Automated prediction of major adverse cardiovascular events

Mario Viti

université PARIS-SACLAY

Prédiction automatique des épisodes cardio-vasculaires adverses majeurs

Automated prediction of major adverse cardiovascular events

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°580 : sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat: Traitement du signal et des images
Graduate School : Informatique et sciences du numérique. Référent : CentraleSupélec

Thèse préparée dans l' unité de recherche Centre de Vision Numérique (Université Paris-Saclay, CentraleSupélec), sous la direction de **Hugues TALBOT**, Professeur, CENTRALE SUPELEC et le co-encadrement de **Nicolas GOGIN** Ingénieur, GE Healthcare

Thèse soutenue à Paris-Saclay, le 23 Novembre 2022, par

Mario VITI

Composition du jury

Membres du jury avec voix délibérative

Nicolas PASSAT Assistant professeur, Université de Reims Cham-pagne Ardenne	Président
Elsa ANGELINI Directrice de recherche, NIHR Imperial Biomedical Research Centre (BRC)	Rapporteur & Examinatrice
Ismail BEN AYED Professor, ETS Montréal (École de Technologie Supérieure)	Rapporteur & Examinateur
Elodie PUYBAREAU Assistant professeur, EPITA	Examinatrice

THESE DE DOCTORAT

NNT : 2022UPASG084

Chapter 1

Clinical Context

1.1 Computed Tomography (CT) Plaque characterization

Coronary artery plaques are asymmetric thickenings of the inner vessel wall due to the accumulation of varying quantities of foamy macrophages, lipids, blood products, smooth muscle cells, collagen, necrotic debris, and calcium (figure 1.1). Imaging techniques have been employed for the clinical investigation of coronary artery plaque findings and assessing their associated cardiovascular risk. The ultimate goal is to identify individuals at increasing risk of Coronary Artery Disease (CAD) for the prediction of Major Adverse Cardiovascular Events (MACE): chronic angina, infartus, and sudden death. Among imaging techniques, Intravascular Ultrasounds (IVUS) and Optical Coherence Tomography (OCT) provide the closest match to the underlying histopathology of coronary plaques. However, invasive nature of the techniques and high cost are limiting factors in practice. Less invasive techniques such as Magnetic Resonance Imaging (MRI) or CT, are more appropriate for prevention and screening purposes. Although CT presents specific limitations, it provides valuable information that makes it a potentially useful technique for early identification and characterization of coronary plaques and the prevention of MACE [1, 2, 3].

1.1.1 Coronary Computed Tomography Angiography (CCTA) coronary plaque detectability

Image quality in cardiac imaging depends on the system's time resolution in addition to its spatial and contrast resolution. The major limiting factor to achieving optimum temporal resolution is cardiac motion. Recent ECG-gated CT data acquisitions are able to cover the entire heart in a single breath-hold. As a result, the ability to breathe and move the diaphragm are no longer severely restricting constraints. It is prudent to lower the heart rate to less than 65 beats/min to obtain optimum image quality. In CCTA, the voxel size is the major determinant of spatial resolution. In the x-y plane, the smallest voxel size determines spatial resolution at about $0.35\,\mathrm{mm}$. Such high resolution necessitates more difficult reconstruction methods, thinner collimators, and smaller detector sizes.

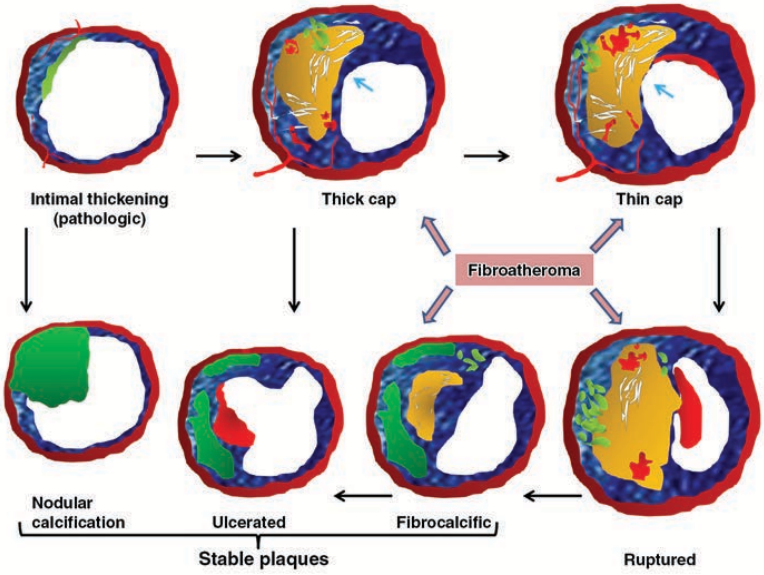


Figure 1.1: Drawing shows morphological characteristics of plaque development: early pathologic vessel wall intima thickening may develop into stable plaques (calcifications) or ruputerd plaques (atheroma). Stable plaques states can develop in rutpture as a consequence of hemorhage [1].

Overall, improving spatial resolution requires advancements in all three dimensions and is not an easy undertaking. In reconstruction, noise resolution filters are frequently used, which has an impact on the underestimation of intensely low values like lipid infiltration and overestimation of intensely high values that cause the blooming effect of coronary calcification. While the x-y axis resolution depends on the size of the matrix and the FOV reconstructed area, the longitudinal spatial resolution (minimum achievable slice thickness) is related to the x-ray collimator and longitudinal dimension of a detector, which is about $0.5\,\mathrm{mm}$. An isotropic representation of the coronary anatomy has interpolated values along one of the dimensions, resulting in an anisotropic voxel that may deform. Contrast resolution in CT is the ability of the system to distinguish low-contrast structures in an image and express a relative attenuation in Hounsfield Units (HU). The most important parameter influencing coronary resolution is noise, which depends on the statistical fluctuation in the number of x-ray photons, patient size, quality of detection, and reconstruction algorithms [4]. Tube voltage (expressed in kV) influence on a reliable depiction of soft-tissue is debated with varying level of HU observed in different studies [5, 6]. More recent CT scanners can simultaneously acquire at two energy levels, these Dual-energy CT have the potential to provide additional data for better differentiation of plaque components. Image quality and plaque representation can also be affected by contrast agent parameters, which demonstrates how iodine dose in low- and high-contrast conditions influences the attenuation of both calcified and non-calcified plaques. Iodine concentration has an impact on lipid-rich and fibrotic

plaque representation as well. Iodine concentration has an impact on lipid-rich and fibrotic plaque representation as well. Instead of using filtered back projection, iterative reconstruction algorithms are used to minimize noise in thin-cut slices. Plaque representation is unaffected by these procedures, however the measurement of tiny calcifications is hampered by the reconstruction filter selection. In CCTA the HU attenuation of different coronary plaque composition has been observed to be $-30\,\mathrm{HU}$ to $60\,\mathrm{HU}$ for lipid plaque, $61\,\mathrm{HU}$ to $149\,\mathrm{HU}$ for fibrous plaque, $150\,\mathrm{HU}$ to $1300\,\mathrm{HU}$ for calcium, these thresholds show good agreement with IVUS [7, 8, 9]. Such discriminants are mostly valid for vessels with coronary lumen above $2\,\mathrm{mm}$. These thresholds, however, have been challenged by several studies and present overlapping intervals ($36\,\mathrm{HU}$ to $91\,\mathrm{HU}$ for lipid $56\,\mathrm{HU}$ to $136\,\mathrm{HU}$ 96 ± 40 HU for fibrous), and can be influenced by all the known factors, the sole HU value is therefore an unreliable characterization feature [1]. Morphological markers that consistently match atherosclerotic plaques include micro- and macrocalcification, positive remodeling, fibrous cap, necrotic core, and napkin ring. [8, 2]. These are better represented in IVUS and OCT. A fibrous cap of $400\,\mu\mathrm{m}$ is hardly detectable using current CT spatial and contrast resolution. While carotid hemorrhage can be detected in CT and MRI with current techniques, it is impossible to detect in coronary structures.The majority of research distinguish between mixed, soft (non-calcified), and calcified plaques. However, there are intra characteristics with different clinical outcomes. Moreover, the early detection can evolve in different types of plaques, as shown in figure 1.1. Overall CT constitutes a non-invasive option for detection and quantification of coronary plaques. However, CT characterization of coronary plaques remains challenging even with optimal image quality. Observable findings in CT include spotty calcifications, napkin ring, lipid core, and positive remodeling.

1.1.2 CCTA and OCT plaque evaluation

OCT is an imaging technique that accesses lesion features and plaque morphology for coronary artery disease by using near-infrared light to produce high-definition images of the artery with high precision. OCT is performed bloodless with a catheter that scatters light impulses for 2D live images acquisition. Although only small incisions are needed to introduce the catheter, the technique is nonetheless very invasive because an artificial stenosis must be created in order to achieve a bloodless condition. OCT is akin to IVUS: sound waves are substituted by light waves thus, resulting in higher resolution images. In [10] sixty-eight plaques were evaluated by CCTA and OCT. Thin-cap fibro atheroma Plaques associated with positive remodeling in CCTA and low plaque attenuation were associated with. CCTA adverse plaque features correlates with OCT findings, while CCTA shows high sensitivity, noisy acquisition still affects the precision (false positives).

1.1.3 Calcium Score Computed Tomography (CSCT) coronary calcium detectability

The variability in the representation of plaques can be mitigated by the fixed image acquisition and the absence of contrast agents. For coronary calcifications, cardiac CT im-

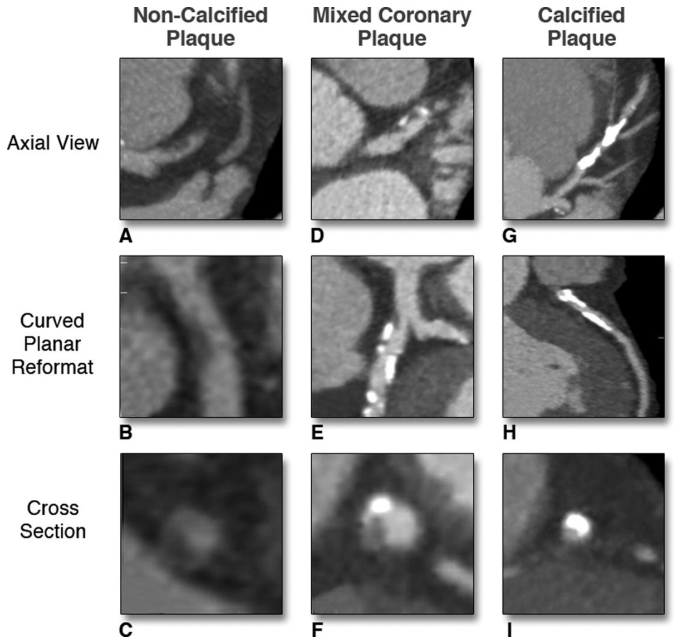


Figure 1.2: Different visualizations of various types of Coronary Plaques by CCTA. Axial visualization of coronary plaques (A, D, and G), curved planar reformatted (B, E, and H), and cross-sectional (C, F, and I) views. The 3 main types of coronary plaques are shown: non-calcified plaque, mixed plaque, and calcified plaques.

ages are acquired at 2.5 or $3.0\,\mathrm{mm}$ slice thickness and an axial resolution of $0.5\,\mathrm{mm}^2$ and a cropped FOV around the heart (cardiac CT) (fig. 1.3). With this parametrization, acquisitions are less noisy, and all voxels above 130 HU are binned to 4 intervals (1=130-199, 2=200-299, 3=300-399, 4=400-) and associated with different calcium concentrations. Calcium deposits quantified using this procedure output the Agatston score [11] which have been consistently associated with cardiovascular risk, and it is currently used as a golden standard for early detection and stratification of CAD.

1.1.4 CT recent advances

The latest achievement in CT image quality and efficiency in radiation dosing is Photon Counting Computed Tomography (PCCT). Because of the shorter acquisition times with respect to solid-state scintillator detector (SSD), PCCT are less subject to artifacts and allow smaller slice thickness, thus potentially increasing the diagnostic confidence for CCTA examinations [12]. A study compared acquisitions with dual layer CT (DLCT) (512 matrix, 0.67 slice thickness) with a PCCT (1024 matrix, 0.2 slice thickness). The quality was first assessed based on phantom test acquisitions, then on patients that underwent CCTA both with DLCT and PCCT acquisitions. Three cardiologists independently assessed the quality score based on their CAD diagnosis confidence in representing coronary plaques, stents, and clarifications quantification. This study concluded that

Figure 1.3: Non-injected CT cardiac examination. This acquisition has fixed parameters of 120 Kvp, a slice thickness from $2.5\,\mathrm{mm}$, $3.0\,\mathrm{mm}$ and an axial reolution of $0.5\,\mathrm{mm}^2$. High intensity calcium can be isolated by masking intensities above 130 HU.

CCTA using PCCT provides higher clinical confidence in assessing CAD and lumen quantification. However, this study did not provide any feedback concerning OCT imaging, meaning that PCCT allows more reliable findings, but the CT visibility of vulnerability factors like early necrosis and hemorrhages have not been assessed directly.

1.1.5 Coronary Artery Calcium (CAC) Prognostic Value

Prospective Multicenter Imaging Study for Evaluation of Chest Pain (PROMISE) is a randomized multicenter controlled retrospective Trial published in 2017 [13]. CAC is an established predictor of MACE in asymptomatic patients. This study aimed to compare anatomical testing (CAC burden estimation) with functional testing (FT) in estimating the prognosis of symptomatic patients. In the PROMISE, patients presenting stable chest pain (or dyspnea) and intermediate obstructive CAD were randomized to FT (stress ECG) or anatomic testing CAC. During this study, 4029 patients underwent CAC estimation, and 4602 patients underwent FT (stress ECG). Both populations were stratified as normal or mildly, moderately, or severely abnormal (for CAC: 0, 1-99, 100-400, and >400 Agatston score (AS), respectively; for FT: normal, mild=late positive treadmill, moderate=early positive treadmill or single-vessel ischemia, and severe=large ischemic region abnormality). Patients were observed over a median follow-up of 26.1 months. Among stable outpatients presenting with suspected CAD, most patients experiencing MACE have measurable CAC at baseline, and fewer than half have any abnormalities on FT.

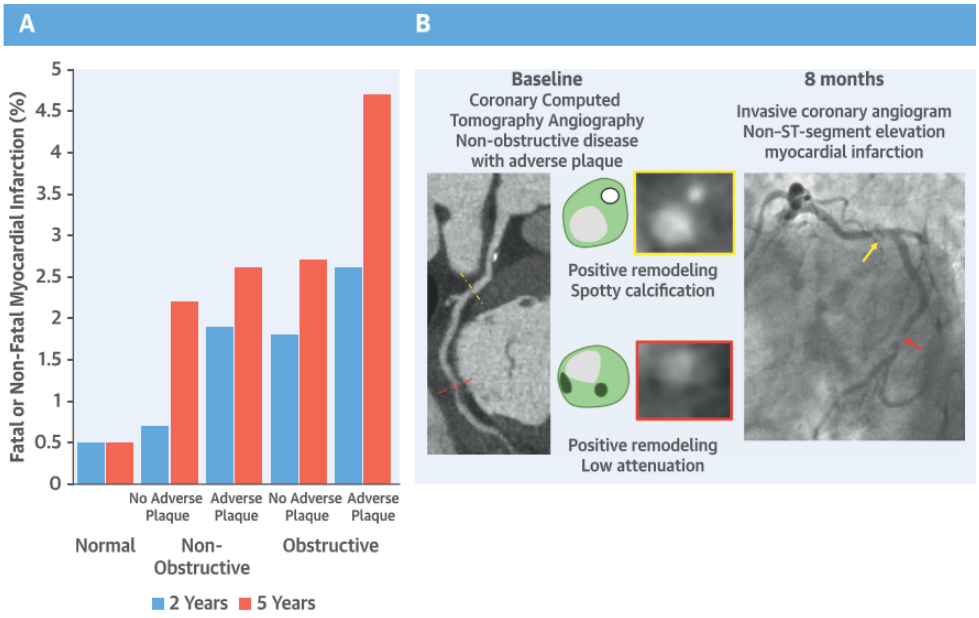


Figure 1.4: (A) Bar graph of the frequency of coronary heart disease death or nonfatal myocardial infarction at 2 and 5 years for patients with normal coronary arteries and non-obstructive or obstructive disease with and without adverse plaque. (B) Coronary computed tomography angiography and invasive coronary angiography images from a patient with non-obstructive coronary artery disease who had a subsequent non-ST-segment elevation myocardial infarction. The red/yellow dotted lines and arrows correspond to the location of the plaques in the red/yellow boxes.

However, an abnormal FT had higher MACE specificity, leading to similar discriminatory capability of both tests. The association of known cardiovascular risk factors and the prognosis of coronavirus 2019 (COVID-19) has been recently emphasized. CAC score is considered a risk modifier in the primary prevention of cardiovascular disease. We hypothesized that the absence of CAC might have an additional predictive value for an improved cardiovascular outcome in hospitalized COVID-19 patients.

1.1.6 CT coronary plaque relevance in predicting MACE

The investigation of patients with suspected CAD has previously focused on functional assessments to identify the presence of myocardial ischemia as a consequential surrogate marker of coronary artery stenosis in proximal coronary regions. In contrast, non-invasive imaging with CCTA has the ability to provide precise structural information about the coronary artery's inner and outer wall and can reliably depict the presence and constituents of atherosclerotic plaque even in the absence of proximal stenosis. Clinical studies have identified an association between high-cardiovascular-risk plaque rupture and adverse plaque characteristics that includes positive remodeling, a large necrotic

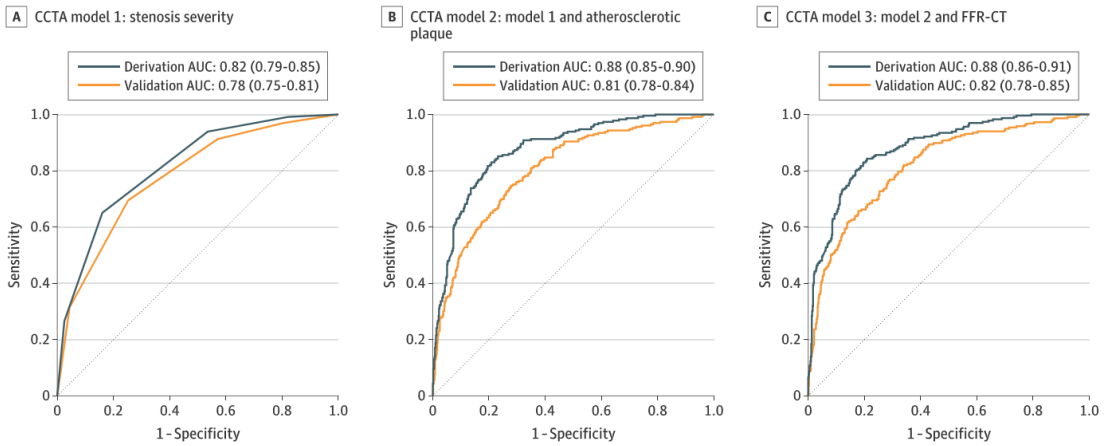


Figure 1.5: Model 2 variables: stenosis severity, number of lesions with mild stenosis, non-calcified plaque volume, high risk plaque features, lumen volume.

core and micro-calcifications, and a thin fibrous cap (high-risk plaque features) [2]. Anatomical and functional analysis implies a comprehensive plaque characterization and the measure of blood fractional flow reserve (FFR) respectively. While both have the final goal of predicting cardiovascular risk in individuals, the anatomical analysis is based on the non-invasive CCTA technique, while the functional analysis is based on an invasive FFR procedure. An empirical study [3] conducted on 612 patients from 23 clinical sites designed to assess whether the anatomical assessment of high-risk plaques could push the productiveness of CAD beyond the FFR functional analysis. The ROC analysis conducted on a binary per-vessel prediction (figure 1.5) shows how a comprehensive anatomical interpretation with CCTA is superior to the functional imaging and diagnosis of FFR.

1.2 Data collection and annotation

1.2.1 GE Healthcare lesion Dataset

The dataset used in this work includes 150 coronary CCTA scans collected from clinical sites in France and Italy. Images were acquired with tube voltage ranging from $100\,\mathrm{kV}$ to $120\,\mathrm{kV}$, a current from $600\,\mathrm{mA}$ to $1000\,\mathrm{mA}$, a pixel spacing from $0.35\,\mathrm{mm}$ $0.42\,\mathrm{mm}$ and $0.6\,\mathrm{mm}$ slice thickness. Each patient underwent both CCTA and anatomical tests [11] and was assigned a Coronary Artery Disease Reporting & Data System (CAD-RADS) [14] score by trained radiologists. Each scan is paired with annotated coronary centerlines: manual annotation was performed internally under clinical supervision. Coronary-artery centerlines are corrected and labeled by experts during medical examination, and visually assessed findings are reported in a clinical review document. Findings are reported

by coronary label, and approximate location is given as proximal medial or distal. A starting and ending point is marked so that each point of the centerline is associated to a finding (Figure 1.6). Each positioned annotation represents an anatomical indicator of CAD, and it is labeled to encode the information of stenosis degree (lumen occlusion) and plaque characterization (composition) reported in table 1.1. Manual annotations were carried out to match the clinical report, and the final annotations were submitted to a trained cardiologist for review. Stratification is performed at the patient level to avoid sharing information among different dataset splits. Each patient is assigned an anatomical indicator annotation. In the case of multiple indicators per patient, the less occurring in the whole annotation distribution is assigned. A subset of 50 patients with 224 annotated segments with plaques (49 soft, 78 mixed, 97 calcified plaques, and 182 non-occlusive, 42 occlusive stenoses) was chosen for testing using CAD-RADS as stratification criteria for random sampling. The remaining 100 patients were split 70-30 using the same stratification strategy. The collection was carried out to obtain a CAD-RADS ≤ 4 uniform distribution: 24, 38, 30, 33, 25 patients with CAD-RADS from 0 to 4 respectively. This sample contains a higher than normal concentration of plaques as CAD-RADS distribution usually follows an exponentially decaying function. For example, in [15] 28 non-calcified annotated plaques were found, while for this study, 104 non-calcified plaques were annotated. Additionally, a set of 75 semi-annotated examinations has been collected. These examinations contain a corrected centerline but not the plaque annotations information.

Comp.\Stenosis	0%	1-24%	25-49%	50-69%	70-99%	LMA >50%	100%
Calcified	00	01	02	03	04	04B	05
Mixed	10	11	12	13	14	14B	15
Soft	20	21	22	23	24	24B	25
Modifiers	S	G	V				
	stent	graft	vulnerable				

Table 1.1: Code table for coronary plaque annotations of findings. Each anathomical indicator of CAD is assigned a tuple of {plaque composition, stenosis degree, modifier}.

1.2.2 Société Française de Radiologie (SFR) challenge Dataset

The dataset provided by the organizers of the SFR data challenge 2020 was composed of three batches of CT acquisitions collected from different clinical sites and provided at different stages of the data challenge. The first batch of 100 CT examinations was made available at the beginning of the data challenge (J1 data set), then a second larger batch of 322 CT examinations was released two days before the end of the challenge (J2 data set). A final test data set of 98 CT examinations (J3 data set) was given on the last day of the data challenge and used as a test set to evaluate in one hour the results submitted by all the participants of the data challenge. All electrocardiogram (ECG)-gated CT examinations were acquired without contrast material by using a fixed tube voltage of $120\,\mathrm{kV}$, a slice thickness of $2.5\,\mathrm{mm}$ to $3.0\,\mathrm{mm}$ and a current ranging from $40\,\mathrm{mA}$ to

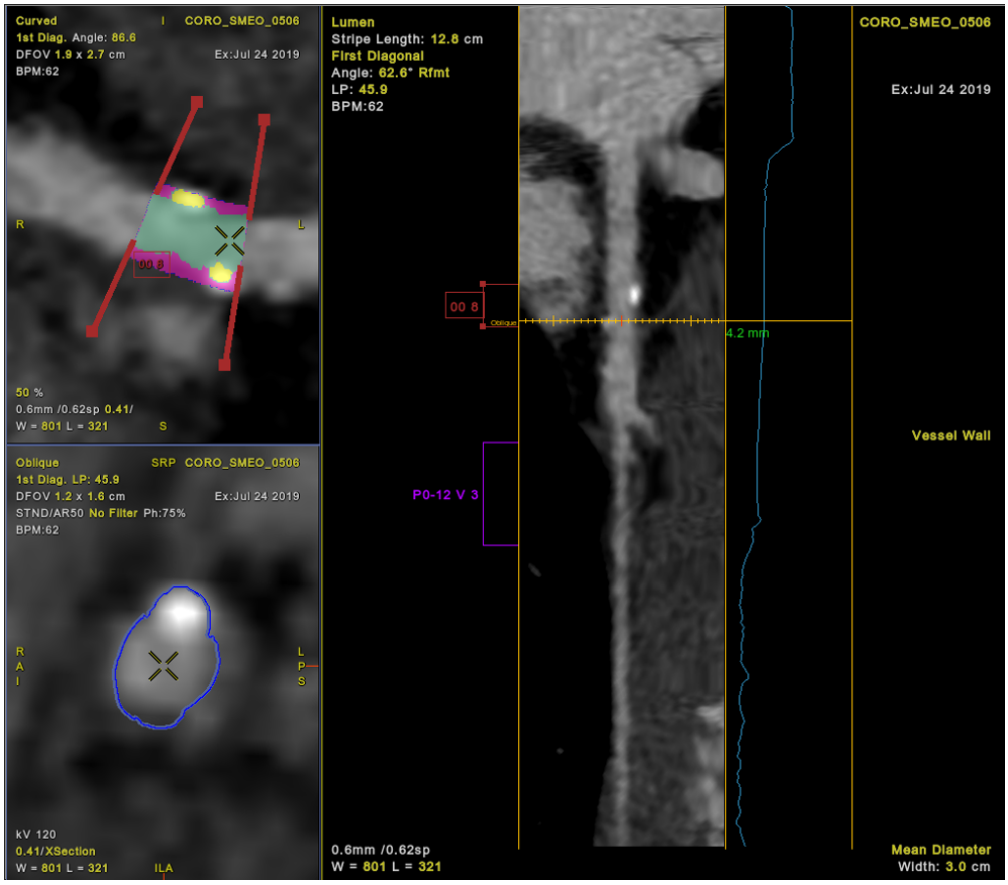


Figure 1.6: Annotation visualization tool developed at GE Healthcare for coronary clinical review.

$200\,\mathrm{mA}$, depending on the weight of the subject and the habits of each center to optimize image quality. The need for informed consent from the patient was waived by the national commission in charge of data and privacy protection (CNIL). Furthermore, the SFR ensured data protection to radiologists wishing to participate in the challenge by integrating an automatic anonymization process into the platform. The challenge organizers ensured a uniform distribution of the examinations between four CT vendors in order to avoid results depending on image characteristics specific to manufacturers. Expert manual annotation of each patient image series was provided as a set with a list of 3D positions of each coronary calcification. Finally, based on the CAC burden computation results, each patient is assigned a standard risk category: A, B, C, D, or E when the AS corresponding to zero, 1–10, 11–100, 101–400, > 400, respectively. Categories A and B thus represented a low level of cardiovascular risk, whereas class C represented the intermediate-risk group and categories D and E the high-risk group. Patients included in the final test data set (J3) were classified as intermediate risk by the clinicians in charge of these subjects from the same preventive medicine center according to the

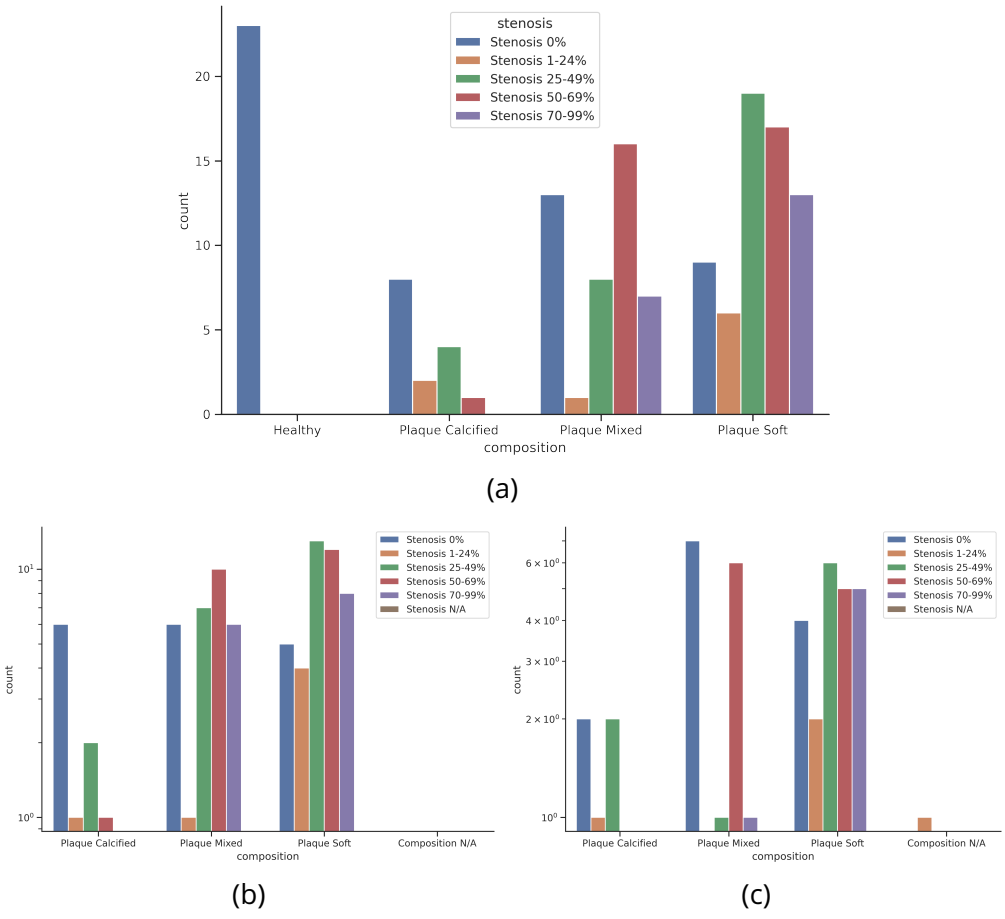


Figure 1.7: Patient level plaque composition distribution with stenosis degree hue for all the dataset, train/valid, and test (a)(b)(c) respectively. Each patient is labeled with an indicator which is composed of plaque characterization and stenosis degree, in case of multiple indicator per patient, the less probable occurring in the whole annotation distribution is assigned. The stratification is based on the CAD-RADS assigned by a trained cardiologist during examination.

AS formula proposed by the European Society of Cardiology. The population presented a mean age of 60 ± 13 (standard deviation) years (range: 43–84 years). However, based on the manually calculated AS, and unbeknownst to the challenge participants, the distribution of the risk class categories of the subjects selected by the challenge organizers was homogeneous.

Dataset		# Patients	Agatston Score risk category					Vendor distribution
			A	B	C	D	E	
J1	Train	100	20%	16%	21%	21%	22%	GE Healthcare 56% Siemens Healthineers 44%
J2	Train	322	30%	6%	24%	18%	23%	GE Healthcare 75% Siemens Healthineers 15% Canon Medical System 10%
AD	Train	361	34%	5%	20%	20%	21%	GE Healthcare 98% Siemens Healthineers 1% Philips Healthcare 1%
J3	Test	98	20%	20%	20%	18%	20%	Siemens Helthineers 100%
OS	Test	40	20%	30%		20%	30%	GE Healthcare Siemens Healthineers Philips Healthcare Toshiba Medical

Chapter 2

State of the art

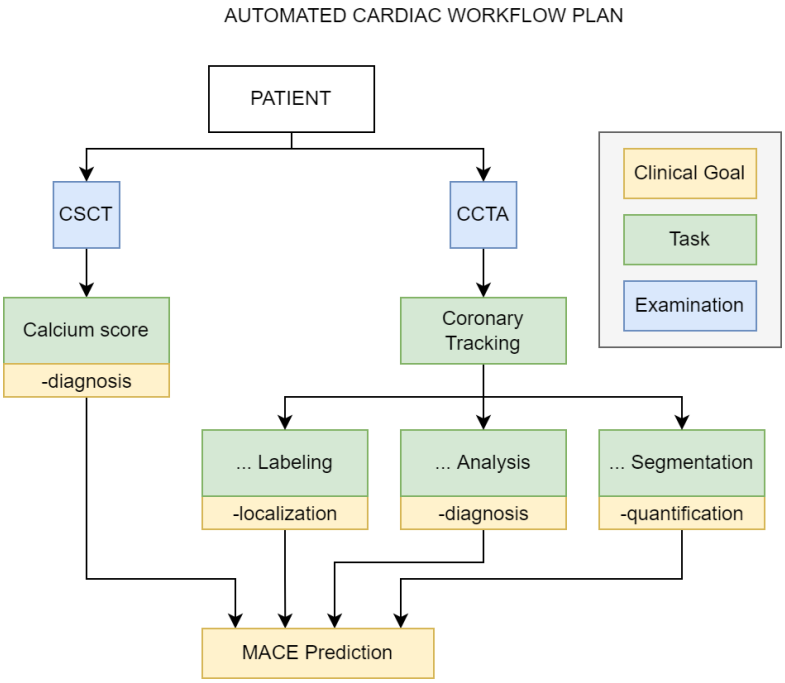


Figure 2.1: Automated Cardiac Workflow Plan: this diagram represents the ideal planning of the tasks involved in Major Adverse Cardiovascular Events (MACE) prediction. A patient with suspect Coronary Artery Disease (CAD) is subject to mainly two examinations: a non-injected low radiation dose cardiac examination and a contrast injected high-quality examination, Calcium Score Computed Tomography (CSCT) and Coronary Computed Tomography Angiography (CCTA), respectively. Tasks can be either enablers of subsequent downstream tasks or have a unique clinical goal. All ultimately converge in the Major Adverse Cardiovascular Events (MACE) prediction.

In this chapter, we will present state-of-the-art deep-learning based methods in car-

diac imaging-based automated Coronary Artery Disease (CAD) prediction 2.1. Lately, algorithms have shifted from knowledge-based to machine-learning-based approaches which are particularly suited for medical imaging applications as these can efficiently exploit expert annotations to design and validate algorithms. The supervised approach allows both clinicians and algorithm designers to have immediate feedback by directly comparing the expert clinical assessment against algorithm prediction. With knowledge-based methods, feedback is not direct; once the algorithm has been designed and executed, the result is subject to review by experts, thus inducing a delay in providing essential feedback to engineers working on the design of the algorithm. A specific clinical task corresponds to a data type (volume, image, sequence, graph) that can be processed using various algorithms. Algorithms are often tailored to process one type of data; thus, a tailored algorithm is required for each clinical task. Neural network architectures are designed to address issues corresponding to a particular data type; these architectures can be seen as building blocks to process more complex data types (e.g., sequences of volumes or graphs with images as nodes). In this chapter, we will also present the neural network architectures that will appear along with the text; most state-of-the-art methods are either direct applications or present adapted architectures. We will explore the following relevant clinical tasks involved in the prediction of Major Adverse Cardiovascular Events (MACE).

Calcium Score Calcium score is a highly validated indicator of MACE in both symptomatic and asymptomatic patients. It is used as an exploratory exam for pretext coronarophaty and overall has the same outcome of ECG stress physical test with a higher throughput for medical institutions (Section 1.1.5). We propose an approach based on ensembling U-net networks to produce a semantic segmentation of the Coronary Artery Calcium (CAC) to quantify the Agatston score (AS) [16].

Coronary Tracking Coronary Computed Tomography Angiography (CCTA) allows for the detection of coronary structures and the extraction of coronary centerline for clinical inspection through advanced visualizations (Section 2.4).

2.1 Neural Networks Architectures

2.1.1 U-net

The U-net [17] architecture has been designed for semantic segmentation and has an encoder decoder structure, symmetrically contracting and expanding operators allow for the hierarchical capture of visual features while skip connections allow for the reconstruction of higher resolution details (Figure 2.2). U-net has rapidly become the go-to architecture to solve any type of semantic segmentation tasks 2D or 3D due to its simplicity of configuration. The main hyper-parameter is the number of blocks or levels in order to tune the field of view of the object of interest at various scales. Each symmetric level is composed by convolutional layers intervaled by batch norm and ReLU activations. For medical images the scale is fixed (no perspective projection) so the number of levels depends on the size of the object of interest and the fixed resolution of the voxel (another hyperparameter).

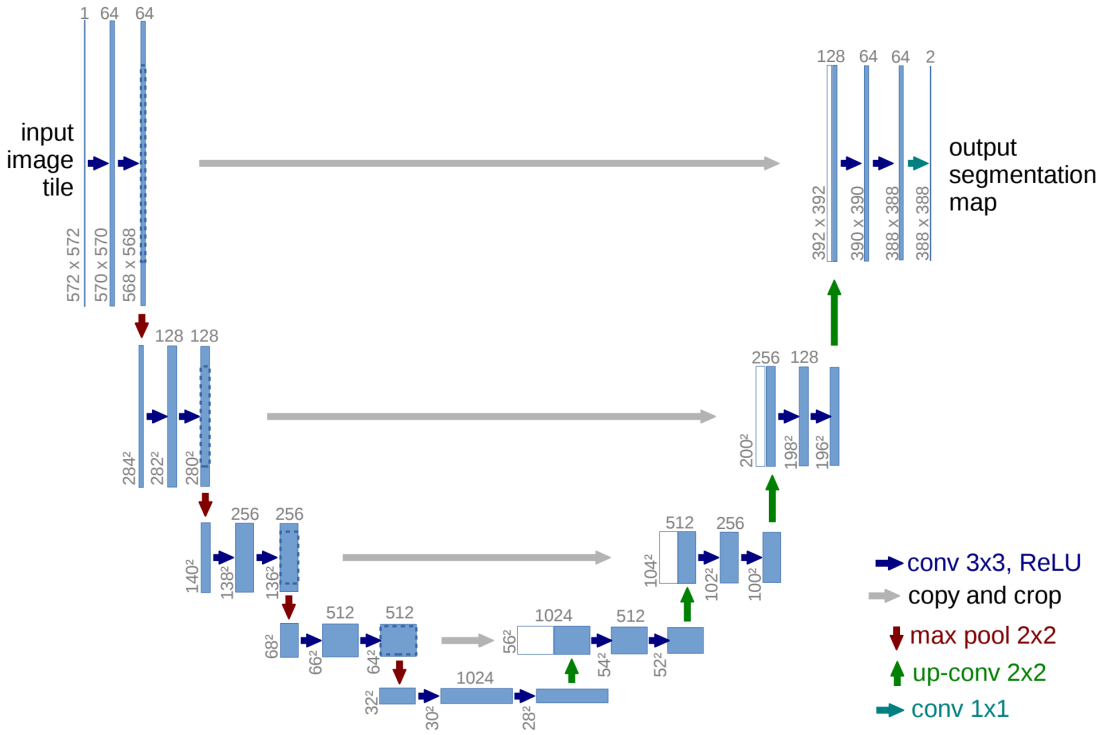


Figure 2.2: U-net architecture example for a 2D segmentation task with a 32×32 pixels in the lowest resolution. Each blue box corresponds to the multi-channel convolution feature map output. The number of channels expands at each level and it is noted on top of the box. White boxes are the skip connection maps.

2.1.2 Recursive Neural Networks

long-short term memory (LSTM)

The recursive neural network (RNN) is an architecture suited for processing sequential data or temporal sequences $X = [x_0, \ldots, x_t, \ldots, x_T]$.

$$
\begin{aligned}
h_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\
h_0 &= 0
\end{aligned} \tag{2.1}
$$

RNN is modeled as temporal sequence h_t, the parameters W_f are fixed in time, thus the recursion. $[.]$ stands for the concatenation operator (Equation 2.1). Its recursive nature implies that gradient based learning method must compute back-propagation with respect to t thus this procedure is referred to as back-propagation through time (BPTT), for long sequences (big T) gradient often balances between exploding (depending on the $\|W_f\|$ or vanishing (depending on $\sigma = 1/(1 + e^{-x})$. The Long-Short-Term-Memory architecture addresses the issue of exploding or vanishing gradient in BPTT [18]. By modifying a recurrent unit with a novel memory cell and a gate unit. In the forward computation memory cells allow to store or overwrite activation corresponding to a given time step depending on the current activation; in the backward computation this implies that gradients can skip time steps thus allowing long term dependencies both forward and backward.

$$
\begin{aligned}
f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) && (2.2)\\
i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) && (2.3)\\
_t &= \tanh\left(W_c[h_{t-1}, x_t] + b_c\right) && (2.4)\\
C_t &= f_t * C_{t-1} + i_t *_t && (2.5)\\
o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) && (2.6)\\
h_t &= o_t * \tanh C_t && (2.7)
\end{aligned}
$$

The forget state 2.2, new information 2.3, and the new memory candidates 2.4 are computed from the previous state and the current input $[h_{t-1}, x_t]$. The new memory C_t is computed as a convex combination of the previous memory state C_t and the new candidate memory $_t$ 2.5. The output is passed through an extra layer before being updated 2.7. More recently [19] proposed a simplified LSTM named gated recurrent unit (GRU). This model has fewer parameters than the original LSTM without loss of performances.

$$
\begin{aligned}
z_t &= \sigma(U_z h_{t-1} + W_z x_t + b_z) && (2.8)\\
r_t &= \sigma(U_r h_{t-1}, W_r x_t + b_r) && (2.9)\\
_t &= \tanh U_h(h_{t-1} * r_t) + (W_h x_t + b_h) && (2.10)\\
h_t &= z_t *_t + (1 - z_t) * h_{t-1} && (2.11)
\end{aligned}
$$

The most noticeable difference is the absence of the memory unit C_t. The memory is overwritten in the hidden state h_t (2.11) with candidate values $_t$ (2.10) by means of the reset gate values z_t (2.9).

2.1.3 Graph (Convolutional) Neural Network

A graph is a couple $G = (X, E)$ with n d dimensional nodes $X \in R^{n \times d}$ and the set of edges $E \subset [1 - n] \times [1 - n]$ which can be represented in an adjacency matrix $A^{n \times n} \mid A_{i,j} = 1$ *if* $(i, j) \in E$. A graph convolutional layer is defined by equation 2.45, where σ is a non-linear activation, A is the adjacency matrix of the graph, D is the degree matrix of the graph, a diagonal matrix holding the number of neighbours for each node. $H_0 = X$ which are the input features, and W^l is a trainable set of parameters.

$$H^{l+1} = \sigma(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^l W^l) \tag{2.12}$$

Each layer is an "hop" in the graph meaning that it aggregates for each node the information from its direct neighbours. Thus each layer explore the graph in a BFS (Breadth First Search) order, meaning that H^k holds information from all the network with k being the degenerancy of the graph which is the smallest k for each each subgraph $G_k \subset G$ has degree at most k. This operation is also known as message passing.

2.1.4 Transformer

Historically sequence data have been processed using GRU and LSTM units. The drawbacks of these methods are the difficulty to properly capture long term relationships and the lack of parallelization capability as each element in the sequence depends on the hidden state processed by its predecessor (and successor if bi-directional). Self attention models have been proposed to model inter data relationship therefore aiming to be more general than convolutional and recursive architectures [20]. Lately vision transformers architectures relying completely on self attention mechanisms [21, 22] have been shown to match or even surpass convolutional neural networks, although the debate is still ongoing in the research community [23].

Multi Head Self Attention (MHSA)

In [20] the authors propose a new transformer architecture equipped with a new attention mechanism. The MHSA module is described as a mapping between query and a set of key-value pairs: query, keys, and values are all output vectors. The attention mechanism is in the scaled dot-product operation which results in the dot-product with the values V with the softmax activated normalized queries Q and keys K outer product (Equation 2.13).

$$\text{ScaledDotProduct}(Q, K, V) = \text{Softmax}(\frac{QK^T}{\sqrt{d_k}}) \tag{2.13}$$

An alternative to the dot-product attention mechanism is the additive attention mechanism in [24] which an attention context vector is computed by a convex combination of the hidden states of a RNN sequence encoder weighted by attention coefficients. The dot-product attention is faster as it can be computed in parallel. While for d_k small both mechanisms perform similarly, when d_k grows the additive attention outperforms the dot-product attention. A workaround is to normalize the values of the outer product QK^T by a factor $1/\sqrt{d_k}$. To illustrate this fact assume that query and key values are sampled $Q, K \sim \mathcal{N}(0, 1)$ than the outer product $QK^T \sim \mathcal{N}(0, d_k)$, thus high values input to the Softmax can make the gradient disappear for certain locations. So far the described computation involves a single self attention unit, multiple heads are used in practice to conjointly attend to information using different sub-spaces in parallel and to reduce the quadratic computational cost of the outer product (Equation 2.14).

$$\begin{aligned} \text{MultiHeadAttention}(Q, K, V) &= [\textit{head}_1, \ldots, \textit{head}_h]W^O \\ \textit{head}_i &= \text{ScaledDotProduct}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{2.14}$$

Where W_i^Q, W_i^K, W_i^V are linear projections onto a sub-space $\mathbf{R}^{d_{model}} \mapsto \mathbf{R}^{d_k}$ and W^O $\mathbf{R}^{hd_k} \mapsto \mathbf{R}^{d_{\text{model}}}$ where $d_{\text{model}} = hd_k$. In practice usually $h = 8$ and $d_{model} = 64$. The MHSA unit is the core layer of the Transformer encoder, the whole encoder consists in MHSA blocks followed by residual connection and normalization layers $X^{t+1} = \textit{Norm}(\textit{MultiHeadAttention}(X^t, X^t, X^t) + X^t)$. The final output is passed through a feed forward network $FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2$.

Positional Encoding

The positional information is not present in the dot-product attention mechanism. This makes it able to process generic information like sets, however in order to exploit the positional information it must be injected in the model. To do so "positional encodings" are added to the input values of the MHSA. These can be learned or fixed. A popular choice for fixed positional embedding are sine and cosine transforms (Equation 2.16)(Figure 2.3).

$$\textit{PE}(\textit{pos}, 2i) = \sin\Big(pos/10000^{2^i/d_{model}}\Big) \tag{2.15}$$

$$\textit{PE}(\textit{pos}, 2i+1) = \cos\Big(pos/10000^{2^i/d_{model}}\Big) \tag{2.16}$$

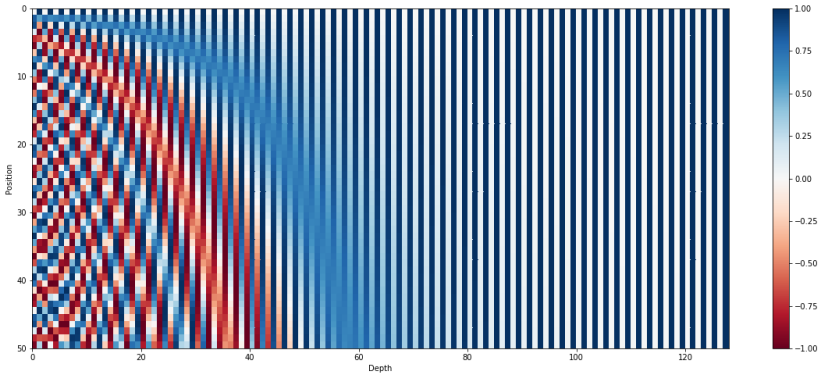


Figure 2.3: The 128-dimensional positional encoding for a sequence with the maximum lenght of 50. The wavelengths form a geometric progression from 2π to 20000π

The *PE* $\in \mathbf{R}^{d_{model} \times L}$ has the same dimension as the input features X_0 therefore these are added before input to the MHSA unit $X_0' = X_0 + PE$.

Attention paradigm

Transformers have the advantage to model long term relationships in one layer. For RNN and convolutional neural network (CNN) in order to attend to different locations multi-layer architecture are necessary. When the sequence length is less than the vector dimension the MHSA is also computationally more efficient than RNN for sequence processing; this is often the case for most NLP translation tasks. The receptive field of a convolutional neural network depends on the number of subsequent convolutional filters that are applied to the image. These are often expensive as the feature maps need to be retained in memory in order to compute back propagataion. In [21] authors apply transformer to images and completely renounce convolutional filter, thus showing that for classification tasks the reliance on convolution is not a necessity while requiring less computation resources to train large model. Another advantage of MHSA is the interpretabilty of the activations maps, these can be used to investigate how each element in a sequence influenced the final output as activation maps are probabilty distribution these compute $P(X^t \mid X^{t_0})$ with $t \geq t_0$ being any time in the sequence while recurrent model compute $P(X^t \mid X^{t-1})$ therefore needs $t - t_0$ steps to compute a long term relationship.

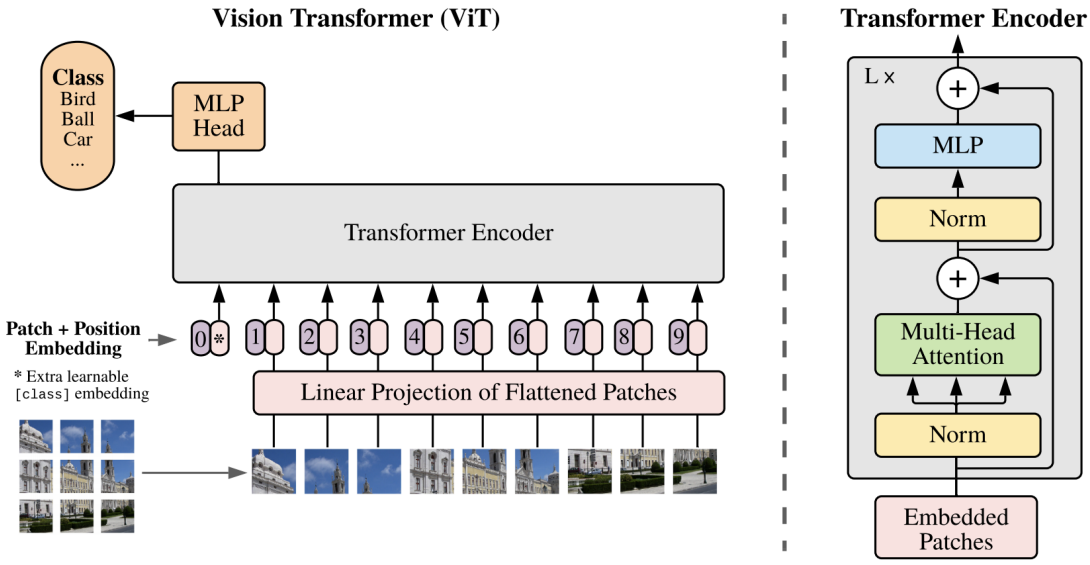


Figure 2.4: Visual Transformer architecture [21]. The image is split into non overlapping patches, each patch is processed individually and mapped to a feature vector to which a learnable positional encoding is added, the obtained sequence of vectors is fed to the Transformer enocoder.

2.2 Calcium Score

2.2.1 Related Works

In the last few years, semantic segmentation architectures were used to predict dense segmentation maps by extending CNN to Fully Convolutional Neural Networks (FCNN) [25]. They were first applied to 2D biomedical imagery with the so-called U-net [26] and later with a straightforward extension to 3D and 4D V-Net [27]. From a methodological point of view, different artificial intelligence methods have been proposed in the automated detection and analysis of CAC score on Computed Tomography (CT). By using public datasets with paired non-enhanced-CT and CCTA obtained for chest examinations with CT annotation, a method based on a combination of 3 CNN where each detected the heart in a different orthogonal plane obtained 72% of accuracy in predicting patients CAC score class by using only the enhanced CT and a per-branch calcification score and a risk prediction evaluation framework [28]. Among the top performing results of CNN, one approach makes use of fuzzy features and atlas-based information in conjunction with Random Forest Models to exploit the available data at best and get accurate quantification and branch-wise location of CAC [29], and only the non-enhanced-CT images are required. To produce a more accurate segmentation of coronary artery [30] another method has been suggested to automatically detect calcified lesions on the non-enhanced CT images, but the segmentation of the aorta, the heart, and coronary arteries was required and obtained by using associated contrast CT images. Because of its relevance in CAD screening in a population of smokers, another deep learning method has been tested on the NLST (National Lung Screening Trial) [31]. This method uses of a 2-stage FCNN prediction [32] to quantify branch-wise calcifications using the sole non-enhanced CT examination. This method, previously trained on manual and segmental labeling of calcified coronary lesion per coronary artery in a subset population of NLST patients, was further tested on multiple cardiac CT protocols as previously mentioned [33]. Finally, a last interesting method still based on CNN [34] has been reported and was trained on the NLST data set, but it uses only of the CAC score information as supervision (weakly supervised) and a spatial transformer as a first preprocessing step. This method is less dependent on human supervision because it only takes into account the CAC score results of the initial examination reports, which allowed the NLST data to be fully exploited for training and evaluation. The method employs two CNNs, one for registration to align the input images to an atlas image made from cardiac CTs and one for direct CAC score prediction using regression. This second CNN operates on 2D slices and does not use the information across slices which, according to the authors, may account for some incorrect identification of CAC near the coronary artery Ostia.

2.3 Coronary Tracking

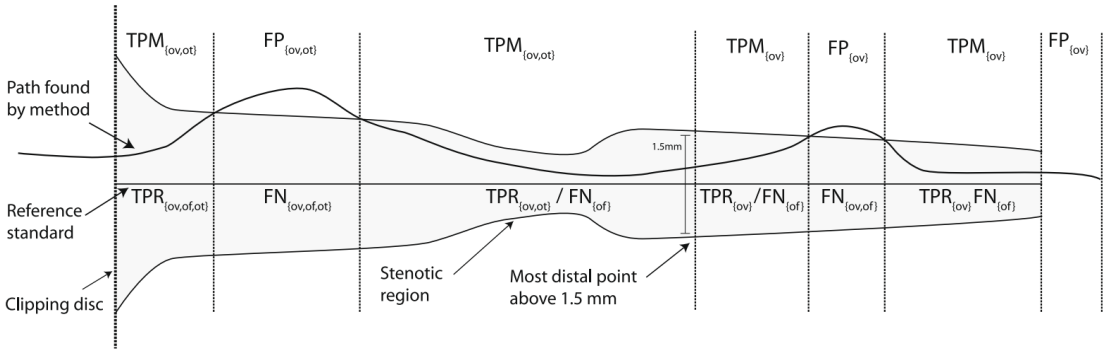


Figure 2.5: The straightened vessel, the ideal centered tracking result and a reference tracking are depicted to illustrate the evaluation terms used as performance metrics. The annotated ground truth is depicted as the reference standard straight line, the prediction as the path found by method line.

An essential step in analyzing the CCTA examinations is the estimation of coronary artery centerlines (tracking). It consists of identifying the 3d curve as the sequence of coordinate points passing through the barycenter of the coronary lumen cross-section, and it is critical to coronary inspection as it enables advanced curved planar reformation (CPR) visualization (sec. 2.4) and automated diagnosis. This task is often carried out manually by cardiologists or automatically by placing points on CCTA slices. In the earliest approaches, automatic and semi-automatic centerline extraction is based on shortest path finding [35] from manually detected seed points (extremities). These methods rely on heuristic-based cost functions that model different scenarios (stenosis, plaques, artifacts). Other approaches aim to obtain the centerline as a byproduct of coronary segmentation. Coronary segmentation has mainly exploited analytical and morphological vesselness filters [36, 37] or has been modeled as an optimization problem [38]. With the advent of machine learning and deep learning, heuristic-based cost functions are being replaced by supervised models that can exploit annotated data directly [39, 40, 41, 42]. The deep-learning state-of-the-art method extracts the centerline by iteratively tracking the coronary vessels using CNN for local Orientation Classification (CNN-OC) [43, 44].

2.3.1 Standardized evaluation of coronary artery centerline extraction algorithms

This section will describe our framework for evaluating coronary CCTA centerline extraction techniques. Consider the coronary artery centerline as the curve that passes through the center of gravity of the lumen cross-section. We define the start point of a centerline as the center of the coronary ostium (i.e., the point where the coronary artery

originates from the aorta) and the endpoint as the most distal point where the artery is still distinguishable from the background. The centerline is smoothly interpolated if the artery is partly indistinguishable from the background, e.g., in case of a total occlusion or imaging artifacts.

Prediction and ground truth correspondence

The evaluation metrics are measured on a point-to-point correspondence between the annotated ground truth and the predicted centerline. To determine this correspondence both centerlines are resampled to 0.2mm and normalized with respect to the arclength to enable accurate comparison. In order to have the same starting position centerlines are clipped to a threshold distance to the ostia. The point-to-point correspondence is computed by minimizing the Euclidean distance between each valid point-to-point correspondence. A valid correspondence for centerline I, consisting of an ordered set of points p_i ($0 \leq i < n$, p_0 is the most proximal point of the centerline), and centerline II, consisting in an ordered set of points q_j ($0 \leq j < m$, p_0, q_0 is the most proximal point of the centerline), is defined as the ordered set of connections $C = \{c_0, \ldots, c_{n+m+1}\}$ where c_k is a tuple $[p_a, q_b]$ representing a connection from p_a to q_b which fulfills the conditions: The first connection c_0 connects the starting points $c_0 = [p_0, q_0]$. The last connection connects the endpoints $c_{n+m+1} = [p_n, q_m]$. If connected $c_k = [p_a, p_b]$ then connections $c_{k+1} = [p_{a+1}, q_b]$ or $c_{k+1} = [p_a, q_{b+1}]$ or $c_{k+1} = [p_{a+1}, q_{b+1}]$. This conditions are imposed so that each point of centerline I is connected to another point of centerline II. The minimum cost path with respect to the Euclidean norm from c_0 to c_{n+m+1} gives the optimal matching.

Evaluation metrics

The centerline extracted from the coronary artery is used for different clinical purposes. Each has its requirements to be fulfilled. Thus, a global metric is not enough to assess whether an approach is suited for a particular clinical use. On the other hand, having tailored metrics for each use makes comparison unfeasible. In order to overcome this issue, a set of metrics have been designed to cover most of the clinical requirements. In particular, these aims to discern between extraction capability and accuracy capability. Accuracy can only be measured when the extraction is successful. Therefore a tracking failure must not be included in the accuracy of the method (figure 2.5).

True positive, false positive and false negative points All metrics are based on the labeling of correspondences as a true positive, false negative, or false positive. A point on the reference prediction is marked as TPR_{OV} if the distance of its correspondent ground truth is less than the coronary estimated radius, it is false negative otherwise, FN_{OV}. A point on the ground truth manual is labeled true positive TPM_{OV} if the distance in its correspondent prediction is less than the coronary estimated radius, it is false positive otherwise FP_{OV}.

Overlap (OV) represents the ability to track the complete vessel. This metric has a global interest when evaluating different methods.

$$OV = \frac{\text{TPR}_{OV} + \text{TPM}_{OV}}{\text{TPR}_{OV} + \text{TPM}_{OV} + \text{FN}_{OV} + \text{FP}_{OV}} \tag{2.17}$$

Overlap until first error (OF) This metric determines how much of a coronary artery has been extracted before making an error. This metric is interesting for image-guided vessel navigation applications. The first error is defined as the first occurrence of FN_{OV} when traversing the ground truth starting from 5mm from the first point. Errors in the first 5mm are not taken into account to relax the constraint on the most proximal segment of the coronary, which is usually large and not usually of critical importance.

$$OF = \frac{\text{TPR}_{OF}}{\text{TPR}_{OF} + \text{FN}_{OF}} \tag{2.18}$$

Overlap with the clinically relevant part of the vessel (OT) This metric is suited to measure the tracking capability of the most relevant part of the vessel. In order to keep stenosis, the lumen radius is computed starting from the most distal point. The vessel is traversed in the proximal direction. When the luminal radius is above $1.5\,\text{mm}$ the remaining part is assumed to be the clinically relevant part. Only points of the prediction and ground truth within the clinical part of the vessel are kept to compute this metric.

$$OT = \frac{\text{TPR}_{OT} + \text{TPM}_{OT}}{\text{TPR}_{OT} + \text{TPM}_{OT} + \text{FN}_{OT} + \text{FP}_{OT}} \tag{2.19}$$

Average inside (AI) This metric measures how accurately the method closely tracked the centerline. It consists of the average distance in mm of the connections inside the lumen radius. Therefore provided that the prediction is inside the radius AI metric is proportionally inverse to the accuracy of the method.

2.3.2 Related works

CNN Orientation Classifier

In [44] the authors propose a CNN-based method to determine the orientation and radius of a coronary artery at a specific location x in the CCTA volume I. The method predicts the coronary's local orientation from an isotropic image patch P sampled at the location x centered around the centerline. The orientation is thus regressed as the most likely given the posterior probability $p(D|P)$ over a set of discrete directions D computed by the classifier (fig. 2.6). All models are trained on eight CCTA examinations from the publicly available MICCAI 2008 Coronary Artery Tracking Challenge (CAT08) dataset containing 32 manually annotated centerlines. Another 24 CCTA with 96 centerlines were used for the test, and the method achieved a 93,7% overlap with the clinically

relevant part of the coronary (luminal radius above $1.5\,\mathrm{mm}$). The UMC Utrecht institution provided an additional test set and an annotated subset of the MICCAI 2014 OrcaScore [45] challenge CCTA examinations. The method achieved an overlap of 92% with the clinically relevant part of the coronary.

CNN tracker The patch size and the model architecture are chosen with a field of view of $9.5\,\mathrm{mm}$ sufficient to cover the coronary lumen at different distalities. The CNN is relatively shallow and employs five 3D convolutional layers and two fully connected layers for classification. The CNN classifier is trained on 3d patches sampled around the coronary centerline for robustness noise is added to the central location, and directions are corrected to predict a direction that points towards the centerline. If on the centerline, the model will predict the orientation. If the model is off the centerline, it will predict a direction towards the nearest centerline. The characteristic of this method is that the training and inference pipeline differ. The orientation classifier model is one of three that compose the algorithmic inference pipeline. Another two models are trained to detect seed points to initialize the tracking. These models are identical to the orientation classifier but for the output layer, which outputs an estimated proximity value $d(x)$.

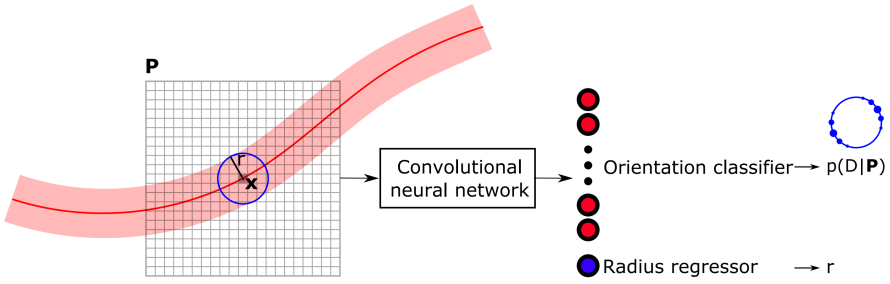


Figure 2.6: Overview of the coronary tracker based on the CNN orientation classifier method. The isotropic patch P is sampled from the CCTA volume I at location x centered about the coronary centerline. The CNN predicts the radius and the direction is computed by the maximum of the posterior probability $D = \max_{d \in D} p(d|P)$

$$d(x) = \begin{cases} e^{a(1-\frac{D_c(x)}{d_M})} - 1 & \text{if } D_c(x) < d_M \\ 0 & \text{otherwise} \end{cases} \tag{2.20}$$

Initialization and termination strategies Where $D_c(x)$ is the distance to the closest centerline and d_M is the max distance ($4\,\mathrm{mm}$) and a is a gain value (6). Two CNN are employed in the tracking framework: one network is trained to identify the Ostia, and the other network identifies coronaries. Because the method is not given any initialization

information but the CCTA volume, first seed points are computed by computing proximity values from randomly sampled points from the CCTA volume. Once both coronary and Ostia seed points are localized as the max of the proximity values $x^* = \max_x d(x)$, the CNN orientation classifier is initialized with patches sampled about the seed points from which to predict the local orientation D. The orientation classifier is iteratively applied to patches at location $x_i = x_{i-1} + \epsilon D$, $x_0 = x^*$. The tracking is guided by stopping criteria based on the measured classifier confidence. At each point the normalized entropy $H(p(D|P)) \in [0, 1]$ of the posterior probability is computed as:

$$H(p(D|P)) = \frac{\sum_{d \in D} -p(d|P) \log p(d|P)}{\log |D|} \tag{2.21}$$

The tracking stops if the entropy of the selected probability distribution crosses a threshold value of $\Theta_H = 0.9$. This may happen when near the ostium or the end of coronary arteries. However, also stenotic areas may result in low entropy values. The termination is determined as a three-step moving average to push the tracking over these areas. A post-processing routine prunes vessels that are too long or with a small estimated radius.

Deep Recursive Bayesian Tracking for Fully Automatic Centerline Extraction of Coronary Arteries in CT Images

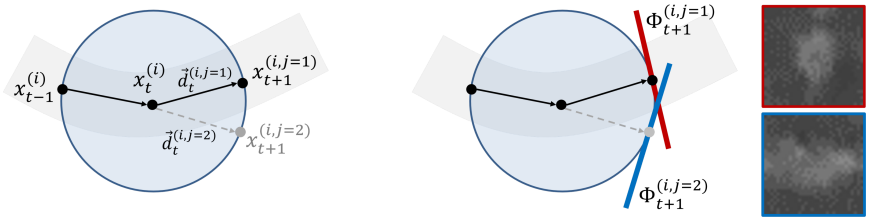


Figure 2.7: Overview of the deep particle filtering (Deep-PF) method; at one position the next direction is estimated by estimating the vesselness value from 2d patches randomly sampled along the centerline.

This study proposes a method combining a CNN and particle filtering method to identify the trajectories from the coronary ostium to each distal end from 3D CT images (Deep-PF). The CNN acts as a robust vesselness estimator from 2D tangent patches representing cross-sections of coronary arteries of circular shapes. The predicted values are integrated as likelihoods in the particle filtering framework for tracking. Thus, 2D tangent patches are assumed to include enough features of coronary arteries, and their processing is computationally efficient. A method based on clustering the particle locations is employed to solve the issue of bifurcations. This method was tested on the CAT08 dataset and showed promising results with respect to existing commercial solutions.

Particle Filtering The particle filtering method is a sample-based method for trajectory estimation. Given the position x_t, the next position is estimated utilizing a direction vector computed as the weighted sum of the directions of a set of particles. For a given location of $x_t^{(i)}$, its estimate is computed as follows:

$$x_t^{(i)} = x_{t-1}^{(i)} + \epsilon \sum_{j=1}^{N_T} w_{t-i}^{(i,j)} \vec{d}_{t-i}^{(i,j)} \tag{2.22}$$

The weights are predicted by a Bayesian model of likelihood and prior. The prior models the probability of transitioning from one sample x_t to x_{t+1}, the likelihood is estimated by an observation Y done at location x_{t+1}.

$$w_{t+1}^{(i,j)} \propto w_t^{(i)} p(Y \mid x_{t+1}^{(i,j)}) p(x_{t+1}^{(i,j)} \mid x_t^{(i)}) \tag{2.23}$$

The likelihood $p(Y \mid x_{t+1}^{(i,j)})$ is computed by a CNN vesselness classifier. 2D Tangent patches I are randomly sampled from points on the half sphere defined by the current direction d_t.

$$\begin{aligned} & p(x_{t+1}^{(i,j)} \mid x_t^{(i)}) \\ & = p(\vec{d}_t^{(i,j)}, \Phi_t^{(i,j)} \mid \vec{d}_{t-1}^{(i,j)}, \Phi_{t-1}^{(i,j)}) \\ & = p(\vec{d}_t^{(i,j)} \mid \vec{d}_{t-1}^{(i,j)}) p(\Phi_t^{(i,j)} \mid \Phi_{t-1}^{(i,j)}) \end{aligned} \tag{2.24}$$

From the factorization it follows that the relation between directions and tangent patches over time is independent. The first factor can be computed as the angle between the two directions $p(\vec{d}_t^{(i,j)} \mid \vec{d}_{t-1}^{(i,j)}) = max(< \vec{d}_t^{(i,j)}, \vec{d}_{t-1}^{(i,j)} >, 0)$ which implicitly imposes a smoothness constraint on the estimated trajectory. The second factor employs the Jensen-Shannon divergence to measure the distance between two distributions as

$$\begin{aligned} D(P\|Q) &= \frac{1}{2} KL(P\|M) + \frac{1}{2} KL(Q\|M) \\ KL(P\|Q) &= \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \\ M &= \frac{1}{2}(P + Q) \end{aligned} \tag{2.25}$$

P,Q are estimated by $\Pi_t^{(i,j)}$,$\Pi_{t-1}^{(i)}$ which are the normalized histograms of the tangent patches $\Phi_t^{(i,j)}$,$\Phi_{t-1}^{(i)}$. A-priori density functions use a mapping function $m(x) = 2x^3 - 3x^2 + 1$, as the natural bounds are $0 \geq D(P\|Q) \leq \ln_2$ the mapping is bounded onto $[0, 1]$, thus suits a density function: $p(I_t^{(i,j)} \mid I_{t-1}^{(i,j)} = m(\lambda D(\Pi_t^{(i,j)} \| \Pi_{t-1}^{(i)}))$.

Clustering To handle bifurcations the computed weights w_{t+1} from the sample set S_t are used as location densities to detect $k \in \{1, 2\}$ clusters $\Omega_t^{1,2} \subset S_t$, with centroids $\mu_t^k = \sum_{i=1}^{\|\Omega^k\|} w_t^k c_t^k$ which are estimated using the density-based spatial clustering of applications with noise (DBSCAN) algorithm. Thus, the direction for the next position can be estimated as $d_t^k = \frac{\mu_t^k - x_t}{\|\mu_t^k - x_t\|}$.

2.3.3 Centerline extraction using minimal cost path

For each vessel a starting s and an ending e point is computed in an agnostic fashion by a tracking algorithm A. An algorithm B will also compute a probability mask $M(x) = P(x \in \textit{Coronary Vessel})$ with $x \in \Omega \subset \mathbf{R}^3$ and $P(x \in \textit{Coronary Vessel})$ the probability of a voxel position x to be inside a vessel. The maximum probability path on the spatial domain (regular grid) Ω according to the probability map M can be modeled as a search on the complete graph G.

$$
\begin{aligned}
&G = (V, E), V = \{x \in \Omega\}, E = \{(x, y) \in V \times V\} \\
&P^* = \max_P \sum_{x \in P} M(x) - \sum_{x_{\rho(i)} \in P} dist(x_{\rho(i)}, x_{\rho(i+1)}) \\
&P = \{x \mid (*, x) \in E\}
\end{aligned} \tag{2.26}
$$

where the ordered list $P = [x_0, \ldots, x_k, \ldots, x_{N-1}]$ and subscript $\rho(i)$ indicates the node x is the ith node in P. An exhaustive search on a complete graph is too expensive, the graph can be simplified by considering only neighbourhood connections.

$$
G_\epsilon =< V, E >, V = \{x \in \Omega\}, E = \{(x, y) \in V \times V \mid dist(x, y) \leq \epsilon\} \tag{2.27}
$$

By adjusting ϵ to 1 or $\sqrt{3}$ we obtain a 6 or a 26 connected neighborhood, respectively. A tree is a Directed Acyclic Graph (DAG); therefore, the minimal cost path from the starting to the ending point should not contain cycles. Alltough these premises, there is no need to model G directed graph because costs increase monotonically (all ways positives). A maximal probability path on a DAG can be computed by any Breadth First Search (BFS) routine. Therefore a faster optimal solution can be found utilizing a A^* on the graph G_ϵ. In [35] the authors proposes a geodesic distance transform on the probability map M using a fast marching algorithm; this approach corresponds to BFS on a regular voxel grid and locally computes the maximal probability path from the start to the end point of the vessel, the computation of a local front of propagation is also resource-efficient.

2.3.4 Skeletonization

The centerline of coronary structures is closely related to the topological skeleton. Although not for centerline extraction, [46] proposed recently a novel topological loss

(CL-Dice), which enforce connectivity in vessels by exploiting the differentiable morphological *Soft-Skeleton*. However, this formulation does not guarantee the connectivity of tubular structures. The skeleton has many applications in studying tubular or quasi-tubular objects. It conserves the original topology, provides 1-*d* dimensionality reduction. In the continuous domain, the skeleton has an ideal set of properties [47]: it should be centered in the object, with the same homotopy-type as the object (in particular connectivity), and be thin (e.g. its area or volume is negligible). The grassfire process [48] is the first historical model that produces a skeleton. In the discrete domain, these ideal properties cannot be guaranteed. In mathematical morphology, the skeleton can be defined through Lantuéjoul's formula (Equation 2.28), which is thin in the sense that each point of the skeleton is a neighbor of the background (non-simple points) and centered with respect to the Euclidean distance transform of the object. The resulting skeleton remains, however, disconnected.

$$S(X) = \bigcup_{i \in \mathbf{N}} S_i(X) = \varepsilon_{\kappa_i}(X) \setminus \gamma_{\kappa_0}[\varepsilon_{\kappa_i}(X)], \tag{2.28}$$

where γ_{κ_0} is the unit ball opening, ε_{κ_i} is the erosion with κ_i an element of a granulometric family of elementary convex structuring elements, i.e. such that $\forall (i \leq j),\ \gamma_{\kappa_i}(\kappa_j) = \kappa_j$

2.3.5 clDice

Loss functions for semantic segmentation are often based on overlap measures: Precision, Recall, Dice, and Jaccard index. While these are suited for generic object segmentation, tubular structure segmentation often requires to conserve connectedness which is not measurable using overlap-based metrics. From an application point of view, connectedness is necessary to preserve the network structure of the segmentation, which is one of the main functional goals of the identification of tubular structures. In [46] the authors proposed a novel topology-preserving loss based on a suited metric for tubular structures segmentation: the centerline Dice (clDice). The topology-preserving metric means that also connectivity is preserved, and these are based on the notion of the skeleton. For tubular structures, the skeleton corresponds to the centerline, which is the union of geometrical centers of the lumen sections. Consider the sets V_P, V_L and their respective skeleta S_P, S_V, the clDice metric can be defined using overlapping metrics on the skeleta.

$$\begin{aligned}
\text{prec}(S_P, V_L) &= \frac{|S_P \cap V_L|}{|S_P|} \\
\text{sens}(S_L, V_P) &= \frac{|S_L \cap V_P|}{|S_L|} \\
\text{clsDice}(V_P, V_L) &= 2 \times \frac{\text{prec}(S_P, V_L) \times \text{sens}(S_P, V_L)}{\text{prec}(S_P, V_L) + \text{sens}(S_P, V_L)}
\end{aligned} \tag{2.29}$$

This metric is topology-preserving as it acts upon the skeleta, which embeds the topology of the object. The skeleta are a homology class of the object (meaning preserve the original topology) but are thin, meaning that each point of the skeleton is simple, and the absence of any point changes the underlying topology (minimal). The loss is thus suited to match homology classes meaning that it is minimal when the 2 functions V_P, V_L have the same connectedness. In [46] authors proposes a theorem that implies that at its maximum clDice necessarily implies homology. The implication does not work for the minimum, however (a simple counter-example is a translation). Thus, this metric does not measure homology but implies homology at its maximum.

2.3.6 Soft-ClDice

Lantuéjoul's formula (Equation 2.28), gives an iterative morphological interpretation of the skeleton which involves erosion and opening operators. A proxy for the morphological operators (erosion and opening) can be implemented by max and min pooling operators which can be included in the computation of a gradient in back propagation.

Algorithm 1 Soft-Skeleton

Input: I, K
  $I' \leftarrow maxpool(minpool(I))$
  $S \leftarrow Relu(I - I')$
  for $i \leftarrow 0$ **to** K **do**
    $I \leftarrow minpool(I)$
    $I' \leftarrow maxpool(minpool(I))$
    $S \leftarrow S + (1 - S) \odot Relu(I - I')$
  end for
Output: S

For continuous values the clDice must be redefined to operate on prediction feature masks. Similarly to the Soft-Dice that is maximized to obtain a semantic segmentation from continous maps, the Soft-clDice here is proposed (Algorithm 2) to take as input a feature map V_P.

Algorithm 2 Soft-ClDice

Input: V_P, V_L
  $S_P \leftarrow$ Soft-Skeleton(V_P)
  $V_P \leftarrow$ Soft-Skeleton(V_L)
  Prec$(S_P, V_L) \leftarrow \frac{\sum_i (S_P \odot V_L)(i) + \epsilon}{\sum_i S_P(i) + \epsilon}$
  Sens$(S_L, V_P) \leftarrow \frac{\sum_i (S_L \odot V_P)(i) + \epsilon}{\sum_i S_L(i) + \epsilon}$
  clDice $\leftarrow 2 \times \frac{\text{Prec}(S_P, V_L) \times \text{Sens}(S_P, V_L)}{\text{Prec}(S_P, V_L) + \text{Sens}(S_P, V_L)}$
Output: clDice

In general for segmentation when a complete ground truth V_L is available, the objective is topological correctness and accurate segmentation thus the loss is composed by two weighted terms.

$$\mathbf{L}_\alpha = \alpha(1 - \text{Soft-Dice}) + (1-\alpha)(1 - \text{Soft-ClDice}), \alpha \in [0, 0.5] \tag{2.30}$$

2.3.7 Topology preserving loss

Topology metrics can be computed by using computational topology tools [49] like Betti numbers and persistence diagrams. These features, however, map discrete binary sets (segmentation mask) onto discrete values and are non-differentiable. In [50] the authors propose a novel topology-preserving loss function defined on the continuous-valued function, which is differentiable and can thus be integrated into an end-to-end deep learning framework.

Topology and persistent homology

Given a spatial domain $\Omega \subset \mathbf{R}^3$, the likelihood function $f(x) : \Omega \mapsto [0, 1]$, and the threshold operator $f^\alpha = \{x \in \Omega \mid f(x) > \alpha\}$. Given a set X its homology class is the set of d-dimensional manifolds that can be computed from X the number of d-manifold of X is called Betti number (d). The set of different segmentations $\subset f^{\alpha_1} \subseteq f^{\alpha_2} \cdots \subseteq f^{\alpha_n} \subseteq \Omega$, for $0 < \alpha_1 < \alpha_1 \ldots \alpha_n \leq 1$ constitutes a threshold set. The topology varies as α varies, the location at which topological changes occurs are also critical points of the function. These locations can be mapped to connection nodes in a topological feature representation space of the segmentation (graph of manifolds, NB: in 2d only connections location occurs as 0-d manifolds of X are 1-d manifold of the complementary space Ω/X). On these graphs one can compute the persistence diagram corresponding to the α at which a certain connection is made or is eliminated (birth and death) which corresponds to the underlying value of f at the critical point (thus time of birth = $f(p)$ with p being a critical point of connection and time of death $1 - f(p)$), from this diagram it is possible to compute a loss by measuring the distance between the time of birth and death to a ground truth. From these locations the error is back-propagated to the feature map thus computing a topological gradient.

$$\begin{aligned} &\min_{\gamma \in \Gamma} \sum_{p \in \text{Dgm}(f)} \|p - \gamma(p)^2\| = \\ &\sum_{p \in \text{Dgm}(f)} (birth(p) - birth(\gamma^*(p)))^2 + (death(p) - death(\gamma^*(p)))^2 \end{aligned} \tag{2.31}$$

Topological Loss gradient

The loss function depends on time of death and birth of topological structures, which in turn depends on the thresholds α values where these changes happen. These topological discontinuities happen at certain locations, which are critical points of the function

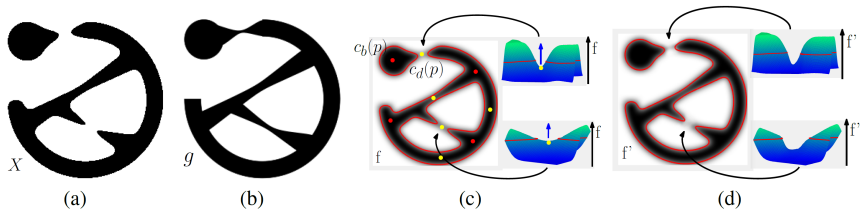


Figure 2.8: Topology of the continuous-valued function f. The darker the higher the value. **(a)** an example of a segmentation obtained by filtering f at a value for which the number of connected components is 2 and the number of holes is 1. **(a)** For another threshold the segmentation obtained has a different topology: 1 connected component and 2 holes. **(c)** The topological point at which the topology changes are also critical points of the underlying function f. **(c)** The highlighted critical point of the underlying function f' have a deeper gap and therefore a lower value it will thus influence the loss differently.

f. Because of the value of birth and death dependency on the function f, the loss can be rewritten as:

$$\sum_{p \in \text{Dgm}(f)} (f(c_b(p)) - birth(\gamma^*(p)))^2 + (f(c_d(p)) - death(\gamma^*(p)))^2 \tag{2.32}$$

where c_b and c_d are functions mapping each point to the critical point of birth or death of topological structures. In a training setting the function f is estimated by a neural network with parameters θ, thus the gradient $\nabla_\theta L_{topo}(f, g)$ is formulated as:

$$\sum_{p \in \text{Dgm}(f)} 2(f(c_b(p)) - birth(\gamma^*(p)))\frac{\partial f(c_b(p))}{\partial \theta} + 2(f(c_d(p)) - death(\gamma^*(p)))\frac{\partial f(c_d(p))}{\partial \theta} \tag{2.33}$$

Thus the error is back-propagated directly to critical points of the function f, which is piece-wise differentiable (neural network). The intuition of this loss is that the negative gradient $\nabla_\theta L_{topo}(f, g)$ which pushes points from the $Dgm(f)$ towards the points in $Dgm(g)$. These points are dependent on the values at critical points $c_b(p)$ and $c_d(p)$. Intuitively a saddle point is pushed "upwards" to connect the bridge. Empirically the topological loss is complementary to the cross-entropy by combating the sampling bias. By focusing on critical points, the relative error induced by mistakes at these locations is magnified while the global cross-entropy counter-balance the over-fitting of solving for critical locations.

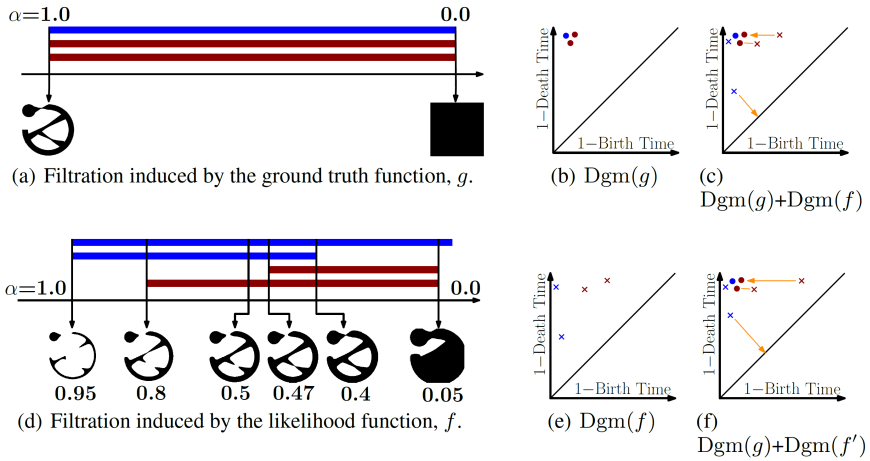


Figure 2.9: The point distances in the $Dgm(f) + Dgm(g)$ and $Dgm(f') + Dgm(g)$ show that the topological loss $L_{topo}(f, g) < L_{topo}(f', g)$.

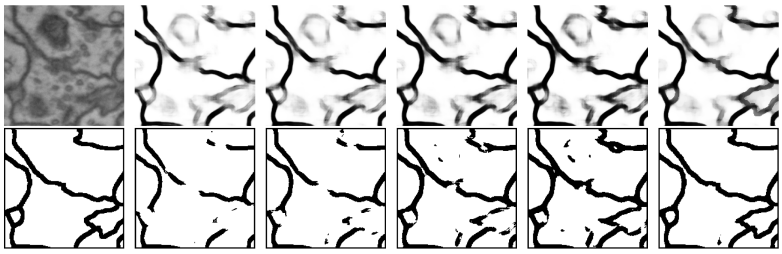


Figure 2.10: For a sample patch from the CREMI dataset, the likelihood maps is shown at different training epochs. The first row corresponds to likelihood maps and the second to the thresholded likelihood maps.

2.4 Curved planar reformation

Non-invasive imaging of the vascular system with computed tomography (CT) has become a well established alternative to invasive intra-arterial angiography. These data, however, may contain many objects of less or no diagnostic interest. This makes volume-rendering (i.e., maximum intensity projection (MIP), ray casting, shaded surface display) without pre-processing often impossible or inaccurate. In addition to that pathological features may superimpose diagnostically relevant information. In the case of a circular vessel wall calcification the true vessel lumen can not be determined by conventional volume rendering. CPR is a way to visualize vascular structures with small diameters. High level information as the vessel's centerline is used to re-sample and visualize the data. By this technique the entire tubular structure is displayed within a single image. Vascular abnormalities, i.e., stenosis, occlusions, aneurysms and vessel wall calcification, are then investigated by physicians.

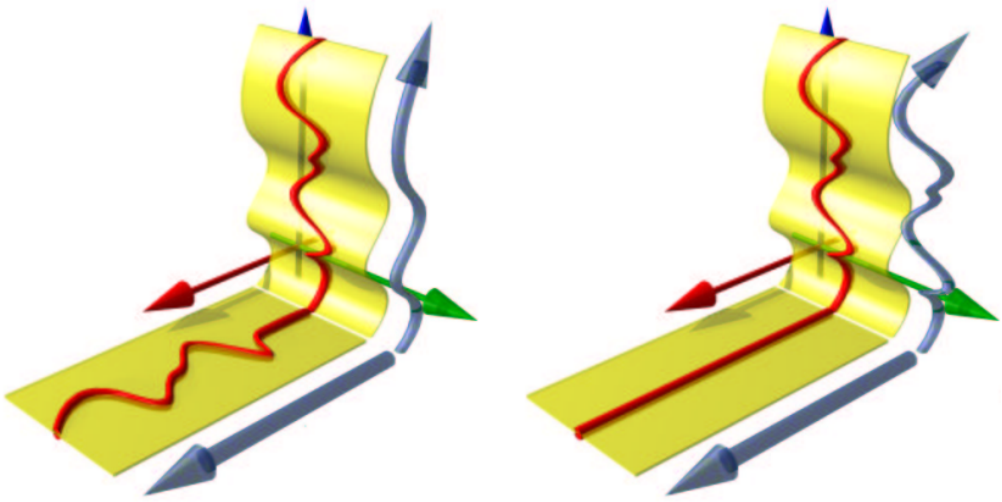


Figure 2.11: Different CPR generation methods: Stretched CPR, and Straightened CPR

2.4.1 Stretched CPR

A normalized VOI (Vector Of Interest) and the coordinates points of the vessel centerline compose a surface which as a planar projection parallel to the VOI and a curved projection. Stretching the curved projection shows the tubular structure without overlapping (Figure 2.12). This type of CPR is referred to as *stretched* CPR. The line composed by one point of the centerline and the VOI is mapped to a 2d surface. The resulting planar mapping is isometric. The point vector $\mathbf{d}_i = P_i\vec{P}_{i+1}$ represents the path direction at position i. The vector $\mathbf{l}$ is the normalized VOI representing the current rotation. The offset in space is therefore computed as Equation 2.34.

$$\Delta_i = \sqrt{\|\mathbf{d}_i\|^2 - < \mathbf{l}, \mathbf{d}_i >^2} \tag{2.34}$$

The image position of the next line is given by $y_{i+1} = y_i + \Delta_i$. The centerline is assumed to be re-sampled to a normalized arclength. This process ensures that tubular structures are visible along the entire distality. By changing the angle of l with respect to the origin one can obtain different visualizations of the surrounding structures or different portions of the vessel wall. The main advantage is the preservation of isometry.

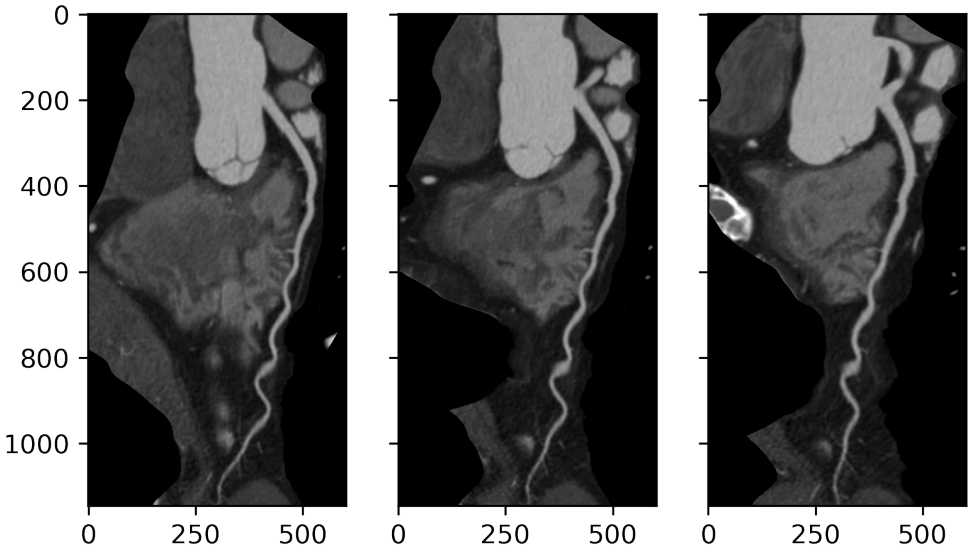


Figure 2.12: A tortuous vessel used as example for stretched CPR. This example shows how a small variation in the angle has little visual changes locally but propagates major changes in the rest of the image.

2.4.2 Straightened CPR

Straightened CPR is a way to visualize vascular structures. The detected centerline is used to visualize the CCTA data, characterize plaque and assess stenosis. It has been also used consistently for automatic coronary analysis as an input for automated diagnosis [51, 15, 52, 53]. CPR and MPR are often used interchangeably depending on the context (another less used name is medial axis reformation MAR). The coronary centerline can be seen as discrete parameterized 3d curve, and arc-length parameterized:

$$\begin{aligned} c(s) = x(s), y(s), z(s), i \in [0, L] \\ ||c'(s)|| = 1 \\ \sum_{i=1}^{n} |c'(s)| = L \end{aligned} \tag{2.35}$$

For each point $c(s)$ a local frame $[t(s), n(s), b(s)] \in \mathbf{R}^{3\times3}$, tangent, normal, and binormal respectively, is used as orthonormal coordinate system to sample voxel intensities along the plane orthogonal to the curve c. While $t(s) = c'(s)$ is unique there are infinite $n(s)$ lying in the plane $t(s) \cdot n(s) = 0$ with $n(s) \times t(s) = b(s)$: depending on the need different formulations have been proposed.

- **VR**: a volume relative (VR) view can be obtained by keeping the closest orthogonal plane to any of the volume planes (axial, coronal, saggital) represented as $I = [e_0, e_1, e_2] \in \mathbf{R}^{3\times3}$: the tangent is obtained simply by: $t(s) \times e_i = n(s)$.
- **Freenet**: The arclength parameterized curve has a tangent second derivative: $c'(s) \cdot c''(s) = \frac{1}{2}\frac{d}{ds}(c'(s) \cdot c'(s)) = \frac{1}{2}\frac{d}{ds}||c'(t)||^2 = \frac{d}{ds}1 = 0$, the curvature $\kappa(s) = ||c''(s)||$ the normal is obtained as the normalized second derivative $n(s) = \frac{c''(s)}{\kappa(s)}$. A preferable strategy is to use $n(s) = \frac{c'(s)\times c''(s)}{||c'(s)\times c''(s)||}$. A frame is undefined if $K(s) = 0$ so only curves that are regular with non-vanishing curvature are considered.

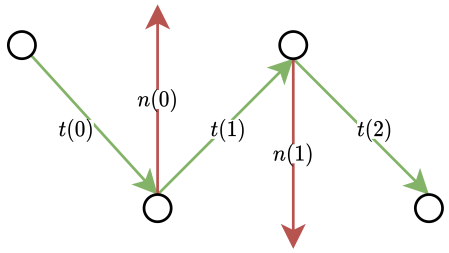


Figure 2.13: Freenet normal flipping depends on the local curvature orientation.

- **RMS**: In [54] the author propose a rotation minimization sweeps (RMS). This formulation assumes $\kappa(s) \neq 0$ and a arclenght parameterized curve $||c'(s)|| = 1$. Ideally $|n(s) \cdot n(s + \Delta s)|$ and $|b(s) \cdot b(s + \Delta s)|$ are maximized, from these assumptions it is possible to arrive to the following DOE.

$$\begin{aligned} n'(s) &= -(c''(s) \cdot n(s))c'(s)/||c'(s)|| \\ b'(s) &= -(c''(s) \cdot b(s))c'(s)/||c'(s)|| \end{aligned} \tag{2.36}$$

  which leads to the following first order iterative scheme $n(s + \Delta s) = n(s) - (c''(s) \cdot n(s))c'(s)/||c'(s)||$ the binormal can be computed at each step $b(s + \Delta s) = n(s + \Delta s) \times t(s + \Delta s)$.

- **RMS-DR**: double reflection (DR) method can be used to solve a vector valued DOE $f'(s) = F(t, f)$ like RMS [55]: it has been empirically proven to have 2nd order approximation error. It can also be implemented in a simple fashion as an iterative method but its guarantees have non trivial geometrical justifications.

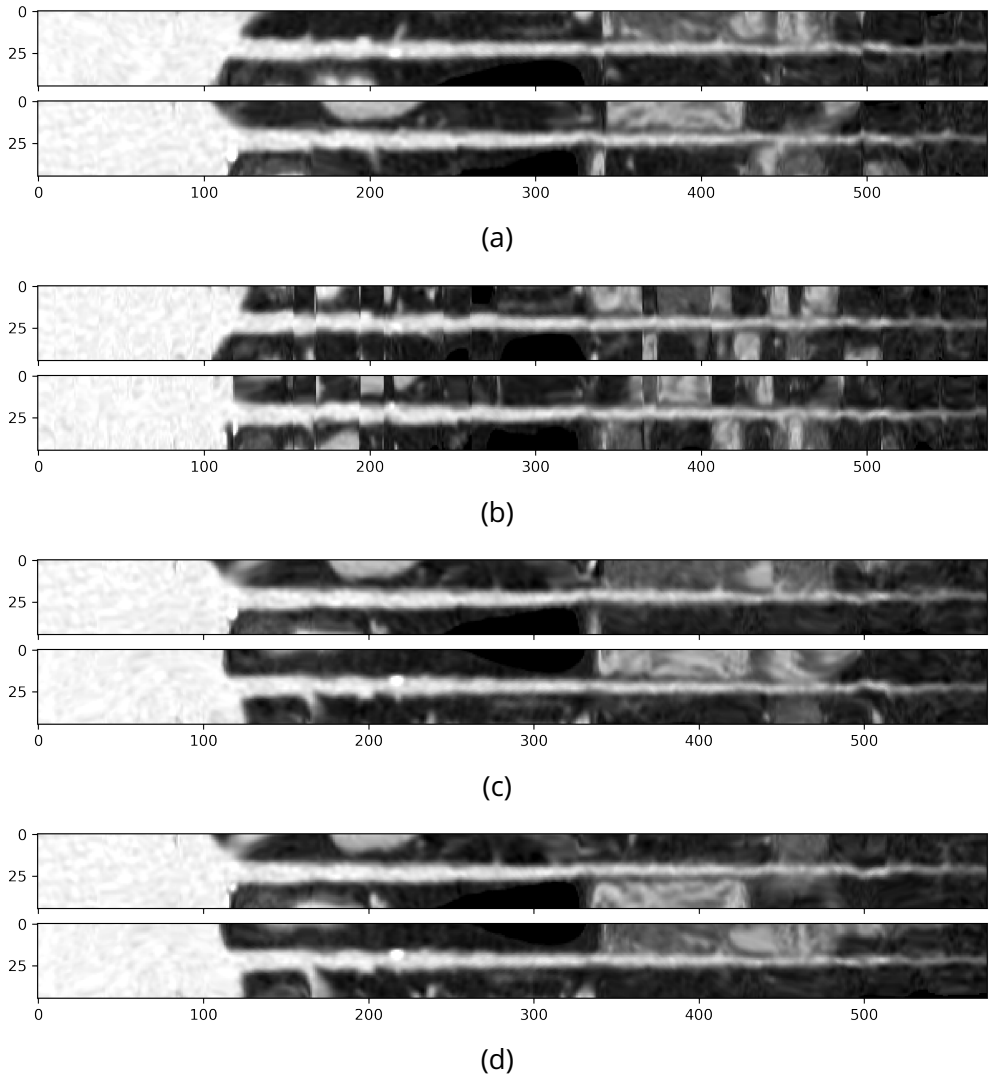


Figure 2.14: A tortuous vessel used as example for multiple frame views results. (a) **VR**: presents a negligible discontinuity. (b) **Freenet**: non negligible discontinuities due to normal flips. (c) **RMS**: no discontinuities. (d) **RMS-DR**: no discontinuities.

2.5 Coronary Labeling

CCTA has been widely used for CAD diagnosis. In a computed-aided diagnosis application, automated coronary labeling facilitates the diagnostic process for cardiologists and radiologists. The most challenging issue in automated coronary labeling is addressing large individual variability inherited in human cardiovascular anatomy [56]. Long, tortuous vessels branching patterns cause a large inter-patient variation. The American Heart Association (AHA) reference model [57] (fig. 2.15) contains a set of main coronary arteries: Left Artery Descending (LAD), Left Circumflex (LCX), Posterior Descending Artery (PDA), Posterior Lateral Branch (PLB), Obtuse Marginal (OM), Ramus Intermedius Branch (RIB), and Diagonals of the left branches among others. Coronary arteries span over different territories of the myocardium and develop in different combinations of coronaries vessels. Coronary trees are therefore often intricate and incomplete due to anatomical variations. The coronary tree branching can be modeled as a directed acyclic graph (tree): a common aortic node (root), the intermediate nodes (bifurcations), and the coronary leaves (vessel branches). Automated coronary labeling consists of assigning the correct anatomical identifier to each coronary segment given a representation of the coronaries computed from a CCTA examination. The path from the vessel branch to the aortic root can determine the complete vessel. This session will present an approach based on deep learning and graph convolutional network (GCN) to solve automated coronary labeling. We will compare the proposed oriented graph convolutional network (ORI-GCN) method against knowledge based (KB) baseline method.

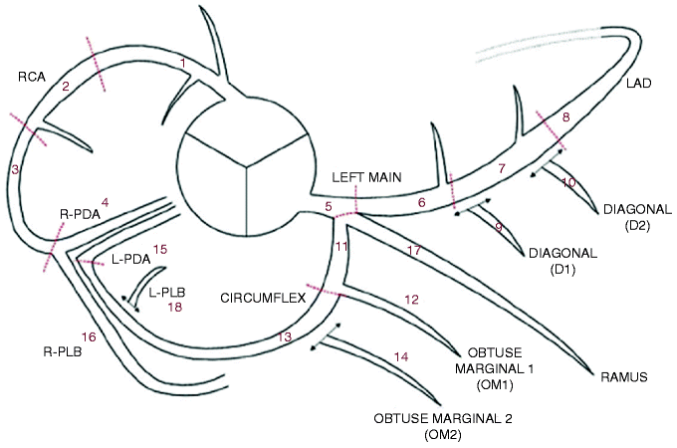


Figure 2.15: The complete coronary tree reference model contains a set of main coronary arteries: LAD, LCX, PDA, PLB, OM, RIB, and Diagonals of the left branches.

2.5.1 Related Works

The coronary tree is a graph structure that is computed by a complete tracking of all coronary branches. Many computer vision based techniques have been developed for

automated coronary labeling [58, 59, 60, 61, 62, 63, 64]. Atlas-based approaches rely on template matching of tree-like structures [58, 62], while another method proposes a machine learning-based approach relying on feature engineering [59, 60, 61]. However, this method may have unsatisfactory performances when the branching structure is particularly complex, motivated by the recent application of deep learning on the graph and as more data become available, current methods employed neural network architectures for automatic coronary labeling [63, 65, 64].

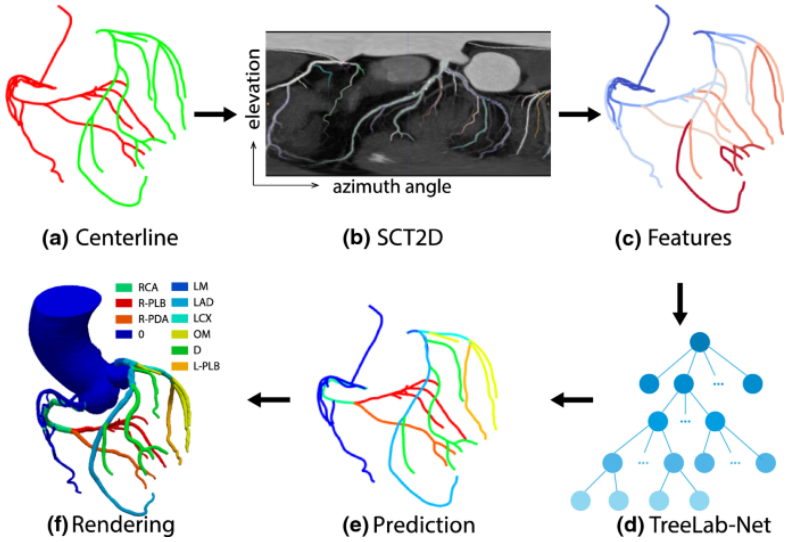


Figure 2.16: Bi-TreeLSTM algorithmic pipeline. The positional features are extracted as position of the coronary segment along the centerline in both Cartesian and polar coordinates, these features are then fed to the TreeLab-Net for prediction.

2.5.2 TreeLab-Net

In [63] the centerlines are first extracted as a preprocessing step. A spherical coordinates transform is performed based on the centerlines CCTA: 3D intensity features are extracted from the volume as well as the 2D spherical coordinates. These features are used as input of a TreeLab-Net to predict the corresponding coronary label (fig. 2.16). The TreeLab-Net is a dynamic architecture that consists of nodes for each coronary segment. Each node in the network includes three main modules: multi layer perceptron (MLP), Bi-TreeLSTM and a softmax layer. The MLP layer encodes each feature vector to obtain a compact representation of each node. The Tree-LSTM encodes these compact representations to obtain the representation of the current node (state) with respect to the predecessor nodes (hidden state). By design LSTM architectures are employed to process sequential data, the Tree-LSTM is a dynamic architecture that has been used to process syntax trees for natural language procesing (NLP) [66]. A Tree-LSTM at node t contains the input features i_t (eq. 2.38), output features o_t (eq. 2.40), memory features

m_t (eq. 2.41), memory cells c_t (eq. 2.42), aggregated hidden state H_t (eq. 2.37), and a set of forget features f_{tl} (eq. 2.39) for each child l of node t and an output hidden state h_t (eq. 2.43).

$$H_t = \sum_{l \in C_t} h_l \tag{2.37}$$

$$i_t = \sigma(W^{(i)} x_t + U^{(i)} H_t + b^{(i)}) \tag{2.38}$$

$$f_{tl} = \sigma(W^{(f)} x_t + U^{(f)} H_t + b^{(f)}) \tag{2.39}$$

$$o_{tl} = \sigma(W^{(o)} x_t + U^{(o)} H_t + b^{(o)}) \tag{2.40}$$

$$m_{tl} = \tanh\Big(W^{(m)} x_t + U^{(m)} H_t + b^{(m)}\Big) \tag{2.41}$$

$$c_t = i_t * m_t + \sum_{l \in C_t} f_{tl} * (1 - m_t) \tag{2.42}$$

$$h_t = o_t * \tanh(c_t) \tag{2.43}$$

where $W^{(i)}, b^{(i)}, W^{(f)}, b^{(f)}, W^{(o)}, b^{(o)}, W^{(m),b^{(m)}}$ are the trainable weights and biases of each layer. The bidirectional extension is obtained by making two computational pathways. For each node the input nodes will be either the children or the predecessors thus merging two computational pathways $h_t = h_t^{\uparrow} + h_t^{\downarrow}$. The model was trained to minimize the classification loss (Equation 2.44).

$$L_\Theta = \frac{1}{N} \sum_{k=1}^{N} y_k \log p_k + \frac{\lambda}{2} \|\Theta\|^2 \tag{2.44}$$

where p_i is the o_t with t being a leaf node and y_i being the ground truth label. The method was validated on 436 annotated subjects using 10 fold cross validation.

2.5.3 CPR-GNN

In [65] the authors propose a conditional partial residual graph convolutional network (CPR-GCN) architecture equipped with a novel discrete partial-residual block. It allows to compose positional and image features information. Local image features are extracted by sampling patches centered around the centerline. These are processed by a 3D-CNN and an LSTM model to output a feature vector y that encodes the segment information matching the vector size of the positional features x. The x and y vectors are input to the model by means of the proposed conditional partial-residual block (fig. 2.17). By analytical derivation, the authors show that spatial information is used as conditions of a partial differential equation influencing the positional features computed via the CPR-GCN layers (eq. 2.46). A graph convolutional residual layer is defined by equation 2.45, where σ is a non-linear activation, A is the adjacency matrix of the graph, D is the degree

matrix of the graph, $H_0 = X$ which are the input features, and W^l is a trainable set of parameters.

$$H^{l+1} = \sigma(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^lW^l) + H^l \tag{2.45}$$

In the CPR-GCN setup, there are two sets of features $x(l)$ and y, which are the positional domain features (spherical and Cartesian coordinates) of a segment and the local image domain CCTA features respectively extracted through a 3D-CNN and a LSTM.

$$\begin{aligned} \nabla H^l(x(l), y) &= \frac{\partial H^l(x,y)}{\partial x}dx + \frac{\partial H^l(x,y)}{\partial y}dy \\ &= \frac{\partial H^l(x,y)}{\partial x}\frac{dx}{dl}dl \\ &= GCN_A(x,y)dl \end{aligned} \tag{2.46}$$

The method was trained to minimize the cross-entropy loss between the predicted value of each node and the ground truth label. The authors experimentally show by using a dataset of 511 annotated subjects with 5 fold cross validation that this conditioning significantly enhances the performances and ultimately surpasses the Bi-TreeLSTM [63].

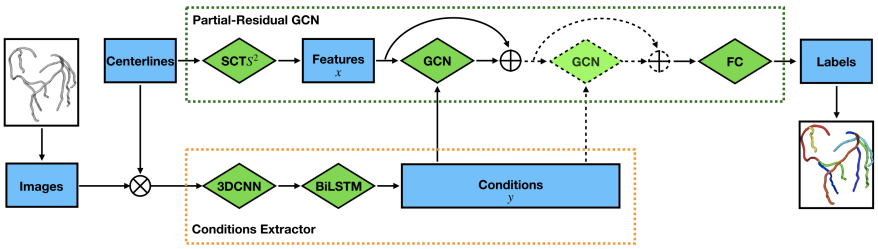


Figure 2.17: CPR-CNN architecture shows how conditioning image features are extracted from the CCTA volume. $\bigotimes$ stands for the sampling operator: patches are sampled from the volume using the centerline positions. $\bigoplus$ stands for the residual connection. Computationally conditioning happens via concatenation of the image features computed by the LSTM and the positional features.

2.6 Coronary Analysis

CT Coronary Analysis provides the means for evaluating the coronary arteries. Diagnosis is guided by the localization of anatomical indicators of CAD: stenosis and plaques. While occlusion remains a highly predictive indicator of MACE, there is growing evidence that the presence and characteristics of coronary atherosclerosis provide additional prognostic information. Stenosis is a morphological characteristic consisting of a localized narrowing of the coronary. As a consequence, blood flow is limited, with potentially harmful consequences on cardiac functions. Plaques are the result of metabolic activities building up lipids and calcium deposits within the coronary vessel wall tissues. Depending on their composition, plaques can cause ischemia or stenosis. CCTA is the only non-invasive imaging technique that reliably depicts the anatomic extent of CAD. In CCTA injected contrast solution highlights morphological characteristics of vessels, furthermore, CT is sensitive to plaque composition (true positives). Calcified plaques display high-intensity Hounsfield Units (HU) representative features, while more complex representations characterize high-risk soft plaques. CT reconstruction noise is detrimental to the specificity of plaque representation (false positives), thus the need for clinical expert visual inspection. However, accurate identification and quantification are burdensome and time-consuming because of X-ray scanners' limited temporal, spatial, and contrast resolutions. Automated CT Coronary Analysis aims to detect anatomical indicators along the coronary centerline and, based on their composition: position, and degree of stenosis, rule out CAD. Deep learning is an obvious candidate for automated CT Coronary Analysis because of the robustness to noise and ability to exploit large amounts of annotated data. In this section, we will present a series of deep learning-based methods for automated CT Coronary Analysis. Neural network models are trained to predict different indicators related to the assessment of CAD. These indicators can be computed during more invasive procedures (fractional flow reserve (FFR)) or provided by expert manual annotation.

2.6.1 Related works

recursive convolutional neural network (RCNN) [15] method CPR coronary volumes to predict coronary plaque characterization and stenosis. The centerline volume is processed as a sequence first by a shallow 3D CNN and a recursive gated recurrent unit (GRU) sequence classifier. In [67] the authors propose to replace the local feature extractor by *Radiomics* features [51] computed from the vessel wall segmentation. These features are then processed together by the same GRU sequence classifier. Subsequently, in [53] the authors proposed to use as input two perpendicular views of the straightened CPR volume to a 2D CNN classifier based on VGG-16 features extractor [68]. The method achieved similar results to the 3D RCNN in predicting revascularization and obstruction for a given input lesion, but with a cost-effective pipeline. However, it relies on both the extracted centerline and the localization of the lesion of interest (starting and ending point). The TR [52] method has been recently proposed as the state of the art for obstructive stenosis ($> 50\%$ occlusion) detection against both [51] and [15].

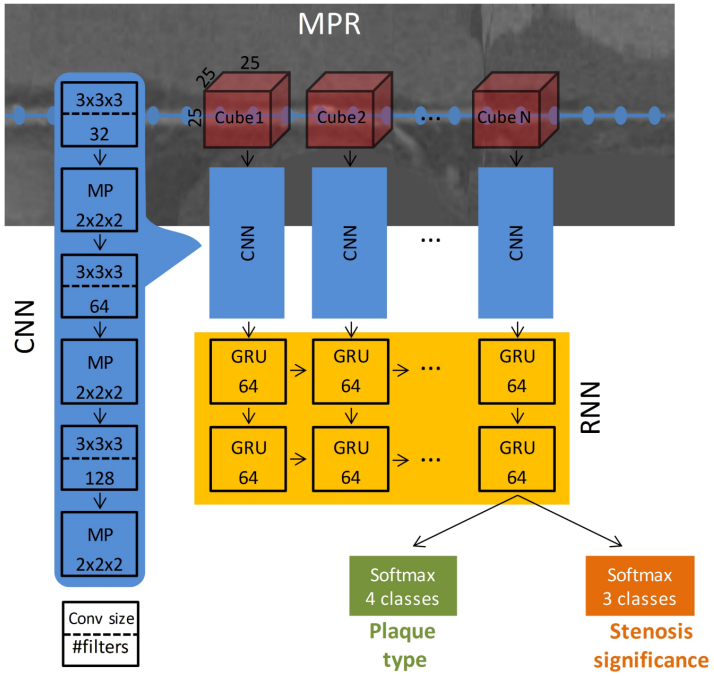


Figure 2.18: The recurrent CNN recursive convolutional neural network (RCNN) architecture [15]. A sequence of overlapping patches are sampled from the CPR volume computed from the CCTA volume by means of the tracked coronary centerline. The patches are processed individually by a shallow CNN (3 convolutional network interleaved with maxpooling) for local feature extraction, these features are arranged in a sequence and fed to a unidirectional gated recurrent unit (GRU) with two recurrent layers, the final output is used to predict the plaque characterization and the stenosis degree.

2.6.2 Coronary plaque and stenosis detection with RCNN

In [15] the authors propose to use a RCNN to process straightened CPR volumes extracted from CCTA volumes to detect stenosis and characterize plaque into 4 classes: no-plaque, calcified, mixed and soft, while stenosis is detected if significant ($\geq 50\%$ luminal narrowing). The straightened CPR volume is processed by extracting overlapping patches sampled with fixed interleave space of $1.5\,\mathrm{mm}$ from an isotropic $0.3\,\mathrm{mm}$. These patches are processed individually by a 3D-CNN for local feature extraction. A small input patch $p \in \mathbf{R}^{25\times25\times25}$ allows for the use of a relatively shallow CNN, which is less prone to over-fitting. The local feature vectors $f \in \mathbf{R}^{\mathbf{128}}$ are then arranged in a temporal sequence $s \in \mathbf{R}^{T,128}$. The temporal sequence is fed to the gated recursive unit GRU with two recurrent layers $GRU_{128} \circ GRU_{128}$. The GRU analyzes the relevant sequential dependencies and maps the sequence to an output vector which is tested against the ground truth for classification (fig. 2.18). The manual annotation labels the entire segment. In order to achieve finer localization, the inference pipeline is modified to process

small overlapping subsequences. Each location is thus assigned the corresponding label predicted by the model sliding over fixed length subsequences. The network is trained and tested on 98 and 65 CCTA examinations, respectively. Each CCTA volume is paired with coronary centerlines and manually defined segments (start and end point). For detection and characterization of coronary plaque, the method achieved an accuracy of 0.77; for stenosis detection, an accuracy of 0.80.

2.6.3 Stenosis detection with transformer (TR)

In [52] the authors propose to use a transformer network (TR) to process straightened CPR volume extracted from CCTA volumes to detect significant stenosis ($\geq 50\%$ luminal narrowing). The CPR volumes are processed by sampling patches at a fixed interleaved interval. These are fed to a shallow 3D CNN with a receptive field inspired by [15] for local feature extraction. The CNN encoder produces a feature vector $x \in \mathbf{R}^C$ for each location along the coronary segment. These feature vectors are then arranged in a sequence to which the positional encoding is added and subsequently fed to a transformer encoder [20]. The transformer encoder is composed by a series of MHSA that outputs a prediction vector $Z_T \in \mathbf{R}^{L\times(1)}$ where L is the segment length and T is the number of MHSA layers (eq. 2.47).

$$
\begin{aligned}
Z_0 &= [x_1 + o_1, \ldots, x_L + o_L]\mathbf{R}^{L\times(C)} \\
Z'_t &= \mathrm{MHSA}(\mathrm{LN}(Z_{t-1})) \in \mathbf{R}^{L\times(C)} \\
Z_t &= \mathrm{FCN}(\mathrm{LN}(Z'_t) + Z_{t-1}) + Z'_t + Z_{t-1} \in \mathbf{R}^{L\times(C)}
\end{aligned}
\tag{2.47}
$$

where L is the length of the sequence Z_0 is obtained by adding the output of the back bone to the trainable order embedding $Z0 = [f_0+o_0, \ldots, f_{L-1}+o_{L-1}]$. The advantage is that the MHSA is bidirectional (all directional) while the GRU used in [15] is not. This is a choice as GRU can be bidirectional. However, the authors did not notice improvements in using the bidirectional GRU. This method was tested on 76 CCTA examination for 609 coronary branches CPR volumes. Each coronary branch was subdivided into segments of equal lengths ($150\,\mathrm{mm}$). The model outperforms the RCNN [15] on the proprietary data in terms of accuracy (0.87 against 0.94) and specificity (0.89 against 0.96) and MCC (0.60 against 0.74).

2.6.4 Coronary Analysis with 2.5D CNN (2.5D)

Both aforementioned methods exploit the CPR straightened 3D representation of the coronary, which is computationally expensive to obtain and process by the subsequent machine learning pipeline. To mitigate this, authors in [53] propose a 2.5D multi-view approach as shown in figure (2.20). This method was trained and tested on 95 patients' CCTA examinations paired with the coronary centerline and 345 manually defined lesions of interest. The evaluation has been carried out by using 5-fold cross-validation. The 2.5 methods trained from scratch obtained an AUC of 0.9 and 0.92 for stenosis and

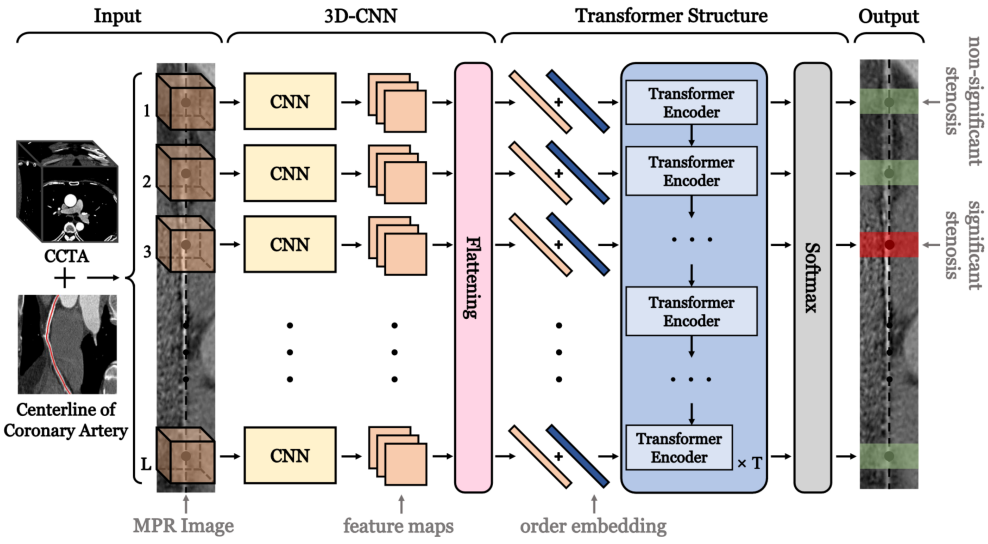


Figure 2.19: The transformer network (TR) architecture [52]. A sequence of overlapping patches is sampled from the CPR volume computed from the CCTA volume through the tracked coronary centerline. The patches are processed individually by a 3D CNN for local feature extraction. These features are arranged in a sequence and fed to a multi-head self-attention MHSA unit with the positional encoding of the coronary segment, and the final output is used to predict the stenosis degree.

revascularization classification, respectively, outperforming both 3D CNN and VGG-16 and ResNet feature extractors.

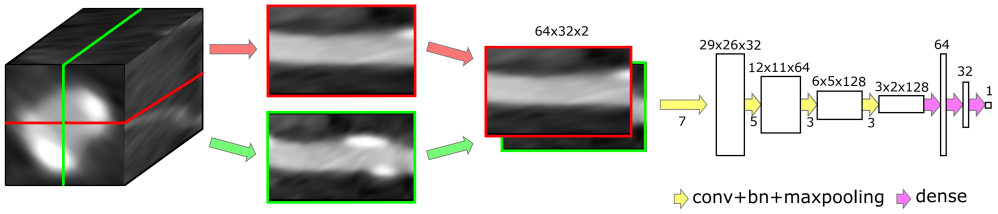


Figure 2.20: Two orthogonal views are extracted from the CPR volume of the lesion of interest. These are concatenated in a 2.5D view and fed as input of a 2D-CNN. This method relies on both the centerline extraction and the localization of the lesion of interest

Chapter 3

Contributions

Coronary Artery Disease (CAD) and Major Adverse Cardiovascular Events (MACE) automatic prediction is a broad-spectrum topic that relies mostly on the analysis of anatomical indicators of CAD. Coronary Computed Tomography Angiography (CCTA) is the only non-invasive imaging solution which is sensible to the depiction of CAD related anatomical indicators (sec. 1.1.1), because of the high through-put this modality is suited for the screening of CAD and it is essential to develop automated and computer-aided solutions by retrospective analysis. The goal of our studies is to explore the feasibility of image-based automated diagnosis of coronary artery diseases by the bias of deep-learning techniques which mostly relies on annotated CCTA data. This is done by evaluating the capability of automated method to predict relevant clinical indicators from Computed Tomography (CT) images that are predictive of CAD. For the algorithmic pipeline design, in this work we chose to focus on a data-driven approach mostly relying on deep learning techniques. There are several tasks involved in the diagnosis of CAD which are often referred to as cardiac workflow. In this chapter we propose our contributions to Calcium score prediction from Calcium Score Computed Tomography (CSCT) and coronary tracking, labeling and analysis from CCTA.

3.1 Calcium Score

CAD is one of the leading causes of mortality in the world. Coronary Artery Calcium (CAC) has been shown to be associated with the presence of CAD and to be a strong and independent predictor of cardio-vascular events and mortality [13]. CAC can be quantified or scored with low dose ECG-gated non enhanced CT acquisition namely CSCT. In clinical practice, the CAC burden is measured by means of the Agatston score (AS). The predictiveness of this metric has been widely validated in numerous studies involving cardiovascular risk [11]. Typically AS is computed using standardized parameters, and notably a tube voltage of $120\,\mathrm{kV}$ and a series of $2.5\,\mathrm{mm}$ to $3\,\mathrm{mm}$ thick slices covering the entire cardiac volume. The AS is computed from 2D axial connected components above 130 Hounsfield Units (HU) that are manually or semi automatically identified as coronary calcification by the radiologist. A manual intervention is often required to validate a segmentation made beforehand by a software. Beyond the time spent, this exposes to measurement variability depending on the software and the experience of the operator. According to several guidelines, this score is a reliable tool to classify patients into 5 classes for risk assessment and to guide follow-up preventive strategy and testing [69]. Several recent studies have reported different experiences in automating the procedure of segmentation and quantification using different deep learning approaches [25, 45, 32, 70]. We propose an approach based on Unet architecture ensembles for CAC burden assessment of CSCT volumes [16] and evaluate it in the context of the data challenge organized by the Société Française de Radiologie (SFR) [71].

3.1.1 Ground truth generation

For each CT volume of the training set, a ground truth was created in the form of a binary mask of calcifications obtained from the provided position of each calcification. This process was performed automatically by selecting the connected components of voxels above 130 HU connected to the calcification 3D positions provided in the annotation. To ensure the quality of the segmentation mask, the AS was then computed from the mask and compared to the original AS risk category. In case of discrepancy between the two risk categories, the masks were reviewed, and manually corrected if needed. The AS is then computed by 2d connected component analysis performed on each slice of the mask using the following formula:

$$\text{Agatston Score} = \sum_i \sum_j a_{ij} d_{ij} \frac{\Delta z}{3}$$

where a_{ij} is the area of the j^{th} 2d connected component (of at least three voxels) on the i^{th} axial slice, d_{ij} is a density factor determined by the maximum attenuation in this connected component $(130{-}199\ \text{HU} : 1, 200{-}299\ \text{HU} : 2, 300{-}399\ \text{HU} : 3, \geq 400\ \text{HU} : 4)$, and Δz is the axial thickness of the acquisition in millimeters. Non matching cases are reviewed by a radiologist.

3.1.2 Proposed method

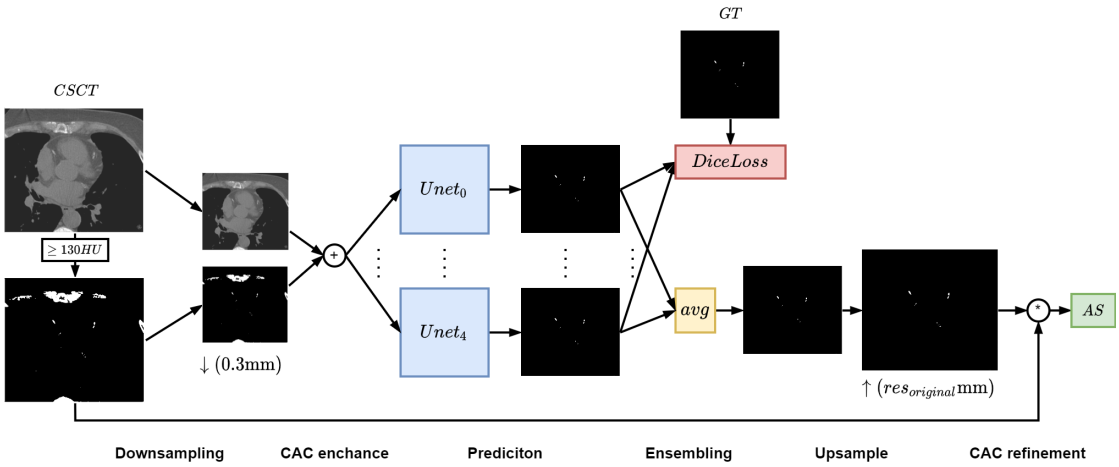


Figure 3.1: Pipeline for AS computation. CSCT volumes are downsampled to isotropic resolution of $2.5\,\text{mm}$ to $3\,\text{mm}$ depending on the original slice thickness, the calcium mask is downsampled using nearest neighbours in order to keep magnify small calcifications. The down-sampled volumes are concatenated and fed to each $Unet_i$ for prediction. The max of each prediction is then up-sampled to the original resolution, The original mask is used to refine CAC prediction. Finally the AS is computed.

We propose an approach based on deep learning to predict the CAC segmentation mask given a CSCT scan as input. The training pipeline is based on the training of individual $Unet_i$ in parallel on the same training set using the generated CAC dense masks (section 3.1.1). The inference pipeline consists into ensembling the different prediction to compute a CAC mask from which to compute the binned AS to asses the risk.

Model architecture

The ensemble consists in identical 3D Unet architectures with a depth of 4 levels and 16 initial root filters. The encoder consist into 4 convolutional blocks, each composed as $\textit{ConvBlock}_c = \textit{Conv3d}_{k,c} \circ \text{Elu} \circ \textit{Conv3d}_{k,c} \circ \textit{BatchNorm}_c$, with $k = 3$. The decoder is symmetrically organized. Each block is connected with skip connections, the pre-processing step of isotropic resolution downsampling allows to use isotropic convolutional kernels.

Inference

Each CSCT volume is first downsampled to an isotropic resolution of either 2.5 mm or 3 mm (depending on the original slice spacing). As an effect of the down-sampling, the information of low-density calcifications could be lost in the low-resolution image. In order to restore this information, intensities above 130 HU in the original image were

enhanced in the down-sampled image. Finally, the processed volume was clamped below -300 HU and above 800 HU, and linearly normalized between 0 and 1. The volumes are concatenated by addition and fed individually to each $Unet_i$ to output prediction masks $pred_i$ $i \in [0-4]$, The final prediction mask is obtained by applying the max (vote union) over each voxel $pred = \sum_i pred_i/5$. The prediction is then up-sampled to match the original resolution of the CSCT volume and the original mask is used to refine the CAC prediction. The output volume is used to compute the AS using the formula (section 3.1.1).

Training

The training of each Unet model is done independently to allow parallelisation, without pre-training, and using a different random initialization. Each model prediction is tested against the ground truth using the Dice loss function. A validation set of 20 exams from J1 was used to select the best epoch during training. The training process was done with TensorFlow version 1.14 in Python 3.6, on a NVIDIA® GV100-32GB GPU. We considered 3 different data subsets for training: J1, J1+J2, J1+J2+AD, and trained independently 5 models for each of these subsets.

3.1.3 Experiments

Testing

For each test volume, the model predicted a calcification mask on which the AS was further calculated. Four metrics were computed to compare the scores predicted for each patient of the test set (J3) to the ground truth scores: category accuracy, C-index, Cohen linearly weighted kappa and two-way intraclass correlation coefficient (ICC) for absolute agreement. The category accuracy gave the percentage of correct risk category in the evaluation set. The C-index or Harrel index measured the proportion of concordant pairs divided by the total number of possible evaluation pairs:

$$C\text{-index} = (\#\text{concordant pairs})/(\#\text{all pairs}) = (\#(i,j) \mid p_i < p_j \wedge t_i < t_j)/(\binom{n}{2})$$

where p_i is the predicted risk score category, t_i is the true risk score category and n is the number of elements to compare. The C-index lies between 0.5 (random prediction) and 1 (all scores risk category correctly predicted). The category accuracy, C-index and Cohen linearly weighted kappa were computed from the AS risk categories, while the ICC was evaluated from the AS. Bland-Altman plots were used to assess agreement between the ground truth AS and the AS predicted by the final ensembled model on each examination of the test set (J3). The final ensembled model trained on the full training set (J1 + J2 +AD) was also evaluated on the orCaScore [45] test set (OS) with the category accuracy, ICC and F1-score defined as: 2 Precision Recall / (Precision + Recall). Unlike the other metrics that were computed from the AS, the F1-score (equivalent to the Dice similarity coefficient) was directly computed from the segmentation masks and therefore evaluated the quality of the CAC segmentation.

Method	Metrics			
	C-index	Categorical Accuracy	Cohen kappa	ICC
1 Unet J1	0.932 ± 0.008	$83.2 \pm 1.0\%$	0.860 ± 0.011	0.963 ± 0.005
5 Unet J1	0.939	84.7%	0.875	0.967
1 Unet J1+J2	0.943 ± 0.005	$84.9 \pm 1.5\%$	0.881 ± 0.009	0.963 ± 0.002
5 Unet J1+J2	0.944	84.7%	0.882	0.964
1 Unet J1+J2+AD	0.949 ± 0.006	$86.4 \pm 1.0\%$	0.893 ± 0.008	0.959 ± 0.009
5 Unet J1+J2+AD	0.951	85.7%	0.894	0.970

Table 3.1: Performance of each model on the test set (J3), 1 Unet versus 5 ensemble Unet. For single models, means and standard deviations across the 5 trained models are reported for each metric.

The performance of the different models on the test set (J3) is presented in tab. 3.1.3 and fig. 3.3, showing the average performance obtained with a single model compared to the performance of the ensembled model for the different training sets considered. The C-index of the ensembled models improved with the size of the training set from 0.939, 0.944 to 0.951 when using J1, J1 + J2 and J1 + J2 +AD data sets respectively. The ensembling of five different models improved the performance in terms of C-index compared to the average performance of a single model (tab. 3.1.3). On the models trained with the full training data set, only a small degradation in accuracy was observed with the ensembled model, while the C-index, Cohen linearly weighted κ and ICC were improved. With those metrics, a large error on the predicted class has more influence than a small one, whereas this predicted class difference is weighted equally with the accuracy measure. Thus, the ensembled model made slightly more errors than the average model, but the magnitude of the errors was smaller. The confusion matrix shows the distribution of errors across the different AS risk categories (Fig. 3.3). With the final model, most of the errors were distributed near the diagonal, corresponding to off-by-one errors. A more detailed view is provided through the Bland-Altman plots in Fig. 3.2 which show the agreement between the ground truth AS and the AS score from the final model. The average difference between the two values is close to zero, showing no significant bias between the manual and the automatic score. However, as the score increases, there is a greater dispersion of differences between the two methods. Thus, in practice, a difference of more than 100 on the score was observed in seven subjects having already a high score > 100, and the differences may be even higher in the few subjects with a score > 10. The largest errors could be reviewed on the CT images and some recurring failure patterns were observed: small low-density calcifications were missed in some examinations (6 errors where A is predicted instead of B), and in a few CT examinations some coronary calcifications were mistaken for calcifications on the mitral valve. Examples of both successful and failed predictions are presented in Fig. 3.6. These error patterns could indicate the type of examinations that should be added to the training set in order to further increase the performance. Finally, the final ensembled model yielded a F1-score of 0.974, an accuracy of 97.5% and an ICC of 0.995 on the 40 CT volumes of the orCaScore test set (OS). This confirmed our model could

generalize well on a wide range of unenhanced-ECG-gated cardiac CT examinations. An average computing time of 1.8 seconds per CT examination was measured for the final ensembled model.

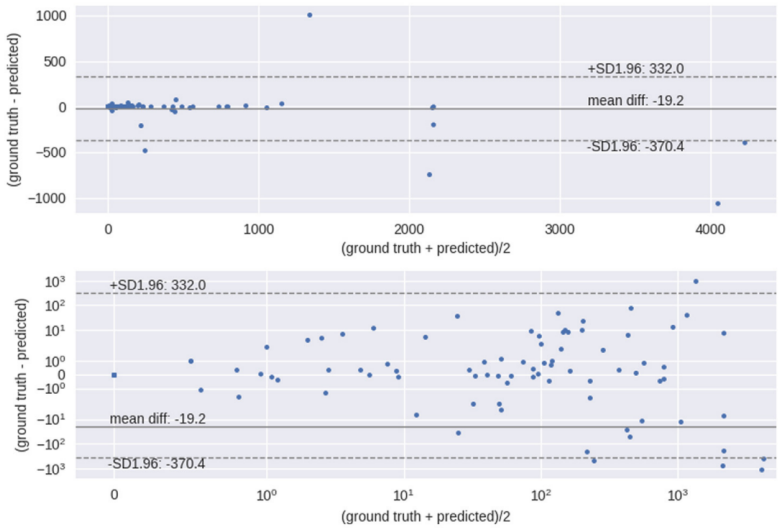
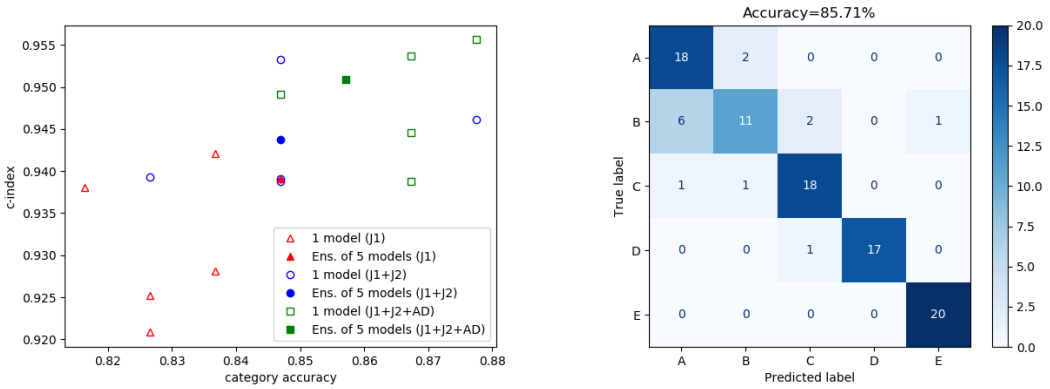


Figure 3.2: Blant-Altman plots show agreement between the ground truth Agatston score predicted by the final ensembled model on each examination of the test set (J3) with linear scale (top) and log scale (bottom). Bland-Altman bias and limits of agreement are indicated.

Extension to multi-branch CAC

The proposed method focuses on predicting the total CAC mask of the non-injected CT volume. A possible extension is to localize the calcium deposits and assign them to a coronary branch. This application allows to quantify the per branch CAC burden and compute a per branch AS. CSCT is not injected with contrast solutions therefore coronary structures are indistinguishable from the sorrounding soft tissue of the miocard. However trained cardiologists are able to estimate, from the sole CSCT volume, the branch of each calcification based on the anatomical context and the region of the miocard where the coronary is supposed to be located. A multi-labeling semantic segmentation approach based on a 3D U-net can be therefore employed to output a prediction mask for each of the main branches.

Method In order to facilitate the task we proposed to use the heart mask to preprocess the volume and register the heart box to a fixed size in order to mitigate the variability of the different branch locations. The UNet architecture has been slightly modified to accommodate for the anisotropic voxel resolution. With a ratio of 1:1:5 the CSCT input is processed by an anisotropic max-pooling layer in the encoding path and an



(a) (b)

Figure 3.3: (a) C-index vs accuracy for different models evaluated on the test set (J3). (b) Confusion matrix for the risk score of the final prediction on the test set (J3).

anisotropic linear un-pooling layer in the decoding path. The model predicts locations relative to the 3 main coronary branches: Left Artery Descending (LAD), Left Circumflex (LCX) and Right Coronary Artery (RCA). Training and evaluation of this method has been conducted on a separate internal GE Healthcare owned dataset as the SFR dataset cannot be used outside the context of the challenge.

Testing For each volume a multi-label semantic segmentation is predicted. The testing data set population is split into AS intervals corresponding to the risk classes. The measurements in Figure 3.6 aim to measure the risk prediction capability and the localization capability. For risk prediction accuracy is re-weighted to take into consideration uniform risk distribution thus obtaining a re-weighted accuracy of 0.95. For localization the overlapping metric of Dice is measured for the 3 main branches (RCA, LAD, LCX) re-weighting is applied in order to obtain a uniform voxel count for each branch thus obtaining a re-weighted dice of 0.94.

3.1.4 Discussion

Our method based on an ensemble of five 3D U-Net models to detect dense segmentation maps has obtained good fully automated detection and quantification of the AS. Indeed, our method was able to correctly classify the cardiovascular risk category in 86% of the 98 subjects of the validation database according to the AS level, quantified manually and conventionally by trained operators. As shown in the confusion matrix of the figure 4, a class discrepancy was observed in only 14 subjects, and since the amplitude of this discrepancy was low and ≤ 1 class on 13 of these subjects, the c-index obtained with our method was 95%. We have also shown here the importance of the volume of the training dataset of the different CNN models taken alone or combined on

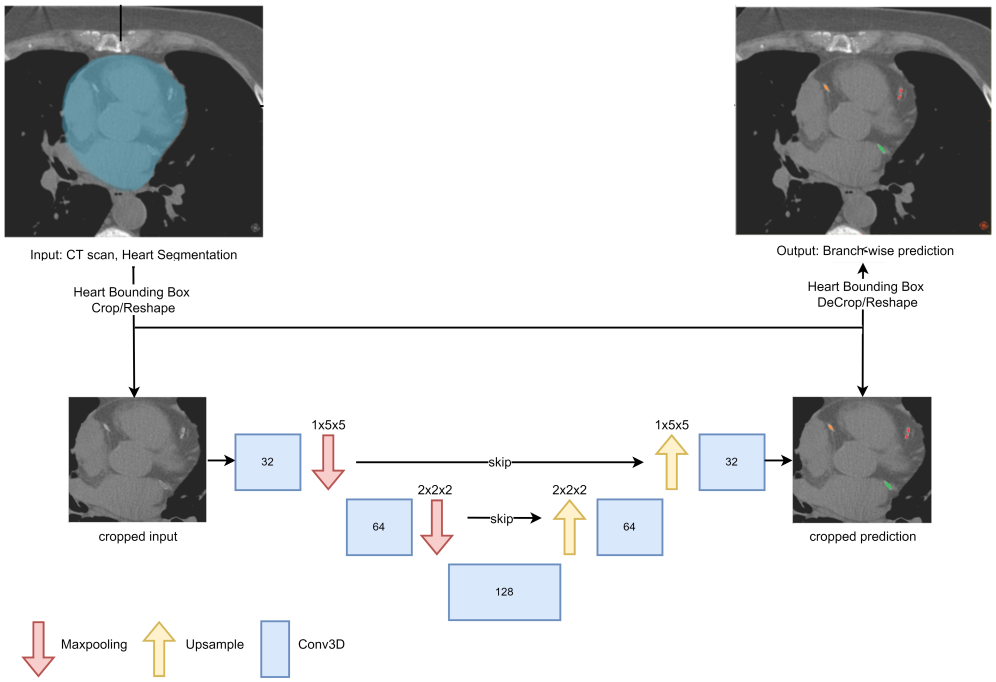
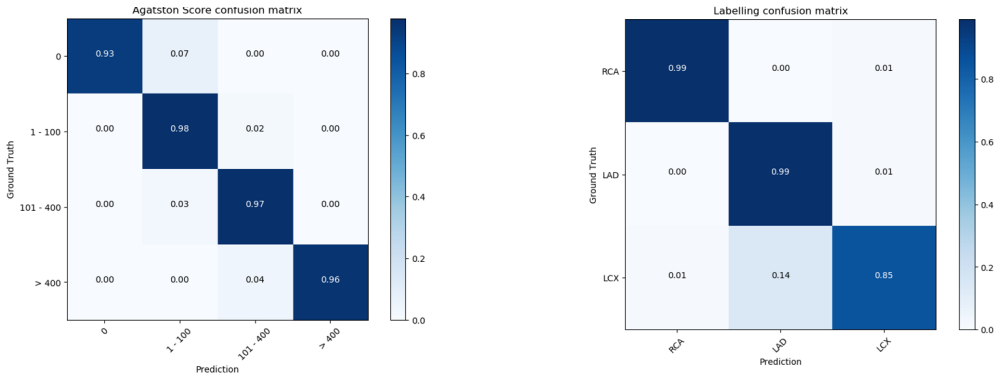


Figure 3.4: Pipeline for multi-branch CAC prediction. The input consist of the CSCT scan paired with the heart segmentation used to crop/reshape the input. The cropped input is fed to the UNet architecture equipped with an-isotropic maxpooling layers matching the original an-isotropic voxel resolution. The output is than upsampled and padded (invert crop) to obtain the final output.

the final result. In these conditions, it is not surprising to see that AI model proposals published in 2018 on a training data base of 1744 ECG gated CT acquisition dedicated for CAC scoring [32] now obtain results very close to those obtained manually by operators trained to detect the coronary calcium score when these databases have been diversified by different types of scanner and when these training databases are much larger [33]. In this large study including 7240 participants with a wide range of CT examination with ECG gated CAC scoring CT , diagnostic chest CT , PET attenuation correction CT and radiation therapy CT, the deep learning solution was tested on 4324 CT examination to obtain a weighted kappa value of 0.90 for all test CT scans and ICC yielded 0.79–0.97 for CAC across the range of different types of CT examinations. From a methodological point of view, different methods have been proposed in the automated detection and analysis of CAC score on CT. A method based on an ensemble of pairs of CNNs obtained 83% of accuracy in predicting patients CAC score risk class by using only the injected CT when the classification requested to the algorithm was only 4 classes instead of 5 [28]. If classes B and C are grouped together in a single class (CAC score from 1 to 100) as it is often suggested in clinical practice in non-diabetic subjects, the accuracy of our solution could also increase from 84 to 94%. Another approach makes use of



(a) (b)

Figure 3.5: (a): confusion matrix of estimated AS category risk. (b) confusion matrix of the estimated CAC for each coronary region.

fuzzy features and atlas-based information in conjunction with Random Forest Models to exploit the available data at best and get accurate quantification and branch-wise location of CAC and only the non-enhanced-CTimages are required [29]. To produce a more accurate segmentation of coronary artery, another method has been suggested to automatically detect calcified lesions on the non-enhanced CT images, but the segmentation of the aorta, the heart and coronary arteries was required and obtained by using associated contrast CT images [30]. Because of its relevance in CAD screening in a population of smokers, another deep learning method has been tested on the NLST (National Lung Screening Trial) [31]. This method uses of a 2-stage FCNN prediction to quantify branch-wise calcifications using the sole non-enhanced CT examination [32]. This method previously trained on manual and segmental labeling of calcified coronary lesion per coronary artery in a subset population of the NLST patients was further tested on multiple cardiac CT protocols as previously mentioned [33]. Finally, a last interesting method still based on CNN [34] was trained on NLST data set using only the CAC score information as supervision. The method employs two CNNs, one for registration to align the input images to an atlas image made from cardiac CT s and one for direct CAC score prediction using regression. This second CNN operates on 2d slices and does not use the information across slices which according to the authors may account for some incorrect identification of CAC near the coronary artery ostia. By contrast our approach is fully 3D and therefore is capable of better using the information across adjacent slices to identify CAC. To maintain a large receptive field while limiting the depth of the CNN, we chose to train our model on low-resolution versions of the CT volumes using a single standard U-Net architecture, which led to a cost-effective training and inferencing process (less than 2 seconds per examination). The performance obtained with our model on the orCaScore test set was very close to the other top performing methods evaluated with this framework [30, 29, 32]. The limitation of our work presented here was that only dedicated non-enhanced-ECG-gated CT scans for calcium score were trained and eval-

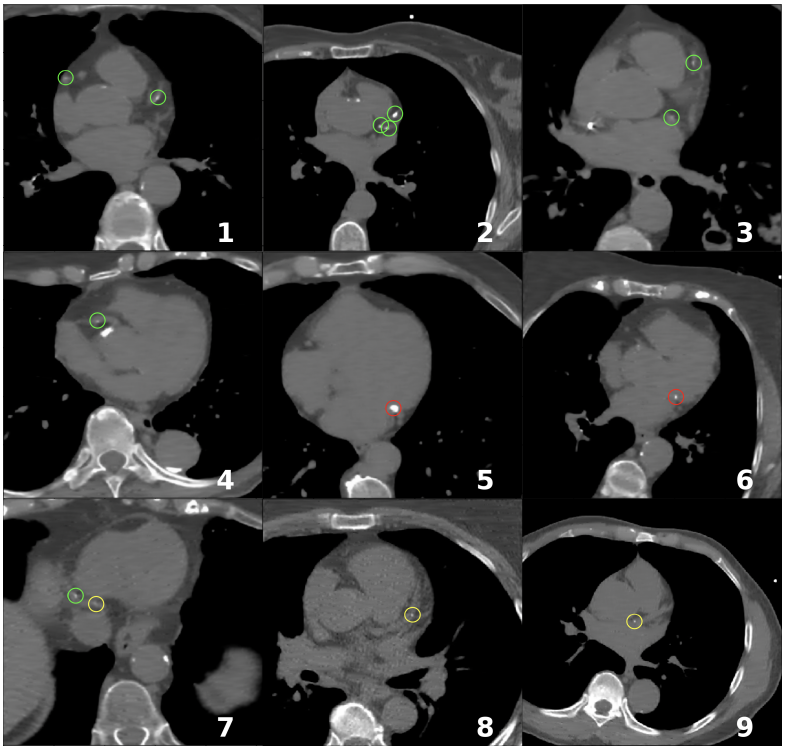


Figure 3.6: Examples of prediction by the final ensemble model. Green circles are true positives, red circles are false positives (false detection), yellow circles are false negatives (missed calcification). 1-4 are examples of correct predictions, 5-9 are examples of incorrect predictions with either false detection or missed detection. Examples 5, 6 and 9 are typical examples of the confusion made between mitral valve and coronary artery calcifications.

uated. Given the value of the CAC score in detecting subjects at high cardiovascular risk who require primary prevention, it would be useful in future studies to test and validate this method on a wider range of CT acquisition performed for routine examinations for lung disease.

3.2 Coronary Tracking

Coronary tracking and centerline extraction from CCTA examinations allow for the assessment of stenosis and the characterization of plaques in patients with suspect CAD [72]. This section proposes applying deep learning techniques to automatic coronary tracking from CCTA volumes. A novel Soft-Persistent Skeletonization (SPS) algorithm that avoids disconnections in the discrete setting, a differentiation scheme of the skeletonization operator, and a MS_Loss that enforces connectedness of the solution are presented. We propose to use the SPS algorithm, the differentiation scheme, and the MS_Loss to train a deep neural network to segment coronary tubular structures using the sole centerline as supervision. Finally, we design an algorithm based on local semantic segmentation, particle filtering, and recursive tracking for coronary centerline extraction. To this end, we collected a dataset of 225 CCTA examination scans paired with manually annotated centerlines. These examinations are split into 100, 55, and 70 for training, validation, and test. The proposed method rivals another baseline deep-learning-based method present in literature and achieves 0.903 overlap (OV), 0.754 overlap until first error (OF), 0.910 overlap with the clinically relevant part of the vessel (OT), and 0.46 average inside (AI) on the test set.

3.2.1 Ground truth generation

Manual centerline extraction is a burdensome and time-consuming task requiring years of experience in the visual inspection of CCTA examination. For our dataset, coronary centerlines are first automatically initialized with GE Healthcare software for automatic coronary segmentation. This segmentation algorithm relies on the manual positioning of the aortic valve central point. Once the point has been manually placed, the centerline is computed automatically. Coronary centerline portions in proximity of lesions (plaques, stenosis, and stents) are at the same time the most clinically relevant and the hardest for an automatic method to estimate correctly; therefore a trained cardiologist refines and corrects the centerline focusing on lesions.

3.2.2 Proposed method

Soft-Persistent-Skeleton

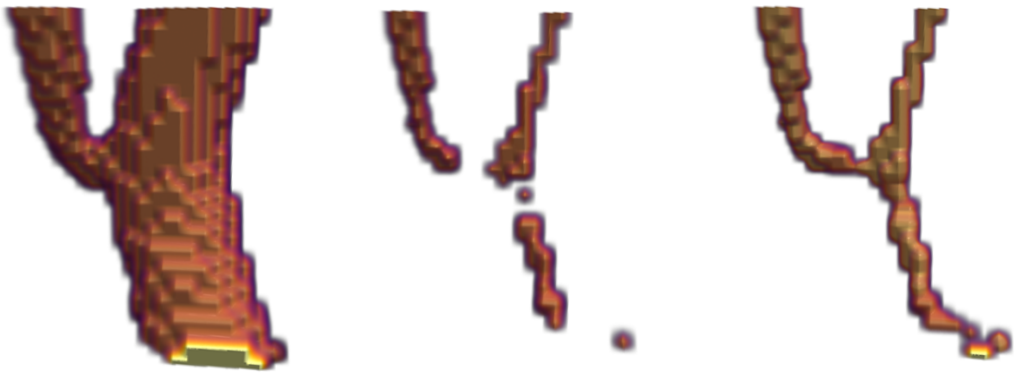
The skeletonization approach proposed in [46] is based on the discretization of Lantuéjoul's formula (2.28). In this case, the algorithm terminates with $S(X)$ such that $\gamma_{\kappa_0}(X) = \emptyset$. In the discrete setting, however, this formulation does not guarantee homotopy of the result (same topology). Thus, it is not an ideal skeletonization procedure. For tubular structures, this usually results in disconnections. We propose an improved Soft-Persistent Skeletonization (SPS) algorithm that enforces connectivity in the discrete setting (fig. 3.7). However, the resulting skeleton S from Algo. 1 may still contain simple points.

For S to be a skeleton, the simple points property must apply at each iteration. In

Algorithm 1 SPS

Input: I, K
  $I' \leftarrow maxpool(minpool(I))$
  $S \leftarrow Relu(I - I')$
  for $i \leftarrow 0$ **to** K **do**
    $S \leftarrow max(I \odot maxpool(S), S)$ ▷ Add connection
    $I \leftarrow minpool(I)$
    if $I = \emptyset$ **then** break ▷ Terminates
    end if
    $I' \leftarrow maxpool(minpool(I))$
    $S \leftarrow S + (1 - S) \odot Relu(I - I')$
  end for
Output: S

this case the connection added by $I \odot maxpool(S)$ depends on $maxpool(S)$ which has a unitary discreet radius. Therefore in the worst-case scenario, the obtained result is included in the unitary dilation of the skeleton. To get a thinner output, the algorithm is reapplied $S \leftarrow \text{SPS}(S, k)$. As the implementation in [46] we used connectivity of 6 for minpool and connectivity of 26 for maxpool.



(a) (b) (c)

Figure 3.7: (a) a tubular synthetic structure I. (b) its skeleton computed with *Soft-Skeleton* [46] until termination. (c) skeleton computed using Alg. 1.

MS_Loss and efficient back-propagation

Given a set of location in a spatial domain $X \subset \Omega$, the skeletonization operator S is anti-extensive $S(X) \subseteq X$. If we were to define this operator on binary functions $X : \Omega \mapsto [0, 1]$ the anti-extensive property translates as a masking operator $S(X) = S(X) \odot X$ thus its derivative has a recursive formulation.

$$S'(X) = (S(X) \odot X)' = S(X) + S'(X) \odot X \tag{3.1}$$

Hereafter we refer to the result of the Soft-Persistent-Skeleton applied on an argument as $\mathsf{SPS}(*)$ and we propose a Dice-like loss suited for centerline segmentation as follows:

$$\mathsf{MS\_Loss}(X, Y) = 1 - \frac{2\mathsf{Prec}(X, Y)\mathsf{Sens}(\mathsf{SPS}(X), Y)}{\mathsf{Prec}(X, Y) + \mathsf{Sens}(\mathsf{SPS}(X), Y)} \tag{3.2}$$

where $X : \Omega \mapsto [0, 1]$, $Y : \Omega \mapsto \{0, 1\}$ are binary valued functions defined on a continous image domain $\Omega \subseteq \mathbf{R}^3$, and $Prec(X, Y) = \sum_i X_i Y_i / \sum_j X_j$ and $Sens(X, Y) = \sum_i X_i Y_i / \sum_j Y_j$ are scalar functions. Equation 3.2 requires the computation of the skeleton employing Algorithm 1. Although being differentiable, it is iterative. Computing gradients with back-propagation requires each iteration step to be conserved in memory, thus the memory footprint scales with the number of iterations K. We propose an approximate differentiation scheme with constant memory footprint. The proposed scheme exploits the centered anti-extensive property of the skeleton which implies that $\mathsf{SPS}(X) \subseteq X$and therefore $\mathsf{SPS}(X) \odot X = \mathsf{SPS}(X)$. By an arbitrary change of variable we consider a copy of the segmentation as a constant factor with respect to the skeletonization operator $\mathsf{SPS}(X) = \mathsf{SPS}(X_{const}) \odot X$ it implies that its derivative is trivial $\mathsf{SPS}(X)' = \mathsf{SPS}(X_{const})$ which is indeed a term of the analytical derivative $\mathsf{SPS}(X)' = \mathsf{SPS}(X) + \mathsf{SPS}'(X) \odot X$. Consider the non linear activation function $\sigma(x) = 1/(1 + e^{-x})$, the neural network feature map Z_θ, and the prediction $X_\theta = \sigma(Z_\theta)$: the back-propagation development with respect to θ using the chain rule follows (Equation 3.4).

$$\begin{aligned}
&Sens(\mathsf{SPS}(X_\theta), Y) \\
&\equiv \frac{\partial Sens(\mathsf{SPS}(X_\theta), Y)}{\partial \theta} \\
&= \frac{\partial Sens(\mathsf{SPS}(X_\theta), Y)}{\partial \mathsf{SPS}(X_\theta)} \frac{\partial \mathsf{SPS}(X_\theta)}{\partial X_\theta} \frac{\partial X_\theta}{\partial \theta} \\
&= \frac{\partial Sens(\mathsf{SPS}(X_\theta)), Y)}{\partial \mathsf{SPS}(X_\theta)} \mathsf{SPS}'(X_\theta) \frac{\partial X_\theta}{\partial \theta} \\
&= \frac{\partial Sens(\mathsf{SPS}(X_\theta), Y)}{\partial \mathsf{SPS}(X_\theta)} \odot \mathsf{SPS}(X_\theta) \frac{\partial X_\theta}{\partial \theta} \qquad (3.3) \\
&= \frac{\partial Sens(X_\theta, Y)}{\partial X_\theta} \odot \mathsf{SPS}(X_\theta) \frac{\partial X_\theta}{\partial \theta} \qquad (3.4)
\end{aligned}$$

By back-propagating only a selection of values formally it implies a change of variable in the differentiation (this is true only for the $Sens$ function and all pixel-wise metrics in which there is no intra-location dependency at one location), this formulation can

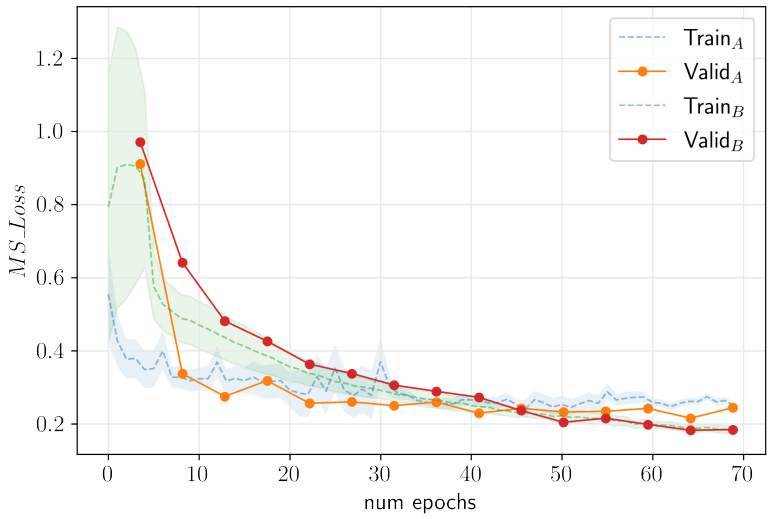


Figure 3.8: Train$_A$, Valid$_A$ no variation of the iterative method (alg. 1), Train$_B$, Valid$_B$ using the masked error back-propagation with same $K = 8$. In terms of training resources. Time: 5h:57m vs 4h:47m; GPU-DRAM: 17.89 GB vs 12.02 GB; for Train$_A$/Valid$_A$ vs Train$_B$/Valid$_B$ respectively (GeForce RTX 3070 NVIDIA®).

be thus interpreted as a masking of the back-propagated local error *Err*: $\frac{\partial Sens(X_\theta, Y)}{\partial X_\theta} \odot \text{SPS}(X_\theta) = \textit{Err} \odot \text{SPS}(X_\theta)$. The practical advantages of this formulation are that there is no need to fine-tune the number of iterations parameter K for the forward computation, and the back-propagation is resource-efficient (Fig, 3.8). To segment fine/sparse tubular structures the model is trained using the MS_Loss on both foreground and background to capture the entire 3d topology (connected components, handles and tunnels) $\mathbf{L}(X_\theta, Y) = \frac{1}{2}\text{MS\_Loss}(X_\theta, Y) + \frac{1}{2}\text{MS\_Loss}(1 - X_\theta, 1 - Y) + \lambda DiceLoss(X, Y)$. MS_Loss complements the $DiceLoss$ by combating the sampling bias, while MS_Loss focuses on the skeletonization points the $DiceLoss$ counterbalances by preventing over-fitting on the sole topology.

Topological interpretation of the gradient

Another interpretation can further validate the intuition behind the approximate differentiation scheme of the skeletonization operator $S'(X) = S(X)$. Non-simple points are locations at which topological changes happen: removing a point necessarily means either the birth of a new connected component (0D manifold) or the death of a tunnel (1D manifold) or a hole in (2D manifold). Let us define the tubularity likelihood as a function of the distance from the center. Its filtration corresponds to the normalized distance transform of the binary segmentation of the tubular structure. Therefore, the topology remains unchanged until the final erosion, which implies that critical points are singular points of the skeleton.

$$
\begin{aligned}
f &= 2d/(1+d) \\
d(p) &= \frac{|d_B(p) - d_C(p) + |}{\max(d_B(p), d_C(p))+}, p \in X \\
d_B(p) &= \min_{x \in \partial X} \|x - p\|_2 \\
d_C(p) &= \min_{x \in S(X)} \|x - p\|_2
\end{aligned} \tag{3.5}
$$

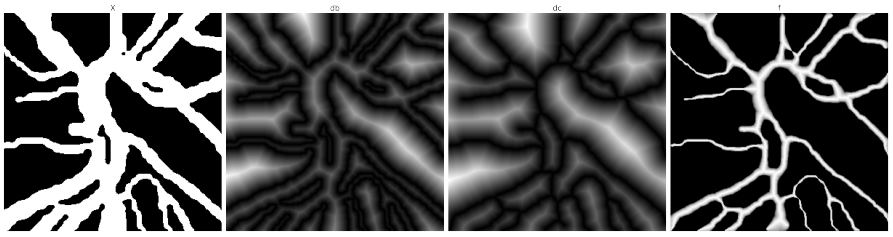


Figure 3.9: Left to right: the image X with its tubular structures, the distance d_b from the border, the distance from the centerline or skeleton d_c, and the likelihood function f.

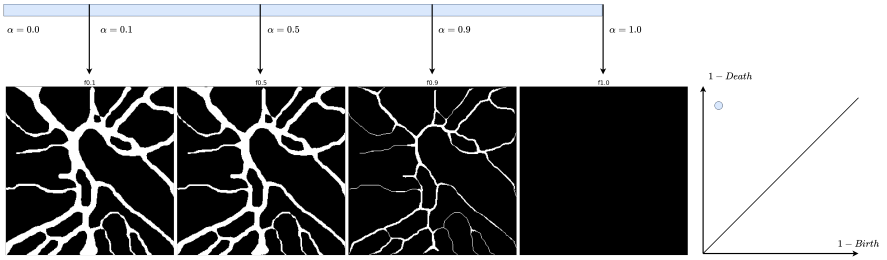


Figure 3.10: $Dgm(f)$ f is defined in Equation 3.5, the filtering $\emptyset \subseteq f^{0.1} \subseteq f^{0.5} \subseteq f^{0.9} \subseteq \Omega$ has a constant topology (all filtering are in a homology) the critical points appear when the filtering corresponds to the skeleton.

Consider the filtering of f as the set of tresholded functions $f^{\alpha} = x \in \Omega \mid f(x) \geq \alpha$ such that $\emptyset \subset f^{\alpha_1} \subset f^{\alpha_2} \cdots \subset f^{\alpha_n} \subset \Omega$, wich implies that $\alpha_i \in [0, 1]$. A persistence diagram Dgm measures at what value of α topological changes occurs (fig. 3.10). Consider the persistence diagram of the function defined in eq. 3.5. Topology variation do not occur for $f^{\alpha<1}$ the topology vanishes into the spatial domain Ω for $f^{\alpha=1}$, in other words the topology is constant. The critical points of this function are the non-simple points of the skeleton which are the locations where topological variations occurs $\{(f(x), 1 - f(x)) \mid x \in S(X)\} \equiv \{p \in dgm(f)\}$. The topological loss is defined as eq.3.6, observe that for a ground truth the filtration has only one value as it is not a continuous function, therefore $birth(p) = 1 - death(p) \forall p \in Dgm(GT)$ where $G : \Omega \mapsto [0, 1]$

and $\nu : Dgm(f) \mapsto Dgm(GT)$. By design of f all points of the skeleton are critical points, the same is true for the GT.

$$
\begin{aligned}
L_{topo} &= \min_{\nu \in N} \frac{1}{2} \sum_{p \in Dgm(f)} \|p - \nu(p))\|_2^2 \\
&= \sum_{p \in Dgm(f)} \frac{1}{2}(birth(p) - birth(\nu^*(p)))^2 + \frac{1}{2}(death(p) - death(\nu^*(p)))^2 \\
&= \sum_{p \in Dgm(f)} \frac{1}{2}(birth(p) - birth(\nu^*(p)))^2 + \frac{1}{2}(1 - birth(p) - (1 - birth(\nu^*(p))))^2 \\
&= \sum_{p \in Dgm(f)} (birth(p) - birth(\nu^*(p)))^2 \\
&= \sum_{x \in S(f^{\alpha<1})} (f(x) - GT(x)))^2 \qquad (3.6)
\end{aligned}
$$

We approximate the function f by means of a non linear activation function $\sigma(x) = 1/(1 + e^{-x})$ and a neural network parameterized by Θ feature map $X_\theta = \sigma(Z_\theta)$ and, therefore $f \sim X_\theta$ which implies that $\{X_\theta(x), 1 - X_\theta(x) \mid x \in S(X_\theta)\} \equiv \{p \in dgm(X_\theta)\}$. The gradient is therefore defined as

$$
\begin{aligned}
&\frac{\partial L_{topo}}{\partial \theta} && (3.7)\\
&= \sum_{x \in S(X_\theta)} 2(X_\theta(x) - GT(x)) \frac{\partial X_\theta}{\partial \theta}(x) && (3.8)\\
&= \sum_{x \in \Omega} (2X_\theta(x) - GT(x)) \odot S(X_\theta)(x) \frac{\partial X_\theta}{\partial \theta}(x) && (3.9)\\
&= \sum_{x \in \Omega} \frac{\partial (X_\theta(x) - GT(x))^2}{\partial X_\theta} \odot S(X_\theta)(x) \frac{\partial X_\theta}{\partial \theta}(x) && (3.10)\\
&= \frac{\partial L_{topo}}{\partial X_\theta} \odot S(X_\theta) \frac{\partial X_\theta}{\partial \theta} && (3.11)\\
& && (3.12)
\end{aligned}
$$

If we substitute the sum of the distances with an overlap measure like sensitivity $Sens(X, Y)_{topo} = \sum_{i \in S(X)} X_i Y_i$ than the gradient would read $\frac{\partial Sens_{topo}}{\partial \theta} = \frac{\partial Sens_{topo}}{\partial X} \odot S(X) \frac{\partial X}{\partial \theta}$. Conversely from Equation 3.12 if we were to use the L_2 norm for the Equation 3.4 we can derive the topological gradient. This result shows how the approximation of the skeleton used in this method can be derived by applying the gradient to a topological loss with a suited likelihood function.

Tracking Strategy

The skeletonization algorithm proposed in [46] is prone to disconnection and unsuited for centerline segmentation. We propose combining the MS-Loss and the resource-efficient scheme for back-propagation to train a U-Net to segment the coronary centerline (MS-Unet). To this end, we collected a dataset of 225 CCTA examination scans paired with manually annotated centerlines. These examinations are split into 100, 55, and 70 for training, validation, and test. For training: as a preprocessing step, each image is resampled to $0.4\,\mathrm{mm}$ isotropic voxels, and a windowing (center=400, width=1000) is applied. The MS-Net is fed with 32 voxels-sided cubic patches in a 64 batch. These patches are sampled around the centerline coordinate position and augmented using random translation, rotation, scaling, and skew. For the inference: the segmentation alone is insufficient: extremities must be identified to isolate a single coronary centerline.

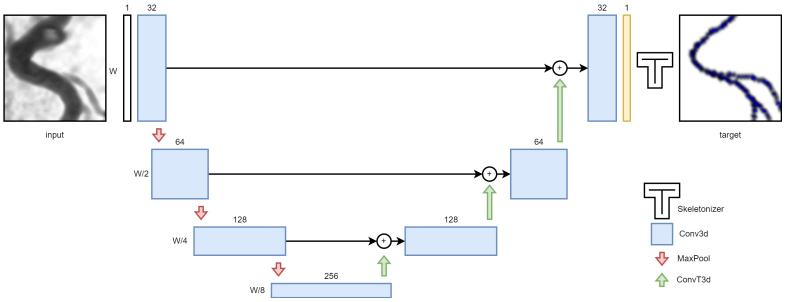
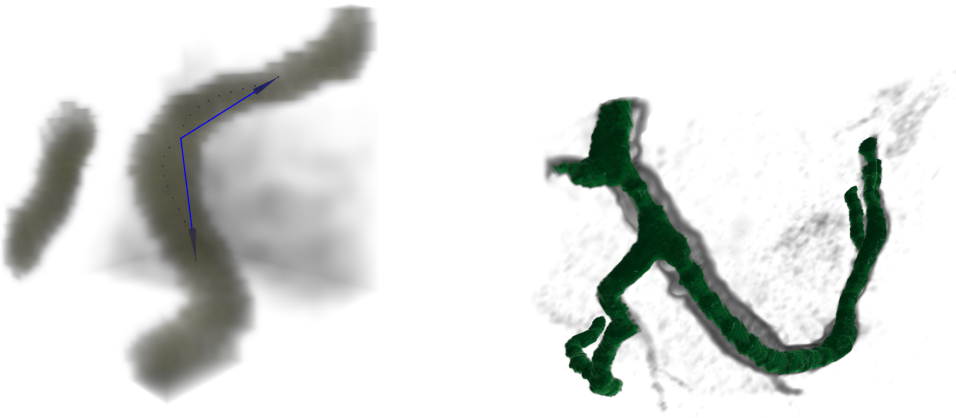


Figure 3.11: MS-Unet Architecture: The Unet takes as input a patch $P \in \mathbf{R}^{24,24,24}$ at a isotropic resolution of 0.4 mm centered about the centerline and outputs a soft segmentation map $\sigma(Z_\theta)$, the Skeletonizer block applies 1 to output the final segmentation.

We adapted the recursive tracking proposed in [42] by replacing the convolutional neural network (CNN) for 2D orthogonal patches classification with a Unet for 3D patches centerline semantic segmentation (MS-Unet): First, the ostia locations are identified by pre-computing the mask of the aorta. Second, the next locations are identified using a particle filtering sampling technique for statistical estimation: samples are obtained by intersecting the MS-Unet semantic segmentation with a fixed-diameter sphere adjusted to the receptive field of the model. The intersected samples are then clustered using density-based spatial clustering of applications with noise (DBSCAN) from which the centroids of the centerline are obtained. This procedure is carried out for a given current location recursively on the next locations until no intersections are found, and the current location is marked as an extremity. To avoid early termination, the receptive field of the MS-Unet is chosen to be large enough to bridge over stenosis. The recursive inference procedure outputs a set of extremities, locations, and the associated volume voxel-wise segmentation. The voxel-wise segmentation is used to find the minimum cost path from the extremities to the corresponding ostium. All positive voxels of the segmentation output constitute the nodes, and the average MS-Unet output value is the weight of the edge between 2 voxels (see Fig. 3.15).



(a) (b)

Figure 3.12: (a) Local tracking strategy. A patch centered on the vessel is fed to the network for semantic segmentation; the intersection between a fixed radius sphere and the segmentation is estimated to compute the next central locations. This procedure is called recursively on the next locations. (b) Semantic segmentation of the coronary vessel before skeletonization. Interestingly the network learns in a unsupervised manner to avoid calcifications in order to obtain a precise centerline

3.2.3 Experiments

Topology and Homology

Given a continuous image domain $\Omega \subseteq \mathbf{R}^3$, a segmentation $I \subseteq \Omega$ has its *d*-dimension topological structure, called *homology* as an equivalence class of *d*-manifolds with the same topology as I [49]. The *homology class* for a *d*-manifold is a countable set under I and its cardinality is the d^{th} Betti number β_d which vanishes above the dimension of the domain Ω. For example for 3-*d* domains, β_0, β_1, β_2 correspond to the number of connected components, tunnels and holes respectively. We propose a simple check for *homotopy* (same topology) between two segmentations I and $S(I)$ is to measure the difference of their Euler number $E : \Omega \mapsto \mathbf{Z}$ as $E_{dist}(A, B) = |\inf(E(A), E(B)) - \sup(E(A), E(B))|$ which is a combination of Betti numbers: for 3-*d* objects, the Euler number is obtained as β_0+β_1-β_2. An ideal skeleton $S(I) \subseteq I \subseteq \Omega$ is homotopic and belong to the same *homology* as I and therefore has same Euler number. Tab. 3.2 holds the results of this empirical check on a set of tubular synthetic structures I(fig. 3.7).

Testing

Our dataset employs 225 examinations: 100 for training 55 for validation, and 70 for the test. The test set was chosen by focusing on the clinical use of coronary tracking; most CCTA examinations have severe calcifications and present artifacts or stents (40 with

Homotopy	β_0	β_1	β_2	E_{dist}(*,I)
I	1.3	0.1	0	0
S(I)	12.5	0	0	10.3
S_p(I)	5.2	0	0	**3.2**

Table 3.2: S is *Soft-Skeleton* as in [46] S_p is Soft-Persistent-Skeleton, each measure is averaged on a set of synthetic 3-*d* tubular structures I.

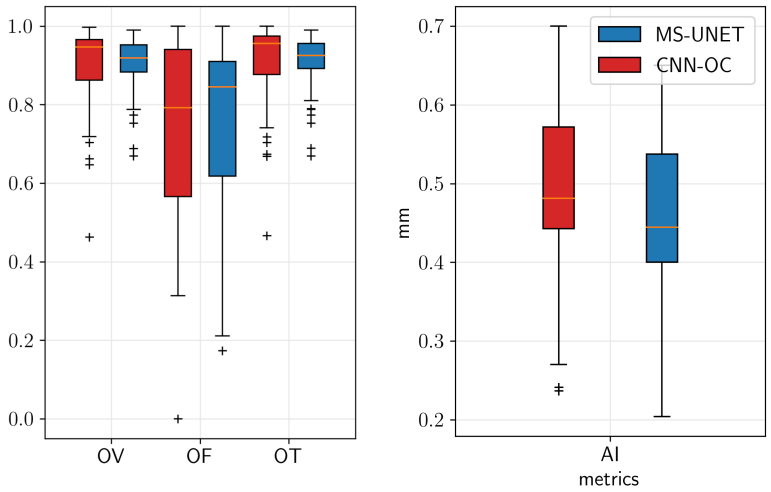


Figure 3.13: Metrics measured on the test set of 70 CCTA patients with manually corrected centerlines. These metrics measure the performances of a centerline tracker. AI metric takes account of all points not only the ones within the radius of the coronary.

Method	OV	OF
CNN-OC [44]	0.898 $\pm$ 0.099 [0.46;0.99]	0.742 $\pm$ 0.219 [0.46;0.99]
MS-Unet (ours)	**0.903** $\pm$ 0.066 [0.66;0.98]	**0.754** $\pm$ 0.212 [0.17;1.00]]

Method	OT	AI
CNN-OC [44]	**0.912** $\pm$ 0.099 [0.46;1.00]	0.485 $\pm$ 0.146 [0.214;0.7]
MS-Unet (ours)	0.910 $\pm$ 0.066 [0.66;0.98]	**0.460** $\pm$ 0.120 [0.204;0.65]

Table 3.3: Metrics on 70 CCTA examinations of the test set compared with 2 deep learning based method: CNN-OC [44] MS-Unet, our proposal. Each metric is computed per patient and presented as $mean \pm std[min; max]$.

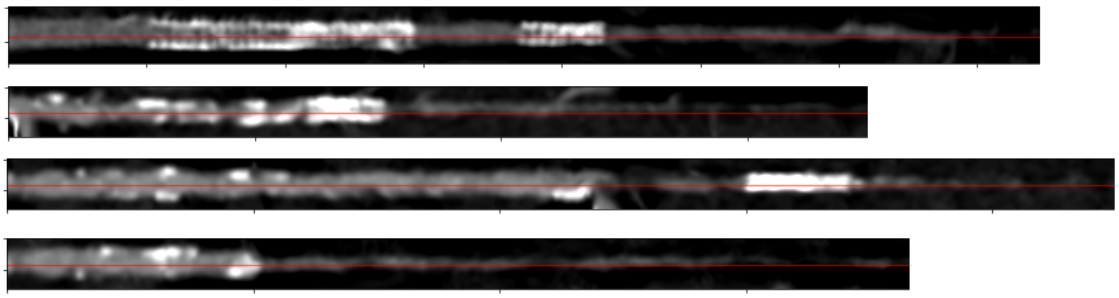


Figure 3.14: A sample of extracted centerlines, rough coordinates are used to compute orthogonal frames to inspect the coronary visualized with multi-planar reconstructions at a sub-voxel resolution of $(0.25 \times 0.25 \times 0.25).mm^3$ [73]

> 400 AS, 12 with stents). The proposed tracking method is evaluated on the test (Figure 3.13)(Table 3.3) against a state-of-the-art deep-learning approach [44][1]. Predicted centerlines are evaluated with a standard evaluation method [74] (see 2.3.1). Total overlap (OV), overlap until first error (OF), and overlap of the extracted centerline with the clinically relevant part of the vessel (radius ≥ 0.75 mm, OT) are computed using true positive (TP), false positive (FP) and false-negative (FN) detections. A TP point lies within the radius of the closest manually annotated point. An FP point does not lie within the radius of any manually annotated point. An FN point is a manually annotated centerline point with no corresponding automatically extracted point. The average inside accuracy metric (AI) measures the average distance between the manually annotated and extracted centerline for automatically extracted points. These metrics aim to measure the tracking capability (fail or pass) and the accuracy of the correctly tracked portions of the vessel. Our method shows less significant low values outliers than the baseline, while median and high values are comparable. The average performances in Table 3.3 are higher for the proposed method, but these are sensible to outliers. Another important point is the hyperparameter tuning: although both algorithms are initialized with the same ostia detection routine, these have different inference strategies (iterative and recursive tracking, step sizes, tracking stop criteria) with different hyperparameters, and exploring all configurations becomes infeasible.

3.2.4 Discussion

We proposed a coronary segmentation method based on deep-learning semantic segmentation of centerlines (MS-Unet) and a recursive tracking inference scheme. To train the MS-Unet we propose a novel skeletonization algorithm based on morphological operations that empirically shows better properties than the straightforward implementation using discrete erosion and opening operators. We also propose a novel back-propagation scheme that is resource efficient and present a theoretical topological in-

[1] implementation available at https://github.com/BubblyYi/Coronary-Artery-Tracking-via-3D-CNN-Classification.

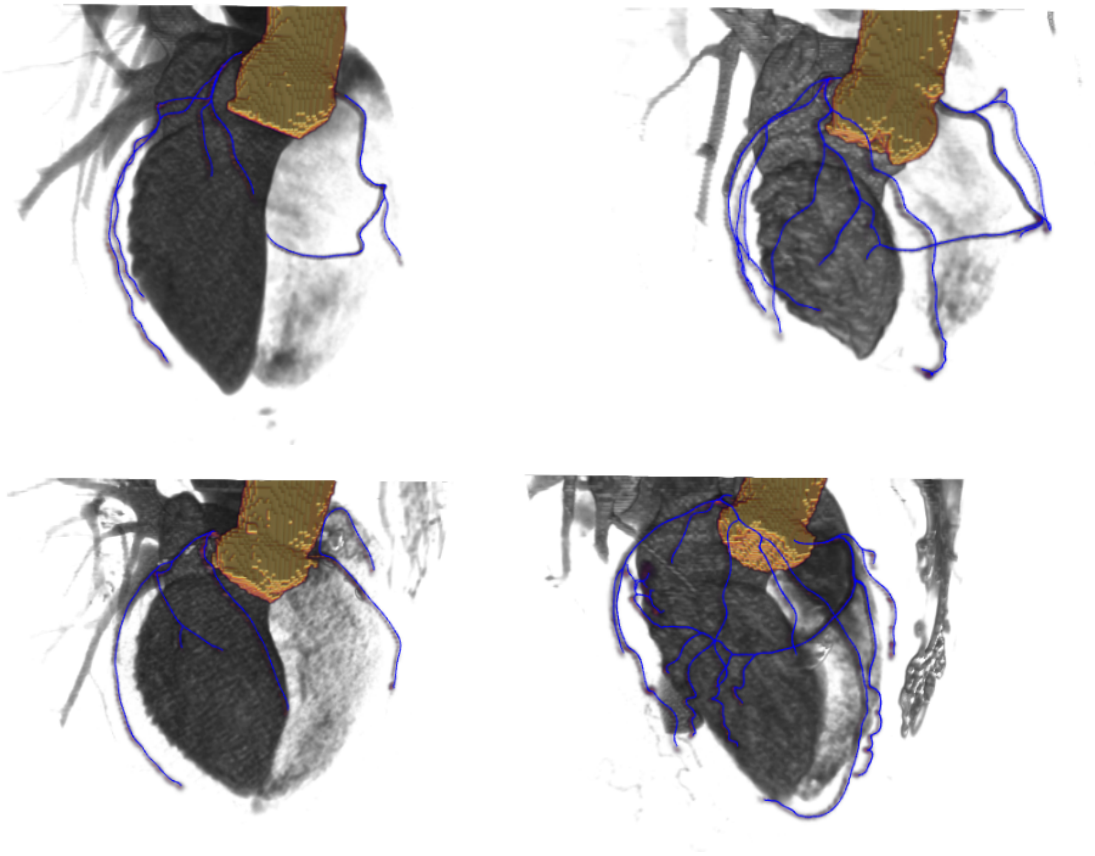


Figure 3.15: Recursive tracking output. Blue: coronary centerline extracted. Yellow: Aorta mask.

terpretation. To infer the centerline we propose to exploit an initial local segmentation to estimate the next location of the centerline recursively, this strategy takes inspiration from particle filtering trajectory estimation. The proposed method was tested on an dataset propriety of GE Healthcare. The performances were measured using a standard coronary tracking evaluation framework composed of four metrics designed to describe tracking capabilities given a ground truth reference. The method was tested against a deep-learning based CNN orientation classifier (CNN-OC) present in literature: 0.898 and 0.903 OV, 0.742 and 0.754 OF 0.912 and 0.910 OT, and 0.485 and 0.460 AI were obtained for the baseline CNN-OC and MS-Unet respectively. The metrics are close however the proposed method has fewer outliers meaning that is less prone to get lost as CNN-OC can sometimes track vessels outside the heart.

3.3 Coronary Labeling

3.3.1 Ground truth generation

The ground truth labels are manually assigned to the leaves detected by a coronary tracking algorithm from the CCTA volume. When this algorithm fails, missing branches are manually added by using manual ROI annotation and region growing. Due to anatomical variability the segmentation algorithm may detect short branching, these are anatomical results of natural revascularization process but are not considered coronary branches. The resulting coronary computed by inspecting the coronary segmentation volume is therefore pruned of terminal contiguous segments with length below a certain threshold fixed in millimeters. Due to the high anatomical variability there might be outliers in branching patterns (more than 4 diagonals, more than 4 marginals). In this case such anomalies are also pruned to avoid adding noise to the training process.

3.3.2 Proposed Method

Knowledge based (KB)

We present a knowledge-based method consisting in detecting anatomical landmarks of the heart; from these the geometry of the branching and the coronary labels are computed inductively. The coronary tree is computed by processing the coronary segmentation and tracking. The anatomical landmarks and the coronary tracking are computed using the CCTA volume by shape analysis algorithms developed at GE Healthcare. First the aortic central coordinate H is identified by manually placing a point inside the aorta, the leaves of the segmentation are automatically computed and the paths from the leaves to the aorta are merged in order to derive the coronary tree composed by center-line points. The merging of the paths allows to detect the bifurcation and finally the coronary segments. Subsequently the heart apex coordinates A and the ostia P_1, P_2 are computed. Once the short axis $\overrightarrow{AH}$ is identified the ostia points are projected on the axial plane. The front and the left of the exam are known, therefore it is possible to discriminate between left and right coronary territory. Once these are detected the successive branches are pruned from the coronary tree graph so that only consecutive segments with a fixed minimum length are kept for labeling. By proceeding hierarchically the Left Main Artery and Left Main Circumflex are identified using respectively the front left and the back left most territory. The same inductive approach follows for the Posterior Descending Artery and the Posterior Lateral branch in the right coronary tree. Finally the diagonals and marginal branches are labeled using the projection on the short axis (height with respect to the apex) of the leaves of the remaining coronary branches.

Oriented GCN (ORI-GCN)

We propose a deep learning based method that takes as input the geometrical features of the coronary tree: oriented graph convolutional network (ORI-GCN). The coronary

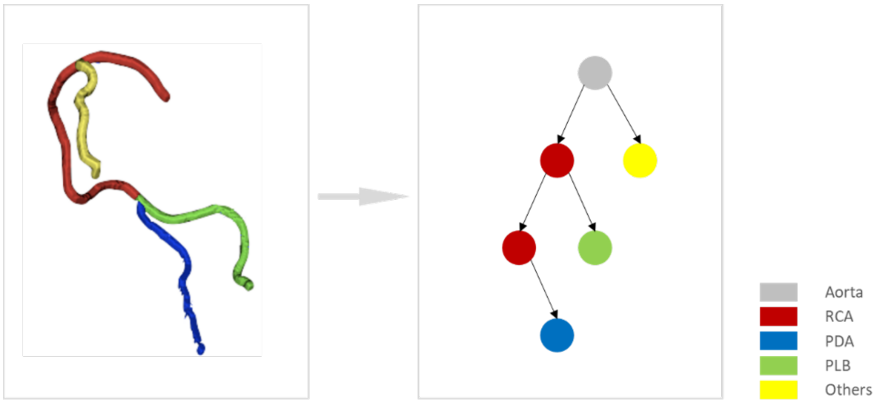


Figure 3.16: The coronary tree extracted from the CCTA coronary segmentation is simplified into contiguous segments (nodes) and the branching patterns of the bifurcations (edges).

tree is simplified so that each node represent a coronary segment (figure 3.16). Each branch is composed by a set of segments connected at bifurcations to other branches. The task consists of predicting the label of the leaf given the coronary nodes geometric features and the tree structure. The geometrical features of each node are computed from the contiguous segment it represents. These features are: starting and ending curvilinear coordinates $\{c_0, c_1\}, c_i \in \mathbf{R}$, and 5 uniformly sampled aortic coordinates $\{p_0, \ldots, p_4\}, p_i \in \mathbf{R}^3$, additionally the branching factor of the segment N_c and the depth in the graph D are added as features. The aortic coordinates are computed by subtracting the aortic center right-anterior-superior (RAS) a_{RAS} RHS coordinates from the RAS coordinate $p_i = p_{RAS} - a_{RAS}$. Although segments do not have the same length the geometrical features are used to capture the shape and the location of each segment in a vector $x \in \mathbf{R}^{17}$, while the tree structure links these segments together (fig.3.16).

$$O_{i,j,d} = \begin{cases} 1, & i \text{ is successor of } j, d \text{ is succ} \\ -1, & i \text{ is predecessor of } j, d \text{ is prec} \\ 0, & \text{else} \end{cases} \tag{3.13}$$

The linking of these features is modeled using the normalized adjacency matrix $\mathscr{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ computed with the adjacency and the degree matrix $A, D \in \mathbf{R}^{n \times n}$ respectively, where n is the number of segments. To model the a-cyclic property we propose an additional orientation matrix O (eq. 3.13): the oriented normalized adjacency matrix is therefore computed as $\mathcal{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \odot O$. Thus the architecture comprises $L = 5$ GCN layers which are modified by mean of the oriented adjacency (eq. 3.14). In practice the O matrix masks contributions of either the predecessors (children) or successors (parents) nodes, the two contributions have individual parameters and are subsequently added and passed to the next layer x^{l+1}.

$$
\begin{aligned}
&x^0 = \{c_0, c_1, p_0, \ldots, p_4\, N_c, D\}, c_i \in \mathbf{R}, (N_c, D) \in \mathbf{N}^2, p_i \in \mathbf{R}^3 \\
&x^{l+1} = \mathsf{Relu}((\mathscr{A} \odot O_{succ}) x^l W_{succ}^{l+1}) + \mathsf{Relu}((\mathscr{A} \odot O_{prec}) x^l W_{prec}^{l+1}) + x^l \\
&\tilde{y} = \mathsf{Softmax}(x^L)
\end{aligned}
\tag{3.14}
$$

where $\tilde{y}$ is the final output prediction and $L = 5$ for the network depth.

3.3.3 Experiments

Training

In this section we propose to explore the advantages of a deep learning based approach with respect to a knowledge-based algorithm. Additionally we propose a slight modifications of the graph convolutional network (GCN) message passing mechanism ORI-GCN that we could observe giving better performances than the classical GCN. The model is evaluated using a large database of annotated coronary trees: 516 examinations are used for training and validation, 318 for test. The model is trained using a cross-entropy loss on the label of each node with a higher weighting on the leave nodes. The output prediction vector $\tilde{y}$ is tested against the ground truth multi-class label (LAD, LCX, Posterior Descending Artery (PDA), Posterior Lateral Branch (PLB), Diagonals (Diag), Obtuse Marginal (OM), Ramus Intermedius Branch (RIB)). As additional supervision the labels are grouped into super labels or territories. The LAD teritory comprises: LAD, Diag, and RIB, the LCX territory with the LCX, OM, and Left Main Artery (LMA), and the RCA territory which is the union of PDA, and PLB.

Testing

The quantitative analysis is carried out globally on the territories by measuring and comparing the accuracy. The test set comprises a total of 318 annotated trees, branch wise not all trees are complete (99% LAD, 99% LCX, 69% PDA, 70% PLB, 15% RIB) and some present anatomical variations (RIB) on which the method is also tested on. A global metric is used to measure the accuracy capability of different methods: knowledge based (KB), ID, GCN, and ORI-GCN. This ablation aims to show the impact on performances of the proposed oriented method ORI-GCN (tab. 3.4). The identity (ID) is equivalent to using fully connected network (FCN) on the feature vector x_0 of each node to predict the label. The GCN consists of using the un-directed adjacency matrix. The discrimination capability is tested by means of a confusion matrix (fig.3.17) and a precision recall measurement for each branch (tab. 3.4).

3.3.4 Discussion

The results on the experiment show that the ORI-GCN has better performances than KB method based on the detection of anatomical landmarks. In terms of discrimination

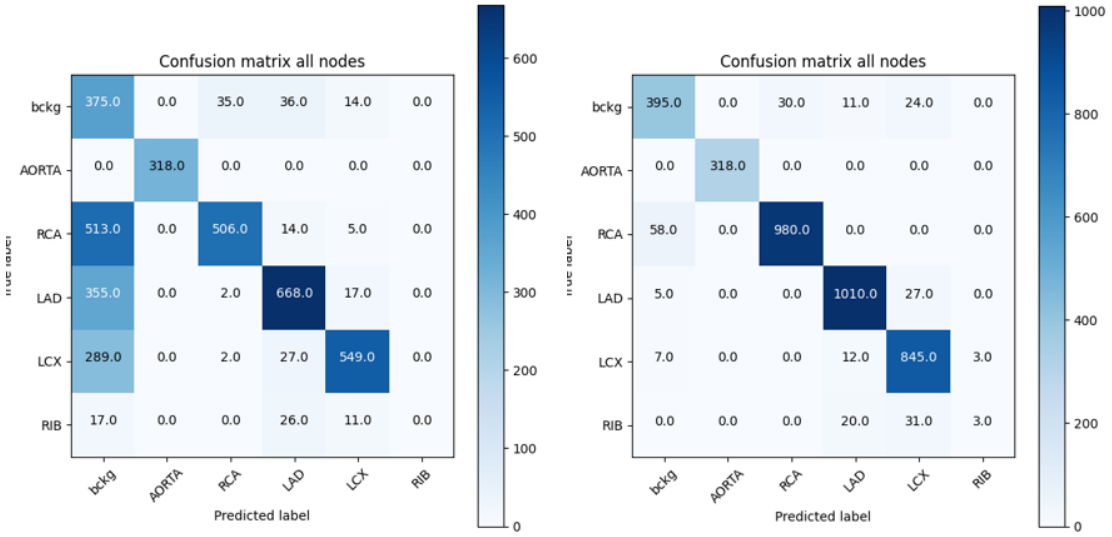
	KB	ID	GCN	ORI-GCN
Branch Accuracy	0.763	0.112	0.745	**0.800**
Territory Accuracy	0.843	0.327	0.921	**0.940**

Table 3.4: Result on territories accuracy: comparison among proposed methods: Knowledge-Based (KB), Identity (ID), Graph Convolutional Network (GCN), Oriented Graph Convolutional Network (ORI-GCN).

		KB		ORI-GCN	
Accuracy		0.763		**0.800**	
label	# branches	Precision	Recall	Precision	Recall
LAD	317	0.869	0.855	**0.946**	**0.943**
LCX	315	0.748	0.724	**0.889**	**0.892**
PDA	220	**0.909**	0.905	0.898	**0.918**
PLB	222	**0.922**	0.910	0.867	**0.910**
Diag	519	**0.758**	0.671	0.752	**0.946**
OM	435	**0.777**	0.520	0.663	**0.931**
RIB	50	N/A	N/A	**0.500**	**0.600**

Table 3.5: Result on main branches: comparing baseline Legacy method and proposed GCN method

capabilities (Figure 3.17) the RIB remains the hardest to detect. This branch is characterized as an anatomical variance that occurs in 15% of the test set which is representative of the normal population [56]; this issue could be solved by artificially augmenting the training dataset. In terms of accuracy capability the ORI-GCN not only it improves the KB approach but also the classical graph neural network GCN approach. To test the impact on the metrics of the proposed approach we also tested a ID approach which consists of using the positional input features x_0 only and discards connections (Table 3.3.3). An issue with this approach is that it does not take into account the hierarchical structure of coronary artery vessels. Each class is treated independently, whereas each label probability is conditioned by the knowledge of the connected branch. A possible solution to this issue is to incrementally predict branches by following the hierarchical structure from the left and right branch down to diagonals and marginal branches. A qualitative analysis of the results shows that some mistakes are more acceptable than others, there might be multiple valid labeling for the same tree as observers may give equally reasonable but different labeling. This work is lacking of a comparison with the state of the art approach [65], the method however has already some advantages as it only relies on positional features, thus realistic data can be generated in a simple procedural way, from a methodological point of view the proposed approach encodes the bi-directional nature of the tree in a similar way to [63] which shows to have an impact on the performances (Table 3.3.3). The whole training does not need any specific computational resources and takes 30 minutes in total and up to 1GB or D-RAM on a capability 5 Nvidia



(a) (b)

Figure 3.17: KB (a) method ORI-GCN (b) performances measured by means of the confusion matrix. Each model is tested on the multiclass task of labeling the Aorta, RCA, LAD, LCX, and RIB. The background label is predicted when none of the label is identified.

GPU against 13 hours for the conditional partial residual graph convolutional network (CPR-GCN) [65].

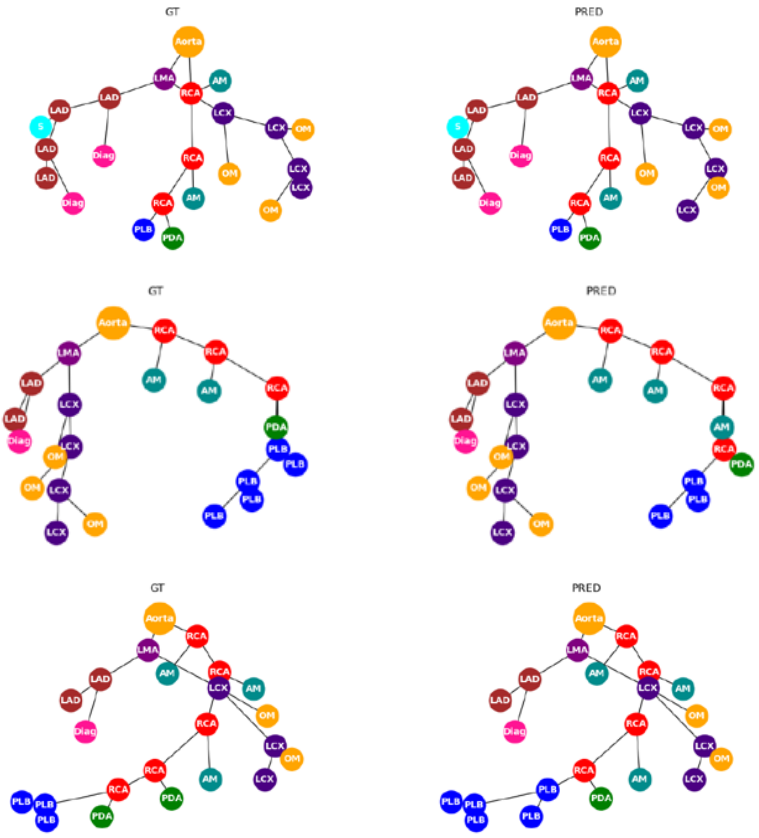


Figure 3.18: Examples of predicted labels and corresponding ground truths.

3.4 Coronary Analysis

CCTA is the only non-invasive imaging technique that reliably depicts the anatomic extent of CAD. While occlusion remains a highly predictive indicator of MACE, there is growing evidence that the presence and characteristics of coronary atherosclerosis provide additional prognostic information. In CCTA calcified plaques display high-intensity HU representative features while more complex representations characterize high-risk soft plaques. As such, accurate identification and quantification is burdensome and time consuming because of the limited temporal, spatial and contrast resolutions of X-ray scanners. Despite the success of deep learning in medical imaging, automatic localization of coronary plaques and especially soft plaques remains a challenging subject in CCTA vessel analysis. In this section, we present our contribution to automated CT coronary analysis, in particular on CAD anatomical indicators which entails the localization and characterization of plaques and stenosis using transformers [75]. We propose a series of deep-learning-based methods for automated localization and characterization of plaque composition and the assessment of stenosis as classification in occlusive and non-occlusive 50 % luminal narrowing; these method performances are measured through a multi-faced analysis of the localization and characterization capabilities observed with respect to different plaque composition and at different scales of interest. While fine segment level is necessary for a thorough quantitative evaluation of coronaropathy, a coarse coronary scale analysis suffices to rule out CAD and thus holds significant clinical information. Automated CT coronary analysis relies on the quality of the extracted coronary centerline. A main concern is the quality of the extracted centerline on which the localization methods rely. In order to test the robustness of different methods to a perturbed centerline, the baseline and the proposed method performances are observed at increasingly pronounced random noise offsets of the centerline coordinates. Additionally we propose an experiment in which coronary plaque characterization is used as a downstream task following localization of the plaque of interest. 2.5D methods represent a valid trade-off in terms of quality of detection and processing time for coronary plaque characterization and stenosis assessment [53]. We propose a slight variance of the 2.5D method present in the literature to evaluate the clinical relevance of the proposed circular pooling curved planar reformation (CPR) (CP-CPR) visualization technique.

3.4.1 Baseline

The baseline consists of 3 recently proposed deep-learning based methods. The recursive convolutional neural network (RCNN) [15], TR [52], and the 2.5D method [53], which are discussed in detail in Sections 2.6.2, 2.6.3 2.6.4, respectively. We are able to refine the RCNN and TR design by observing that the 3D-CNN receptive field matches the interleaving space of the patch sampling. We thus avoid redundant computation with a smaller memory footprint. Another improvement consists into the addition of the EOF (End Of Phrase) token, it is a standard natural language processing (NLP) practice to use a EOF token $s' = [s_0, \ldots, s_{T-1}, \text{EOF}] \in \mathbf{R}^{T+1,128}$ which is usually a constant valued

vector.

3.4.2 Ground truth generation

The straightened CPR volume (sec. 2.4) is annotated along the centerline by manually placing a starting and an ending point. These points correspond to the beginning and end of findings present in a clinical report filled by an expert cardiologist, issued during patient examination. Findings localization is roughly described in terms of proximality and distality; the manual annotator refines these notations by delineating the finding on the coronary centerline. The manual annotation task is performed internally by non-clinical GE Healthcare personnel, and a trained cardiologist is then issued the annotated exam for review, which are either included or excluded. As a proof of concept (POC) we built a custom annotation tool for the visualization of the annotation included the CP-CPR view which presents the circular maximum intensity projection below the centerline, and the circular minimum intensity projection above it. This allows the user to perceive at a glance the presence of calcium deposits or low-attenuation plaque without having to rotate the view about the centerline (3.19).

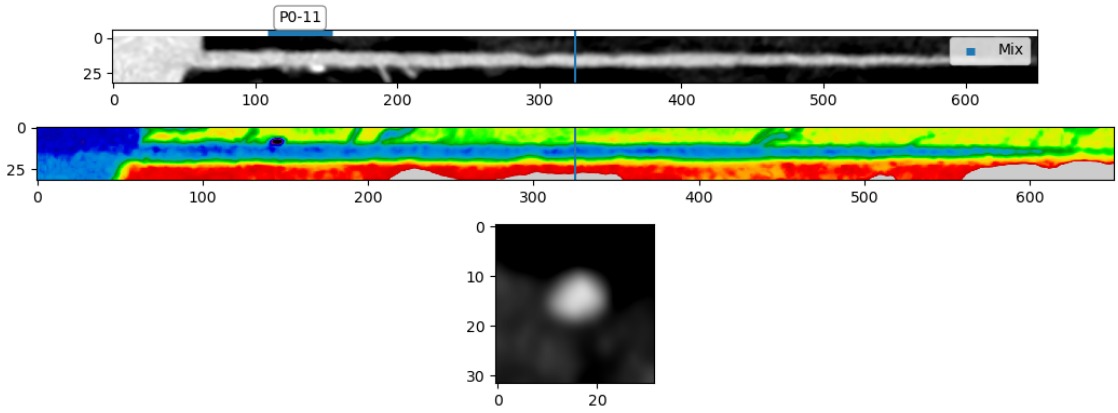


Figure 3.19: Circular pooling CPR (CP-CPR) used as an additional view in a custom-build annotation tool to review coronary plaque annotations, RGB allows to encode the whole HU scale better than gray-scale values, thus results in better contrast reducing the need for windowing. Usually the straightened CPR view needs to be rotated around the centerline for a complete inspection, CP-CPR view allows to localize high intensity and low attentions at glance without rotating the CPR volume about the centerline

3.4.3 CCTA data

We briefly recall the different types of data format employed in the following experiments that can be generated by pairing CCTA volume with the coronary centerline coordinates and measure the performance variability of different methods with respect to data format.

Straightened CPR

Axial, sagittal, and coronal multi planar reformation (MPR) provide insufficient clinical information to diagnose coronary diseases. CPR has proven a practical solution and in particular, straightened CPR has been consistently used as a standard view to diagnose coronaropathies, however, this technique is sensitive to the quality of the extracted coronary centerline, which relies on the accuracy of automated coronary segmentation and tracking algorithms. A series of orthogonal frames, from which voxels are sampled using trilinear interpolation, are computed for each point of the centerline sampled with constant interleaving. This visualization has 3 main parameters; the interleaving step in mm and the 2D orthogonal frame voxel resolution mm^2, the size of the sample grid. The resolution of choice is $0.3\,\mathrm{mm}$ for the interleaving step and $0.3\,\mathrm{mm}^2$ for the voxel resolution and a sampling grid of 28×28 to capture the whole coronary lumen ($5\,\mathrm{mm}$ diameter on average).

Circular Pooling CPR (CP-CPR)

The CPR straightened view can be further processed to obtain a 2.5D visualization by pooling intensities on a circular trajectory (e.g. maximum intensity projection (MIP)). A $L \times H$ image from a $L \times 2H \times 2H$ CPR volume is obtained by means of eq. 3.15 as shown in fig. 3.20. The result is a rotation invariant 2D representation of the vessel (CP-CPR). The morphological features of the vessel are pooled by means of the function used, for example stenosis is defined as a the minimal local narrowing of the coronary vessel, thus minimum and maximum circular pooling coincide in absence of stenosis, however false stenosis due to incorrect centerline estimation is exacerbated. The dimensional reduction comes at the cost of a loss of spatial coherence, most noticeably the displacement of branching (fig. 3.20).

$$\begin{aligned} &\mathrm{CircPool}(f, X)(i, j) = f_\theta X(i, cos(\theta)j, sin(\theta)j)) \\ &i, j \in [0, L] \times [0, H] \\ &X \in \mathbf{R}^{L \times 2H \times 2H} \\ &f \in \{min, max, mean\} \end{aligned} \tag{3.15}$$

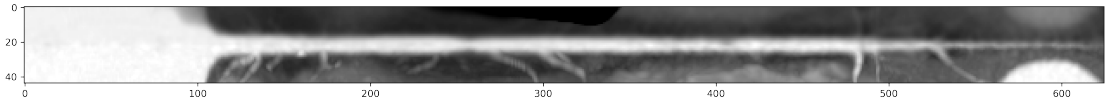


Figure 3.20: An example of circular pooling, top part of the image displays the minimum circular intensity HU and the bottom part the maximum HU. Notice the branching displacement on the bottom part of the image, which is an artifact of the projection.

CCTA sampled patches

CCTA is acquired with an anisotropic voxel with high axial resolution and up to $0.6\,\mathrm{mm}$ slice thickness. In order to sample the coronary centerline uniformly, the coordinates are normalized with respect to the arc length. Patches centered on the centerline interpolated points are sampled using trilinear interpolation for accounting for both centering and anisotropic resolution.

Graph patch

A set of patches can be either sampled from the centerline along a contiguous segment of fixed size or from a sub-graph region of the coronary tree, or graph patch. Given the graph $\mathcal{G} = (V, E), V = v_1, \ldots, v_n, E \in V \times V$ a *k*-order neighborhood of a node v_i is $\mathcal{N}_i = \{v_j \in V | d(v_i, v_j) \leq k\}$ If we were to consider an image as a 2D grid a *k*-hop neighbourhood is what is usually referred to as a patch: we will therefore here refer to a *k*-order neighborhood as a graph patch. In order to sample a graph patch for a given location, the *k*-order neighborhood is computed, and the corresponding locations are then used to sample patches from the volume. In our case, a graph patch is suited to seamlessly model bifurcations and contiguous segments.

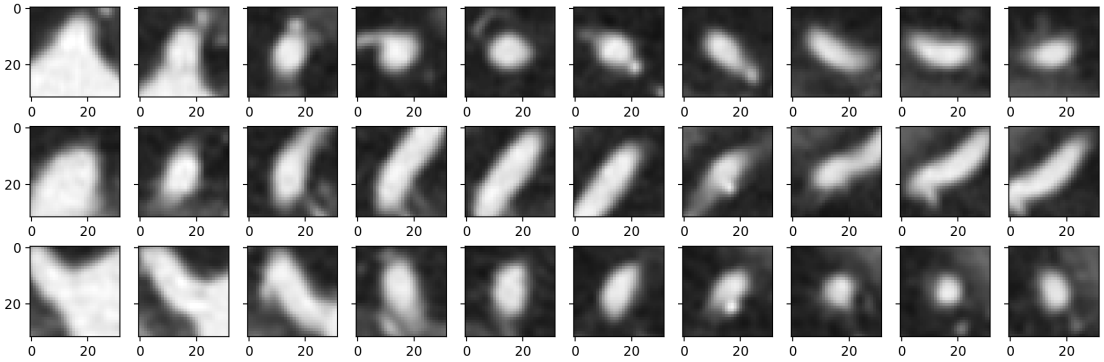


Figure 3.21: Three orthogonal views of a sequence of patches sampled with trilinear interpolation from the coronary resampled centerline.

3.4.4 Proposed method

CNN 2.5D Circular Pooling CPR (2.5D-CP)

In [53], the authors propose to sample orthogonal planes from the straightened CPR volume to predict revascularization and assess stenosis. Although this method requires both the starting and the ending position of the lesion of interest, this approach can be applied to the case of coronary plaque characterization by simply replacing the 3D-CNN with a multi-plane 2.5 D input and a 2D CNN. This approach, however, achieves in practice better results with more than four and up to eight different views. This is probably

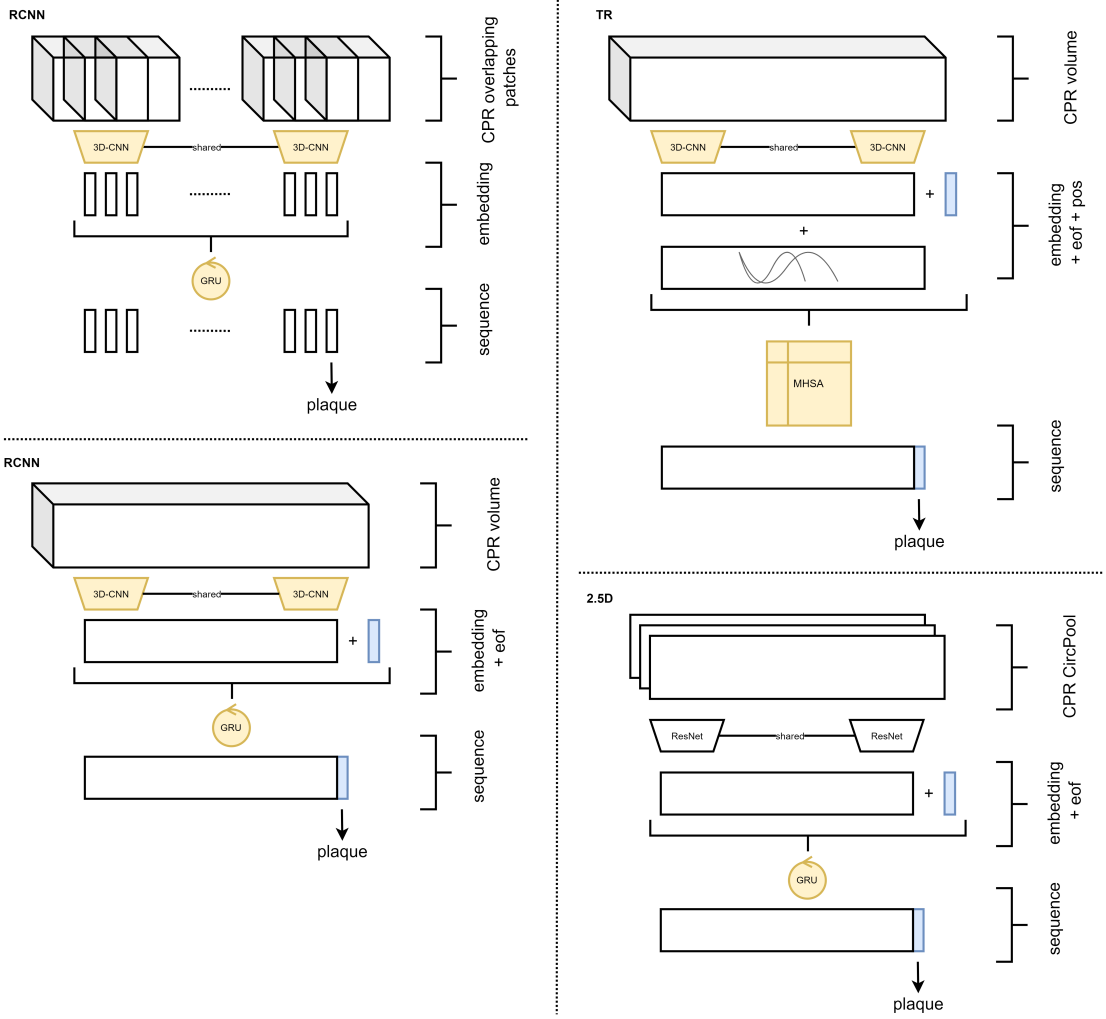


Figure 3.22: The genealogy of the explored architectures. We also improved on the baseline by slightly modifying the implementation using strided convolutions and eof tokens. From top left to bottom left: original RCNN, RCNN with eof, RCNN with straigthened CPR volume. From top right to bottom right: the Multi Head Self Attention (MHSA) with straightened CPR volume, 2.5D input.

due to the fact that the indicator of interest may not be in the sampled plane therefore inducing a bias in training. Furthermore, the concatenation of sampled planes from the straightened CPR volume is not rotation invariant; we propose to use a 2D CNN with the 2.5D approach by exploiting the CP-CPR of an entire coronary segment, which allows data reduction by pooling intensity following circular trajectories around the coronary centerline. This method will be tested for plaque localization and characterization (without and with the starting and ending point of the lesion, respectively).

CNN PW (patch-wise)

The input format consists of patches sampled from the CCTA volume at centerline coordinates. The patches are sampled using an isotropic resolution of $0.35\,\mathrm{mm}$ and a patch size of $25 \times 25 \times 25$. Each patch sampled is individually processed by the same shallow 3D CNN and an FCN in sequence to output a local prediction.

Distance Based CNN (DB-CNN)

We propose to a modified CNN by replacing convolutional kernels with distance-based kernels. The distance-based kernel is designed to be rotational invariant. Because patches are sampled in the system of coordinates of the CCTA volume, the coronary orientation can be different for each patch. This issue is targeted by designing a rotational invariant kernel (figure 3.23).

$$\begin{aligned} &(I * w)[\mathbf{x}] = \sum_{d \in D} w_d \sum_{y \in S_d} I[\mathbf{x} - \mathbf{y}] \\ &S_d = \{\mathbf{y} \in \mathrm{R}^n \mid \|\mathbf{y}\|_2 = d\} \\ &D = [0, 1, \sqrt{2}] \\ &\text{n.b. } \|\mathbf{y}\|_2 = \|\mathbf{R}\mathbf{y}\|_2, \mathbf{R} \in SO(n) \end{aligned} \tag{3.16}$$

By this parameter arrangement, these can be regrouped by associativity and the number of multiplications depends on the parameter d and not on the size of the kernel (Equation 3.16). Once the convolution is defined, the convolutional operator can be embedded in a convolutional layer for learning. This layer operates on 3d patches with fewer parameters.

CNN+ATT

We propose to replace the FCN by an attention mechanism Multi Head Self Attention (MHSA) block to embed the graph patch neighbourhood information (CNN+ATT). 3D patches $28 \times 28 \times 28$ with an isotropic resolution of $0.35\,\mathrm{mm}$ are sampled from the CCTA volume along a previously extracted centerline without straightened CPR (figure 3.25) and arranged in a graph patch. Each cube is individually processed for local feature extraction by a 3D-CNN comprising of 4 convolutional blocks interleaved with max-pooling

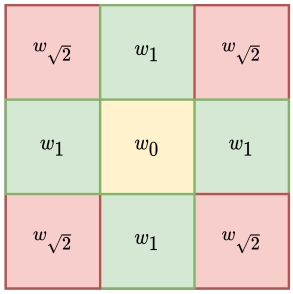


Figure 3.23: A 2d example of distance-based kernel. The parameters are shared among the equidistant positions from the center of the kernel.

to achieve spatial reduction: $28^3 \mapsto 14^3 \mapsto 7^3 \mapsto 3^3$. Each convolutional block has 2 convolutional layers with residual connections and non-linear activation and batch normalization. The last convolutional layers is pooled by a global average pooling layer to obtain a fixed-size feature vector (256 channels). The Self-Attention aggregation block has two MHSA blocks with residual connections in sequence followed by a fully connected layer for classification. The number of layers is tuned experimentally on the evaluation set to be as small as possible without performance loss. Positional context information is given to the model by computing positional embeddings from the graph patch structure of the centerline (fig. 3.25). Using the self-attention technique from [20], the layer operating on a graph patch $\mathcal{N}_i = \{v_j \in V | d(v_i, v_j) \leq k\}$ is defined as for eq. 3.17:

$$
\begin{aligned}
v_i^0 &= v_i + pos_i \\
Q_i^l &= \sigma(W_Q v_i^{l-1}) \\
K_i^l &= \sigma(W_K \mathcal{N}_i^{l-1}) \\
V_i^l &= \sigma(W_V \mathcal{N}_i^{l-1}) \\
v_i^{l+1} &= \text{Softmax}(Q_i^l {K_i^l}^\top / \sqrt{d_k}) V_i^l
\end{aligned}
\tag{3.17}
$$

where W_K, W_Q, W_V are trainable parameters and σ is a non-linear activation function, where we overload the row vector multiplication to operate on sets as the result of $v^{l \geq 1}$ is the same independently of the ordering of $\mathcal{N}_i$.

Positional Embeddings

Positions pos_i are computed locally by exploiting the directed tree structure using a signed hop count from the center node ($+1$ if distal, -1 if proximal); this scalar function is then mapped to vectors of dimension d_k by using sin functions at multiple frequencies [20] to obtain positional embeddings. In order to avoid information for the center node getting diluted in the aggregation step, only the center node value vector V_i is used to compute the output, while the whole neighborhood query Q and key K vector

are used to compute the self-attention matrix. There are many options for positional embeddings in graphs: a simple hop-count suffices for this application as it holds the desirable property of a constant representation for all sorts of configurations, unlike eigenvector decomposition [76] and it is order-invariant, unlike learnable embeddings. For example consider 2 simple graph patches g_0 and g_1 of order $k = 2$; their representation should be held constant for the predecessor nodes (0,1), however using a non-linear transform like spectral embeddings does not hold the desired property (Figure 3.24).

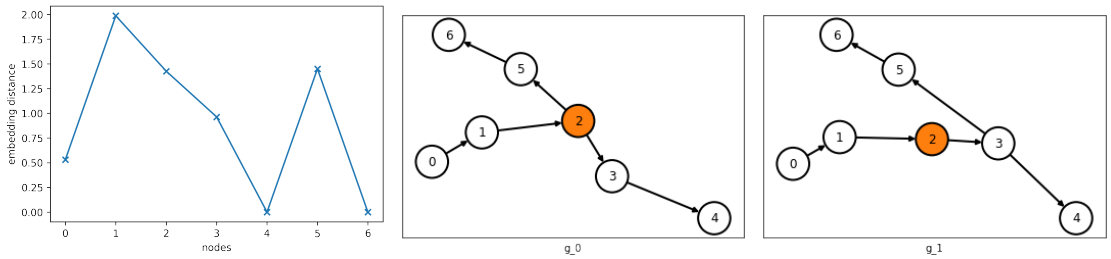


Figure 3.24: For two examples of graph patches g_0, g_1 the embedding distance is computed as $D_{\text{emb}}(g_0, g_1, v) = \sum_{f=0}^{F}(T(L(g_0))[v, f] - T(L(g_1))[v, f])^2, v \in [0, \ldots, n-1]$ where $L(g) : (N \times E) \mapsto \mathbf{R}^{n \times n}, n = \#N$ of the graph g, and T is the embedding transformation $T : \mathbf{R}^{n \times n} \mapsto \mathbf{R}^{n \times F}$.

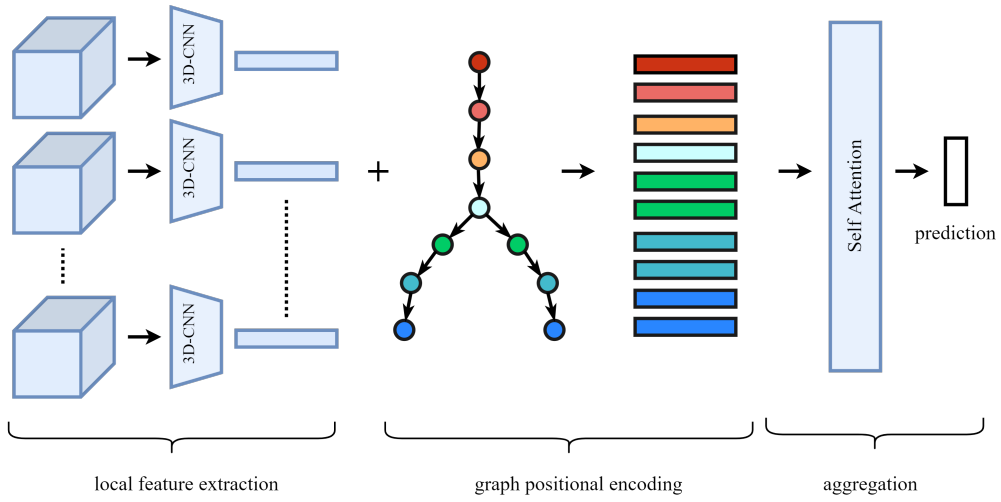


Figure 3.25: 3D patches are extracted from the CCTA volume by sampling points from the centerline; for each patch, a 3D-CNN computes a local features vector, and the graph distance is used. The centerline is extracted as a preprocessing step.

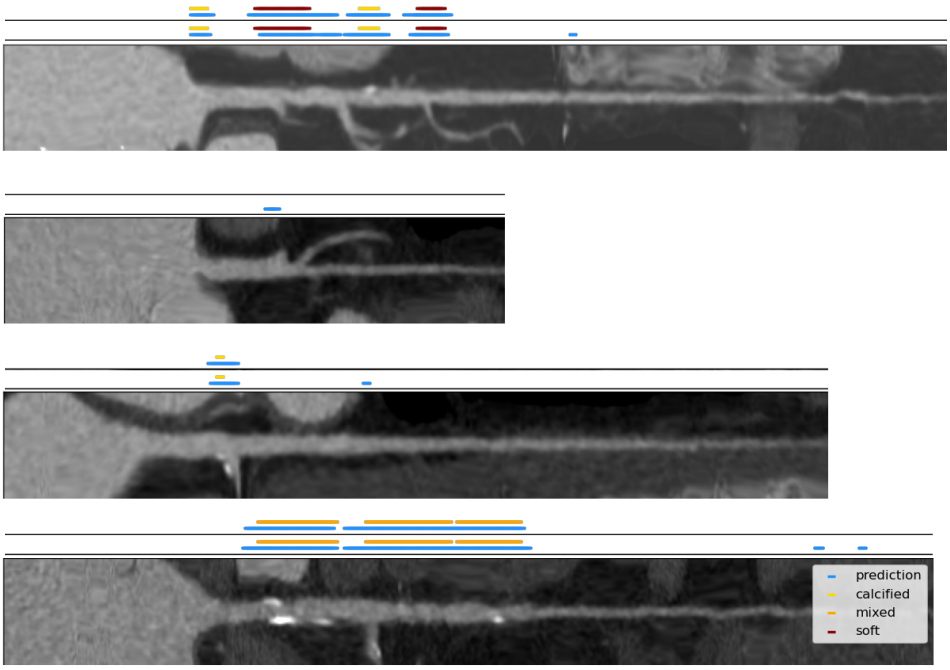


Figure 3.26: A sample of the inference result displayed along the coronary with CPR reconstruction. For each picture, the top and bottom bars represent the ground truth prediction (calcified, mixed, and soft) overlapped with the prediction of CNN+ATT and standard CNN, respectively. These examples show the False Positive reduction provided by CNN+ATT. From the top row down, the first example shows False Positives in correspondence of the LMA bifurcation and in low contrast distal portions. The same for the following examples with False positives in correspondence of bifurcations and low contrast distalities. This example is in agreement with the specificity metrics of the CNN method and CNN+ATT method shown in table 3.4.5.

3.4.5 Experiments

Testing Segment level plaque localization

The test set is composed of ~ 500 manually annotated segments. In order to avoid the bias of manually selected segments, each segment is of fixed length ($30\,\mathrm{mm}$) and can contain either one or multiple labels. In order to simplify the comparison, each model is tested on the binary task of plaque localization; this also avoids the issue of inter-observer variance in the assessment of plaque characterization. The localization is then observed with respect to all plaque classes (Table 3.4.5). A ROC analysis of the CNN+ATT proposed method is then performed in isolation against the baseline (RCNN [15], TR [52]) (Figure 3.27).

Method	Archi	Data	No-Plaque				Calc			
			sens.	spec.	prec.	f1	sens.	spec.	prec.	f1
Baseline	RCNN [15]	orig.	0.81	0.88	0.81	0.84	0.87	0.97	0.81	0.84
Baseline	RCNN [15]	ours	0.83	0.72	0.69	0.75	0.83	0.94	0.91	0.86
Baseline	TR [52]	ours	0.86	0.52	0.59	0.70	0.86	**0.97**	**0.96**	**0.88**
Proposed	2.5D-CP	ours	0.89	0.60	0.77	0.68	0.89	0.60	0.77	0.68
Proposed	CNN	ours	**0.93**	0.70	0.79	0.74	**0.93**	0.87	0.62	0.72
Proposed	dbCNN	ours	0.90	0.67	0.61	0.63	0.90	0.82	0.59	0.68
Proposed	CNN+ATT	ours	0.88	**0.83**	**0.79**	**0.83**	0.88	0.82	0.79	0.86
			Mix				No-Calc			
			sens.	spec.	prec.	f1	sens.	spec.	prec.	f1
Baseline	RCNN [15]	orig.	0.81	0.92	0.84	0.85	0.32	0.88	0.25	0.28
Baseline	RCNN [15]	ours	0.83	**0.96**	0.91	0.86	0.83	0.65	0.57	0.61
Baseline	TR [52]	ours	0.86	0.94	**0.92**	**0.91**	0.86	0.58	0.48	0.52
Proposed	2.5D-CP	ours	0.89	0.83	0.64	0.72	0.89	0.36	0.42	0.39
Proposed	CNN	ours	**0.93**	0.79	0.63	0.71	**0.93**	0.45	0.44	0.45
Proposed	dbCNN	ours	0.90	0.65	0.58	0.51	0.90	0.40	0.33	0.36
Proposed	CNN+ATT	ours	0.88	0.83	0.79	0.85	0.88	**0.70**	**0.66**	**0.75**

Table 3.6: Segment level plaque localization.

Testing Coronary level plaque characterization

The CNN+ATT method is tested in isolation against the baseline for plaque characterization at the coronary level. Each coronary is assigned a label based on the highest risk factor associated with the plaque characterization (Soft, Mixed, Calcified, Healthy). Each coronary is split into segments, and the prediction associated with the highest risk factor is assigned to the whole coronary and tested against its label. As most coronaries have plaques, only the correct label assigned is considered, and performance is measured with the C-Index or concordance score and confusion matrix (Tab. 3.4.5).

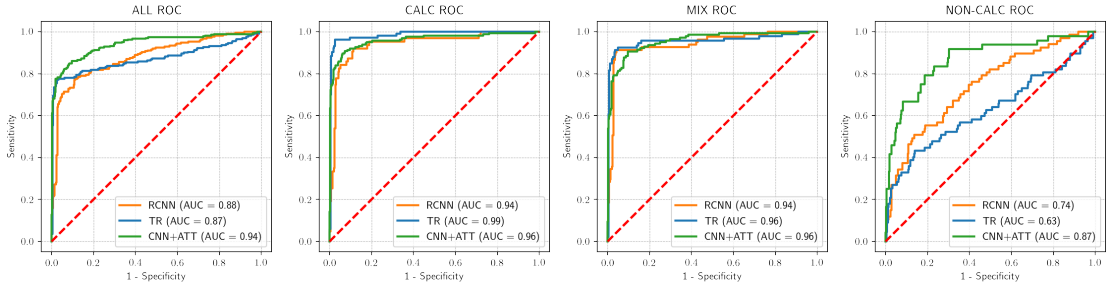


Figure 3.27: Segment level ROC curves: localization ROC analysis is stratified with respect to the coronary plaque composition. All models are comparable on the calcified-plaque and mixed-plaque class, while on the class soft-plaque the proposed method shows better performances.

Coronary Level	PREDICTED			
RCNN [15]	Healthy	Calc	Mix	No-Calc
Healthy	59	5	3	6
Calc	8	4	12	10
Mix	20	6	29	12
No-Calc	15	3	11	14
C-Index	0.700			
TR [52]	Healthy	Calc	Mix	No-Calc
Healthy	**60**	2	1	10
Calc	3	10	14	7
Mix	21	13	23	10
No-Calc	20	10	13	10
C-Index	0.697			
CNN+ATT	Healthy	Calc	Mix	No-Calc
Healthy	52	2	9	10
Calc	3	**11**	10	10
Mix	6	5	**50**	6
No-Calc	12	3	10	**18**
C-Index	**0.720**			

Table 3.7: Coronary level plaque characterization.

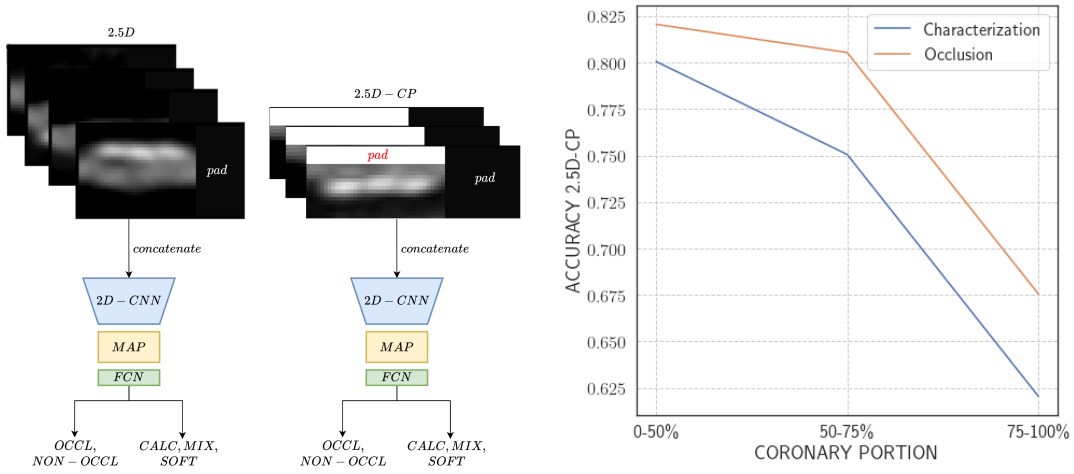
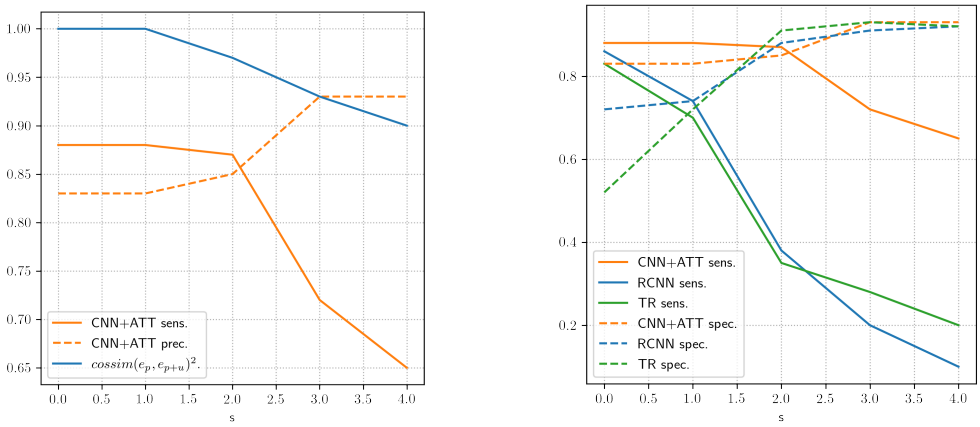


Figure 3.28: (a) 2.5D and 2.5D-CP data format and architecture. The same CNN takes a different input to predict plaque characterization and stenosis degree. (b) Accuracy as a function of coronary portion. The more distal lesions are included, the lower the accuracy on both characterization and occlusion detection. The accuracy measure is for the 2.5D-CP model on the entire test set (75-100%) for only the lesions that occur in the coronary portion up to 3/4 of the total length (50-75%) and finally for half of the coronary (0-50%).



(a) (b)

Figure 3.29: (a) On the x-axis, the centerline perturbation term s in voxels. The y-axis is shared by sensitivity, specificity, and a local features e alignment metric, which together represent the consistency of intermediate and output representation of the inputs as the centerline is disturbed around its original location. (b) The comparison among different models with different input data format. CNN+ATT take isotropic patches arranged on a graph patch, RCNN and TR takes straightened CPR segments. The perturbation on the centerline propagates to the CPR reconstruction which propagates through the network at detriment to the localization capabilities.

Centerline robustness

One of the major concerns about this method is the robustness to the centerline extraction quality. In order to test how results are affected by centerline quality, our method underwent evaluation using increasingly noisy centerline coordinates to ensure that the content perceived by the model remains unchanged under random shifts. The consistency of intermediate representation local feature extracted by the 3D-CNN is measured using cosine alignment metrics. Each coordinate point p around the centerline is perturbed by a uniform noise $u \sim U(-s, s)^3$. Embeddings e_p consistency is measured using the cosine similarity as noise increases cossim(e_p, e_{p+u}) (Figure 3.29). The results suggest that the method is consistent up to a threshold of $2.25\,\mathrm{mm}$ shift from the original ideal centerline; this result is reasonable considering that the average coronary radius is about $2.5\,\mathrm{mm}$.

2.5D-CP plaque characterization

Plaque characterization is a downstream task of localization. In [53] the author proposes to characterize plaques from the identified starting and ending position with the use of 2D orthogonal views sampled from the CPR straightened volume. We propose to repeat this experiment and compare the extraction of the 2.5D views with the CPR

straightened view. Therefore, we observe only how the model performances differ when the input is switched. The 2D CNN architecture is fixed and can therefore be arbitrary. We chose a shallow 2D CNN architecture composed of 3 convolutional blocks: Conv ∘Relu ∘MaxPool. Instead of padding or stretching the input, the final layer consists of a Maximum Adaptive Pooling MAP layer outputting a feature vector $z \in \mathbb{R}^{128}$ which is fed to a fully connected network (FCN) to output the classification vector $y \in \mathbb{R}^5$. This experiment is designed to measure the characterization capability of 2D CNN.

Training Each model takes a segment of fixed length (30 mm) and outputs a multi-class label as input. The dataset is highly unbalanced; therefore, sampling is biased towards the minority classes in order to obtain a uniform distribution of plaque labels. The alternating procedure proposed by [15] in which batches are alternating did not improve the characterization task. Each method is trained on the same dataset of coronary lesions but without the healthy segments; therefore, the model predicts three classes: calcified, mixed, or soft. The input lesion is stretched to the mean lesion length, so the input is of fixed length. However, this method makes the view non-isotropic; the view changes as the length changes; therefore, the method is not scale-invariant (the same lesions at different lengths are processed differently). In order to account for this issue, the views are padded to a maximum size, and the maximum activation along the length dimension is taken for classification.

Testing lesion level plaque characterization Plaque characterization can be seen as a downstream task of plaque detection. Thus once a plaque's starting and ending point are localized a plaque can be further discriminated into sub-classes for characterization. In our case, there are three main classes: calcified, mixed, and soft plaques. For this task, we propose to compare two methods that rely on the CPR volume and, in particular, on the extraction of 2.5D views. The models are evaluated using a multi-class confusion matrix; each entry corresponds to a lesion. Lesion classification accuracy is also measured with respect to distality, which correlates with coronary diameter. The model is tested on the set of lesions present in the test set: 49 soft, 78 mixed, 97 calcified plaques, and 182 non-occlusive, 42 occlusive stenosis.

3.4.6 Discussion

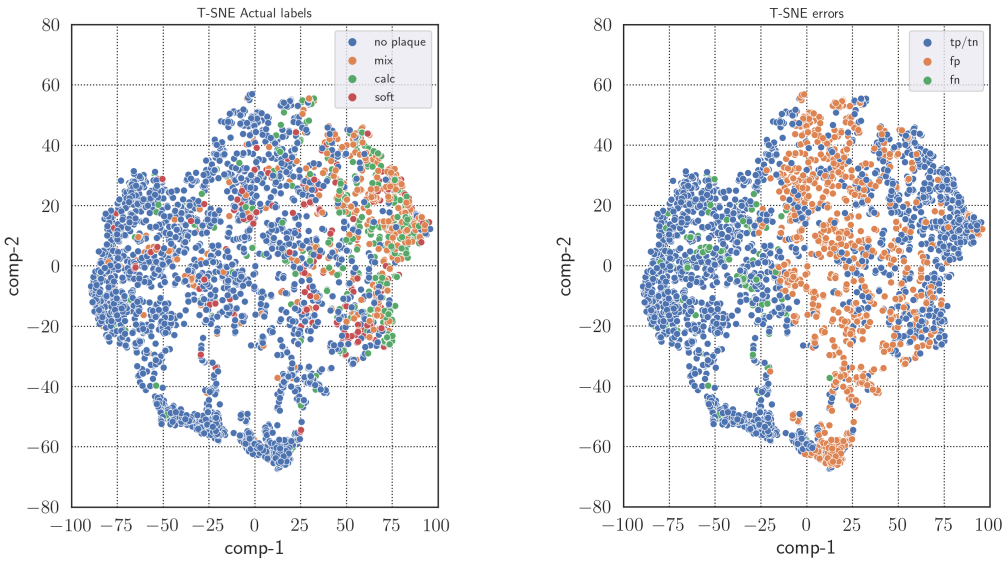
In this section, we presented a set of plaque localization and characterization methods given an input CCTA image paired with coronary centerline coordinates. Coronary plaques can be subdivided into four main classes: healthy (no-plaque), calcified, mixed, and soft. Both localization and characterization have entangled features; the localization of a soft plaque necessitate the knowledge of its characteristics; localization is a binary classification task that is simpler to analyze than a multi-faced classification problem. A model that performs well in localization will necessarily perform as suitable for characterization, while the converse is not true. As shown in previous section, characterization is achievable if the localization of a plaque is given. The main issue is, therefore,

Lesion Level	Predicted				
2.5D[53]	CALC	MIX	SOFT	OCCL	NON-OCCL
CALC	**85**	23	4		
MIX	12	**38**	19		
SOFT	0	18	**26**		
OCCL				30	80
NON-OCCL				12	102

Lesion Level	Predicted				
2.5D-CP	CALC	MIX	SOFT	OCCL	NON-OCCL
CALC	82	24	7		
MIX	15	35	20		
SOFT	0	19	22		
OCCL				**41**	71
NON-OCCL				2	**111**

Table 3.8: Plaque characterization and stenosis degree assessment at the lesion-level. A confusion matrix comparison between two methods (2.5D against 2.5D-CP).

the variability of the healthy (no-plaque) class, which can be indicated by many factors: missing annotations, noise, and anatomical variability. Plaque characterization can thus be simplified to reduce false positives while keeping true negatives constant. As shown in figure 3.26 the CNN+ATT method reduces false positives concerning a simple CNN local method. The localization capability measurements are summarized in Table 3.4.5. The proposed CNN+ATT method shows better performances across the No-plaque and the Non-Calcified (Soft) plaques which are the hardest to identify. When other methods present higher specificity, the CNN+ATT shows higher sensitivity. For example, for mixed plaques the RCNN has a sensitivity of 0.83 and a specificity of 0.96 against the CNN+ATT with 0.88 and 0.83 for sensitivity and specificity, respectively. We hypothesize that this issue relates to the binary decision boundary being biased towards negatives. A ROC analysis corroborates this observation as the AUC is higher for Mixed plaques in isolation: 0.96 and 0.94 for CNN+ATT and RCNN, respectively (fig 3.27). The same does not hold for the other methods when observing the metrics on the non-calcified and overall plaque detection. Both the ROC curve and the metrics agree for both classes in isolation. The coronary level analysis (tab. 3.4.5) shows that the CNN+ATT has better discriminative capabilities than the baseline. By correctly identifying patients with soft plaques, the corresponding risk can be better assessed, and a bias towards negatives and non-calcified plaques provides more informative clinical support overall. As an additional experiment, we measured the potential clinical use of the CP-CPR for coronary plaque characterization and stenosis assessment (Table 3.8). While CPR orthogonal views seem better suited for plaque characterization, the CP-CPR holds significant information for stenosis assessment. The major issue, however, remains in false-negative detection. Complex cases lie between healthy and soft plaque (Figure 3.30). The inter-operator variance in the annotation might also have been an issue as multiple annotators with different skills have been employed. Hard/corner cases can be identified using the T-



(a) (b)

Figure 3.30: (a) T-SNE on deep-features computed by the CNN encoder highlighted by true positive and true negatives, false positives and false nagatives. (b) false negatives count with respect to class.

SNE analysis of deep feature map outputs. The figure suggests that false positives are distributed almost evenly amongst correctly predicted segments.

3.5 Coronary Segmentation

The severity of the plaque build-up, also known as plaque burden, necessitates a multi-faceted evaluation of coronary artery plaque content, volume, distribution, and lumen stenosis. To this purpose the segmentation of the coronary and the atherosclerotic plaque have fundamental clinical applications. Coronary artery segmentation can be mono-phase (the whole coronary) or bi-phase (inner lumen and outer wall) thus comprising the plaque-burden, unlike lumen coronary segmentation, which only focuses on delineating the high-intensity lumen at the inner wall border from the low-intensity wall tissue and surroundings, the segmentation of coronary artery wall and plaques from CCTA images is substantially more challenging. There are various approaches to coronary segmentation. In this section we explore an approach that necessitates the tracking of the vessel to be performed as a pre-processing step. This approach consists of modeling the local appearance of a bright tubular structure. Frameworks for bi-phase coronary segmentation often employ multiple modules from region growing phases, to vesselness mesaurements, the level set method [77] and mesh operations for visual refinement. We propose to analyse a simple level set method for preliminary vessel segmentation and propose a numerical scheme for fast narrow band updates on volumetric data.

3.5.1 Proposed method

Masked gradient descent

The Chan-Vese [78] method is an iterative algorithm that proposes the following update function of a level set (equation 3.18).

$$\phi_t = H'_\epsilon(\phi)\{div(\frac{\nabla\phi}{|\nabla\phi|}) - \lambda((c_1 - f)^2 - (c_2 - f)^2)\} \tag{3.18}$$

Here, the function H is the Heaviside function. In practice a smooth Lipschitz approximation is often used such that $H_\epsilon(x) \to H(x)$ for $\epsilon \to 0$ (eq. 3.19).

$$\begin{aligned} H_\epsilon(x) &= \frac{1}{2}(1 + \frac{2}{\pi}\text{arctan}(\frac{x}{\epsilon})) \\ H'_\epsilon(x) &= \frac{1}{\pi}\frac{\epsilon}{\epsilon^2 + x^2} \end{aligned} \tag{3.19}$$

The level set function ϕ like the Eikonal function of the phase function $\Omega \subset D$ has a non null gradient by design and it is zero in $\partial\Omega$. Therefore $div(\frac{\nabla\phi}{|\nabla\phi|})$ is always defined. In [79] the level set function is restricted to $u \in BV[0, 1]$ therefore the segmentation can be formulated as a constrained convex optimization problem (Equation 3.20).

$$\begin{aligned}&\min_{0\leq u\leq 1}\int_D |\nabla u(x)| + \lambda\int_D \{(c_1 - f(x))^2 - (c_2 - f(x))^2\}u(x)dx = \\ &\min_u \int_D |\nabla u(x)| + \lambda\int_D \{(c_1 - f(x))^2 - (c_2 - f(x))^2\}u(x) + \alpha\nu(u)dx\end{aligned} \tag{3.20}$$

The first constrained problem has the same set of solutions of the second problem where $\nu(\xi) = max(0, 2|\xi - \frac{1}{2}| - 1), \alpha \geq \frac{\lambda}{2}||s(x)||_{L^\infty}$. The update function is therefore expressed in equation 3.21.

$$u_t = div(\frac{\nabla u}{|\nabla u|}) - \lambda((c_1 - f))^2 - (c_2 - f)^2) - \alpha\nu'(u) \tag{3.21}$$

We can observe that the update formula is not defined over the whole spatial domain $\{x \mid \nabla u(x) \neq 0\} \subset D$. To circumvent this issue the minimized energy is reformulated with a smoothing constant ϵ (equation 3.22).

$$\min_{0\leq u\leq 1}\int_D |\nabla u(x) + \epsilon| + \lambda\int_D \{(c_1 - f(x))^2 - (c_2 - f(x))^2\}u(x)dx \tag{3.22}$$

We can observe that for a function $u \in BV[0, 1]$, regions with $\nabla u = 0$ are flat constant regions (either 0 or 1). For any continuous function $f(a) = 0, f(a + \epsilon) = 1 \implies \nabla f(a) \neq 0 \iff \epsilon \to 0$. We can therefore apply a constraint on the gradient descent by masking a set of variables (Equation 3.23).

$$\begin{aligned}u_t(x) = div(\frac{\nabla u(x)}{|\nabla u(x)|}) - \lambda((c_1 - f(x)))^2 - (c_2 - f(x))^2) \\ x \in \{x \mid \nabla u(x) \neq 0\}\end{aligned} \tag{3.23}$$

The set $\{x \mid \{H'_\epsilon(\phi(x)) > 0\} = \{x \mid \nabla u(x) \neq 0\}$ for $\epsilon \to 0$. This implies that the local maxima found by the previous method are a subset of the global maximum solution. Furthermore we can observe that if we were to choose any u_0 as initialization: the connected component of the global solution which is connected to u_0 is the local minimum found by means of (Equation 3.23).

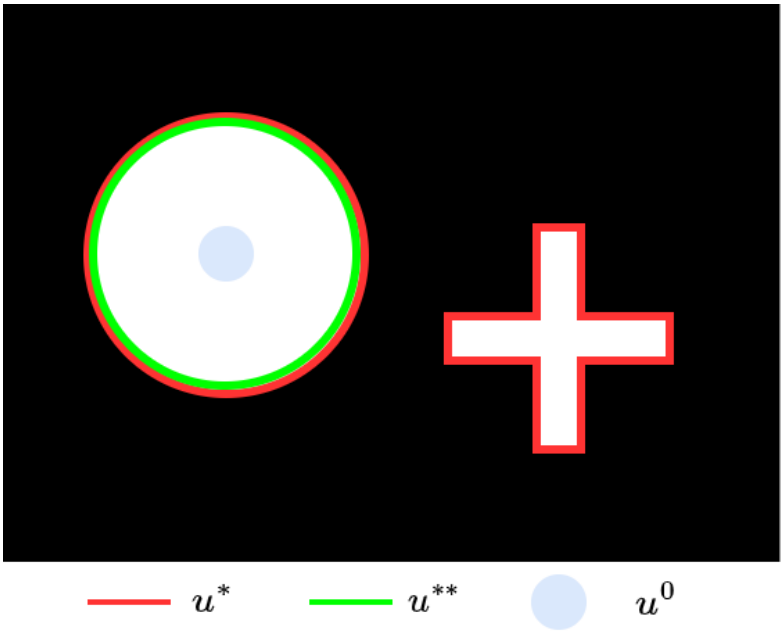


Figure 3.31: the set $\{x|u^{**}(x) = 1\} \subset \{x|u^*(x) = 1\}$, the local minimum u^{**} is connected to the initialization u^0. This can be seen as a limitation or an advantage depending on the application.

An even simpler method is to use the MBO model [80] with the heat equation as regularizer which has a simpler centred discretization by means of the Laplacian kernel (Equation 3.24).

$$\begin{aligned} u_t(x) = \Delta u(x) - \lambda((c_1 - f(x)))^2 - (c_2 - f(x))^2) \\ x \in \{x \mid \Delta u(x) \neq 0\} \end{aligned} \tag{3.24}$$

The conditions $\{x \mid \nabla u(x) \neq 0\} \subset \{x \mid \Delta u(x) \neq 0\}$ and $\{x \mid \Delta u(x) \neq 0\} - \{x \mid \nabla u(x) \neq 0\}$ are the non-simple points of $\{x \mid u(x) > 0\}\}$ which do not occur in our application. Narrow band updates is a well known strategy to optimize the computation of the level set method by front evolution [81, 82]. The narrow band here is presented as a masked iterative scheme in which only a part of the variables are updated with no need to re-compute the level set function for numerical stability. A distantly related problem is the optimization of highly non convex function involving neural networks, it is good practice to stop early the iterative optimization in order to avoid over-fitting, we can draw similarities with this approach: although in theory the solution is sub-optimal in practice it allows to segment the connected component that intersects the initial u^0 this is particularly suited for coronary segmentation when the initial segmentation is the centerline.

Local Level set Method

The proposed level set method exploits the tracking center-line in 2 ways. First the voxelized center-line is used as initial segmentation for the method, secondly small patches are sampled around the center-line thus allowing to re-sample at a subvoxelic resolution. Another advantage is that the level set method can run on a sparse set of small patches rather than the whole volume. Additionally coronary plaque detection methods developed in this work can be used as guide to parameterize the segmentation. When calcified, mixed, or soft plaques are detected, the patches extracted at the location where the plaque has been detected can be segmented using a better suited strategy.

3.5.2 Experiments and Discussion

There is no reason to believe that local minima are more relevant to this image processing task [83]. Furthermore the narrow band computed by the masked gradient descent is computationally convenient, especially for sparse segmentation in general and tubular 3D structure in particular. For coronary segmentation a classic strategy is to initialize the segmentation with the extracted center-line of the tracking (manual or automatic).

We propose a qualitative analysis of this method as there is no reference manual segmentation available for the measurement of the necessary metrics to carry out a quantitative analysis. What can be observed is that the convex formulation with the constrained binary function is not suited for the problem at hand. Changes in topology are in general a desirable property of a segmentation algorithm, in our case the centerline and the final segmentation belong to the same homology class (objects with the same topology). Level set methods can be seen as a Bayesian model in which the prior term is total variation, the methodology based on the variational model (solution of the Euler-Lagrange equation) is however considered obsolete and inefficient. More recent methods based on dual formulation of the minimization problem and proximal operators [84] constitute a more modern approach which however do not incorporate the topological constraint of connectedness to the initialization. What would have been a major addition to this work is the comparison with directional total variation minimization for the segmentation of tubular structures proposed in [85, 86], but the lack of ground truth coronary segmentation for quantitative comparison remains as an issue. However, due to time constraints, we could not explore this hypothesis maybe to be explored in future works.

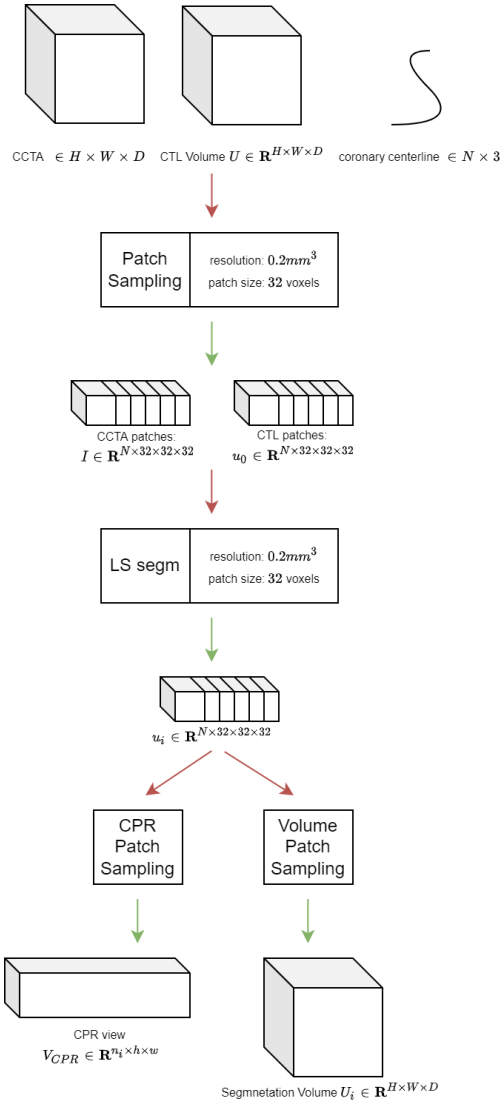


Figure 3.32: Pipeline of the segmentation algorithm: patches are sampled from the CT volume at the coronary centerline locations, from these local statistics are computed from the values at the centerline, the level set is executed in parallel on each patch with a fixed number of iterations (10), the whole segmenation is obtained by averaging all local patch segmentations.

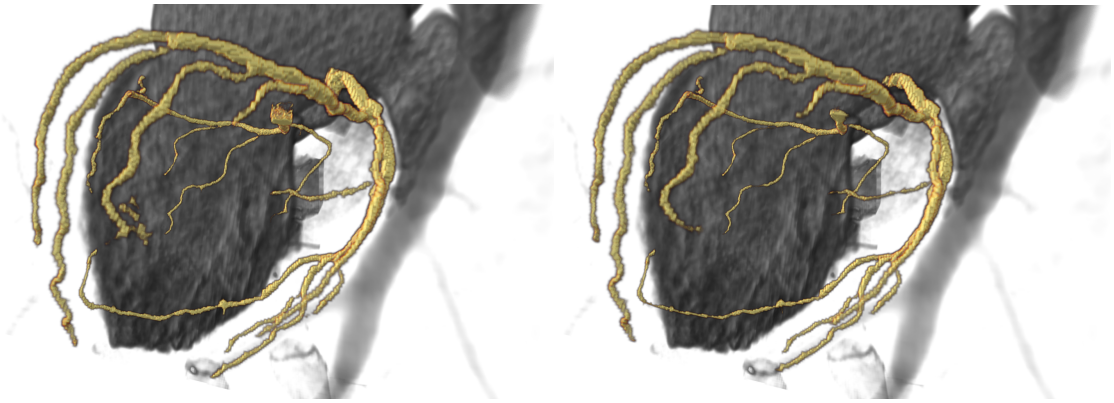


Figure 3.33: Left: Level set segmentation according to the convex formulation (equation 3.21), Right: Level set segmentation according to the narrow band formulation (equation 3.24).

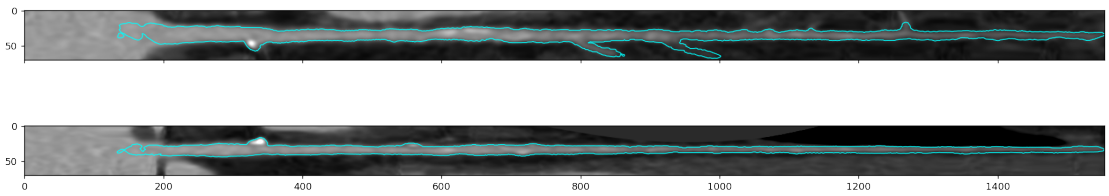


Figure 3.34: Level set segmentation according to the narrow band formulation (equation 3.24).

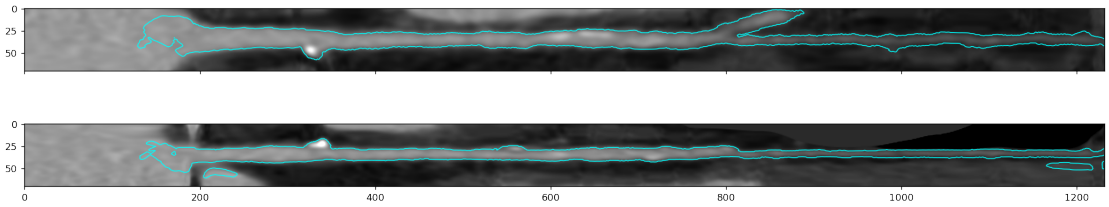


Figure 3.35: Level set segmentation according to the convex formulation (equation 3.21).

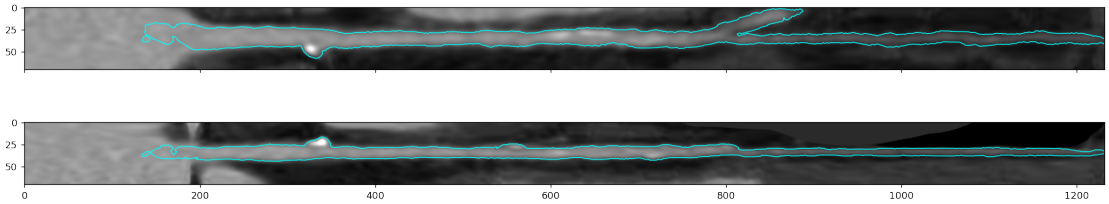


Figure 3.36: Level set segmentation according to the narrow band formulation (equation 3.24).

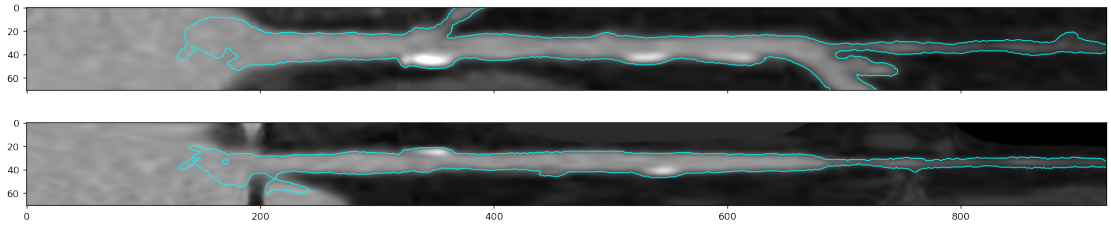


Figure 3.37: Level set segmentation according to the convex formulation (equation 3.21).

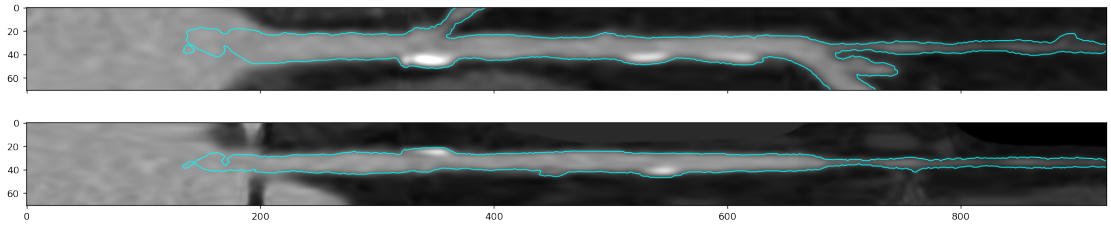


Figure 3.38: Level set segmentation according to the narrow band formulation (equation 3.24).

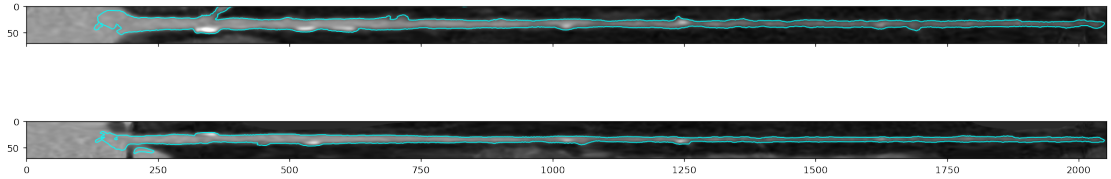


Figure 3.39: Level set segmentation according to the convex formulation (equation 3.21).

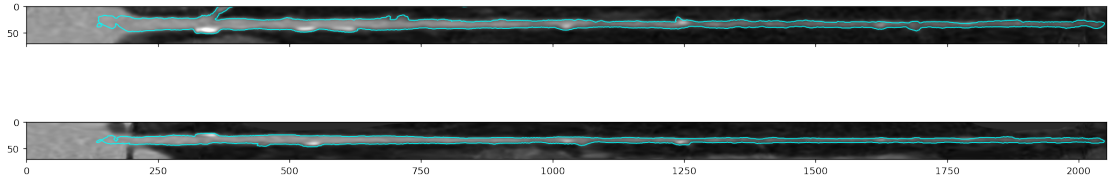


Figure 3.40: Level set segmentation according to the narrow band formulation (equation 3.24).

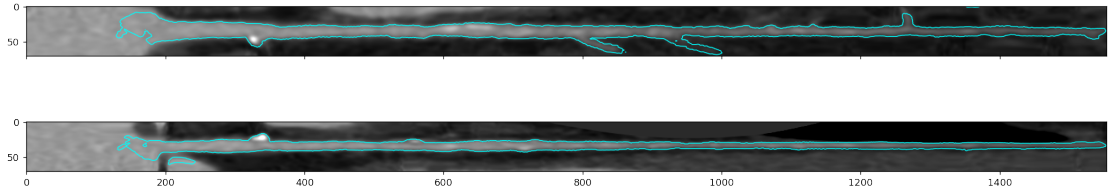


Figure 3.41: Level set segmentation according to the convex formulation (equation 3.21).

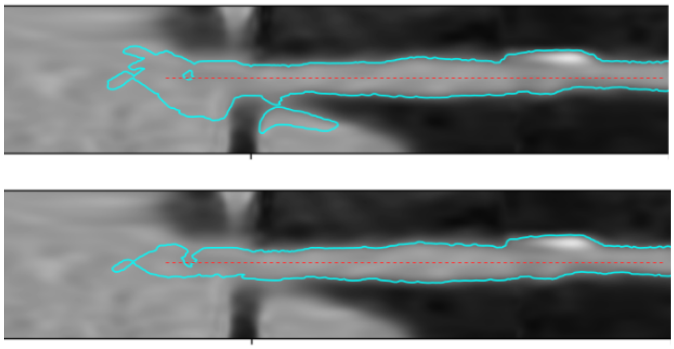


Figure 3.42: An example of how the convex method global optimum is not as relevant as the local optimum found with the proposed descent method. Top: 10 iterations of the global method initialized at the center line. Bottom: 10 iterations of the proposed method initialized at the centerline. A portion of another injected structure is visible and it is integrated in the segmentation while in the proposed method only the correct coronary tubular structure is segmented.

Chapter 4

Final remarks

4.1 Conclusion

CCTA has the potential to become the imaging modality of choice for automated screening of CAD. Its biggest limitation is the presence of artifacts noise that impedes precision and thus reliability of the clinical inference. As the automated diagnosis tools are limited by spatial and temporal resolution constructors are improving CT resolution (Photon Counting Computed Tomography (PCCT)) opening new possible clinical applications. Any performance can be increased by adding more data and the bugbear of catastrophic forgetting can be solved by using bigger models, thus the achievement of expert-level automated diagnosis is only limited by today's computational limits and by the lack of expert annotation. As the application of deep learning replaces clinical tasks it is important that a structured dialogue is established between clinicians and computer scientists. Medicine is an empirical science and as such lacks absolute agreement among its community members. An uncertainty which is reflected by a high inter and intra observer variance in the annotations which can be only mitigated by averaging collective observations at the cost of a major logistic bottleneck. On the other end of the spectrum there are performance measurements, which are provided to clinicians to evaluate an automated diagnostic tool. It is the responsibility of the computer scientist to have a deeper understanding of the clinical requirements in order to provide tailored measurements and the most fitting performance metrics for a fair evaluation and a meaningful feedback for the clinicians. There's no blueprint for such structured dialogue because of the high combinatorial of clinical requirements and possible automated system. Initiatives like the SFR challenge and MICCAI's grand challenges aims to solve precisely that by providing solid foundation for clinicians and computer scientists to agree upon. However we could observe a lack of structure when dealing with coronary CCTA imaging related clinical tasks. In most cases the amount and quality of the data provided are too scarce to fit the strict requirements provided by the clinical evaluation frameworks. In the current data-driven historical period CCTA data for coronary analysis seems to be one of the few exception as it is becoming scarcer: sadly enough one of the major contributor to the CCTA coronary segmentation community [45] is no

longer maintained. The scope of this thesis was to explore how deep learning can be beneficial to the whole pipeline of CCTA coronary analysis. While performances are not still satisfying, we can confidently state that with more focus on providing high quality annotation and by defining the scope of the application (segmentation, detection, classes for classification) deep learning based applications could become an effective diagnostic tool to support clinicians evaluate CAD. Another major issue which has not been discussed is the scalability of neural networks. Industrial solutions existing today are mostly oriented towards 2D image processing which present a much lower dimensionality with respect to CT imagery. To reach human level (and beyond) performances, models with tens of millions of parameters are put into production and are constantly retrained with new data (continuous learning). The computational resources needed to train these model are the mono-pole of a few industrial actors, thus medical imaging although being promising might not still be ready for easily-available real-world diagnosis. The middle ground is limiting the application to a set of known cases in which the anatomical variability is known and well identifiable (e.g. calcium score or bony structure segmentation). The issue of balancing learning segmentation of tubular structures regardless of their size and or section thus ensuring connectivity have been addressed by formulating a topological interpretation of the skeletonization operator gradient. This issue is shared among all medical imaging modalities and anatomies: for example liver vessels segmentation. In addition the use of weak supervision by means of the sole centereline is a major practical advantage for all tubular structure segmentation issues and as a byproduct a whole segmentation can be also obtained. Another unexplored field which may be beneficial to the whole diagnosis pipeline is uncertainty measuring. Because of the high intra-variability of plaque visual features these can be detected with varying confidence. This could be addressed by stacking multiple prediction only for those instances falling into a low confidence zone. Such low-confidence zone can be identified by using geometric projection on low dimensional spaces (embeddings) using deep features.

4.2 Resume

This work proposes to explore the vast and multi-faced topic of automated prediction of Major Adverse Cardiovascular Events (MACE). An automated imaging workflow based on deep-learning techniques has been developed by focusing on the clinical tasks linked to identifying anatomical indicators of Coronary Artery Diseases (CAD). Patients accepted at ER with CAD suspicions, symptomatic and asymptomatic, undergo examinations that can be more or less invasive. One of the objectives of this work is to exploit imaging examination solutions that are minimally invasive, thus minimizing the stress on both the patient and the clinical site. Calcium Score Computed Tomography (CSCT) is a non-invasive low-dose exploratory examination aiming to quantify early CAD indicators of Coronary Artery Calcium CAC burden based on the computation of the Agatston Score (AS). In this work, we propose a robust and efficient approach to quantifying the CAC burden based on semantic segmentation provided by an ensemble of U-net neural net-

work architectures. We could verify that a minimum resolution of 2.5 mm isotropic CT scan is required for the accurate computation of AS score, allowing fast inference and thus little computational requirements for clinical site workstations. Coronary Computer Tomography Angiography (CCTA) is the sole minimally invasive imaging modality that allows the visualization of morphological features of coronary arteries. The injection of iodine solution in the bloodstream highlights coronary arteries, thus allowing for the identification of anatomical indicators of CAD. We propose a coronary analysis workflow that first extracts the coronary centerline (coronary tracking), thus enabling subsequent tasks linked to identifying coronary lesions (plaques and stenosis). To this end, we developed a tailored loss function that can learn from the annotated centerline alone to identify tubular structures and a recursive tracking strategy to compute the entire coronary vessel tree. The mapping or labeling of the coronary vessel tree holds important clinical information as the knowledge of the main coronary branches guides the expert investigation of coronaropathy. In this work, we show that a learning approach based on Graph Neural Networks (GNN) supplemented with a tree structure apriori can automatically identify the coronary branches of the vessel tree even when the tree is incomplete or highly irregular due to anatomical variations that may occur in the patient population. The vessel tree paired with the CCTA volume is then used for coronary analysis. We propose a transformer neural network architecture that holds the tree structure's local information to predict the localization of coronary plaques and their characterization. The proposed method is more robust than the ones in the literature as it is not based on advanced Curved Planar Reformatted (CPR) views that are sensitive to the centerline extraction quality. CSCT and CCTA volumes paired with expert manual annotation have been collected in different clinical sites during the whole duration of the thesis. Experts provide the requirements and validation for clinical applications. Supervised deep learning is particularly suited for this workflow as it allows to learn, scale optimally, and validate directly from manual annotation. Overall this work provides evidence to support the automation of clinical tasks linked to the prediction of CAD and MACE by exploiting noninvasive imaging data.

4.3 Resume (FR)

Ce travail propose d'explorer le vaste sujet, aux multiples facettes, de la prédiction automatique des événements cardiovasculaires indésirables majeurs (MACE). Un flux de travail automatisé basé sur des techniques d'apprentissage profond a été développé en se concentrant sur les tâches cliniques liées à l'identification des indicateurs anatomiques des maladies coronariennes (coronary artery diseases CAD). Les patients acceptés aux urgences avec des suspicions de coronaropathie, symptomatiques et asymptomatiques, subissent des examens qui peuvent être plus ou moins invasifs. L'un des enjeux majeurs de ce travail est d'exploiter des solutions d'examen d'imagerie qui sont peu invasives. La tomographie numerique du score calcique (Calcium Score Computed Tomography CSCT) est un examen exploratoire non invasif à faible dose de radiations qui vise à quantifier les indicateurs précoces de la charge calcique des artères coronaires (CAC) en se

basant sur le calcul du score d'Agatston (AS). Dans ce travail, nous proposons une approche robuste et efficace pour quantifier le fardeau du CAC basé sur la segmentation sémantique fournie par un ensemble d'architectures U-net. Nous avons pu vérifier qu'une résolution minimale de 2,5 mm de tomodensitométrie isotrope est nécessaire pour le calcul précis du score AS, ce qui permet une inférence rapide et donc des besoins de calcul réduits pour les stations de travail des sites cliniques. La Angiographie CT des Coronaires (Coronary CT Angiorgray CCTA) est la seule modalité d'imagerie peu invasive qui permet de visualiser les caractéristiques morphologiques des artères coronaires. L'injection d'une solution iodée dans le sang met en évidence les artères coronaires, ce qui permet d'identifier les indicateurs anatomiques de la maladie coronarienne. Nous proposons un flux de travail d'analyse coronaire qui extrait d'abord la ligne centrale coronaire (suivi coronaire), permettant ainsi les tâches ultérieures liées à l'identification des lésions coronaires (plaques et sténoses). À cette fin, nous avons développé une fonction de cout adaptée qui permets d'apprendre de la ligne centrale annotée seule pour identifier les structures tubulaires et une stratégie de suivi récursif pour calculer l'arbre entier des vaisseaux. L'étiquetage de l'arbre des vaisseaux coronaires contient des informations cliniques importantes, car la connaissance des principales branches guide l'examen expert de la coronaropathie. Dans ce travail, nous montrons qu'une approche d'apprentissage basée sur les réseaux neuronaux par graph (GNN) complétée par une structure arborescente apriori peut identifier les branches coronaires de l'arbre des vaisseaux, même lorsque l'arbre est incomplet ou très irrégulier en raison des variations anatomiques qui peuvent se produire dans la population de patients. L'arbre vasculaire associé au volume CCTA est ensuite utilisé pour l'analyse coronaire. Nous proposons une architecture de réseau neuronal transformateur qui retient les informations locales de l'arborescence pour prédire la localisation des plaques coronaires et leur caractérisation. La méthode proposée est plus robuste que celles de la littérature car elle n'est pas basée sur des vues avancées Curved Planar Reformatted (CPR) qui sont sensibles à la qualité de l'extraction de la ligne centrale. Des volumes de CSCT et CCTA associés à des annotations manuelles d'experts ont été collectés dans différents sites cliniques pendant toute la durée de la thèse. Les experts fournissent les exigences et la validation pour les applications cliniques. L'apprentissage profond supervisé est particulièrement adapté à ce flux de travail, car il permet d'apprendre, de mettre à l'échelle de manière optimale et de valider directement à partir de l'annotation manuelle. Dans l'ensemble, ce travail fournit des preuves pour soutenir l'automatisation des tâches cliniques liées à la prédiction du CAD et MACE en exploitant les données d'imagerie non invasives.

4.4 Compliance with Ethical Standard

4.4.1 Human rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments

involving humans.

4.4.2 Informed consent and patient details

The authors declare that neither this report nor the collected data contain any personal information that could lead to the identification of the patient, data has been collected upon agreement with the clinical sites and patients.

Bibliography

[1] Farhood Saremi and Stephan Achenbach. Coronary Plaque Characterization Using CT. *American Journal of Roentgenology*, 204(3):W249–W260, March 2015.

[2] Michelle C. Williams, Alastair J. Moss, Marc Dweck, Philip D. Adamson, Shirjel Alam, Amanda Hunter, Anoop S.V. Shah, Tania Pawade, Jonathan R. Weir-McCall, Giles Roditi, Edwin J.R. van Beek, David E. Newby, and Edward D. Nicol. Coronary Artery Plaque Characteristics Associated With Adverse Outcomes in the SCOT-HEART Study. *Journal of the American College of Cardiology*, 73(3):291–301, January 2019.

[3] Wijnand J. Stuijfzand, Alexander R. van Rosendael, Fay Y. Lin, Hyuk-Jae Chang, Inge J. van den Hoogen, Umberto Gianni, Jung Hyun Choi, Joon-Hyung Doh, Ae-Young Her, Bon-Kwon Koo, Chang-Wook Nam, Hyung-Bok Park, Sang-Hoon Shin, Jason Cole, Alessia Gimelli, Muhammad Akram Khan, Bin Lu, Yang Gao, Faisal Nabi, Ryo Nakazato, U. Joseph Schoepf, Roel S. Driessen, Michiel J. Bom, Randall Thompson, James J. Jang, Michael Ridner, Chris Rowan, Erick Avelar, Philippe Généreux, Paul Knaapen, Guus A. de Waard, Gianluca Pontone, Daniele Andreini, Mouaz H. Al-Mallah, Yao Lu, Daniel S. Berman, Jagat Narula, James K. Min, Jeroen J. Bax, Leslee J. Shaw, and CREDENCE Investigators. Stress Myocardial Perfusion Imaging vs Coronary Computed Tomographic Angiography for Diagnosis of Invasive Vessel-Specific Coronary Physiology: Predictive Modeling Results From the Computed Tomographic Evaluation of Atherosclerotic Determinants of Myocardial Ischemia (CREDENCE) Trial. *JAMA Cardiology*, 5(12):1338, December 2020.

[4] Andrew N. Primak, Cynthia H. McCollough, Michael R. Bruesewitz, Jie Zhang, and Joel G. Fletcher. Relationship between Noise, Dose, and Pitch in Cardiac Multi–Detector Row CT. *RadioGraphics*, 26(6):1785–1794, November 2006.

[5] Maiken Glud Dalager, Morten Bøttcher, Søren Dalager, Gratien Andersen, Jesper Thygesen, Erik Morre Pedersen, and Hans Erik Bøtker. Imaging Atherosclerotic Plaques by Cardiac Computed Tomography In Vitro: Impact of Contrast Type and Acquisition Protocol. *Investigative Radiology*, 46(12):790–795, December 2011.

[6] Jun Horiguchi, Chikako Fujioka, Masao Kiguchi, Hideya Yamamoto, Yun Shen, and Yasuki Kihara. In vitro measurement of CT density and estimation of stenosis re-

lated to coronary soft plaque at 100kV and 120kV on ECG-triggered scan. *European Journal of Radiology*, 77(2):294–298, February 2011.

[7] Damini Dey, Tiziano Schepis, Mohamed Marwan, Piotr J. Slomka, Daniel S. Berman, and Stephan Achenbach. Automated Three-dimensional Quantification of Noncalcified Coronary Plaque from Coronary CT Angiography: Comparison with Intravascular US. *Radiology*, 257(2):516–522, November 2010.

[8] H. Brodoefel, A. Reimann, M. Heuschmid, I. Tsiflikas, A. F. Kopp, S. Schroeder, C. D. Claussen, M. E. Clouse, and C. Burgstahler. Characterization of coronary atherosclerosis by dual-source computed tomography and HU-based color mapping: a pilot study. *European Radiology*, 18(11):2466–2474, November 2008.

[9] Mark J. Boogers, Alexander Broersen, Joëlla E. van Velzen, Fleur R. de Graaf, Heba M. El-Naggar, Pieter H. Kitslaar, Jouke Dijkstra, Victoria Delgado, Eric Boersma, Albert de Roos, Joanne D. Schuijf, Martin J. Schalij, Johan H.C. Reiber, Jeroen J. Bax, and J. Wouter Jukema. Automated quantification of coronary plaque with computed tomography: comparison with intravascular ultrasound using a dedicated registration algorithm for fusion-based quantification. *European Heart Journal*, 33(8):1007–1016, April 2012.

[10] R. Nakazato, H. Otake, A. Konishi, M. Iwasaki, B.-K. Koo, H. Fukuya, T. Shinke, K.-i. Hirata, J. Leipsic, D. S. Berman, and J. K. Min. Atherosclerotic plaque characterization by CT angiography for identification of high-risk coronary artery lesions: a comparison to optical coherence tomography. *European Heart Journal - Cardiovascular Imaging*, 16(4):373–379, April 2015.

[11] Arthur S. Agatston, Warren R. Janowitz, and Hildner. Quantification of coronary artery calcium using ultrafast computed tomography. *Journal of the American College of Cardiology*, 15(4):827–832, March 1990.

[12] Salim A. Si-Mohamed, Sara Boccalini, Hugo Lacombe, Adja Diaw, Mohammad Varasteh, Pierre-Antoine Rodesch, Riham Dessouky, Marjorie Villien, Valérie Tatard-Leitman, Thomas Bochaton, Philippe Coulon, Yoad Yagil, Elias Lahoud, Klaus Erhard, Benjamin Riche, Eric Bonnefoy, Gilles Rioufol, Gerard Finet, Cyrille Bergerot, Loic Boussel, Joel Greffier, and Philippe C. Douek. Coronary CT Angiography with Photon-counting CT: First-In-Human Results. *Radiology*, 303(2):303–313, May 2022.

[13] Matthew J. Budoff, Thomas Mayrhofer, Maros Ferencik, Daniel Bittner, Kerry L. Lee, Michael T. Lu, Adrian Coles, James Jang, Mayil Krishnam, Pamela S. Douglas, and Udo Hoffmann. Prognostic Value of Coronary Artery Calcium in the PROMISE Study (Prospective Multicenter Imaging Study for Evaluation of Chest Pain). *Circulation*, 136(21):1993–2005, November 2017.

[14] Harvey S. Hecht, Paul Cronin, and Michael J. Blaha. 2016 SCCT/STR guidelines for coronary artery calcium scoring of noncontrast noncardiac chest CT scans: A report

of the Society of Cardiovascular Computed Tomography and Society of Thoracic Radiology. *Journal of Cardiovascular Computed Tomography*, 11(1):74–84, January 2017.

[15] Majd Zreik, Robbert W. van Hamersvelt, and Jelmer M. Wolterink. A Recurrent CNN for Automatic Detection and Classification of Coronary Artery Plaque and Stenosis in Coronary CT Angiography. *IEEE Transactions on Medical Imaging*, 38(7):1588–1598, July 2019.

[16] Nicolas Gogin, Mario Viti, Luc Nicodème, Mickaël Ohana, Hugues Talbot, Umit Gencer, Magloire Mekukosokeng, Thomas Caramella, Yann Diascorn, Jean-Yves Airaud, Marc-Samir Guillot, Zoubir Bensalah, Caroline Dam Hieu, Bassam Abdallah, Imad Bousaid, Nathalie Lassau, and Elie Mousseaux. Automatic coronary artery calcium scoring from unenhanced-ECG-gated CT using deep learning. *Diagnostic and Interventional Imaging*, 102(11):683–690, November 2021.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351, pages 234–241. Springer International Publishing, Cham, 2015.

[18] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997.

[19] Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar, 2014. Association for Computational Linguistics.

[20] Ashish Vaswani, Noam M. Shazeer, and Niki Parmar. Attention is all you need. *ArXiv*, abs/1706.03762, 2017.

[21] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[22] Srinadh Bhojanapalli, Ayan Chakrabarti, Daniel Glasner, Daliang Li, Thomas Unterthiner, and Andreas Veit. Understanding Robustness of Transformers for Image Classification. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10211–10221, Montreal, QC, Canada, October 2021. IEEE.

[23] Tan Yu, Xu Li, Yunfeng Cai, Mingming Sun, and Ping Li. S ^{2} -MLP: Spatial-Shift MLP Architecture for Vision. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3615–3624, Waikoloa, HI, USA, January 2022. IEEE.

[24] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv:1409.0473 [cs, stat]*, May 2016. arXiv: 1409.0473.

[25] Thomas Weikert, Marco Francone, Suhny Abbara, Bettina Baessler, Byoung Wook Choi, Matthias Gutberlet, Elizabeth M. Hecht, Christian Loewe, Elie Mousseaux, Luigi Natale, Konstantin Nikolaou, Karen G. Ordovas, Charles Peebles, Claudia Prieto, Rodrigo Salgado, Birgitta Velthuis, Rozemarijn Vliegenthart, Jens Bremerich, and Tim Leiner. Machine learning in cardiovascular radiology: ESCR position statement on design requirements, quality assessment, current applications, opportunities, and challenges. *European Radiology*, 31(6):3909–3922, June 2021.

[26] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).

[27] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571, Stanford, CA, USA, October 2016. IEEE.

[28] Jelmer M. Wolterink, Tim Leiner, Bob D. de Vos, Robbert W. van Hamersvelt, Max A. Viergever, and Ivana Išgum. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Medical Image Analysis*, 34:123–136, December 2016.

[29] Felix Durlak, Michael Wels, Chris Schwemmer, Michael Sühling, Stefan Steidl, and Andreas Maier. Growing a Random Forest with Fuzzy Spatial Features for Fully Automatic Artery-Specific Coronary Calcium Scoring. In Qian Wang, Yinghuan Shi, Heung-Il Suk, and Kenji Suzuki, editors, *Machine Learning in Medical Imaging*, volume 10541, pages 27–35. Springer International Publishing, Cham, 2017.

[30] Guanyu Yang, Yang Chen, Xiufang Ning, Qiaoyu Sun, Huazhong Shu, and Jean-Louis Coatrieux. Automatic coronary calcium scoring using noncontrast and contrast CT images: Automatic coronary calcium scoring. *Medical Physics*, 43(5):2174–2186, April 2016.

[31] The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine*, 365(5):395–409, August 2011.

[32] Nikolas Lessmann, Bram van Ginneken, Majd Zreik, Pim A. de Jong, Bob D. de Vos, Max A. Viergever, and Ivana Isgum. Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks With Dilated Convolutions. *IEEE Transactions on Medical Imaging*, 37(2):615–625, February 2018.

[33] Sanne G. M. van Velzen, Nikolas Lessmann, Birgitta K. Velthuis, Ingrid E. M. Bank, Desiree H. J. G. van den Bongard, Tim Leiner, Pim A. de Jong, Wouter B. Veldhuis, Adolfo Correa, James G. Terry, John Jeffrey Carr, Max A. Viergever, Helena M. Verkooijen, and Ivana Išgum. Deep Learning for Automatic Calcium Scoring in CT: Validation Using Multiple Cardiac CT and Chest CT Protocols. *Radiology*, 295(1):66–79, April 2020.

[34] Bob D. de Vos, Jelmer M. Wolterink, Tim Leiner, Pim A. de Jong, Nikolas Lessmann, and Ivana Isgum. Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. *IEEE Transactions on Medical Imaging*, 38(9):2127–2138, September 2019.

[35] Karl Krissian, Hrvoje Bogunovic, Hrvoje Bogunovic, Jose Maria Pozo, Maria Cruz Villa-Uriol, and Alejandro Frangi. Minimally Interactive Knowledge-based Coronary Tracking in CTA using a Minimal Cost Path. *The MIDAS Journal*, July 2008.

[36] Ro Frangi, W.J. Niessen, Koen Vincken, and Max Viergever. Multiscale vessel enhancement filtering. *Med. Image Comput. Comput. Assist. Interv.*, 1496, 02 2000.

[37] Odyssee Merveille, Hugues Talbot, Laurent Najman, and Nicolas Passat. Curvilinear Structure Analysis by Ranking the Orientation Responses of Path Operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(2):304–317, February 2018.

[38] Odyssee Merveille, Benoit Naegel, Hugues Talbot, and Nicolas Passat. n D Variational Restoration of Curvilinear Structures With Prior-Based Directional Regularization. *IEEE Transactions on Image Processing*, 28(8):3848–3859, August 2019.

[39] Amos Sironi, Engin Turetken, Vincent Lepetit, and Pascal Fua. Multiscale Centerline Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(7):1327–1341, July 2016.

[40] Alexandru Dorobanțiu, Valentin Ogrean, and Remus Brad. Coronary Centerline Extraction from CCTA Using 3D-UNet. *Future Internet*, 13(4):101, April 2021.

[41] Ruochen Gao, Zhihui Hou, Jun Li, Hu Han, Bin Lu, and S. Kevin Zhou. Joint Coronary Centerline Extraction And Lumen Segmentation From Ccta Using Cnntracker And Vascular Graph Convolutional Network. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1897–1901, Nice, France, April 2021. IEEE.

[42] Byunghwan Jeon. Deep Recursive Bayesian Tracking for Fully Automatic Centerline Extraction of Coronary Arteries in CT Images. *Sensors*, 21(18):6087, September 2021.

[43] Geert Litjens, Francesco Ciompi, Jelmer M. Wolterink, Bob D. de Vos, Tim Leiner, Jonas Teuwen, and Ivana Išgum. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC: Cardiovascular Imaging*, 12(8):1549–1565, August 2019.

[44] Jelmer M. Wolterink, Robbert W. van Hamersvelt, Max A. Viergever, Tim Leiner, and Ivana Išgum. Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier. *Medical Image Analysis*, 51:46–60, January 2019.

[45] Jelmer M. Wolterink, Tim Leiner, Bob D. de Vos, Jean-Louis Coatrieux, B. Michael Kelm, Satoshi Kondo, Rodrigo A. Salgado, Rahil Shahzad, Huazhong Shu, Miranda Snoeren, Richard A. P. Takx, Lucas J. van Vliet, Theo van Walsum, Tineke P. Willems, Guanyu Yang, Yefeng Zheng, Max A. Viergever, and Ivana Išgum. An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework: Evaluation of cardiac CT-based automatic coronary calcium scoring. *Medical Physics*, 43(5):2361–2373, April 2016.

[46] Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P. W. Pluim, Ulrich Bauer, and Bjoern H. Menze. clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16555–16564, Nashville, TN, USA, June 2021. IEEE.

[47] Laurent Najman and Hugues Talbot, editors. *Mathematical Morphology: From Theory to Applications*. John Wiley & Sons, Inc., Hoboken, NJ, USA, February 2013.

[48] H. Blum. A transformation for extracting new descriptors of shape. In W. Wathen-Dunn, editor, *Models for Perception of Speech and Visual Form*. MIT Press, Cambridge, MA, 1967.

[49] Herbert Edelsbrunner and J. Harer. *Computational topology: an introduction*. American Mathematical Society, Providence, R.I, 2010. OCLC: ocn427757156.

[50] Xiaoling Hu, Li Fuxin, Dimitris Samaras, and Chao Chen. Topology-Preserving Deep Image Segmentation, 2019.

[51] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, and Sara Carvalho. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446, March 2012.

[52] Xinghua Ma, Gongning Luo, Wei Wang, and Kuanquan Wang. Transformer Network for Significant Stenosis Detection in CCTA of Coronary Arteries. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, volume 12906, pages 516–525. Springer International Publishing, Cham, 2021.

[53] Felix Denzinger, Michael Wels, and Katharina Breininger. Deep Learning Algorithms for Coronary Artery Plaque Characterisation from CCTA Scans. In Thomas Tolxdorff, Thomas M. Deserno, Heinz Handels, Andreas Maier, Klaus H. Maier-Hein, and Christoph Palm, editors, *Bildverarbeitung für die Medizin 2020*, pages 193–198. Springer Fachmedien Wiesbaden, Wiesbaden, 2020.

[54] Fopke Klok. Two moving coordinate frames for sweeping along a 3D trajectory. *Computer Aided Geometric Design*, 3(3):217–229, November 1986.

[55] Wenping Wang, Bert Jüttler, Dayue Zheng, and Yang Liu. Computation of rotation minimizing frames. *ACM Transactions on Graphics*, 27(1):1–18, March 2008.

[56] Sunil Kini, Kostaki G. Bis, and Leroy Weaver. Normal and Variant Coronary Arterial and Venous Anatomy on High-Resolution CT Angiography. *American Journal of Roentgenology*, 188(6):1665–1674, June 2007.

[57] Maurice Termeer, Javier Oliván Bescós, Marcel Breeuwer, Anna Vilanova, Frans Gerritsen, Eduard Gröller, and Eike Nagel. Patient-specific coronary artery supply territory AHA diagrams. *Journal of Cardiovascular Magnetic Resonance*, 11(S1):P103, 1532–429X–11–S1–P103, January 2009.

[58] Akin Akinyemi, Sean Murphy, Ian Poole, and Colin Roberts. Automatic labelling of coronary arteries. *2009 17th European Signal Processing Conference*, pages 1562–1566, 2009.

[59] Bui Huy Hoang, Masahiro Oda, Zhengang Jiang, Takayuki Kitasaka, Kazunari Misawa, Michitaka Fujiwara, and Kensaku Mori. A study on automated anatomical labeling to arteries concerning with colon from 3D abdominal CT images. page 79623R, Lake Buena Vista, Florida, March 2011.

[60] Tetsuro Matsuzaki, Masahiro Oda, Takayuki Kitasaka, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. Automated anatomical labeling of abdominal arteries and hepatic portal system extracted from abdominal CT volumes. *Medical Image Analysis*, 20(1):152–161, February 2015.

[61] Qing Cao, Alexander Broersen, Michiel A. de Graaf, Pieter H. Kitslaar, Guanyu Yang, Arthur J. Scholte, Boudewijn P. F. Lelieveldt, Johan H. C. Reiber, and Jouke Dijkstra. Automatic identification of coronary tree anatomy in coronary computed tomography angiography. *The International Journal of Cardiovascular Imaging*, 33(11):1809–1819, November 2017.

[62] Mehmet A. Gülsün, Gareth Funka-Lea, Yefeng Zheng, and Matthias Eckert. CTA Coronary Labeling through Efficient Geodesics between Trees Using Anatomy Priors. In Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger, and Robert Howe, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, volume 8674, pages 521–528. Springer International Publishing, Cham, 2014.

[63] Dan Wu, Xin Wang, Junjie Bai, Xiaoyang Xu, Bin Ouyang, Yuwei Li, Heye Zhang, Qi Song, Kunlin Cao, and Youbing Yin. Automated anatomical labeling of coronary arteries via bidirectional tree LSTMs. *International Journal of Computer Assisted Radiology and Surgery*, 14(2):271–280, February 2019.

[64] Nils Hampe, Jelmer M. Wolterink, Carlos Collet, R. Nils Planken, and Ivana Išgum. Graph attention networks for segment labeling in coronary artery trees. In Bennett A. Landman and Ivana Išgum, editors, *Medical Imaging 2021: Image Processing*, page 50, Online Only, United States, February 2021. SPIE.

[65] Han Yang, Xingjian Zhen, Ying Chi, Lei Zhang, and Xian-Sheng Hua. CPR-GCN: Conditional Partial-Residual Graph Convolutional Network in Automated Anatomical Labeling of Coronary Arteries. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3802–3810, Seattle, WA, USA, June 2020. IEEE.

[66] Kai Sheng Tai, Richard Socher, and Christopher D. Manning. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. *arXiv:1503.00075 [cs]*, May 2015. arXiv: 1503.00075.

[67] Felix Denzinger, Michael Wels, Nishant Ravikumar, et al. Coronary Artery Plaque Characterization from CCTA Scans Using Deep Learning and Radiomics. In Dinggang Shen, Tianming Liu, and Terry M. Peters, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, volume 11767, pages 593–601. Springer International Publishing, Cham, 2019.

[68] M. Cimpoi, Subhransu Maji, and A. Vedaldi. Deep filter banks for texture recognition and segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3828–3836, 2015.

[69] Paul Valensi, Patrick Henry, Franck Boccara, Emmanuel Cosson, Gaetan Prevost, Joseph Emmerich, Laura Ernande, Dany Marcadet, Elie Mousseaux, François Rouzet, Ariane Sultan, Jean Ferrières, Bruno Vergès, and Eric Van Belle. Risk stratification and screening for coronary artery disease in asymptomatic patients with diabetes mellitus: Position paper of the French Society of Cardiology and the French-speaking Society of Diabetology. *Diabetes & Metabolism*, 47(2):101185, March 2021.

[70] Sanne G. M. van Velzen, Nikolas Lessmann, Birgitta K. Velthuis, Ingrid E. M. Bank, Desiree H. J. G. van den Bongard, Tim Leiner, Pim A. de Jong, Wouter B. Veldhuis, Adolfo Correa, James G. Terry, John Jeffrey Carr, Max A. Viergever, Helena M. Verkooijen, and Ivana Išgum. Deep Learning for Automatic Calcium Scoring in CT: Validation Using Multiple Cardiac CT and Chest CT Protocols. *Radiology*, 295(1):66–79, April 2020.

[71] N. Lassau, I. Bousaid, E. Chouzenoux, J.P. Lamarque, B. Charmettant, M. Azoulay, F. Cotton, A. Khalil, O. Lucidarme, F. Pigneur, Y. Benaceur, A. Sadate, M. Lederlin, F. Laurent, G. Chassagnon, O. Ernst, G. Ferreti, Y. Diascorn, P.Y. Brillet, M. Creze, L. Cassagnes, C. Caramella, A. Loubet, A. Dallongeville, N. Abassebay, M. Ohana, N. Banaste, M. Cadi, J. Behr, L. Boussel, L. Fournier, M. Zins, J.P. Beregi, A. Luciani, A. Cotten, and J.F. Meder. Three artificial intelligence data challenges based on CT and MRI. *Diagnostic and Interventional Imaging*, 101(12):783–788, December 2020.

[72] Jonathon Leipsic, Suhny Abbara, Stephan Achenbach, Ricardo Cury, James P. Earls, GB John Mancini, Koen Nieman, Gianluca Pontone, and Gilbert L. Raff. SCCT guidelines for the interpretation and reporting of coronary CT angiography: A report of the Society of Cardiovascular Computed Tomography Guidelines Committee. *Journal of Cardiovascular Computed Tomography*, 8(5):342–358, September 2014.

[73] Armin Kanitsar, Dominik Fleischmann, and Rainer Wegenkittl. Diagnostic Relevant Visualization of Vascular Structures. In Georges-Pierre Bonneau, Thomas Ertl, and Gregory M. Nielson, editors, *Scientific Visualization: The Visual Extraction of Knowledge from Data*, pages 207–228. Springer-Verlag, Berlin/Heidelberg, 2006.

[74] Michiel Schaap, Coert T. Metz, Theo van Walsum, Alina G. van der Giessen, Annick C. Weustink, Nico R. Mollet, Christian Bauer, Hrvoje Bogunović, Carlos Castro, and Xiang Deng. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical Image Analysis*, 13(5):701–714, October 2009.

[75] Mario Viti, Hugues Talbot, and Nicolas Gogin. Transformer Graph Network for Coronary Plaque Localization in CCTA. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, Kolkata, India, March 2022. IEEE.

[76] Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *CoRR*, abs/2012.09699, 2020.

[77] Ahmed M. Ghanem, Ahmed H. Hamimi, Jatin R. Matta, Aaron Carass, Reham M. Elgarf, Ahmed M. Gharib, and Khaled Z. Abd-Elmoniem. Automatic Coronary Wall and Atherosclerotic Plaque Segmentation from 3D Coronary CT Angiography. *Scientific Reports*, 9(1):47, December 2019.

[78] Tony Chan and Luminita Vese. An active contour model without edges. In Mads Nielsen, Peter Johansen, Ole Fogh Olsen, and Joachim Weickert, editors, *Scale-Space Theories in Computer Vision*, pages 141–151, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[79] Tony F. Chan, Selim Esedoglu, and Mila Nikolova. Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, January 2006.

[80] Steven J. Ruuth. Efficient Algorithms for Diffusion-Generated Motion by Mean Curvature. *Journal of Computational Physics*, 144(2):603–625, August 1998.

[81] Julien Mille. Narrow band region-based active contours and surfaces for 2D and 3D segmentation. *Computer Vision and Image Understanding*, 113(9):946–965, September 2009.

[82] R. Malladi, J.A. Sethian, and B.C. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, February 1995.

[83] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, November 1992.

[84] Antonin Chambolle and Thomas Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.

[85] Odyssée Merveille, Olivia Miraucourt, Stéphanie Salmon, Nicolas Passat, and Hugues Talbot. A variational model for thin structure segmentation based on a directional regularization. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4324–4328. IEEE, 2016.

[86] Odyssée Merveille, Benoît Naegel, Hugues Talbot, and Nicolas Passat. n d variational restoration of curvilinear structures with prior-based directional regularization. *IEEE Transactions on Image Processing*, 28(8):3848–3859, 2019.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Prédiction automatique des épisodes cardio-vasculaires adverses majeurs
Mots clés: 3 à 6 mots clefs (version en français)

Résumé: Dans ce projet de recherche financé en contrat CIFRE avec GE Healthcare, on cherche a prédire les épisodes cardio-vasculaire adverses majeurs (ECAM), c'est à dire typiquement les embolies et les anévrismes dans l'aorte et les artères coronaires, qui donnent lieu a une respectivement à une interruption catastrophique du flux sanguin vers le coeur et donc un infarctus, ou à une hémorragie interne. Les deux types d'épisodes sont extrêmement graves. Lorsqu'un patient est hospitalisé pour une alerte reliée à ces épisodes, il va subir un examen scanner X, injecté ou non, plus ou moins invasif. Un objectif majeur de cette recherche est d'utiliser au mieux l'information obtenue sous forme d'images 3D ainsi que l'historique du patient pour éviter de soumettre le patient à des examens inutiles, invasifs ou dangereux, tout en garantissant le meilleur résultat clinique. Les méthodologies proposées reposeront sur des techniques d'analyse et traitement d'image, de vision par ordinateur et d'imagerie médicale qui seront développée en partenariat entre GE Healthcare et le laboratoire Centre de Vision Numérique (CVN) de CentraleSupélec.

Title: Automated prediction of major adverse cardiovascular events
Keywords: 3 à 6 mots clefs (version en anglais)

Abstract: This research project is expected to be financed by a CIFRE scholarship in collaboration between GE Healthcare and CentraleSupelec. We are seeking to predict Major Adverse Cardiovascular Events (MACE). These are typically embolism and aneurisms in the aorta and the coronary arteries, that give rise respectively to interrupted blood flow to the heart and so a risk of infarctus, or major hemorrhage. Both are life-threatening. When a patient is brought to hospital for an alert (angina, etc), they will undergo an X-ray CAT scan, which can be more or less invasive. A major objective of this research is to utilize as well as possible the available information in the form of 3D images together with patient history and other data, in order to avoid needless, invasive, irradiating or dangerous exams, while simultaneously guaranteeing optimal care and the best possible clinical outcome. The proposed methodologies include image analysis, image processing, computer vision and medical imaging procedures and methods, that will be developed in partnership between GE Healthcare and the CVN lab of CENTRALE SUPELEC.

Contents

List of Figures

List of Tables

Acronyms

AHA American Heart Association

AI average inside

AS Agatston score

BFS Breadth First Search

CAC Coronary Artery Calcium

CAD Coronary Artery Disease

CAD-RADS Coronary Artery Disease Reporting & Data System

CAT08 MICCAI 2008 Coronary Artery Tracking Challenge

CCTA Coronary Computed Tomography Angiography

CNN convolutional neural network

CPR curved planar reformation

CPR-GCN conditional partial residual graph convolutional network

CSCT Calcium Score Computed Tomography

CT Computed Tomography

DAG Directed Acyclic Graph

DBSCAN density-based spatial clustering of applications with noise

FCN fully connected network

FFR fractional flow reserve

GCN graph convolutional network

GRU gated recurrent unit

HU Hounsfield Units

IVUS Intravascular Ultrasounds

KB knowledge based

LAD Left Artery Descending

LCX Left Circumflex

LMA Left Main Artery

LSTM long-short term memory

MACE Major Adverse Cardiovascular Events

MHSA Multi Head Self Attention

MIP maximum intensity projection

MLP multi layer perceptron

MPR multi planar reformation

MRI Magnetic Resonance Imaging

NLP natural language procesing

OCT Optical Coherence Tomography

OF overlap until first error

OM Obtuse Marginal

ORI-GCN oriented graph convolutional network

OT overlap with the clinically relevant part of the vessel

OV overlap

PCCT Photon Counting Computed Tomography

PDA Posterior Descending Artery

PLB Posterior Lateral Branch

PROMISE Prospective Multicenter Imaging Study for Evaluation of Chest Pain

RCA Right Coronary Artery

RCNN recursive convolutional neural network

RIB Ramus Intermedius Branch

RNN recursive neural network

SFR Société Française de Radiologie