fleagle, M. M. Hunt, F. F. Hiratzka, R. E. Kerber, M. L. Marcus, S. M. Collins, and D. J. Skorton, "Videodensitometric analysis of human coronary stenoses: validation in vivo by intraoperative highfrequency epicardial echocardiography," *Circulation*, vol. 77, no. 2, pp. 328–366, 1988.

[16] L. L. Leape, R. E. Park, T. M. Bashore, J. K. 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, pp. 106–113, 2000.

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

[18] 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. 12, pp. 1793–1800, 2013.

[19] J. Brieva, M. Galvez, and C. Toumoulin, "Coronary extraction and stenosis quantification in x-ray angiographic imaging," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, September 1-5, 2004.

[20] M. Fatemi, S. Mirhassani, and E. Ghasemi, "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, no. 9, pp. 27–33, 2011.

[21] 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, no. 8, 2020.

[22] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *European Conference on Computer Vision*, pp. 818–833, 2014.

[23] 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," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018.

[24] 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," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019.

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

[26] 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 (British Cardiac Society)*, vol. 104, no. 14, pp. 1156–1164, 2018.

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

[28] E. S. Holmboe and S. J. Durning, "Assessing clinical reasoning: moving from in vitro to in vivo," *Diagnosis*, vol. 1, no. 1, pp. 111–117, 2014.

[29] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," *Multimedia Tools and Applications*, 2021.

[30] A. C. Neves, I. Gonzalez, J. Leander, and R. Karoumi, "A new approach to damage detection in bridges using machine learning," *nternational Conference on Experimental Vibration Analysis for Civil Engineering Structures*, pp. 73–74, 2017.

[31] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological review*, vol. 65, p. 368, 1958.

[32] D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural networks*, vol. 32, pp. 333–338, 2012.

[33] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning vol. 1." MIT press Cambridge, 2016.

[34] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," pp. 1–11.

[35] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," *Computer Science, Mathematics*, pp. 270–279, 2018.

[36] E. S. Olivas, J. D. M. Guerrero, M. M. Sober, J. R. M. Benedito, and A. J. S. Lopez, "Handbook of research on machine learning applications and trends: Algorithms, methods and techniques."

[37] Z. Li, M. Donga, S. Wend, X. Hub, P. Zhouc, and Z. Zeng, "Clu-cnns: Object detection for medical images," *Neurocomputing*, vol. 350, pp. 53–59, 2019.

[38] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *Computer Vision – ECCV*, pp. 21–37, 2016.

[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *The 3rd International Conference on Learning Representations (ICLR)*, 2015.

[40] NEMA, "Digital imaging and communications in medicine (dicom)," *The National Electrical Manufacturers Association (NEMA)*, 2008.

[41] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An introduction to statistical learning: with applications in r," in *An Introduction to Statistical Learning: with Applications in R.* Springer, 2017, pp. 120–130.

[42] B. D. Ripley, "Pattern recognition and neural networks." Cambridge ; New York : Cambridge University Press, 2007, pp. 350–355.

[43] http://mscoco.org/dataset/ detections leaderboard, "Coco: Common objects in context." 2016.

Ringraziamenti

Per arrivare alla conclusione di questo percorso è stato essenziale il supporto di alcune persone che desidero ringraziare. Ringrazio, innanzitutto, il Professor Emanuele Frontoni, relatore di questa tesi per avermi dato la possibilità di lavorare a questo progetto, in un ambito del tutto nuovo per me, permettendomi di ampliare le conoscenze e le competenze. Desidero esprimere la mia riconoscenza ai correlatori di questa tesi: al Professor Lorenzo Scalise per la disponibilità e professionalità dimostrata in questi mesi; a Sara per le grandi capacità organizzative e per le sue preziosissime indicazioni e suggerimenti. A Maria Chiara, Francesca e Maria Chiara che da dietro le quinte sono state delle perfette direttrici dei lavori: grazie di cuore per la dedizione, la disponibilità e la pazienza dimostratami sin dall'inizio di questo lavoro.

Immensa gratitudine va alla mia famiglia: ai miei genitori per spronarmi ogni giorno a fare del mio meglio, per avermi insegnato che non esistono difficoltà che non possa superare con lo studio, la dedizione e l'impegno; alle mie sorelle Elisabetta, Emanuela ed Elena per essere sempre presenti e pronte a darmi il loro sostegno, per aver creduto in me. Ai miei splendidi nipotini Vittoria e Mattia, non so come farei senza di voi, siete il mio sole ogni giorno, vi adoro! A Birillo, il mio cucciolino, sempre pronto a festeggiare il mio ritorno a casa e a subire le mie coccole.

Grazie di cuore alle mie "Fantastiche": a Greta, perché ormai da 11 anni è la mia fedele compagna di banco e non solo, non potevamo chiedere conclusione migliore! A Federica, per il suo grande cuore, per avere le parole giuste al momento giusto, per avermi ricordato di non mollare mai. Ad Alessia, per la

sua solarità e affettuosità, sempre pronta a festeggiare i miei successi. A Lucia, con cui non ci siamo mai perse, anche a km di distanza, l'affetto e la stima non sono mutati.

A Guido, Luca, Nicolas, Marta per essere stati compagni, ma soprattutto alleati, di studio. Mi mancheranno tanto le risate e le serate in vostra compagnia. Vi porto nel cuore. Ad Agostino, per aver sopportato tutte le mie lamentele e i miei cambi di umore. Grazie per il sostegno, per i tuoi modi scherzosi di fare, ma soprattutto per avermi sempre ricordato le mie capacità e potenzialità.

Grazie veramente di cuore a tutti! Vi voglio tanto bene!