Deep learning for automatic detection and quantification of the disease areas of the myocardium from DE-MRI after myocardial infarction

Zhihao Chen

UBFC
UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ

UFC
UNIVERSITÉ
DE FRANCHE-COMTÉ

THÈSE DE DOCTORAT

DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

PRÉPARÉE À L'UNIVERSITÉ DE FRANCHE-COMTÉ

École doctorale n°37

Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Informatique

par

ZHIHAO CHEN

Deep learning for automatic detection and quantification of the disease areas of the myocardium from DE-MRI after myocardial infarction

Apprentissage profond pour la détection et la quantification automatiques des zones pathologiques du myocarde à partir de l'IRM DE après un infarctus

Thèse présentée et soutenue à Belfort, le 19 Octobre 2021

Composition du Jury :

JODOIN PIERRE-MARC	PR à l'Univ. de Sherbrooke	Rapporteur
PASSAT NICOLAS	PR à l'Univ. de Reims Champagne-Ardenne	Rapporteur
KASTLER BRUNO	PR à l'Univ. de Paris	Examinateur
COUTURIER RAPHAËL	PR à l'Univ. Bourgogne Franche-Comté	Directeur de thèse
SALOMON MICHEL	MCF HDR - Univ. Bourgogne Franche-Comté	Codirecteur
LALANDE ALAIN	MCF-PH HDR - Univ. Bourgogne Franche-Comté	Codirecteur

N° 0 0 1

ABSTRACT

Deep learning for automatic detection and quantification of the disease areas of the myocardium from DE-MRI after myocardial infarction

Zhihao Chen
University Bourgogne Franche-Comté, 2021

Supervisors: Raphaël Couturier, Michel Salomon and Alain Lalande

Myocardial Infarction (MI) has become one of the most common cardiovascular diseases. An MI occurs when the blood flow decreases or stops in a part of the heart which can cause damage to the myocardium due to ischemia. The MI can lead to persistent microvascular obstruction (PMO) even after the reperfusion therapy of an occluded coronary artery, which affects a significant proportion of patients. To assess the MI and the extended PMO, Delayed Enhancement MRI (DE-MRI) has become the gold standard. However, to precisely quantify the pathology, physicians have to first manually draw the delineations of different myocardial tissues on the short-axis MRI slices then the volume of normal and pathological tissues will be calculated to evaluate the severity of the attack. The manual annotation procedures are highly time-consuming and subjective, which reinforces the potential interest in an automatic evaluation approach for the MI assessment. Current automatic delineation methods on DE-MRI are mostly probabilistic approaches such as the Mixture Models while more and more deep learning applications achieve state-of-the-art performances compared to the probabilistic approaches. Therefore, in this thesis, the different aspects of deep learning-based approaches have been investigated for the automatic MI evaluation from DE-MRI.

To develop the deep learning-based MI assessment pipeline, three major aspects consisting of the pre-processing, the deep learning models, and the post-processing have been investigated or proposed. The pre-processing aims at preparing more consistent and clearer inputs for deep learning models using image normalizations, cropping, etc. To design well-adaptive deep learning models, many parts (building blocks, loss functions, segmentation models, etc.) were proposed or compared. Moreover, the post-processing

methods including morphological treatments and prior information-based filters were applied to the coarse segmentations to eliminate false positive contours. Apart from the automatic assessment by segmentation, merging the MRI and textual patient features, a machine learning-based approach was proposed to obtain quantitative estimates of MI using only the clinical and pathological features for a quick cardiological assessment in the cardiac emergency department.

KEYWORDS: Deep learning, MRI (Magnetic resonance imaging), Clinical characteristics, Automatic detection, Image segmentation, Myocardial infarction.

RÉSUMÉ

Apprentissage profond pour la détection et la quantification automatiques des zones pathologiques du myocarde à partir de l'IRM DE après un infarctus.

Zhihao Chen
Université Bourgogne Franche-Comté, 2021

Encadrants: Raphaël Couturier, Michel Salomon et Alain Lalande

L'infarctus du myocarde (IM) est devenu l'une des maladies cardiovasculaires les plus courantes. L'infarctus se produit lorsque le flux sanguin diminue ou s'arrête dans une partie du cœur, ce qui peut endommager le myocarde en raison de l'ischémie. L'infarctus peut entraîner une obstruction microvasculaire persistante (PMO) même après le traitement de reperfusion d'une artère coronaire occluse, ce qui touche une proportion importante de patients. Pour évaluer l'infarctus et l'obstruction microvasculaire persistante, l'IRM avec rehaussement tardif (IRM-RT) est devenue la référence. Cependant, pour quantifier précisément la pathologie, les médecins doivent d'abord tracer manuellement les délimitations des différents tissus myocardiques sur les tranches d'IRM à petit axe. Ensuite, le volume des tissus normaux et pathologiques sera calculé pour évaluer la gravité de l'attaque. Les procédures d'annotation manuelle sont très chronophages et subjectives, ce qui renforce l'intérêt potentiel d'une approche d'évaluation automatique pour l'évaluation de l'IM. Les méthodes actuelles de contourage automatique en IRM-RT sont principalement des approches probabilistes telles que les modèles de mélange de distributions gaussiennes, alors que de plus en plus d'applications d'apprentissage profond atteignent des performances de pointe par rapport aux approches probabilistes. Par conséquent, dans cette thèse, les différents aspects des approches basées sur l'apprentissage profond ont été étudiés pour l'évaluation automatique du MI à partir de l'IRM-RT.

Pour développer le pipeline d'évaluation de l'IM basé sur l'apprentissage profond, trois aspects majeurs, à savoir le prétraitement, les modèles d'apprentissage profond et le post-traitement, ont été étudiés ou proposés. Le prétraitement vise à préparer des

entrées plus cohérentes et plus claires pour les modèles d'apprentissage profond en utilisant des normalisations d'images, des recadrages, etc. Pour concevoir des modèles d'apprentissage profond bien adaptés, de nombreux éléments (blocs de construction, fonctions de perte, modèles de segmentation, etc.) ont été proposés ou comparés. En outre, les méthodes de post-traitement, y compris les traitements morphologiques et les filtres basés sur l'information préalable, ont été appliquées aux segmentations grossières afin d'éliminer les contours faussement positifs. En plus de l'évaluation automatique par segmentation, en fusionnant l'IRM et les caractéristiques physiologiques du patient, une approche basée sur l'apprentissage automatique a été proposée pour obtenir des estimations quantitatives de l'IM en utilisant uniquement les caractéristiques cliniques et pathologiques pour une évaluation cardiologique rapide dans le service des urgences cardiaques.

Mot clé : Apprentissage profond, IRM (Imagerie par résonance magnétique), Caractéristiques cliniques, Détection automatique, Segmentation de l'image, Infarctus du myocarde.

ACKNOWLEDGEMENTS

Completing a Ph.D. is a challenging and rewarding journey, and I am deeply grateful to the following individuals and groups who have supported me along the way.

I begin by expressing my sincere appreciation to my three esteemed Ph.D. advisors, Raphaël Couturier, Alain Lalande, and Michel Salomon. Their invaluable contributions have been pivotal in shaping my academic growth, providing insightful guidance, constructive feedback, and unwavering support. Raphaël's humor, Alain's medical training, and Michel's good wine are just a few examples of the little things that have made my journey all the more enjoyable.

I am also grateful to my girlfriend Qiling Wang, whose unwavering love, encouragement, and support have been a constant source of strength and inspiration throughout my Ph.D. journey. Despite the challenges of COVID, her patience, understanding, and unwavering belief in me have kept me going, and I couldn't have completed this work without her support.

My parents, Yishen Chen and Liping Wu, have also been my pillars of support. Their endless love, encouragement, and sacrifices have been the driving force behind my academic pursuits, and I am deeply grateful for their unwavering support.

Finally, I extend my sincere appreciation to the members of the thesis defense committee for accepting the responsibility of evaluating and providing feedback on my work. Their insightful comments, suggestions, and questions have been instrumental in refining and strengthening my research.

To all of these individuals and groups, I offer my deepest gratitude for their unwavering support, encouragement, and guidance throughout my Ph.D. journey.

Contents

List of abbreviations

CE . . . . . Cross Entropy

CMRI . . . . . Cardiac Magnetic Resonance Imaging

cTn . . . . . Cardiac Troponin

DICOM . . . . . Digital Imaging and Communications in Medicine

DL . . . . . Deep Learning

ECG . . . . . Electrocardiogram

FCN . . . . . Fully Convolutional Network

FWHM . . . . . Full-Width at Half-Maximum

GAN . . . . . Generative Adversarial Network

GPU . . . . . Graphics Processing Unit

HSV . . . . . Hue, Saturation, Value

IRB . . . . . Inverted Residual Blocks

JSON . . . . . JavaScript Object Notation

KL . . . . . Kullback-Leibler

LV . . . . . Left Ventricle

LVEF . . . . . Left Ventricular Ejection Fraction

MAE . . . . . Mean Absolute Error

MI . . . . . Myocardial Infarction

ML . . . . . Machine Learning

MLP . . . . . Multi-Layer Perceptron

MRI . . . . . Magnetic Resonance Imaging

NifTi . . . . . Neuroimaging Informatics Technology Initiative

NLP . . . . . Natural Language Processing

NMR . . . . . Nuclear Magnetic Resonance

n-SD . . . . . n-Standard Deviations

NT-proBNP . . . N-terminal Prohormone of Brain Natriuretic Peptide

PIM . . . . . Percent of Infarcted Myocardium

PMO . . . . . Persistent Microvascular Obstruction

PSIR . . . . . Phase Sensitive Inversion Recovery

ReLU . . . . . Rectified Linear Unit

ROI . . . . . Region of Interest

RGB . . . . . Red, Green, Blue

SD . . . . . Standard Deviations

SE . . . . . Squeeze-Excitation

SGD . . . . . Stochastic Gradient Descent

SSFP . . . . . Steady-State Free Precession

SVM . . . . . Support-Vector Machine

ST . . . . . ST segment (that connects the QRS complex and the T wave in electrocardiogram)

STRM . . . . . Signal Threshold versus Reference Mean

TTE . . . . . Transthoracic Echocardiography

URL . . . . . Upper Reference Limit

1

INTRODUCTION

This thesis focuses on the automatic assessment of myocardial infarction (MI) from multivariate medical diagnoses with machine learning methods. Delayed-enhancement Magnetic Resonance Imaging (DE-MRI) and 12 pieces of physiological, clinical and paraclinical features are selected as the experimental data. To process the medical data collected in the University Hospital of Dijon (CHU Dijon), deep learning-based segmentation methods are employed on DE-MRI and machine learning-based classification and regression models are performed on patient features. The goal of the automatic assessment is to provide qualitative and quantitative evaluation of the severity of the myocardial infarction, offering physicians a quick and subjective reference of the pathology. The presented research was carried out in the Department of Computer Science and Complex Systems (DISC in French) of FEMTO-ST laboratory, in collaboration with ImViA (Imaging and Artificial Vision) laboratory, CHU Dijon and CASIS company. More precisely, it took place and was funded in the context of the ADVANCES (Automatic Detection of Viable myocArdiac segmeNts Considering dEep networkS) project, which was successfully submitted to the Second ISITE-BFC call for proposals.

1.1/ MYOCARDIAL INFARCTION

Myocardial infarction has become one of the most common cardiovascular diseases [64]. Myocardial infarction occurs as a result of myocardial distress leading to the death of myocardial tissue. In the clinical context, myocardial infarction is usually due to thrombotic occlusion of a coronary artery, most commonly caused by the rupture of an atherosclerotic plaque. Ischaemia induces perturbations in the myocardium and leads to a rapid depression of cardiac functions, and then in case of prolonged ischemia, necrosis of the myocardial tissue may occur. If the revascularization is delayed or fails, the extensive damage can lead to persistent microvascular obstruction (PMO), also known as the no-reflow phenomenon [58]. Therefore, emergency revascularization therapy to restore perfusion is crucial as soon as the disease is diagnosed.

1.2/ MEDICAL DIAGNOSES

A variety of medical diagnosis methods can be proposed to detect or evaluate the extent of myocardial infarction. DE-MRI is a powerful predictor of myocardial viability after coronary artery surgery, suggesting an important role for this technique in clinical viability assessment [29]. From cardiac MRI, the viability can be evaluated thanks to the assessment of left ventricular end-diastolic wall thickness, the evaluation of contractile reserve, and the extent and the transmural nature of the infarction [17, 35]. Indeed, DE-MRI can precisely indicate the severity of the MI, especially in the area of cardiac necrosis. The infarct area has usually a higher intensity than the normal myocardium due to the difficulty in draining the contrast agent in time. The PMO can be characterized by the low-intensity area wrapped by the infarct and touching the endocardium [39].

However, the accuracy is often in conflict with the time required for different diagnosis techniques, which remains a therapeutic emergency. Although DE-MRI is the gold standard for the diagnosis and the evaluation of myocardial infarction [56], the MI diagnosis with DE-MRI cannot be widely applied in the emergency department because of its required time for the acquisition and post-processing. In current practice, simple tools such as ECG, troponin assay and echocardiography are used to validate the emergency diagnosis of MI. ST segment analysis on ECG (especially in case of ST persistent elevation), the intensity of troponin elevation and LVEF (left ventricular ejection fraction) assessment from transthoracic echocardiography (TTE) have been shown highly correlated with MI [51, 76, 104, 140]. Given these facts, when patients arrive in the emergency department complaining of chest pain, generally a series of indicators will be first listed with the help of the above-mentioned simple tools. If the examinations reveal the possibility of MI, the DE-MRI could be achieved in the next few days to have a more accurate evaluation of the myocardial impairment, after the acute phase and the early therapeutic management including revascularization and medications. Until the ultimate diagnosis based on the MRI exam is available, physicians mainly rely on the obtained physiological, clinical and paraclinical characteristics to determine the severity of a patient's condition and to give sound treatment advice.

1.3/ MACHINE LEARNING-BASED MYOCARDIAL INFARCTION ASSESSMENT

In light of the above-mentioned facts, automatic assessment methods are proposed on the basis of the type of the experimental data.

The MI evaluation on DE-MRI with its manual annotations is first studied. To quantify the

severity of MI, deep learning-based segmentation approaches are proposed, consisting of data preprocessing, U-Net-like [69] segmentation models and post-processing. The data preprocessing aims at preparing consistent and high-contrast input MRI slices for the following segmentation networks. The segmentation networks are based on U-Net while adaptive configurations are proposed and compared, including the building blocks and networks, the in-layer normalization, the gradient optimization, the loss function and the framework. The post-processing then eliminates false segmentation areas in raw segmentation masks according to prior information. The final output of the segmentation proposal is the corresponding mask, indicating the category (cavity, normal myocardium, infarct, PMO) of tissue at each area of the segmented DE-MRI slice. The proposal is evaluated on a private DE-MRI dataset. Individual tests are performed on subsets consisting of pathological slices or mixed normal and pathological slices.

The classification of MI is secondly executed on multivariate data, combining the DE-MRI and patient features. The DE-MRI is encoded to the same dimension of the patient features, then the encoded DE-MRI is concatenated to patient features and the concatenation is fed into the classification models to predict if a patient suffers from MI. The goal of this study is to evaluate the impact of additional data sources and the fusion technique of multivariate medical diagnoses.

The severity quantification of MI is finally performed on the 12 pieces of patient features. This proposal also associates the DE-MRI and the patient features but the DE-MRI only participates in the training stage as the ground truth of the quantification. Machine learning-based regression models take the patient features as the input and try to fit the quantification of the severity of the MI. The severity of the MI is calculated from the annotations of the DE-MRI, hence the DE-MRI is not necessary for the inference stage. The proposal aims at obtaining a precise evaluation of MI before the cardiac MRI is available in case of emergency, and it could be the first work that achieves the qualitative prediction of the MI from only patient features.

1.4/ OUTLINE

This thesis is divided into the following chapters:

- Chapter 2 - Clinical context: myocardial infarction. This chapter introduces the clinical context associated with MI. The cardiovascular system will be introduced, the causes and the definition of MI will be then given. The mechanism of conventional MI diagnosis techniques will be also presented.
- Chapter 3 - Challenges for automatic myocardial infarction assessment and available data. After presenting the clinical context of MI, this chapter explains the mo-

tivation of the work and cites the major difficulties and limits of current automatic MI assessment approaches on different medical diagnoses. The experimental data collected in CHU Dijon, which are composed of two datasets: a private one and a public one, are also described.

- Chapter 4 - Related works: Segmentation models and automatic myocardial infarction assessment from MRI. This chapter introduces different aspects of the related works concerning the MI segmentation on cardiac MRI, from the conceptions of statistical models and deep learning models to the segmentation applications on cardiac MRI. Nevertheless, the explicitly studied and evaluated state-of-the-art techniques are presented in the next chapters of contributions.

- Chapter 5 - Contribution 1: Automatic myocardial infarction assessment from MRI with 2D deep learning models. This chapter details the first contribution of this thesis, namely the proposed DE-MRI segmentation approach for MI assessment with 2D deep learning models. This proposal consists of the image preprocessing method, an adaptive U-Net-based model with proposed loss function and well-validated configurations and the prior information-based post-processing. Experiments are performed on all-pathological slices then on slices of all patients, and their performance is compared with a statistical model and the manual annotation variance.

- Chapter 6 - Contribution 2: Automatic myocardial infarction assessment from MRI with 3D deep learning models. 3D models are then proposed to interpret both the intra-slice and inter-slice information in DE-MRI. The backbone of the 3D models refers to the best 2D model introduced in the previous chapter. Three 3D models that have different receptive fields are proposed and a loss function for multi-class segmentation is proposed. The comparative experiments are performed on slices of all patients between the proposed 3D models and the 2D baseline model. Additionally, in this proposal, the persistent microvascular obstruction is individually segmented.

- Chapter 7 - Related works: Automatic myocardial infarction assessment from multivariate diagnoses with machine learning. This chapter introduces the related works concerning the approaches and applications of the MI assessment from multivariate data, mainly the cardiac MRI and the patient features. The objective of these assessment works is the quantification or classification of the MI.

- Chapter 8 - Contribution 3: Automatic myocardial infarction assessment from multivariate diagnoses with machine learning. This chapter describes the third contribution which consists of a MI assessment based on patient features and DE-MRI. Two approaches that make use of multivariate data in distinct ways are proposed.

The first approach predicts the presence of MI from the fused data of the DE-MRI and patient features. The second approach takes the MI volume calculated from the DE-MRI as the ground truth and the MI quantification is made only from the patient features. Besides the infarct, the volume of the persistent microvascular obstruction is also predicted, which was very difficult without the diagnosis of cardiac imaging.

- Chapter 9 - Conclusions and perspectives.

2

CLINICAL CONTEXT: MYOCARDIAL INFARCTION

2.1/ INTRODUCTION

In this chapter, the clinical context related to the myocardial infarction will be presented, including the anatomy and the physiology of the heart, the mechanism of the myocardial infarction and its clinical diagnosis. Understanding the anatomy, physiology and pathology helps to improve the design of automatic assessment models for the disease prediction and the segmentation of the myocardial infarction. The introduction about the fundamental cardiology is mainly based on references from the books Anatomy & Physiology: OpenStax [108] and Medical Terminology for Health Care Professionals [114].

This chapter is organized as follows:

- Cardiovascular system: the heart. The anatomy and the functionalities of the heart will be introduced. This section provides a global view of the heart.
- Myocardial infarction. The coronary circulation will be presented, followed by the pathophysiology of myocardial infarction.
- Medical diagnosis of myocardial infarction. Myocardial infarction diagnoses that are commonly used in medicine will be given, including imaging, clinical and physiological features.

2.2/ CARDIOVASCULAR SYSTEM: THE HEART

The human heart suggests a powerful engine: it keeps the body continually supplied with blood. This section will introduce the fundamental anatomy and functions of the heart.

2.2.1/ LOCALIZATION OF HEART

The human heart is located within the thoracic cavity on the inside of the lungs, in the space known as the mediastinum. The posterior surface of the heart is close to the vertebrae, and its anterior surface sites deep into the sternum and costal cartilage. The great veins, the superior and inferior venae cavae, the great arteries, the aorta as well as the pulmonary artery, are attached to the superior surface of the heart, which is known as the base. The base can be found at the level of the third costal cartilage and the inferior tip of the heart called apex lies left of the sternum between the junction of the fourth and fifth ribs [108]. Figure 2.1 shows the coronal and sagittal views of the heart in the thoracic cavity.

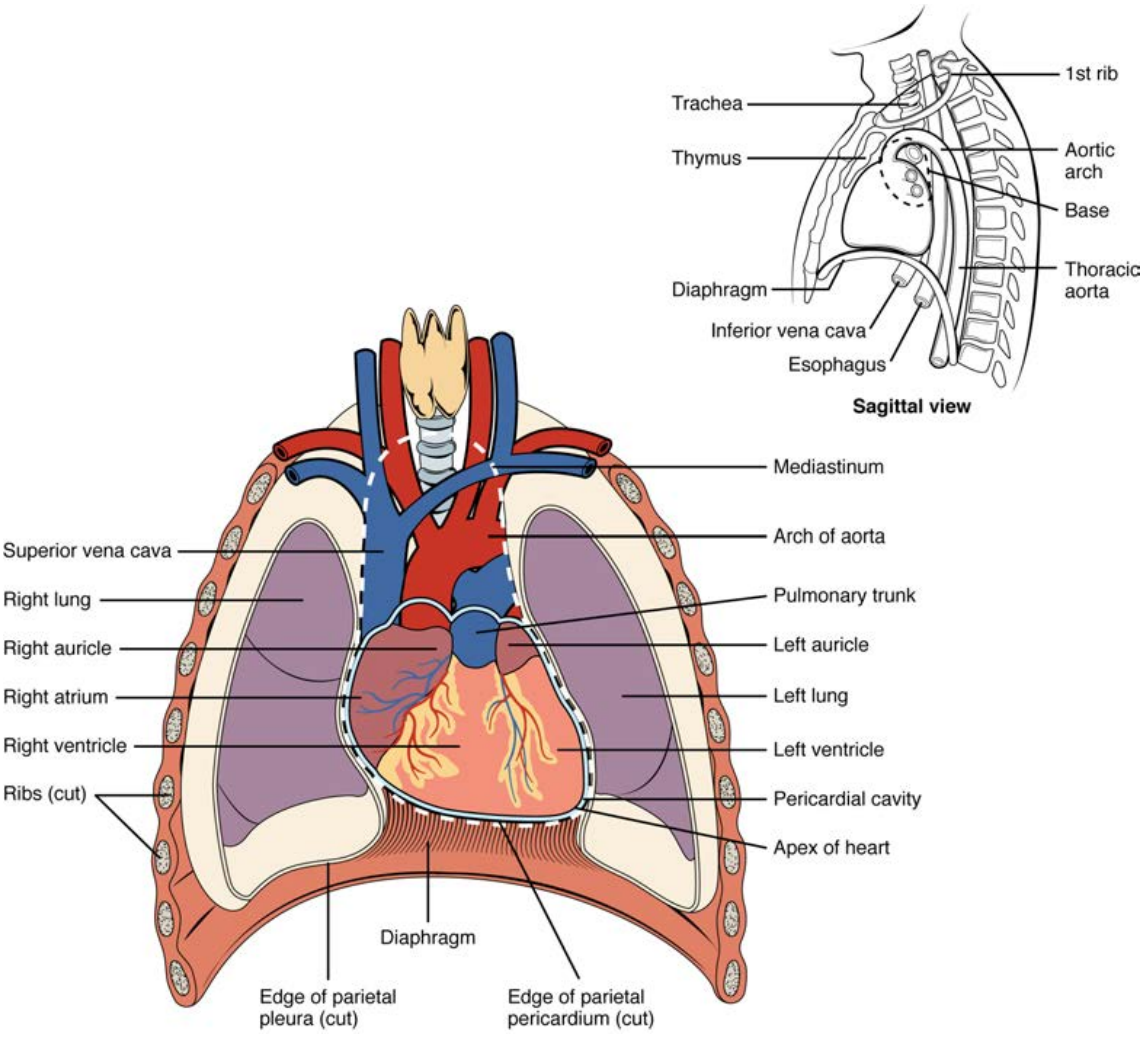


Figure 2.1: The coronal and sagittal views of the heart. (See [108])

The size of the heart has been proven to be proportional to body surface area [12]. A typical heart has approximately the size of a fist, that is on average 12 cm in length, 8 cm in width and 6 cm in thickness. The heart weight differs between sexes: it weighs

approximately 250-300 grams for females and 300-350 grams for males. Meanwhile, the heart can be observed as enlarged for some people. This tendency can be attributed to aerobic sports: the heart responds to activity by expanding the size of individual cells without increasing the number of cells, a phenomenon known as hypertrophy. However, an enlarged heart can result from pathologies, such as hypertrophic cardiomyopathy. The cause of such anomalies is unclear but they could trigger sudden death [108].

2.2.2/ BLOOD CIRCULATION THROUGH CHAMBERS

The four chambers of the human heart are the two atria and the two ventricles. The left and the right atria on the upper of the heart receive blood from organs and contract to push blood into the lower chambers, the right ventricle and the left ventricle. The ventricles are the primary pumps of the heart, blood is then propelled from the ventricles into the arteries.

The extracardiac organs are linked by two distinct circuits, the pulmonary and systemic circuits. The pulmonary circuit transports blood within the lungs so that carbon dioxide can be expelled and oxygen can be picked up. The systemic circuit transports the oxygen-rich blood to all tissues and returns back deoxygenated blood to the heart. The heart connects the two circuits. Deoxygenated blood is transported to the right atrium from the superior and the inferior vena cava and is then propelled to the lungs by the right ventricle via pulmonary arteries. The left atrium receives oxygenated blood from the lungs via pulmonary veins and is then propelled to all tissues by the left ventricle via the aorta. Figure 2.2 illustrates the flow of deoxygenated and oxygenated blood between the pulmonary and systemic circuits.

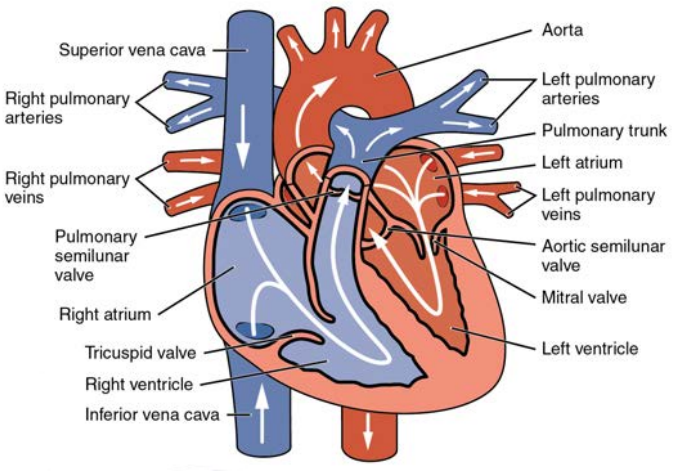


Figure 2.2: Blood flows inside the heart. In blue the deoxygenated blood and in red the oxygenated blood. The upper chambers receive and the lower chambers propel the blood. (Modified from [108])

2.2.3/ ANATOMY OF THE HEART

Knowing that the heart consists of four chambers, more details will be introduced about the internal anatomical structures of the heart.

SEPTA OF THE HEART

The heart is divided by septa (septum in singular) to form the four individual chambers. The septa are physical extensions of the myocardium lined with the endocardium. The interatrial septum separates the two atria and the interventricular septum separates the two ventricles. The interventricular septum is significantly thicker than the interatrial septum to support the higher pressure in ventricles when they contract.

The septum between the atria and ventricles is called the atrioventricular septum. Four valves are presented in the atrioventricular septum to allow a one-way flow of blood between chambers and blood circuits. Due to the weakened structure caused by the presence of the valves, a high-density connective tissue forms and anchors the valves, termed the cardiac skeleton. The cardiac skeleton is formed by four rings surrounding the valves and serves as the point of attachment for the valves. The anatomical structure of septa and valves can be refereed in Figure 2.3.

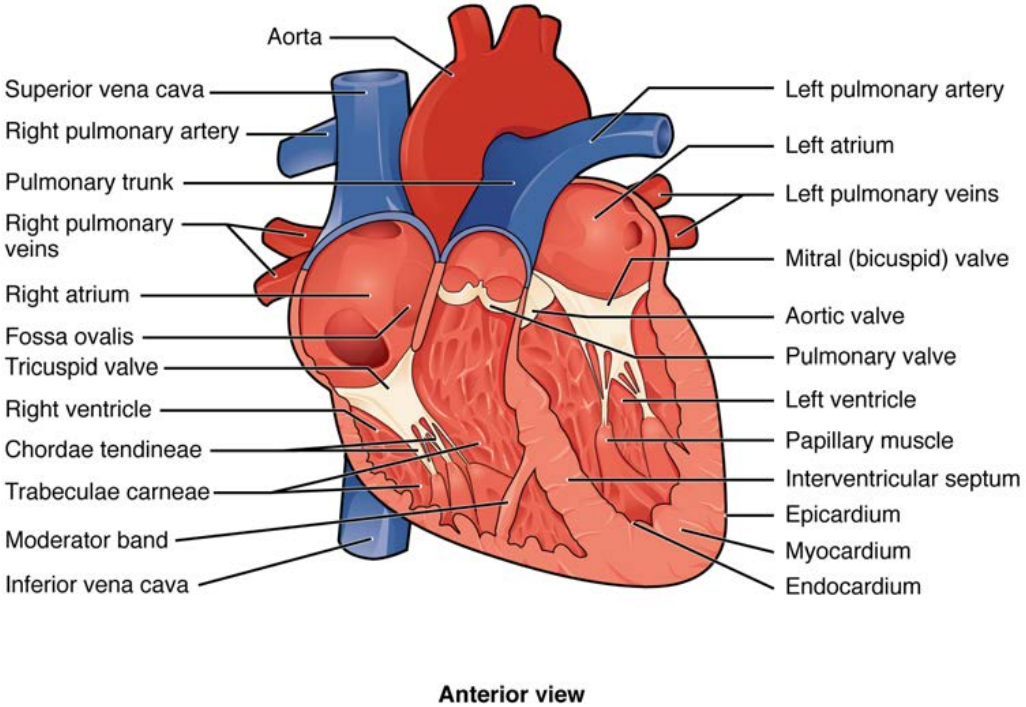


Figure 2.3: Internal anatomical structure of the heart. This anterior view shows the four chambers, the major vessels as well as the valves, and the septa. (See [108])

THE LEFT VENTRICLE

Since in this work, we focus on the assessment of the myocardial infarction in the left ventricle, among the four chambers only the left ventricle is detailed. The left ventricle is the major pumping chamber, dedicated to the systemic circuit. Although the left side and the right side pump the same quantity of blood, the myocardium of the left ventricle is substantially thicker than the one of the right ventricle (Fig. 2.3). Below the opening, several connective tissue called chordae tendineae attaches to the flaps of the mitral valve. The chordae tendineae are composed of collagenous fibers, elastic fibers and endothelium. The other side of the chordae tendineae connects to the papillary muscles. Three papillary muscles are located in the left ventricle, called the anterior, posterior, and septal muscles. When the left ventricle contracts, the pressure within the cavity rises so that the blood flow tends to flow toward the aorta and the atrium. Meanwhile, the blood circuit should be unidirectional. The papillary muscles, along with the chordae tendineae, prevent the mitral valve from inverting or prolapsing during systole: the papillary muscles contract, generating tension on the chordae tendineae during the systole so that the mitral valve is closed. Moreover, on the inner side of the ventricle, rounded or irregular muscular columns called trabeculae carneae protrude from the ventricular wall. One important function of the trabeculae carneae is similar to the papillary muscle: their contraction pulls the chordae tendineae to avoid the backflow. The anatomy of chordae tendineae, papillary muscles and trabeculae carneae can be found in Figure 2.4.

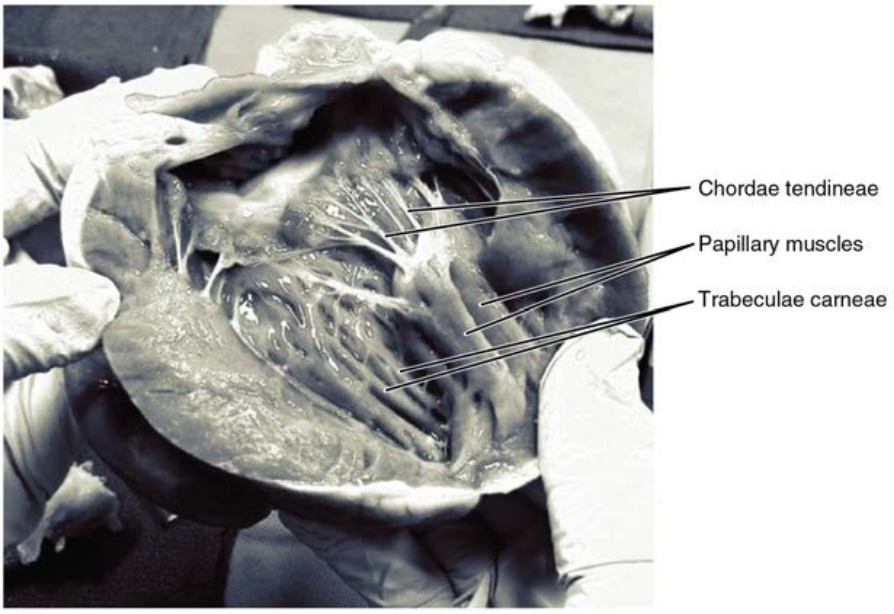


Figure 2.4: Posterior view of the ventricles. The papillary muscles are attached to the tricuspid valve on the right as well as the mitral valve on the left via chordae tendineae. (Credit: modification of work by “PV KS”/flickr.com)

2.2.4/ CARDIAC CYCLE

The cardiac cycle is defined as the period between the beginning of the contraction of the atria and the end of the ventricular relaxation. A cardiac cycle consists of two sub-periods, the systole and the diastole. During systole, the heart contracts and pumps blood into circulation, and during diastole, the heart relaxes and the chambers are filled with blood. Figure 2.5 shows a cardiac cycle including atrial systole, ventricular systole, atrial diastole and ventricular diastole.

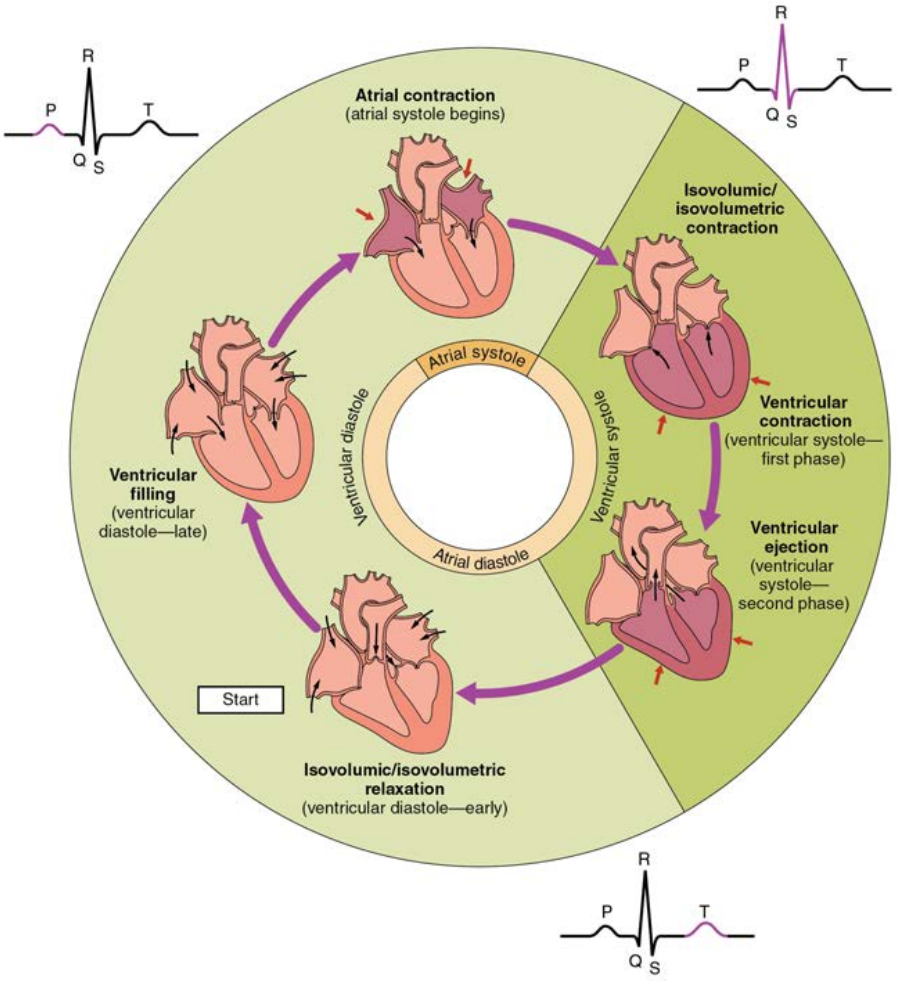


Figure 2.5: A period of the cardiac cycle. The cardiac cycle begins with atrial systole at the end of the ventricular diastole, and then progresses to ventricular systole, atrial diastole, and finally ventricular diastole when the cycle begins again. Correlations to the ECG are highlighted. (See [108])

2.2.5/ BLOOD PRESSURES

The blood, like all fluids, follows the pressure gradients that it moves from high-pressure regions to the ones of lower pressure. In a heart cycle, when all the chambers are relaxed

(early ventricular diastole), blood flows from the veins into the atria. At this stage, due to the higher pressure in the atria, the blood will initially move passively from the atria into the ventricles. When the action potential triggers the muscles in the atria to contract, the atrial pressure will be further increased and the blood will be actively pumped into the ventricles, then the tricuspid and the mitral valves close and the cardiac cycle passes to ventricular systole. The ventricular contraction further increases the pressure in the ventricles so that the blood will be ejected into the pulmonary trunk and the aorta.

Blood pressure changes during a cardiac cycle, which is one of the vital signs. Blood pressure often refers to the aortic pressure. The systolic blood pressure is the highest pressure in a cardiac cycle and the diastolic pressure is the lowest.

2.3/ MYOCARDIAL INFARCTION

Myocardial infarction (MI) has become one of the most common cardiovascular diseases in the emergency department [64]. In 2015, about 15.9 million cases of MI were reported worldwide [84]. The management of MI is also complex. According to a Statistical Brief of the year 2011, MI was one of the top five most expensive conditions during inpatient hospitalizations in the US, with a cost of about $11.5 billion for 612,000 hospital stays [33]. MI occurs as a result of myocardial distress leading to a death of myocardial tissue, usually caused by the lack of blood flow and oxygen to a region of the heart. When a coronary artery is blocked, it frequently results in a heart attack. Therefore, the coronary circulation and the mechanism of MI are introduced in this section.

2.3.1/ CORONARY CIRCULATION

The heart is majorly composed of cardiac muscle cells, hence a reliable supply of oxygen and nutrients are critical to the cardiomyocytes. The coronary circulation is dedicated to providing the necessary supplies for cardiomyocytes and removing wastes from them. The coronary circulation shares the same rhythm as the heartbeat: the coronary circulation reaches a peak when the myocardium is relaxed and ceases while it is contracting.

In coronary circulation (Figure 2.6), the coronary arteries transport oxygenated blood to the myocardium and other components of the heart within the two main coronary arteries. The left and right coronary arteries are supplied from the aorta. The aortic sinuses in the wall of the aortic root give rise to the coronary arteries. The epicardial coronary arteries are the vessel branches of the coronary arteries that remain on the surface of the heart and follow the heart sulcus.

The left coronary artery transports blood to the left side of the heart and to the interven-

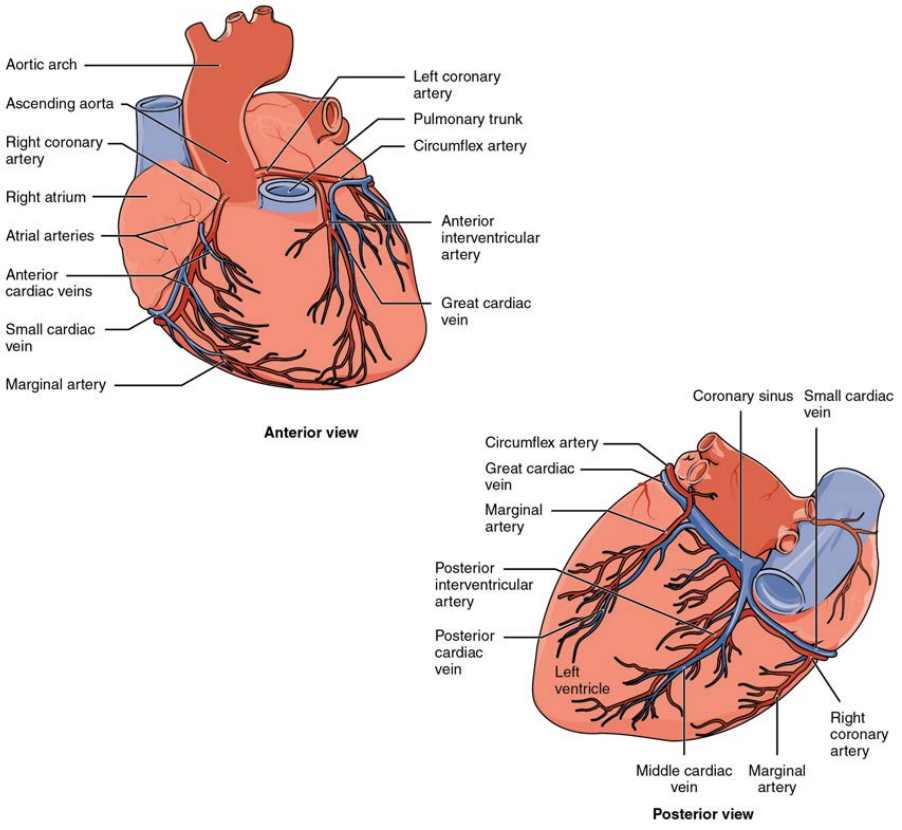


Figure 2.6: Anterior and posterior views of prominent coronary surface vessels.

tricular septum. From the left coronary artery, the circumflex artery is arisen and follows the heart sulcus to left, and eventually gives rise to one or more left marginal arteries. The second major branch arising from the left coronary artery is called the anterior interventricular artery. The right coronary artery's major branch is the posterior interventricular artery. It runs along the posterior part of the interventricular sulcus towards the apex of the heart. The anterior and posterior interventricular arteries give rise to numerous minor branches that interconnect with each other, forming anastomoses. In anastomoses, the vessels interconnect and blood from different branches may circulate to a specific region.

The coronary veins, which run parallel to the main surface arteries, return deoxygenated blood from the heart to the right atrium. It can be concluded that venous blood flow in the coronary veins is of three types: draining into the coronary sinus, direct draining into the right atrium, and direct draining into four chambers (venae cordis minimae). Most venous blood is circulated to the coronary sinus, which is a large, thin-walled vein on the posterior surface of the heart, located in the atrioventricular sulcus and draining directly into the right atrium. The great cardiac vein, middle cardiac vein, small cardiac vein and posterior vein of the left ventricle are the principal coronary veins connected to the coronary sinus. The anterior cardiac veins parallel the small cardiac arteries and drain the anterior surface of the right ventricle into the right atrium bypassing the coronary sinus.

2.3.2/ MECHANISM OF MYOCARDIAL INFARCTION

According to the Fourth Universal Definition of Myocardial Infarction (2018) [116], MI is clinically defined as the presence of acute myocardial injury detected by abnormal cardiac biomarkers in the setting of evidence of acute myocardial ischaemia. Pathologically, MI occurs when myocardial cells die due to prolonged ischaemia. Patients who suffer from typical MI may have the symptoms of fatigue, chest pain or malaise in the days or weeks preceding the event, and typical ST-elevation myocardial infarction (STEMI) may occur more rapidly without showing any symptoms.

Although most of MI results from blocking, the mechanism can be different. Therefore, MI is divided into five types by the Fourth Universal Definition of Myocardial Infarction:

- Type 1 (spontaneous MI): MI caused by atherothrombotic coronary artery disease. Atherosclerosis and thrombosis are the culprits of the MI type 1. Atherosclerotic plaque disruption or intraluminal thrombus in one or more of the coronary arteries decreases or even obstructs the blood flow and thereby results in myocyte necrosis. Figure 2.7 shows plaque disruption and thrombus in a coronary artery.

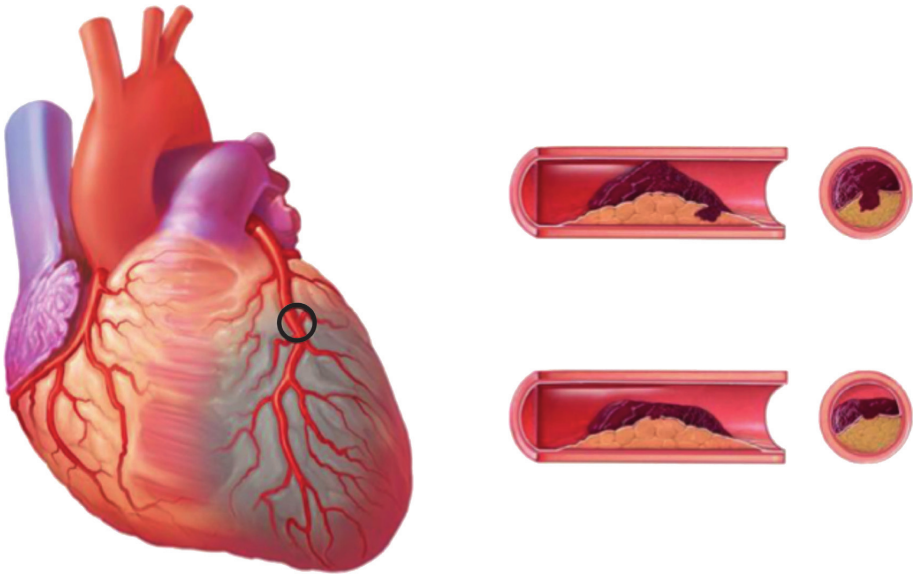


Figure 2.7: Cause of MI type 1: Plaque disruption and thrombus in a coronary artery. (See [116])

- Type 2 (MI related to an ischemic imbalance): a mismatch between oxygen supply and demand leads to ischaemic myocardial injury. The culprit diseases of MI type 2 may be coronary endothelial dysfunction, coronary artery spasm, coronary artery embolus, tachyarrhythmias/bradyarrhythmias, anaemia, respiratory failure, hypertension, or hypotension.

- Type 3 (MI resulting in death when biomarker values are unavailable): patients suffering from MI type 3 can manifest a typical presentation of myocardial ischaemia or infarction, including ischaemic electrocardiogram (ECG) changes or ventricular fibrillation, while the elevation of biomarker values can not be detected before the death of patients. The relative biomarker is not available because the patient succumbs before values are measured or is struck by sudden death with evidence of MI by autopsy.

- Type 4a (MI related to percutaneous coronary intervention): elevation of biomarker values (cardiac troponin (cTn) is recommended) to more than 5 times the 99th percentile of the upper reference limit (URL) in patients with normal baseline values ($<$ 99th percentile URL) or a rise of values over 20% if the baseline values are elevated but stable or falling. In addition, any of the following are required: (1) symptoms suggestive of myocardial ischemia; (2) new ischemic ECG changes or new bundle branch block; (3) angiographic loss of patency of a major coronary artery or a side branch or persistent slow flow or no flow or embolization; or (4) demonstration of the new loss of viable myocardium or new regional wall motion abnormality by cardiac imaging [127].

- Type 4b (MI related to stent thrombosis): MI associated with stent thrombosis as detected by coronary angiography or autopsy in the setting of myocardial ischemia in combination with a rise and/or fall of cardiac biomarkers with at least one value above the 99th percentile URL.

- Type 5 (MI related to coronary artery bypass grafting): elevation of cardiac biomarker values more than 10 times the 99th percentile URL in patients with normal baseline cTn values. In addition, either (1) new pathologic Q waves or new BBB, (2) angiographic-documented new graft or native coronary artery occlusion, or (3) evidence of new loss of viable myocardium or new regional wall motion abnormality by cardiac imaging is required.

2.4/ MEDICAL DIAGNOSIS OF MYOCARDIAL INFARCTION

MI is diagnosed when either of the following two criteria is met [116]:

- Detection of a significant change of cardiac biomarker values and with at least one of the findings including symptoms of ischemia, new or presumed new significant ST-segment-T wave (ST-T) changes or new left bundle branch block, development of pathological Q waves, evidence of new myocardium damage with imaging, or identification of an intracoronary thrombus by angiography or autopsy.

- Cardiac death with symptoms suggestive of myocardial ischaemia, with presumed new ischaemic changes or injury on ECG or new bundle branch block, but death occurs before cardiac biomarker levels are obtained or before cardiac biomarker values increase.

Therefore, a variety of medical diagnoses can be performed according to the patient's conditions and symptoms:

- Electrocardiography.
- Echocardiography.
- Cardiac MRI.
- Clinical and physiological features via blood tests.

2.4.1/ ELECTROCARDIOGRAPHY

ELECTRICAL ACTIVITY OF THE HEART

Cardiac muscle cells consist of two types: myocardial contractile cells and myocardial conducting cells. Contractile cells conduct impulses and contract to pump blood throughout the body. Myocardial conducting cells initiate and propagate the action potential throughout the heart.

An individual cardiac muscle cell can generate its own electrical impulses followed by contraction. The autorhythmicity of a myocardial cell allows it to initiate an electrical potential at a fixed rate that spreads rapidly from cell to cell. When more cells are joined and beat together, the pace is set by the cell with the highest inherent rate. The impulse spreads from the faster to the slower cells to trigger a contraction. The conduction system of the heart includes the sinoatrial node, the atrioventricular node, the atrioventricular bundle, the atrioventricular bundle branches, and the Purkinje cells (Figure 2.8).

The sinoatrial node is a group of cells located in the wall of the right atrium, establishing the normal cardiac rhythm. The sinoatrial node has the highest inherent rate of depolarization and initiates an electrical impulse, which is known as the pacemaker of the heart. This impulse travels across the atria via internodal routes from the sinoatrial node to the atrial myocardial contractile cells and the atrioventricular node. Between the two nodes, the impulse takes around 50 ms. Apart from the internodal routes, the impulse can be also conducted from the right atrium to the left atrium via the interatrial band. Figure 2.9 illustrates the initiation of the sinoatrial node impulse and its transmission to the atrioventricular node.

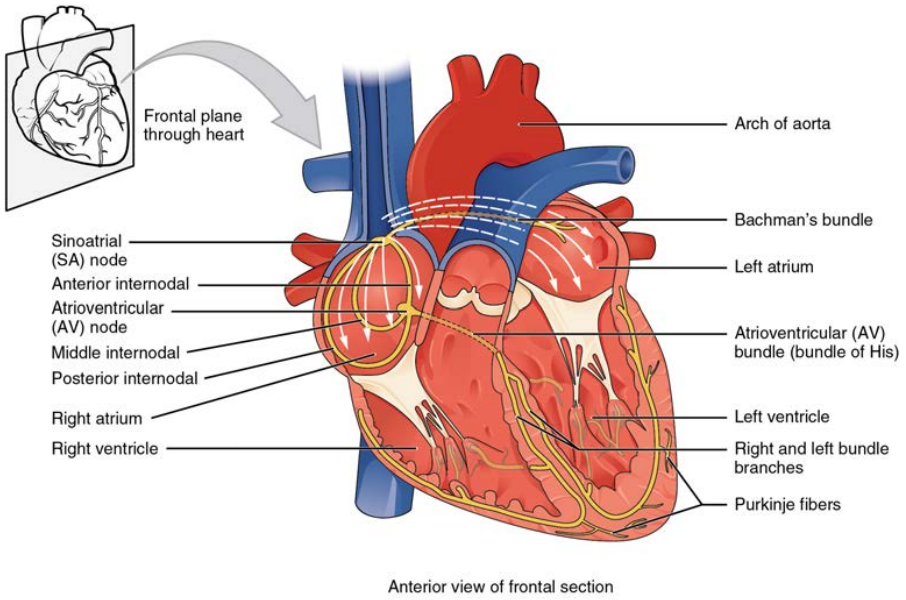


Figure 2.8: Conduction system of the heart. Specialized conducting components of the heart include the sinoatrial node, the internodal pathways, the atrioventricular node, the atrioventricular bundle, the right and left bundle branches, and the Purkinje fibers. (See [116])

The muscular contraction is triggered by the wave of depolarization, beginning from the right atrium. As the impulse spreads across the superior portions of both atria and then down through the contractile cells, the contractile cells begin contraction from the superior to the inferior portions of the atria and pump blood into the ventricles.

The atrioventricular node is the channel for excitation to pass from the atria to the ventricles under normal conditions. Its anterior end emits the atrioventricular bundle and is positioned in the deep endocardial surface of the Koch triangle in the right atrium. Its primary role is to quickly convey excitement from the sinoatrial node to the ventricles, ensuring that the atrial contraction is followed by the onset of ventricular contraction.

The atrioventricular bundle, or bundle of His, begins at the AV node and travels across the interventricular septum before splitting into two atrioventricular bundle branches, generally referred to as the left and right bundle branches. There are two fascicles on the left bundle branch. The left bundle branch provides blood to the left ventricle, while the right bundle branch provides blood to the right ventricle. The left bundle branch is significantly larger than the right because the left ventricle is much larger than the right. The right bundle branch supplies the right papillary muscles and is situated in the moderator band. Because of this connection, each papillary muscle receives the impulse at about the same moment, causing them to contract simultaneously just before the remainder of the myocardial contractile cells in the ventricles.

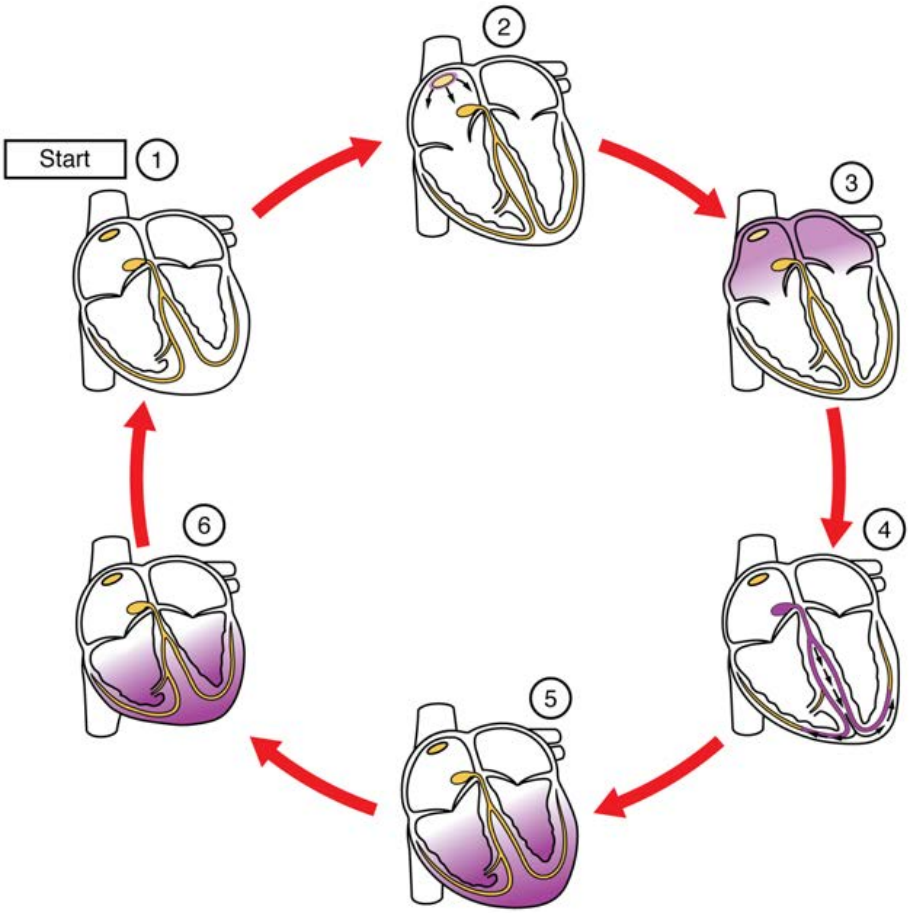


Figure 2.9: Cardiac conduction initiated by the sinoatrial node. (1) The rest of the conduction system is at rest, including the sinoatrial (SA) node. (2) The action potential is started by the SA node and travels through the atria. (3) After reaching the atrioventricular node, a 100 ms delay before the impulse is conveyed to the atrioventricular bundle, allowing the atria to finish pumping blood. (4) The impulse goes through the atrioventricular node after the delay. The moderator band connects the bundle and bundle to the Purkinje fibers, as well as the right papillary muscle. (5) The impulse travels to the ventricle's contractile fibers. (6) The contraction of the ventricles begins. (See [116])

ELECTROCARDIOGRAM

The electrical signal of the heart can be captured by placing surface electrodes on the body. The tracing of the heart's electrical signal is the ECG. Conventional ECG uses 3, 5 or 12 leads while diagnosis is more accurate if a greater number of leads is employed.

The ECG signal has five prominent points: the P wave, the QRS complex, and the T wave as shown in Figure 2.10. The small P wave represents the depolarization of the atria, *i.e.* the membrane potential of atrial myocardium rapidly shifts from negative to positive. The atria begin contracting approximately 25 ms after the start of the P wave. The large QRS complex represents the depolarization of the ventricles followed by the ventricular systole. Due to the larger size of the ventricular myocardium, the depolarization produces a much

stronger electrical signal. When the QRS reaches the peak of the R wave, the ventricles begin to contract. The QRS may be conveniently divided into two phases (isovolumic contraction and ventricular ejection), approximately lasting a total of 270 ms. The T wave represents the repolarization of the ventricles followed by the ventricular relaxation, that is, as the positive charge moves out of the cell the potential within the cell decreases and approaches its resting potential once more (Figure 2.5).

The diagnosis of MI depends on the segments and intervals on ECG tracing. Segments are defined as the regions between two waves. Intervals include one segment plus one or more waves (Figure 2.10). An enlarged Q wave may indicate a MI. An elevation of the ST segment (the segment that connects the QRS complex and the T wave in electrocardiogram) above baseline is often seen in patients with an acute MI, and may appear depressed below the baseline when hypoxia is occurring. Therefore, besides the MI type definition in Section 2.3.2, another MI classification scheme can be identified according to the ECG exam: ST-elevated MI (STEMI) and non-elevated MI (non-STEMI).

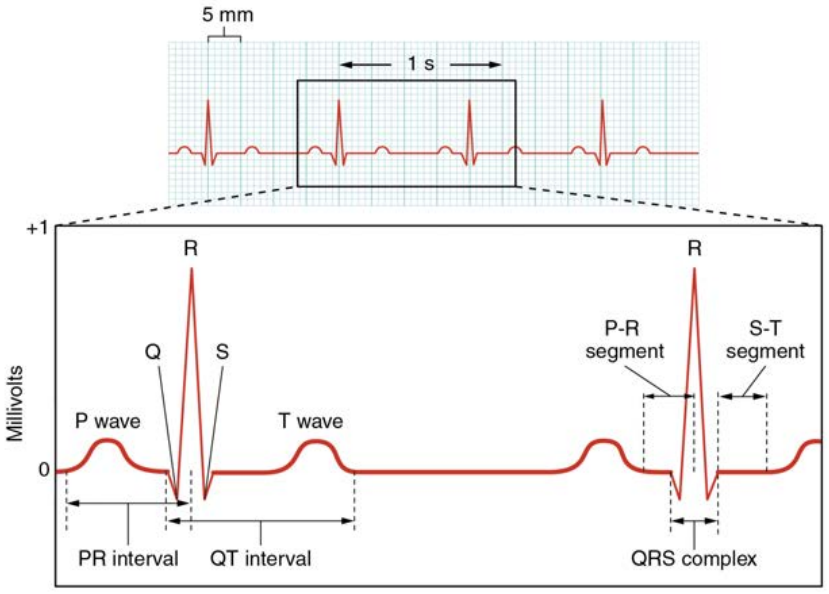


Figure 2.10: Illustration of waves, intervals and segments on a normal ECG tracing. (See [108])

2.4.2/ ECHOCARDIOGRAPHY

Echocardiography is the standard ultrasound or Doppler ultrasound exam of the heart. Depending on the device, maneuver and patient conditions, several different types of echocardiography can be performed such as transthoracic echocardiography, transesophageal echocardiography, stress echocardiography, three-dimensional echocardiography and fetal echocardiography. Transthoracic echocardiography is the most common type that consists of a transducer placed on the patient's chest. The transducer emits

ultrasound toward the heart and the reflected ultrasound is interpreted to form images of the heart. The signal acquisition is performed at high frequency therefore the echocardiogram can be visualized in kinetic format. More analyses can then be conducted within the images that describe the size and shape of the heart, pumping capacity, and the location and extent of any tissue damage.

From the echocardiogram, one important measurement that can be assessed is the ejection fraction. The ejection fraction expresses the quantity of blood the left ventricle pumps out during each contraction. A normal left ventricle ejection fraction (LVEF) may be between 50 and 65 percent [60], and Maceira *et al.* [32] calculated the mean LVEF being 67% (± 4.6%) in a sample of 120 subjects with no known risk factors or history of cardiac disease, and normal physical examination and ECG. Damage to myocardium may reduce the heart's performance. The partial loss of the heart function can be reflected in the ejection fraction value. However, ejection fraction abnormality may be associated with various pathologies, not only MI but also, for example, cardiomyopathy.

2.4.3/ CARDIAC MRI

MECHANISM OF MRI

MRI is a non-ionizing radiation medical imaging technology, based on nuclear magnetic resonance (NMR). Knowing that certain atomic nuclei are able to absorb radio frequency energy when placed in an external magnetic field, the resultant evolving spin polarization can induce a radio frequency (RF) signal in a radio frequency coil and thereby be detected [15]. NMR results from specific magnetic properties of certain atomic nuclei. The abundance of certain nuclei differentiates an object's internal structure, with hydrogen being the most common nucleus employed to generate signals of the body.

The decay of RF-induced NMR spin polarization, termed relaxation, is another factor that differentiates tissues. At equilibrium, nuclear spins process randomly around the direction of the applied field, but suddenly become in a phase when any resulting specific RF wave is orthogonal to the field. The RF pulses cause the spin-state population to be perturbed from its equilibrium value. The return of the longitudinal component of the magnetization to its equilibrium value is known as spin-lattice relaxation, while the loss of phase coherence of the spins is known as spin-spin relaxation. The longitudinal (spin-lattice) relaxation time is called T_1 and the transverse (spin-spin) relaxation time is called T_2. Higher contrast can be achieved by using suitable weighting on the basis of the tissue to be examined. Generally, T_1 weighting is better for identifying fatty tissue and for post-contrast imaging, while T_2 weighting is useful for detecting edema and inflammation. Moreover, the MRI sequence is a particular set of pulse sequences and pulsed field gradients. Different sequences may result in particular image appearances [161].

To perform a clinical MRI study, the patient is first positioned within an MRI scanner where the magnetic field is constantly strong (up to 3T for medical purposes). The gradient coils cause an overlapping gradient magnetic field so that the spatial position can be determined. Then an oscillating magnetic field is triggered to the patient at the resonance frequency. The excited atoms emit an RF signal, which is measured by a receiving coil. The contrast between different tissues is determined by the rate at which excited atoms return to the equilibrium state (T_1 weighting), or by the rate at which excited atoms reach equilibrium or go out of phase with each other (T_2 weighting).

During the image acquisition, each time only one slice can be captured by MRI and a slice refers to the resonance signal from a narrow 3D space. Therefore, to model the heart, the position of the heart should be first determined, then the signal of the spacing covering the heart will be intercepted slice by slice with a small gap between slices.

CARDIAC MRI MODALITIES

Different cardiac MRI (CMRI) modalities can be employed depending on the purpose of diagnosis. Cine CMRI can capture the movement of the heart. Despite cine CMRI specializing in reflecting the movement of the heart and examining the cardiac function, MI can be also characterized by DE-MRI. DE-MRI forms static images, meanwhile, the contrast between normal and pathological tissues is hyperenhanced with a contrast agent. The contrast agent is usually gadolinium-based. Normal, ischemic, and fibrotic or necrotic myocardial tissues wash out differently the contrast agent. Normal tissue shows lower delayed accumulation due to its higher washout rate, whereas ischemic or infarcted tissue retains higher post-enhancement due to its lower washout rate. Necrotic tissue exhibits the slowest rate of absorption and little contrast agent can be accumulated. The tissues' washout characteristics differentiate their vitality when the acquisition is appropriately timed, which is usually 7 to 10 minutes after the injection of the contrast agent. The typical pulse sequence for myocardial delayed enhancement is a segmented inversion-recovery–prepared fast gradient-echo sequence. The imaging occurs in a breath-hold while some free-breathing techniques can be employed for patients who have difficulty on breath-holding.

One challenge of the MRI acquisition is the motion caused by respiratory and contraction of the myocardium [128]. Facing such a challenge, the gating methods should be employed during the image acquisition. Firstly, to avoid the shift due to the contraction, the cardiac self-gating [28] or ECG gating approaches synchronize the heartbeat rhythm. The gating can be prospective (*e.g.* DE-MRI) or retrospective (*e.g.* cine CMRI). The prospective gating captures images by each R-wave and stops once the estimated number of cardiac phases has been gathered. This gating method, therefore, produces a short

period of no image acquisition and the achieved images refer to desired phases. The retrospective gating triggers image acquisition in a continuous way throughout the whole cardiac cycle. The image segments from the various R-R intervals are interpolated onto a computed average length R-R interval so that the motion in an entire cardiac cycle can be assessed. Secondly, to avoid respiratory motion, patients are usually required to hold their breath. Alternatively, for people who have difficulty holding their breath, respiratory gating can be performed together with cardiac gating. Moreover, fast pulse sequences ensure multiple echoes during a single heart cycle such as Steady-state free precession (SSFP) and its variant of balanced gradients, commercially termed TrueFISP (True fast imaging with steady-state precession) by Siemens [20].

Cine CMRI can be performed with or without the injection of a contrast agent, while DE-MRI should be acquired around 10 minutes after the injection. The two modalities can be performed one after another during the same exam to assess both the heart function and the extent of MI. The essential difference is the cardiac gating strategy, the prospective triggering or the retrospective triggering. The cine CMRI consists of a series of static images that covers a full cardiac cycle to reflect the heart motion and the employed sequence is usually SSFP. DE-MRI is usually performed with Phase Sensitive Inversion Recovery (PSIR) sequence that restores the signal polarity, thus avoiding loss of contrast and providing a consistent image appearance without polarity artifacts for cases where the inversion time is set too early. The phase-sensitive reconstruction method dramatically reduces the variation in apparent infarct size which is observed in the magnitude images as the inversion time is changed [24]. Figure 2.11 shows examples of cine CMRI and DE-MRI.

CARDIAC IMAGING PLANES

To better visualize the anatomy of the heart via cardiac MRI, the imaging planes should be correctly adjusted. The cardiac imaging planes are standard orientations for displaying the heart, in reference to the long axis of the left ventricle.

The initial planes are orthogonal, symmetrically separating the body (Figure 2.12). However, the initial planes are not optimized to visually assess the MRI. Figure 2.13 shows MRI at sagittal, coronal and transverse planes covering a part of the heart. To facilitate the visualization, the MRI acquisition should be adjusted to the cardiac imaging planes consisting of the short-axis, the horizontal long-axis and the vertical long-axis as shown in Figure 2.11 and Figure 2.14.

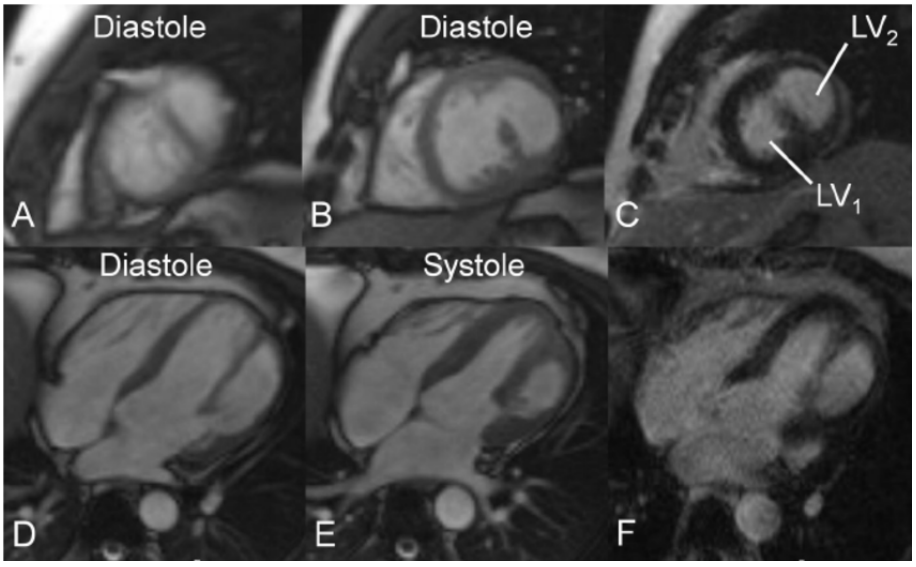


Figure 2.11: Examples of cine CMRI (A, B, D and E) and DE-MRI (C and F). A, Cine CMRI, short-axis view at the apical portion of the left ventricle showing both cavities. B, Cine CMRI, short-axis view at the middle portion of LV showing both cavities. C, DE-MRI, short-axis view at the middle portion of LV without scar/fibrosis. D, Cine CMRI four-chamber view at diastole showing both cavities. E, Cine CMRI, four-chamber view at systole showing the thickening of the lateral wall of the LV 2 . F, DE-MRI, four-chamber view without scar/fibrosis (See [40]).

2.4.4/ Clinical and physiological features

Cardiac troponin (cTn) is part of the troponin complex, which binds to actin in fine muscle filaments and holds the actin-myosin complex in place. Damage to the heart muscle causes cTn to be introduced into the blood, hence measuring changes in the concentration of cTn in the blood can reflect the extent of an acute heart disease. An acute myocardial infarction was diagnosed in cases of a cTn value increase or decrease pattern with at least 1 value 30 $>$ ng/L (99th percentile upper reference limit) together with myocardial ischemia. It is important to recognize that cTn is not entirely specific for myocardial damage secondary to infarction [47]. Elevation of cTn without signs or symptoms indicating overt cardiac ischemia is thought as myocardial injury [82].

The European Society of Cardiology and the American College of Cardiology collaborated to report that myocardial injury detected by abnormal biomarkers in the setting of acute myocardial ischaemia should be labeled as MI [19]. Furthermore, elevated cTn value is frequently encountered and associated with an adverse prognosis [82]. Therefore, according to the Fourth Universal Definition of MI, cTn value is recommended as the primary biomarker, which is crucial for identifying non-fatal MI.

N-terminal prohormone of brain natriuretic peptide (NT-proBNP) is a prohormone cleaved from the molecule to release brain natriuretic peptide. The plasma concentrations of

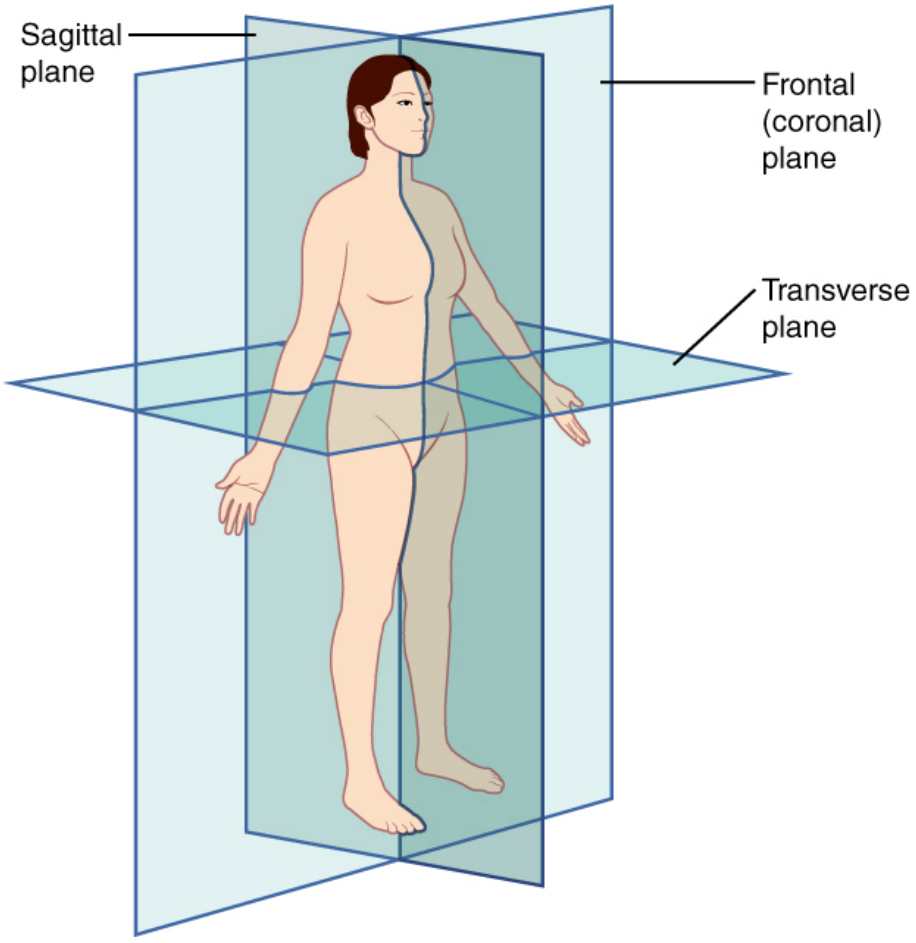


Figure 2.12: Initial anatomical planes for medical imaging (See [108]).

NT-proBNP is typically increased in patients with asymptomatic or symptomatic left ventricular dysfunction and is associated with coronary artery disease and myocardial ischemia [27]. Although NT-proBNP is not the priority biomarker for MI assessment, its value change provides an additional indicator for a more comprehensive analysis of suspected MI.

Besides the biomarkers, the Killip classification system is also used to evaluate the risk of mortality for individuals with an acute MI. Killip evaluates the functional severity based on clinical evidence of heart failure or shock. Four levels are ranked according to the following criteria as the results of treatment of 250 patients with established acute myocardial infarction [1]:

- Killip I: no heart failure. No clinical signs of cardiac decompensation.
- Killip II: heart failure. Diagnostic criteria include rales, S_3 gallop and venous hypertension.
- Killip III: severe heart failure. Frank pulmonary edema.

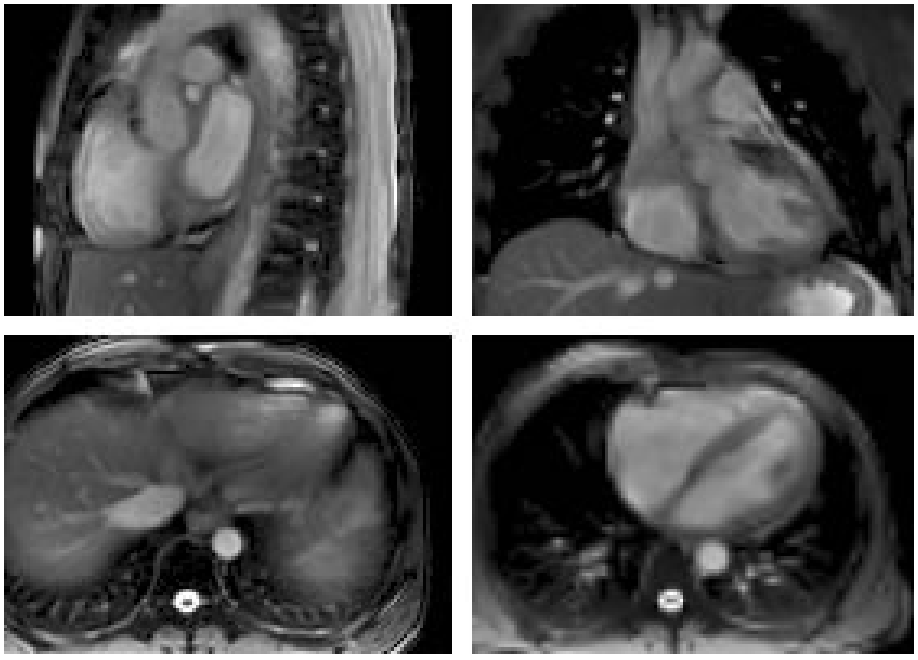


Figure 2.13: Cardiac MRI at initial planes. In the first row the sagittal (left) and coronal (right) planes, at the second row two slices at the transverse plane.

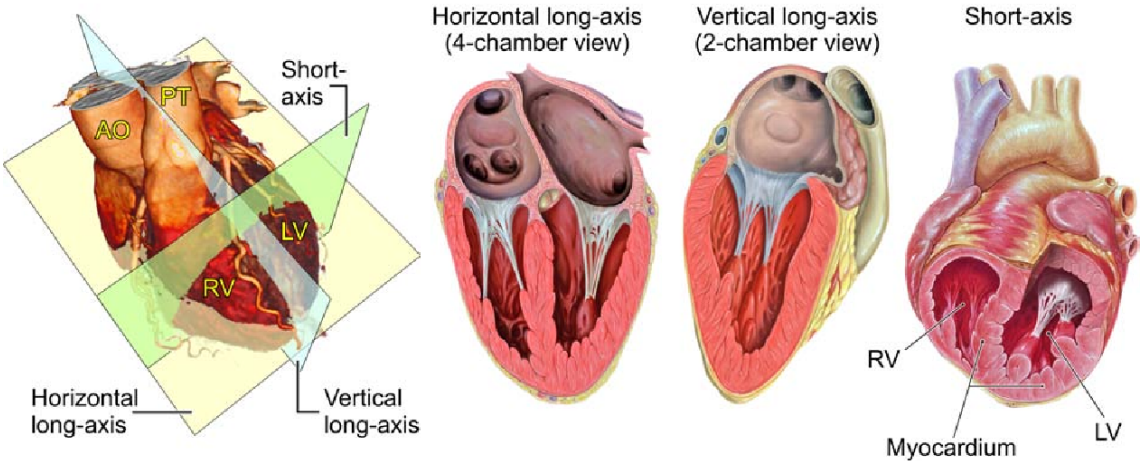


Figure 2.14: Cardiac imaging planes and associated standard views. (a) The three standard cardiac imaging planes are located along the major axes of the heart: horizontal long-axis, vertical long-axis and short-axis. AO: aorta, PT: pulmonary trunk, LV: left ventricle, RV: right ventricle. (b) The images acquired along the major axes show all 4 heart chambers, 2 selected chambers (one ventricle and the associated atrium), or a frontal view of the myocardium (short-axis view) [63].

- Killip IV: cardiogenic shock. Signs include hypotension (systolic pressure of 90 mm. Hg or less) and evidence of peripheral vasoconstriction such as oliguria, cyanosis and diaphoresis. Heart failure, often with pulmonary edema, has also been present in the majority of these patients.

2.5/ CONCLUSIONS

This chapter introduced the fundamental of heart and myocardial infarction with its first-line diagnosis techniques. MI is usually associated with coronary diseases and it has been one of the most common cardiovascular pathologies. According to guidelines of cardiology societies, biomarkers and imaging diagnoses are recommended to evaluate the MI, of which the cardiac troponin and the DE-MRI are considered to be among the most robust diagnostic techniques.

3

Challenges for automatic MI assessment and available data

3.1/ Introduction

The increasing number of patients suffering from MI requires significant clinic resources for the diagnosis, therapy and follow-up treatment. Therefore, this thesis focuses on the automatic assessment of MI from different medical diagnosis modalities. With the help of machine learning algorithms, this work will hopefully improve the diagnosis accuracy and reduce the workload of the doctors. Particularly in less medically developed countries, well-proven machine learning algorithms can be effective in improving regional medical conditions.

According to the Fourth Universal Definition of Myocardial Infarction (2018) [116], the European Society of Cardiology and the American College of Cardiology [19], biomarkers, imaging and ECG are recommended to diagnose MI. Therefore, inspired by the recommendation of cardiac societies and current clinical practice, the proposed automatic assessment of MI in this work is based on DE-MRI and patient features obtained from ECG, transthoracic echocardiogram (TTE) and blood tests and clinical information. Therefore, in this chapter the challenges to be overcome using the recommended medical diagnosis data for the automatic MI assessment will be elaborated. Meanwhile, the experimental datasets consisting of DE-MRI and patient features will be detailed.

3.2/ Myocardial infarction assessment from MRI

Cine MRI and DE-MRI are the common CMRI to assess MI. Cine MRI reconstructs the heart motion within a single heartbeat cycle, while DE-MRI captures static images with higher contrast between normal and pathological tissues. This work particularly concentrates on DE-MRI as the diagnostic imaging.

Since the heart pumps blood mainly by the contraction of the ventricles, and the left ventricle has a significantly thicker myocardium than the one in the right, the imaging assessment of MI can be better performed on the left ventricle myocardium. In CHU Dijon, about 80% cardiac imaging diagnoses examine the area of the left ventricle. According to the differential perfusion and washout characteristics on the basis of tissue viability, the concentration rate of the contrast agent in different tissues indicates the state of myocardium. When the image acquisition is performed at an appropriate time, the normal myocardium will show low agent concentration because of its fast washout rate, the infarct will show high agent concentration because of its delayed perfusion and the more significant accumulation of the agent in the expanded space between fibrotic myocardium, and the persistent microvascular obstruction (PMO) will show low agent concentration because the agent can not pass through the scar tissue [34, 35].

The severity evaluation of MI on DE-MRI depends on the surface of the infarct and PMO compared to normal myocardium. The correct delineation is thus crucial to precisely quantify the severity. In current clinical practice, the delineation is manually performed by physicians. This procedure is tedious and time-consuming. Furthermore, the accuracy depends on the expert's experience and can be biased by some subjective understandings. To reduce the bias caused by subjective factors and free physicians from repetitive work, an automatic segmentation method should be proposed to delineate myocardial tissues according to the uptake of contrast agents. The automatic delimitation indicates the position and size of pathological tissues and therefore both the qualitative and quantitative evaluations of MI can be obtained.

Although DE-MRI is recognized as the gold standard imaging tool for MI assessment [29], the interpretation of the image is still challenging. The challenges can result from the vertical position of the slice, the fatty tissue wrapping the epimyocardium, the presence of PMO, *etc.* Figure 3.1 shows five selected DE-MRI slices on short-axis orientation, covering completely left ventricle. The details of each slice are:

1. Slice A is close to the apex. Therefore only a small part of the right ventricle appears in this slice (blue arrow).
2. Slice B involves an infarct that connects the cavity (blue arrow).
3. Scar tissues in slice C have a broken shape: On the upper side, the scar tissues and the normal myocardium intersperses.
4. Slice D involves an important PMO area.
5. Slice E involves an artifact (blue arrow).

Apart from the difficulties in the images described in Figure 3.1, the movement due to breathing or heart contractions may also render the images noisy, despite the use of

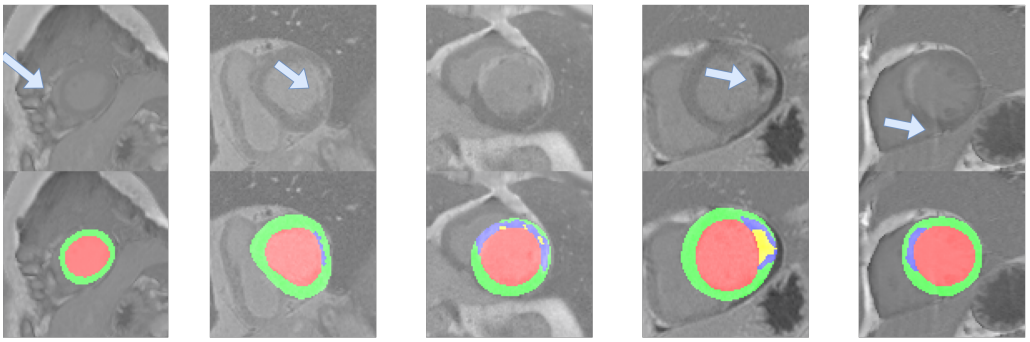


Figure 3.1: Five challenging slices of short-axis DE-MRI with manual annotations of different tissues. Blue arrows specifically highlight difficult areas (low contrast, presence of artifact, etc.). Cardiac cavity in red, normal myocardium in green, myocardial infarction in blue and PMO in yellow. Details can be found in the text.

gating techniques based on ECG or the requirement of breath holding. Moreover, the breath held at different positions may introduce a minor shift between slices, requiring additional inter-slice alignment to reconstruct the 3D model of the heart.

One more constraint to the development of an automatic assessment model is the annotated data if a supervised model is to be employed. Since high-accuracy segmentation approaches are mostly based on supervised deep learning models, the quality and quantity of MRI data with manual ground truth annotations is essential. Publicly available datasets usually provide up to a thousand CMRI slices with annotations such as MS-CMR2019 (https://zmiclab.github.io/projects/mscmrseg19/) and MyoPS 2020 (https://zmiclab.github.io/projects/myops20/) [120, 85] while more than tens of thousands of images are available in the datasets dedicated for human environment applications such as MS COCO and KITTI datasets [61, 53].

3.3/ MYOCARDIAL INFARCTION ASSESSMENT FROM PATIENT FEATURES

Although the DE-MRI provides the imaging gold standard for MI assessment, the application of DE-MRI is constrained by its processing time, material availability and cost compared to ECG or TTE. Furthermore, when a patient complains about non-traumatic chest pain, simple tools such as ECG, troponin assay and echocardiography are first performed as the pre-hospital or emergent diagnosis of MI. If the examinations reveal the possibility of MI, the DE-MRI could be achieved in the next days to have a more accurate evaluation of the myocardial impairment, after the acute phase and the early therapeutic management including revascularization and medications. Until the ultimate diagnosis

based on the MRI exam is available, physicians mainly rely on the obtained physiological, clinical and paraclinical characteristics to determine the severity of a patient's condition and to give sound treatment advice.

In light of the clinical practices, the automatic assessment of MI from patient features obtained before the performance of DE-MRI is of great value. As defined by the Fourth Universal Definition of Myocardial Infarction (2018), the cTn value compared to the 99th percentile of URL is a crucial indicator for non-fatal MI. The cTn assay is thus one important patient feature to be taken into account. ST segment analysis on ECG (especially in case of ST persistent elevation) and LVEF (left ventricular ejection fraction) assessment from TTE have also been shown highly correlated with MI [51]. Apart from these, physiological features such as BMI and age may be related to MI, which can be additional risk factors for the MI assessment.

The MI assessment from the patient features can be qualitative or even quantitative. The qualitative assessment classifies if a patient suffers from MI and the quantitative assessment predicts the extent of the disease area. The data is always one challenge for the automatic assessment approach. In addition to the data volume, the preparation of the ground truth is also tough. Although the follow-up of patients can establish the presence of MI, the severity of acute MI should be evaluated by imaging tools or autopsy. Datasets incorporating both the patient features and the extent of MI is rare.

Since the patient features are majorly scalar or Boolean, the richness of the clinical information they contain may be poorer than DE-MRI. Even an acquisition or recording error of a single piece of feature could inverse the diagnosis conclusion hence the patient features may be less robust compared to DE-MRI.

Although the patient features are easier to interpret, the fusion of DE-MRI and patient features could improve the robustness of automatic assessment but introduce technical challenges in designing multi-variate predictive models and balancing the weight from different data sources.

3.4/ EXPERIMENTAL DATA

Private datasets Despite the existence of some publicly available datasets involving DE-MRI, many of them only contain manual annotations of ventricles while the proposal of the work is to assess MI. Therefore, only two datasets incorporating MI annotations were employed as the experimental data. The first dataset is private and was collected in the University Hospital of Dijon (CHU Dijon), consisting of 160 cases with short-axis DE-MRI covering the left ventricle. Each case has on average 7 slices and the dataset has 1,201 slices in total. The MRI acquisition was done on 1.5 T and 3 T magnet

(Siemens Medical Solution, Erlangen, Germany) with a phased thoracic coil. T1-weighted Phase Sensitive Inversion Recovery (PSIR) images were acquired around 10 minutes after the injection of a gadolinium-based contrast agent (Gd-DTPA, Magnevist, Schering-AG, Berlin, Germany). The mean pixel spacing is 1.55 ± 0.20 mm^2 and the slice interval is generally between 11-13 mm. The acquired images have on average a resolution of 200×200 $pixel^2$ and the left ventricle is always smaller than 80×80 $pixel^2$.

In addition to the DE-MRI, the manual annotation on the left ventricle was also provided by a physician having more than 15 years of experience. The annotation data provide a total of 5 types of tissue masks, including the cavity, normal myocardium, myocardial infarction, PMO and the other tissue out of the epicardium. To evaluate the variability of the manual expert annotation, inter- and intra-observer variation studies were performed on 30 randomly selected cases. The intra-observer variation study was done by the same expert two months after the first segmentation, the inter-observer variation study was performed by another expert. Table 3.1 indicates the variance of inter- and intra-observer variation studies.

Table 3.1: Segmentation results of inter- and intra-observer variation studies on 30 randomly selected cases. The Dif refers to the volume difference between the tested sample and the ground truth sample.

Method	**Myocardium**			**Infarction**			**PMO**		
	Dif(mL)	HD(mm)	Dice	Dif(mL)	PIM(%)	Dice	Dif(mL)	PIM(%)	Dice
Inter-observer	13.57	2.79	0.8220	6.34	5.29	0.6609	2.12	1.65	0.6014
Intra-observer	8.49	2.25	0.8443	5.16	4.35	0.7231	0.75	0.58	0.7214

The MRI data are saved in the Digital Imaging and Communications in Medicine (DICOM) [25] format and all personal information in DICOM header has been anonymized. The annotation files are saved separately in JavaScript Object Notation (JSON) format. The contours that delineate each tissue area were drawn with the QIR software (CASIS company, Quetigny, France). A contour is registered as a set of sequential points, formed by connecting adjacent points. The coordinates of the points are saved as floating point numbers due to the manual annotation is usually done on enlarged images.

Public datasets The second employed dataset is publicly available, it is called EMIDEC. The overall dataset consists of 150 exams. Each exam is divided into two parts, a DE-MRI exam composed of a series of short-axis slices and the associated clinical information [142]. This dataset was improved from the previous private dataset hence some MRI cases are shared by both datasets. Along with MRI, the physiological, clinical and paraclinical characteristics are provided. These patient features consisting of twelve indicators that are potentially related to the acute MI were selected to compose the patient features. All the patients included in the dataset presented symptoms sug-

gestive of MI during admittance to an emergency department. The number of infarcted and non-infarcted cases was not balanced, which reproduced the clinical background in the cardiac emergency department. Physiological, clinical and paraclinical data were recorded during the arrival of the patients. Table 3.2 shows the characteristics of the selected features and the independent t-test [50] is performed to evaluate if the difference is significant between healthy and pathological subjects for each feature. Table 3.3 lists the percentage of infarcted myocardium (PIM) and the present rate of MI and PMO among pathological subjects.

Table 3.2: Characteristics of pathological and non-pathological patients (according to the DE-MRI).

Patient feature	Non-pathological subjects(n=50)	Pathological subjects(n=100)	p-value
Sex	38 females and 12 males	23 females and 77 males	0.000
Age	66 ± 14 years	59 ± 12 years	0.004
Tobacco(yes, no, former smoker)	18%, 22%, 60%	44%, 21%, 35%	0.001
Overweight[1]	62%	53%	0.296
Arterial hypertension	58%	31%	0.002
Diabetes	20%	10%	0.126
History of coronary artery disease	4%	12%	0.065
ECG(ST elevation)	30%	80%	0.000
Troponin(ng per mL)	7.68 ± 12.91	101.04 ± 101.35	0.000
Killip max(1,2,3,4)	76%, 22%, 2%, 0%	83%, 12%, 2%, 3%	0.916
LVEF[2](percentage)	$49.62 \pm 13.49\%$	$47.74 \pm 13.17\%$	0.423
NTProBNP[3](pg per mL)	2136 ± 3696	1314 ± 2109	0.154

[1] If BMI > 25
[2] Left Ventricular Ejection Fraction, calculated from transthoracic echocardiography
[3] N-terminal pro-B-type natriuretic peptide

Table 3.3: Proportion of scar tissues among pathological subjects. PIM refers to the percent of infarcted myocardium

Pathological tissue	PIM	Presence (%)
MI (PMO inclusive)	0.1825±0.1152	100
PMO	0.0330 ±0.0360	51

Unlike the first dataset, all slices in one MRI case have been aligned according to their gravity center obtained from the epicardium so that the potential misalignment has been corrected. Moreover, the MRI and manual annotations have been standardized into two separate Neuroimaging Informatics Technology Initiative (NifTi) files that have strictly the same image shape. Due to the NifTi format, each tissue area is represented by a mask map other than sequential contour points.

In both datasets, all data are fully anonymized and handled within the regulations set by the local ethical committee. The ethical committee of the University Hospital of Dijon checked the compliance of the dataset in accordance with the Declaration of Helsinki. The retrospectively collected data were completely untraceable. Therefore, according to French law and the ethical committee of the University Hospital of Dijon, neither the ethics committee approval nor informed written consent was required.

3.5/ CONCLUSIONS

In this chapter, the challenges for automatic assessment of MI from MRI and patients features are elaborated and the available datasets are presented. The major challenges that come from the DE-MRI are the images of low contrast between normal and scar tissues. Moreover, the identification of the PMO is even more challenging due to its similar density to normal myocardium tissue. With the multivariate data, two problems can be identified: the simplicity of the patient features limits deeper data exploration, and the difference between the patient features and the DE-MRI requires an efficient data encoding method.

4

Related works: segmentation models and automatic MI assessment from MRI

4.1/ Introduction

Related works about the automatic MI segmentation will be presented separately for segmentation models and applications. Segmentation models can be grouped by statistical models and deep learning models. Statistical models have been widely employed on tasks of medical image segmentation for a long time, while deep learning approaches are relatively more recent although the number of their applications is expanding rapidly.

Statistical models usually distinguish groups of pixels in an image according to some predefined rules of the feature distribution. Such models are normally non-supervised and computationally efficient. However, such models are sensitive to noise if the noise shares similar distribution as a target group of pixels.

Deep learning-based segmentation models have been more frequently performed after the proposition of the eminent work of Fully Convolutional Network [67] and then U-Net [69]. Fully Convolutional Network introduces the deconvolution to construct the segmentation in a learnable way with pyramid architecture, and U-Net further adds the skip connection to restore the details in the encoding side for the segmentation. In the context of U-Net, optimizations are studied such as loss functions, building blocks, auxiliary networks, *etc.* to specifically improve the segmentation results on different medical images and target tissues.

This chapter is organized as follows:

- Statistical models for segmentation. The common statistical models for segmentation tasks are introduced.

- Deep learning models for segmentation. Different aspects of deep learning-based segmentation models are introduced, mainly around the improvements on U-Net-based approaches, including preprocessing techniques, network conceptions and related applications on cardiac MRI.
- Conclusions.

4.2/ STATISTICAL MODELS FOR SEGMENTATION

The statistical models for the segmentation task are usually rule-based, including threshold-based approaches like Full-Width at Half-Maximum (FWHM) [23] or n-Standard Deviations (n-SD) (or Signal Threshold versus Reference Mean (STRM)), and clustering-based approaches like Gaussian Mixture Models (GMM) [37].

Threshold-based segmentation approaches differentiate intensity according to a threshold. The thresholding intensity is crucial to the segmentation accuracy. The appropriate threshold can be estimated by analyzing the signal distribution in the Region of Interest (ROI), *i.e.* the LV myocardium. FWHM value can be employed to differentiate between the LV myocardium and the whole MRI slice, or between normal myocardium and myocardial scar inside of the LV myocardium. In the intensity distribution of an ROI, FWHM is the difference between two values of the independent variable, at which the dependent variable is equal to half its maximum value (Figure 4.1). n-SD is another thresholding strategy. n-SD sets the bandwidth from the mean value to several standard deviations (SD). 2, 3, 4, 5, and 6 SDs are the usual threshold range, corresponding to the methods 2-SD, 3-SD, 4-SD, 5-SD and 6-SD.

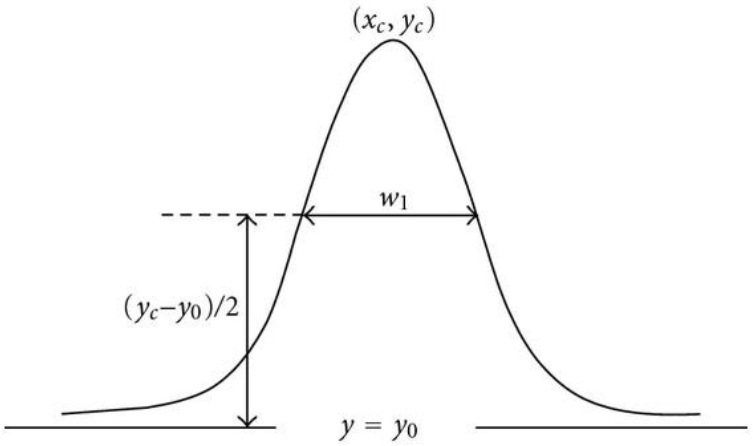


Figure 4.1: Definition of FWHM bandwidth in a distribution histogram. y_0 is the starting value and y_c is the maximum value in the distribution. w_1 is the FWHM bandwidth. Credit: Origin 8, User Guide, OriginLab Corporation.

A GMM can be seen as a mixture model of K single Gaussian models. The term mixture model represents the presence of sub-populations within an overall population. In simple terms, the mixture model represents the probability distribution of the observations in the aggregate, which is a mixture of K sub-distributions. The mixture model does not require the observations to provide information about the sub-distributions to calculate the probability that the observations are in the overall distribution. A mixture model is called GMM if one assumes that all the sub-distributions conform to Gaussian distributions, or normal distributions [148]. The probability density function of univariate Gaussian distribution is:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{4.1}$$

where σ denotes the standard deviation (SD) and μ denotes the mean or expectation of the observed data. GMM is one of the most common mixture models since Gaussian distributions are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known, although other distributions can be assumed in a mixture model. To identify all the sub-distributions, Maximum Likelihood Estimation can be employed if each sub-distribution is assumed a univariate Gaussian distribution. In practice, Maximum Log-Likelihood often replaces Maximum Likelihood Estimation since the log function amplifies the output variance for an input between 0 and 1 and is monotonic. To solve multivariate GMM, iterative methods should be used such as Expectation-Maximization algorithm (EM) [3].

Within DE-MRI, the MI can be segmented by GMM because the signal of MI is considered as a sub-distribution that has its independent distribution characteristics. The DE-MRI can be described by the GMM of univariate Gaussian distributions since DE-MRI is saved as greyscale images. However, the segmentation of MI by univariate GMM can not refer to the morphological information of MI since the spatial information is ignored.

4.3/ DEEP LEARNING MODELS FOR SEGMENTATION

A deep learning model for segmentation tasks covers a series of techniques, typically including data preprocessing and data augmentation methods, the conception of convolutional neural networks (CNN) and post-processing methods. For medical image segmentation tasks, nnU-Net [153] has been proven a successful adaptive framework for the automatic analysis of a variety of different types of medical images. This subsection lists and describes the various relevant methods commonly used for the task of DE-MRI segmentation.

4.3.1/ DATA AUGMENTATION

A significant proportion of deep learning models is supervised learning. The supervised learning models learn to fit the mapping of a target from training data. The amount of training data directly affects the performance of supervised models. A reasonable data augmentation method can equivalently expand the size of the training set. Furthermore, the data augmentation can potentially overcome overfitting in training. In many practices with deep neural networks, to avoid overfitting in training, the prediction performance on validation set should be supervised. The training should stop once the metrics on the validation set no longer decrease or even increase. The decision that stops the training when it tends to overfit is known as early stopping. The comparison between the overfitting and the desired convergence in training is shown in Figure 4.2. The data augmentation may allow for better timing of the early stopping because the augmented data will represent a more comprehensive set of possible data points, thus minimizing the distance between the training and validation set, as well as any future testing sets [132].

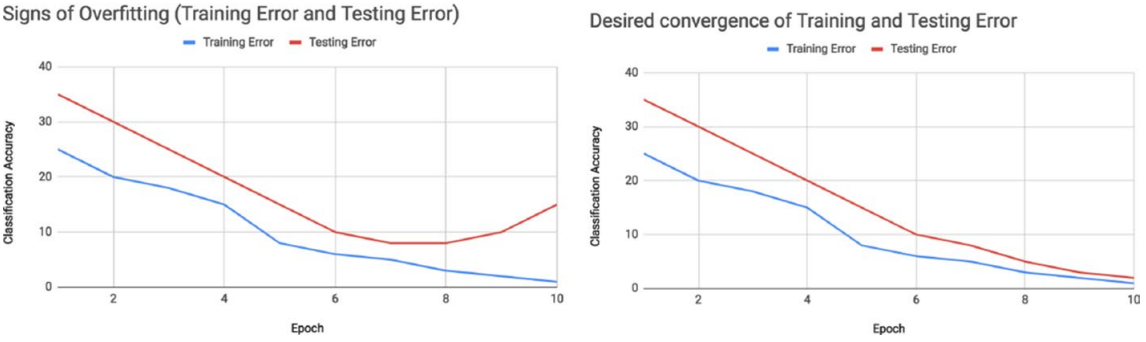


Figure 4.2: Signs of overfitting compared to desired training convergence. The plot on the left shows an inflection point where the validation error starts to increase as the training rate continues to decrease. The increased training has caused the model to overfit the training data and perform poorly on the testing set relative to the training set. In contrast, the plot on the right shows a model with the desired relationship between training and testing error. Credit: C Shorten *et al.* [132]

For medical imaging especially MRI, basic image manipulations involving flipping, rotation, cropping, translation, interpolation and noise injection can be applied to achieve extended data volume. The flipping can be performed on either a horizontal or vertical axis although horizontal flipping is more common. The rotation is done by rotating the image right or left on an axis between 1° and 359°. Horizontally and vertically shifting image is useful to avoid positional bias. The interpolation resizes the image to a specific shape using linear interpolation for example. Noise injection introduces a matrix of random values to the original image [112]. The noise matrix of MRI is usually Rice distributed to simulate the on-site noise of the image acquisition. For example, on DE-MRI, Camarasa *et al.* [147] performed rotations, elastic deformations, and flips on slices to randomly produce supplementary training data while Feng *et al.* [150] forced the model to ignore

the specificity for different orientation features by the rotations only. Lourenço *et al.* [155] adjusted the original semantic information by adding stochastic noise, applying k-space corruption, small image rotations, intensity scalings, and smooth non-rigid deformations.

Manipulations concerning the color space and color transformation are also common data augmentation methods such as the adjustment of contrast and white balance, sharpening and color augmentations in RGB (Red, Green, Blue) or HSV (Hue, Saturation, Value) space, while they are not applicable for the grayscale MRI.

Mixing images together is a counter-intuitive approach to data augmentation. The mix-up strategy averages relevant images to create a new image. Employing an image fusion by affine transformation instead of image averaging, Zhou *et al.* [160] proposed a mix-up strategy that fuse two adjacent DE-MRI slices [119]. The mix-up strategy constructs virtual training examples of both DE-MRI and its annotation mask as follows:

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \tag{4.2}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \tag{4.3}$$

where x_i and x_j are raw input vectors, y_i and y_j are one-hot label encodings, $\tilde{x}$ and $\tilde{y}$ is the pair of artificially created data. λ is a coefficient belonging to $[0, 1]$. Based on this approach, Zhou *et al.* made a targeted improvement to make the generated images closer to a blend of two adjacent images. The proposed mix-up formula for the MRI augmentation is:

$$\tilde{x} = \lambda x_i + (1 - \lambda) T x_j \tag{4.4}$$

where T denotes an affine transformation, and the mask data augmentation was performed accordingly with a similar formula. Given the greater focus on the ROI (Region Of Interest corresponding to the myocardium), the affine transformation T tries to fit the transformation from the foreground area (LV+Myocardium) in a randomly chosen slice x_i to the foreground area in another randomly chosen slice x_j. In the affine transformation, the scaling factor, *i.e.* the linear map is $[s, s]^\top$ where $s = l^i / l^j$, l^i and l^j are the average distances from the foreground pixels to the foreground center for the slice i and the slice j, respectively. The translation offset is $[c_x^i - c_x^j, c_y^i - c_y^j]^\top$ where c_x and c_y denote the coordinates of the foreground area centre. Thus, the matrix of T is:

$$\begin{pmatrix} s & 0 & c_x^i - s \cdot c_x^j \\ 0 & s & c_y^i - s \cdot c_y^j \\ 0 & 0 & 1 \end{pmatrix} \tag{4.5}$$

Although data augmentation has been widely used for different image tasks, its safety should be discussed in the case of DE-MRI. The safety of a data augmentation method refers to its likelihood of preserving the label post-transformation [132]. The created data

should maintain the same semantic characteristics as the original data, no new feature should be produced. However, the definition of such a characteristic is vague. Despite the vagueness, some basic rules can be followed according to the anatomical prior of the heart. For example, the flipped images never exist in the original dataset while the DE-MRI on the short-axis may be saved as horizontal or vertical (Figure 4.3). To correct the image size, cropping can be thought safe if all the ROI is preserved. Meanwhile, transformations such as interpolations may be dangerous because the voxel spacing becomes significantly heterogeneous among MRI slices. Shifting can be safe since originally the position of patients may be slightly different during the examination. When the noise injection is performed, the intensity and distribution type of the introduced noise should be controlled. Nevertheless, experiments should be more convincing for the evaluation of a data augmentation method. In response to this issue, Cubuk *et al.* [122] proposed an automatic evaluation approach in which an algorithm is designed to find the best policy such that the neural network yields the highest validation accuracy on a target dataset.

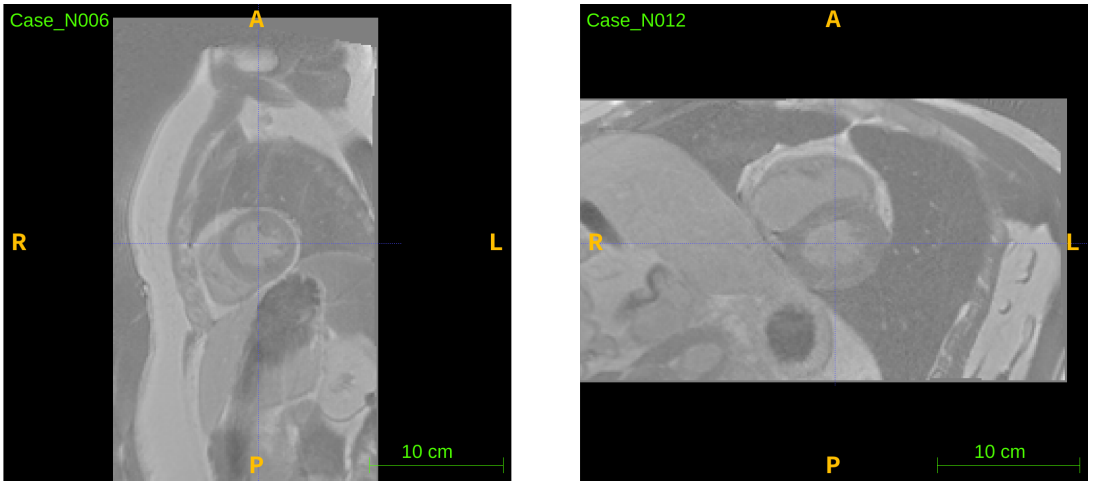


Figure 4.3: Vertical and horizontal views of DE-MRI in short-axis orientation. One vertical (left) and one horizontal (right) slice are randomly selected from EMIDEC training dataset. The DE-MRI may be acquired in different orientations but should be never symmetrical within the same machine and modality.

4.3.2/ ENCODER AND DECODER MODELS

Fully Convolutional Network The segmentation task is equivalent to the sum of classification tasks for each pixel. However, beyond the classification on a single pair of input-label, the segmentation task should refer to spatial information between pixels, that is, the classification result of each pixel depends on both local and global features. To this end, the milestone work of this task Fully Convolutional Network (FCN) proposes an encoder-decoder structure that combines the image encoding and segmentation reconstruction [67].

The encoding part of FCN just reproduces the usual conception of image classification networks. According to the evaluation between AlexNet [48], VGG [70] and GoogLeNet [71], the three well-validated classification architectures at the time, the 16 weight layers implementation of VGG (VGG-16) is selected as the encoding part of FCN. Figure 4.4 shows the architecture of VGG-16.

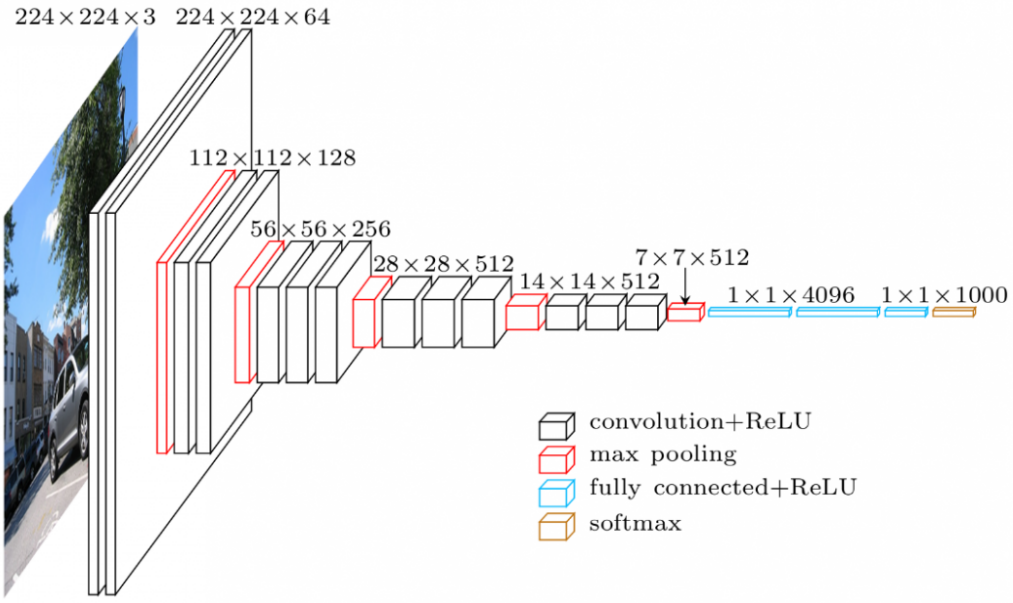


Figure 4.4: VGG-16: 16 weight layers implementation of VGG. Credit: [111]

Given that the classifiers such as VGG provide a scalar value as their output, FCN retains only the convolutional part of VGG to get dense prediction as shown in Figure 4.5 and the fully connected layers are viewed as convolutions with kernels that cover their entire input regions. However, the output dimensions are typically reduced by subsampling. The classification nets subsample to keep filters small and computational requirements reasonable. This coarsens the output of a fully convolutional version of these nets, reducing it from the size of the input by a factor equal to the pixel stride of the receptive fields of the output units [67]. To obtain the segmentation mask that has the same shape as the input image, the dense predictions can be obtained from coarse outputs by stitching together outputs from shifted versions of the input. To do so, the generated new input is the shifted original input with zero padding so that the center of each output equivalently corresponds to a patch of the original input. This operation is intuitively not efficient due to the redundant computation on the overlap between each patch. Another way to connect coarse outputs to dense pixels is an interpolation. In a learnable way, a reversed convolution termed backward convolution (or deconvolution) upsamples the coarse outputs. Passing by a backward convolution of stride f achieves an upsampled output of factor f.

The depth of downsampling layers determines the equivalent receptive field of the output. A deeper downsampling network results in a smaller size dense prediction that

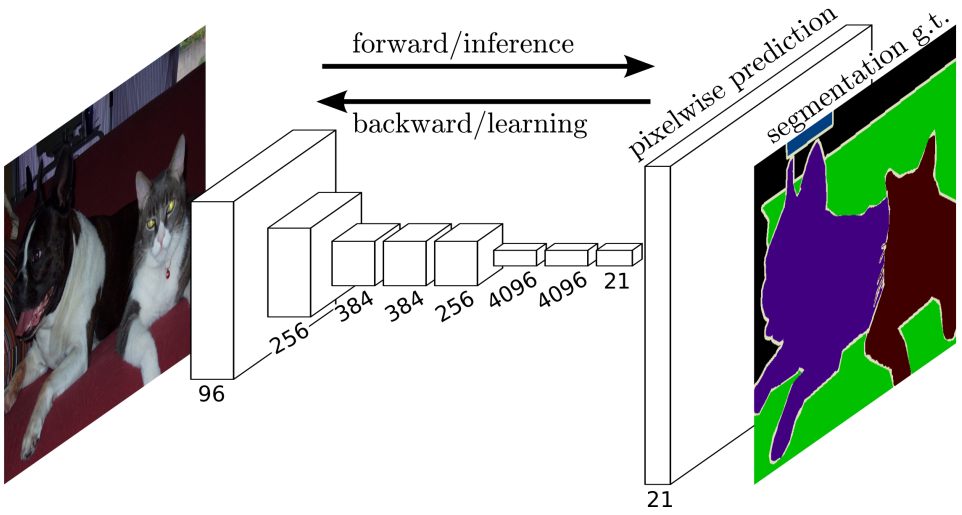


Figure 4.5: Image segmentation by downsampling and $1x1$ convolutions instead of fully connected layers. FCN produces pixel-wise dense prediction with only convolutional blocks.

better interprets high-level semantic information (*i.e.* global features), while a shallower downsampling network retains more detail in the dense prediction. A pyramid structure that fuses interpretations on images of different scales with a multiscale convolutional network was proved efficient [52]. This conception was lately integrated for the object detection task [92]. To refer to pyramid features for finer segmentation, coarse segmentations from the last three pooling layers are aggregated together as shown in Figure 4.6. The coarse segmentations from the layers $conv7$ and $pool4$ are upsampled by 2x and 4x respectively. These upsamplings are initialized with bilinear interpolations and the parameters are learnable. The qualitative result of FCN justifies that refining fully convolutional nets by fusing information from layers with different strides improves segmentation detail (Figure 4.7).

U-Net Inspired by FCN, Ronneberger *et al.* optimized the multi-scale information aggregation procedure so that the proposed network termed U-Net was more suitable for medical image segmentation tasks [69]. The major amelioration results from the introduction of the skip connection and the symmetric encoding-decoding structure (Figure 4.8). In FCN each upsampling branch is independent. Between each upsampling scale, they only interact at the last level, which means the intermediate deconvolution kernels are not shared while they may refer to each other. To tackle this issue, U-Net creates a single symmetric upsampling branch in which all the low-scale feature maps are visible to higher-scale features. Such conception reuses the intermediate deconvolution kernels and decreases the number of model parameters. Furthermore, to better retain the ordinary information, skip connections are introduced between each pair of downsampling

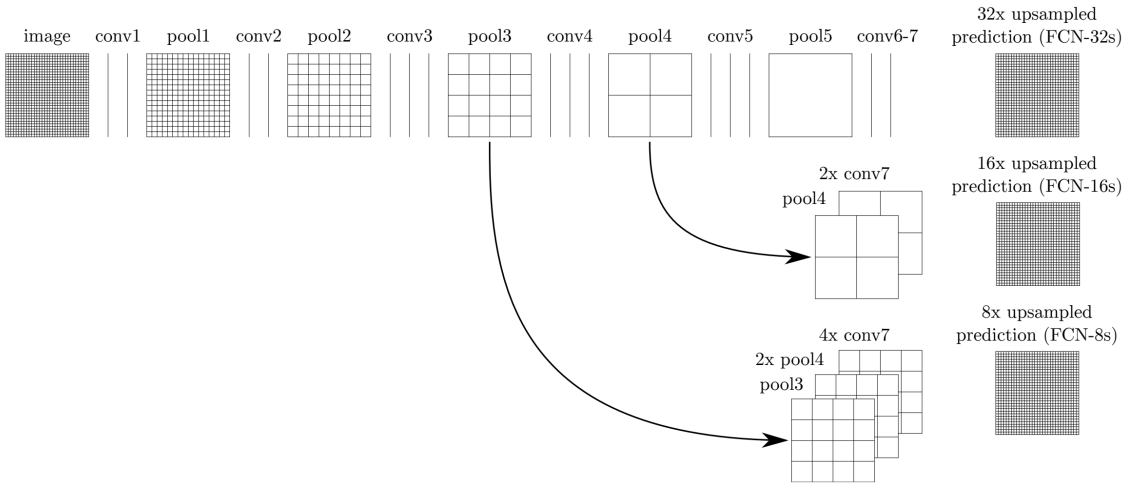


Figure 4.6: Upsampling and aggregation of fine and coarse segmentation by FCN. The aggregation learns to combine coarse, high-layer information with fine, low-layer information. Pooling and prediction layers are shown as grids that reveal relative spatial coarseness, while intermediate layers are shown as vertical lines. First row (FCN-32s): the single-stream net upsamples stride 32 predictions back to pixels in a single step. Second row (FCN-16s): Combining predictions from both the final layer and the pool4 layer, at stride 16, lets the net predict finer details while retaining high-level semantic information. Third row (FCN-8s): Additional predictions from pool3, at stride 8, provide further precision. Credit: [67]

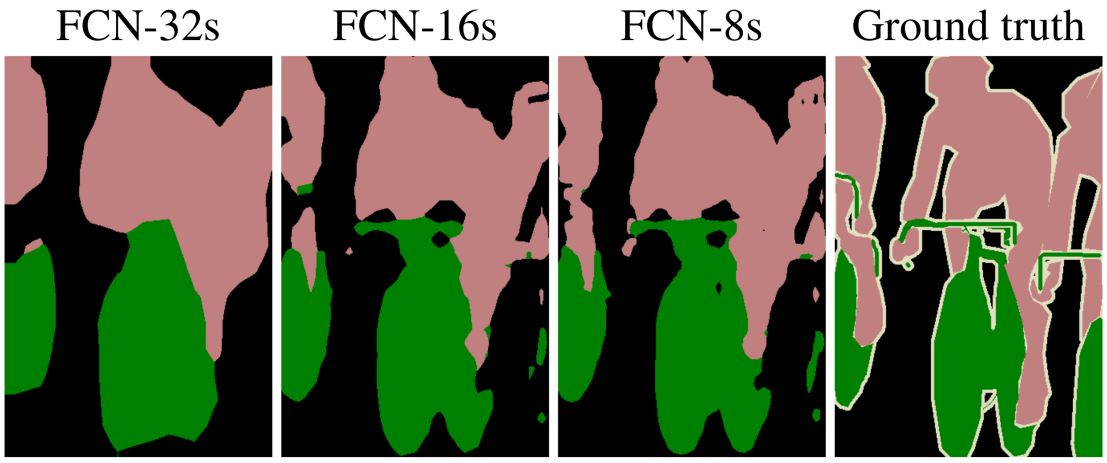


Figure 4.7: Effect of multi-scale output fusion. The first three images show the output from our 32, 16, and 8-pixel stride nets shown in Figure 4.6. Visually the FCN-8s outperforms the two other aggregations with less fused scales. Credit: [67]

feature maps and upsampling feature maps of the same shape. The skip connections copy downsampling feature maps and they are summed to upsampling feature maps. This conception avoids the loss of detailed information from high-scale downsampling feature maps.

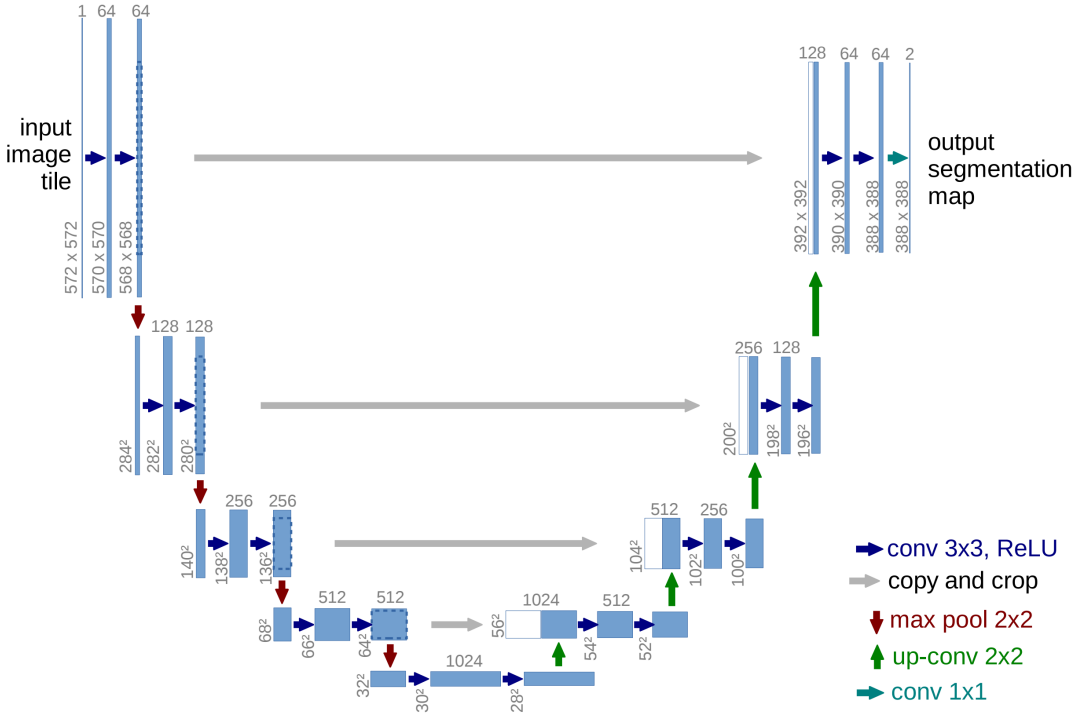


Figure 4.8: Architecture of U-Net. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Credit: [69]

U-Net variants After the publication of U-Net in 2015, a number of variants were then proposed mainly for structural improvements while the symmetrical encoding-decoding structure was maintained. Much work focuses on the more recent convolutional blocks. For example, FusionNet [80] replaces the conventional convolution blocks of U-Net with residual blocks [77]. CE Net [125] introduces the dense atrous convolution and a residual multi-kernel pooling to retain more high-level semantic information. V-Net replaces 2D convolutions with 3D convolutions so that the volumic segmentation can be obtained. Some other work proposes additional branches to gain more details. For example, M-Net [94] adds multi-scale inputs and deep supervision into the U-net architecture. Pohlen *et al.* propose a two-stream branch consisting of a pooling stream and a residual stream [95]. Fourure *et al.* extend the unidirectional skip connections to a grid so that downsampling feature maps can be perceived by upsampling feature maps of different shapes. Moreover, the transmitted information via skip connection is fused with the corresponding feature maps by addition in the vanilla U-Net [88]. In many U-Net variants, the addition operation at the skip connection is replaced by concatenation so that the number of feature maps is increased. The concatenation may be more appropriate if the number of feature maps is different between the downsampling side and the upsampling side, while both of addition and concatenation should have the same effect if the number of feature

maps is identical between the two sides.

Autoencoder The autoencoder [31] is a type of artificial neural network. Autoencoder aims at learning the representation of a set of data. A typical application of the autoencoder is to distinguish the common features and the noises in a dataset for further abnormality detection.

An autoencoder normally has two parts, the encoding and the decoding parts (Figure 4.9), which are quietly similar to U-Net. However, in opposition to U-Net, the autoencoder usually has no skip connection between the encoder and the decoder, and the training of an autoencoder is supervised by the input itself, namely in an unsupervised manner. Therefore, when data are fed to an autoencoder, the encoding part will first reduce the data dimension and generate the data coding (or termed code). The data coding is the abstract representation of the input data and the encoder decides which features are to be retained and squeezed. Then the data coding will be interpreted by the decoder part and the output will be compared with the input data for the backpropagation. The encoder is thus responsible for the re-interpretation of the squeezed and extracted code. The difference between the output and the input can be considered as noise other than the data distribution.

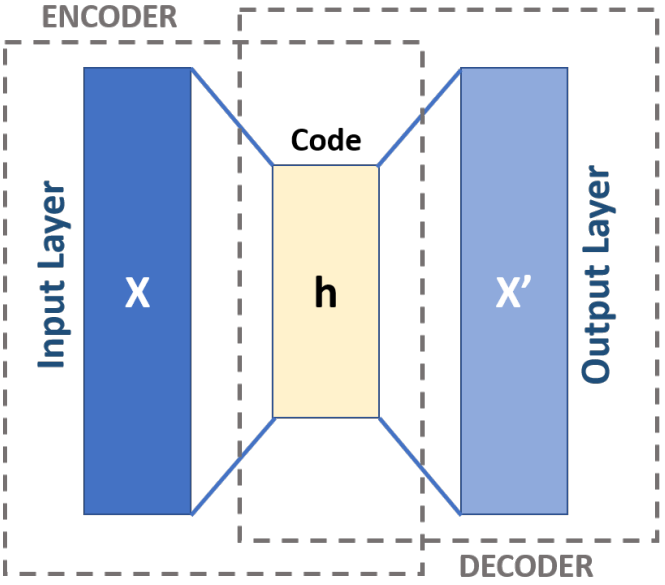


Figure 4.9: Basic architecture of autoencoder. A basic autoencoder normally consists of the encoder and the decoder. The compressed data by the encoder is the code of the input data. Credit: Michela Massi (2019).

In response to some drawbacks of the basic autoencoder, some variants have been proposed. Many variants add regularisation terms in order to capture more important features and optimize the representation. For example, Vincent *et al.* [42] proposed the de-

noising autoencoder. The principle of the denoising autoencoder is to artificially corrupt the training data so that the autoencoder can be more robust and potentially overcomes the overfitting. In practice, in the training stage, noise is first added to the training data. The output of the autoencoder is then compared to the original, non-corrupted data and the backpropagation is based on this difference. The sparse autoencoder [45] is another variant that encourages sparsity of the autoencoder. The purpose of the sparse autoencoder is to obtain a sparse data coding that potentially improves the performance on classification task [54]. The sparsity of the network can be viewed as a small number of the hidden units that are allowed to be active at the same time so that the code is sparse and only the unique features can be interpreted. To do so, a regularisation of the units' sparsity is introduced. Since whether a unit (neuron) is activated or not can be seen as obeying the Bernoulli distribution (activated or not), the Kullback-Leibler (KL) divergence between the expected Bernoulli random variable and the Bernoulli random variable of the units can be employed as a regularisation term. The KL divergence between the expected Bernoulli random variable ρ and the Bernoulli random variable of a unit $\hat{\rho}_j$ can be formulated as:

$$D_{KL}(\rho||\hat{\rho}_j)) = \rho log\frac{\rho}{\hat{\rho}_j} + (1-\rho)log\frac{1-\rho}{1-\hat{\rho}_j} \tag{4.6}$$

The application of autoencoder in the task of MI segmentation can be the input denoising. The potential noise in DE-MRI is first eliminated by means of a denoising autoencoder. Another application is the refinement of the segmentation mask obtained from a segmentation network by learning the representation of morphological features. For example, Yue *et al.* [135] adds an autoencoder as an annex network to refine the shape of the segmented LV myocardium. The autoencoder is designed to learn the shape prior of the myocardium. The difference between the coarse and the refined segmentation masks is quantified by L2 distance, namely the square root of the sum of the absolute squares of all vector elements. The L2 distance of the shape prior autoencoder is then weighted as a part of the final loss function for the segmentation because the autoencoder is also differentiable. Figure 4.10 shows the shape prior autoencoder annexed to the U-Net-based segmentation network.

4.3.3/ BUILDING BLOCKS OF CNN

AlexNet U-Net adopts AlexNet [48] as the backbone CNN. AlexNet won the ImageNet challenge in 2012 by showing 10.8 percentage points of top-5 error lower than the one of the runner-up. AlexNet achieved such performance due to the depth of the model thanks to the computational acceleration of graphics processing units (GPUs). Figure 4.11 shows the architecture of AlexNet. In AlexNet, the first convolutional layer filters the 224×224×3 input image with 96 kernels of size 11×11×3 with a stride of 4 pixels. The

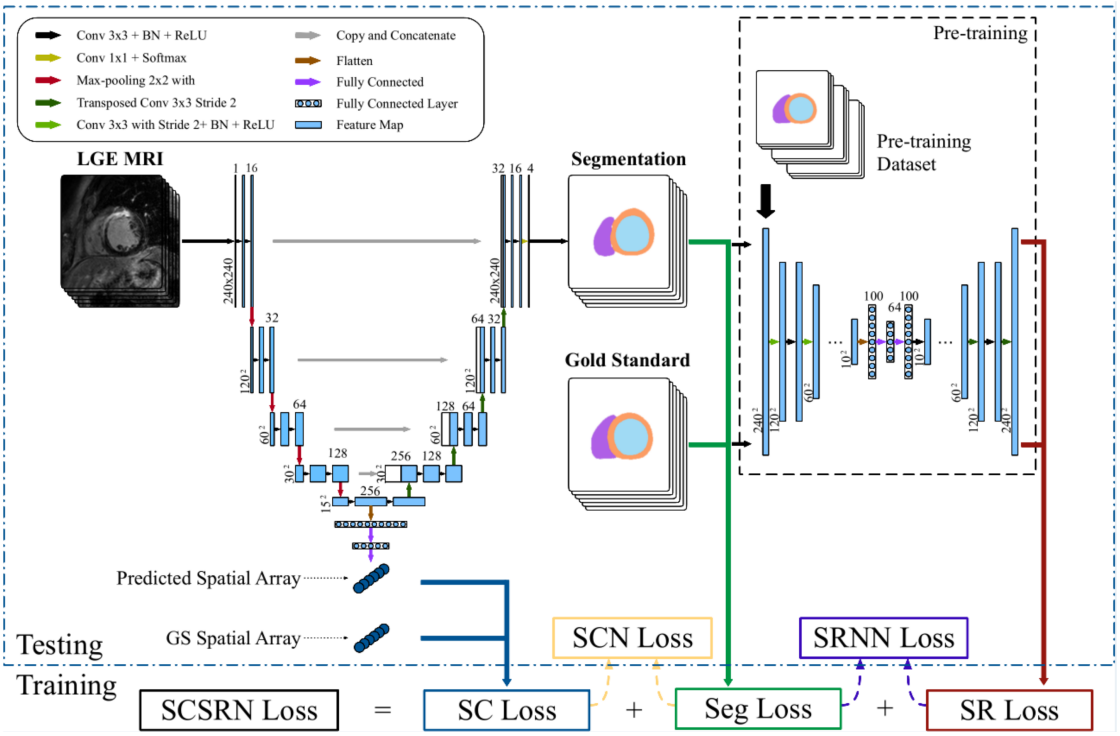


Figure 4.10: Shape prior autoencoder annexed to a U-Net-based segmentation network. In the black dashed box on the right is the shape prior autoencoder. The weighted L2 distance between the ground truth segmentation and the segmentation obtained from the U-Net is the shape loss termed SR Loss. Credit: [135]

second convolutional layer takes as input the (response-normalized and pooled) output of the first convolutional layer and filters it with 256 kernels of size 5 × 5 × 48. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The third convolutional layer has 384 kernels of size 3 × 3 × 256 connected to the (normalized, pooled) outputs of the second convolutional layer. The fourth convolutional layer has 384 kernels of size 3 × 3 × 192, and the fifth convolutional layer has 256 kernels of size 3 × 3 × 192. The fully-connected layers have 4096 neurons each. Such convolution-pooling-normalization structure has become the routine conception for the following works.

ResNet Since the vanilla U-Net [69] was proposed in 2015, a lot of recently proposed building blocks outperformed the AlexNet employed by U-Net. ResNet is one of the most revolutionary work after AlexNet or VGG. ResNet introduces the identity mapping into its building block the residual block. The identity mapping copies the block's input and then is added to the output of the convolution branch. The transformation of a residual block can be represented by:

$$y = \mathcal{F}(x, \{W_i\}) + x \tag{4.7}$$

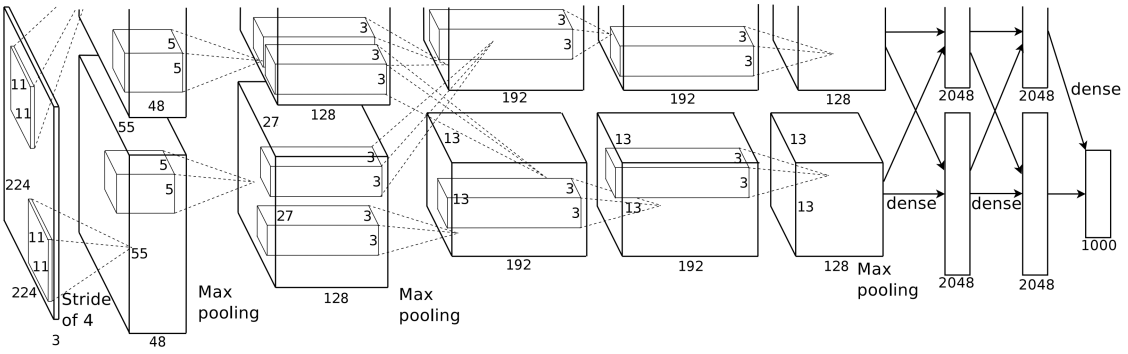


Figure 4.11: Architechure of AlexNet. Credit: [48]

where x is the input feature map, $\{W_i\}$ is the set of transformations thus $\mathcal{F}(x, \{W_i\})$ represents the residual mapping to be learned. The residual block in ResNet has two implementations depending on the depth of the network. Figure 4.12 shows the two implementations for shorter and longer ResNet. For example, in the case of the left implementation of Figure 4.12, the transformation can be formulated as $\mathcal{F} = W_2\sigma(W_1x)$ in which σ denotes Rectified Linear Unit (ReLU) activation function. The operation $\mathcal{F} + x$ is performed by a shortcut connection and element-wise addition.

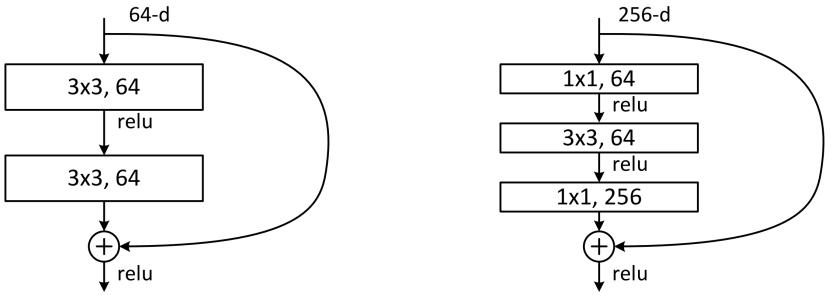


Figure 4.12: Two implementations of residual block. Left: a shorter building block for ResNet34. Right: a “bottleneck” building block for ResNet-50/101/152. Credit: [77]

ResNet won first place on the ILSVRC 2015 classification task. Its success comes from the mitigation of the degradation problem. The degradation problem results from the deeper network architecture. In former CNNs, semantic information may be omitted due to the excessive depth. The identity mapping provides an option in both forward and backward propagations so that such information can be retained. The authors argued that the degradation problem can not be confused with the gradient vanishing [16]. During the backpropagation, layers' weights are updated in order from back to front according to the chain rule. The gradient received by a layer is the derivative of all the former layers' composite. Therefore, the gradients in the backpropagation exponentially decrease and the gradient vanishing results in very slow training of the front-end layer. Recent work includ-

ing ReLU and batch normalization [65] *etc.* have been proven efficient to solve gradient vanishing. To justify if the residual block mitigates the degradation instead of the gradient vanishing, the authors designed a plain implementation that has the same architecture of ResNet including ReLU and batch normalization except that the identity mappings were removed. Experiments showed that in both plain and residual implementations, the residual implementation outperformed the plain implementation. Moreover, compared with former CNNs, the plain implementation achieved competitive results, suggesting that ReLU and batch normalization works to some extent. Although these explanations may be empirical, the application of residual block in many CNN implementations does bring better performance without adding additional computation.

DenseNet DenseNet [89] can be viewed as a successor of ResNet. The major novelty of DenseNet is the more dense shortcuts between layers in a building block. As in Figure 4.13, each layer obtains additional inputs from all preceding layers and passes on its own feature maps to all subsequent layers. Each layer is thus receiving a collective interpretation from all preceding layers. Compared with ResNet, the concatenation is used by DenseNet instead of the addition. DenseNet outperforms ResNet on many classification datasets while the network parameters are fewer.

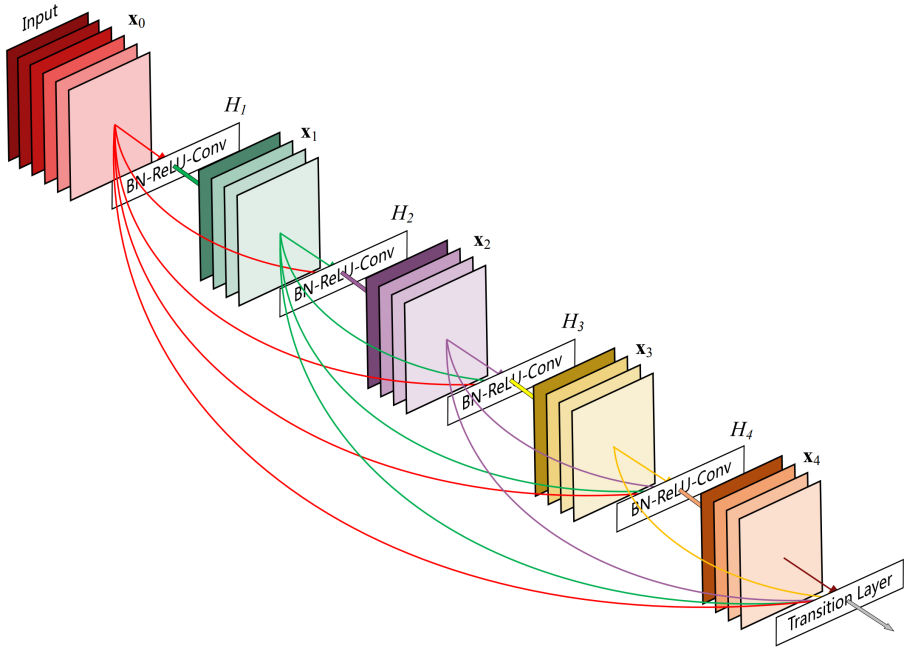


Figure 4.13: A 5-layer dense block. Each layer takes all preceding feature maps as input. Credit: [89]

ResNeXt Xie *et al.* proposed an aggregated version of ResNet called ResNeXt [103]. The core idea of ResNeXt is the split-transform-merge architecture, that is the transfor-

mation of a building block can be presented as:

$$\mathcal{F}(x) = \sum \mathcal{T}_i(x) + x \tag{4.8}$$

where $\sum \mathcal{T}_i$ denotes all parallel splits of transformations and $+x$ denotes the identiy mapping. The split-transform-merge architecture can be traced back as far as the proposal of the first version of InceptionNet [72] (Figure 4.14) while the branches in InceptionNet are complicated and are designed by artificial means. ResNeXt replaces the different branches in InceptionNet with more identical 3x3 convolutions - 1x1 convolutions - 3x3 convolutions (Figure 4.15). One important design in ResNeXt is the number of channels in each transformation branch. Unlike ResNet and InceptionNet, ResNeXt significantly reduces the number of the 1x1 convolution kernels between the adjacent 3x3 convolution layers. The fewer feature maps in the middle of each branch force a branch to learn a specific feature representation. Such conception keeps the independence of features and the network parameters will not increase.

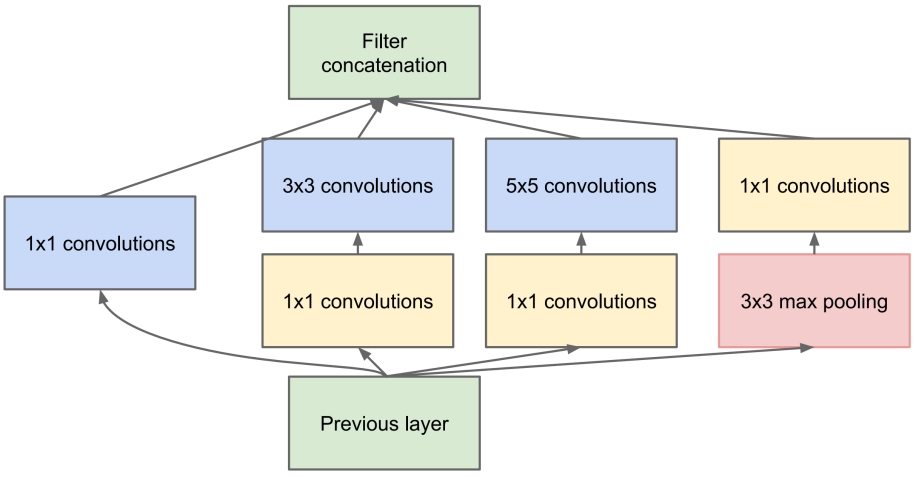


Figure 4.14: Inception module with dimension reductions. Credit: [72]

ResNeXt is reported as more efficient than InceptionNet with the same parameter quantity. However, the computation complexity is slightly heavier due to an excessive number of convolution layers despite that such phenomena may be optimized from the underlying drivers of GPUs.

Attention mechanism and gating blocks The attention mechanism [102] has become a popular topic from serial data as Natural Language Processing (NLP) to computer vision tasks. The attention in neural networks mimics cognitive attention: valuable information should be enhanced and redundant information will be faded out. The attention can be applied to relatively concrete data such as the skip connections, or inside a convolu-

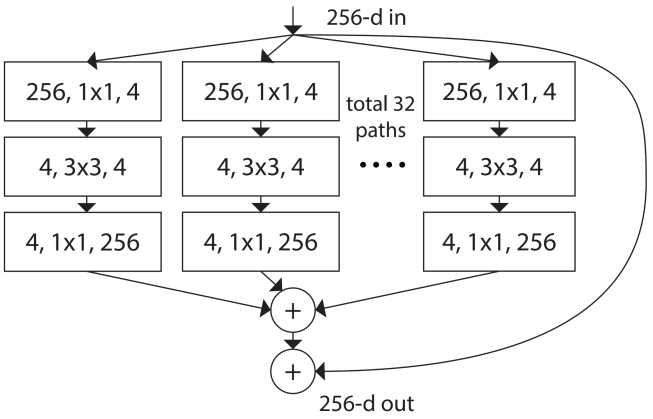


Figure 4.15: A block of ResNeXt with cardinality = 32. This architecture roughly has the same complexity as the building block for ResNet-50/101/152 (Figure 4.12 right). Credit: [103]

tional block for more abstract gating such as Squeeze-Excitation (SE) block and Inverted Residual Blocks (IRB).

The attention U-Net [131] introduces the attention gates to self-supervise the attention on feature maps. The attention highlights prominent image regions and suppresses task-independent feature recognition. In attention U-Net, the attention gates are inserted in skip connections between the encoder and the decoder (Figure 4.16). Downsampled features propagated through the skip connections are filtered by the attention gates. To extract the information, the signal from the downsampling layer is defined as the input features and the coarser signal from the last upsampling layer is defined as the gating features. To obtain the attention vector α, the gating and input signals are first transformed by two 1x1 convolutions respectively. Then the two transformed signals are summed and pass through ReLU, 1x1 convolution, sigmoid and resampler in succession. The obtained vector α represents the attention vector in which $\alpha_i \subset (0:1]$. A higher value means higher attention should be paid to the corresponding region. Finally, the original input signal is element-wisely multiplied with α hence regions are enhanced or faded out. The schematic of the attention gate is shown in Figure 4.17. It should be notified that in the authors' attention gate schematic, the functionality of the resampler is unclear. Supposing that an additional upsampling is applied on the gating signal out of the schematic and therefore the transformed gating signal and input signal can be element-wisely summed, no more resampler should be necessary due to the same feature map dimension between the coefficient vector and the input signal.

SE block [110] is another building block that partially refers to the attention mechanism. The SE block aims at better modeling the interdependencies between channels of the convolutional features. To this end, in the SE block, feature maps were first squeezed into

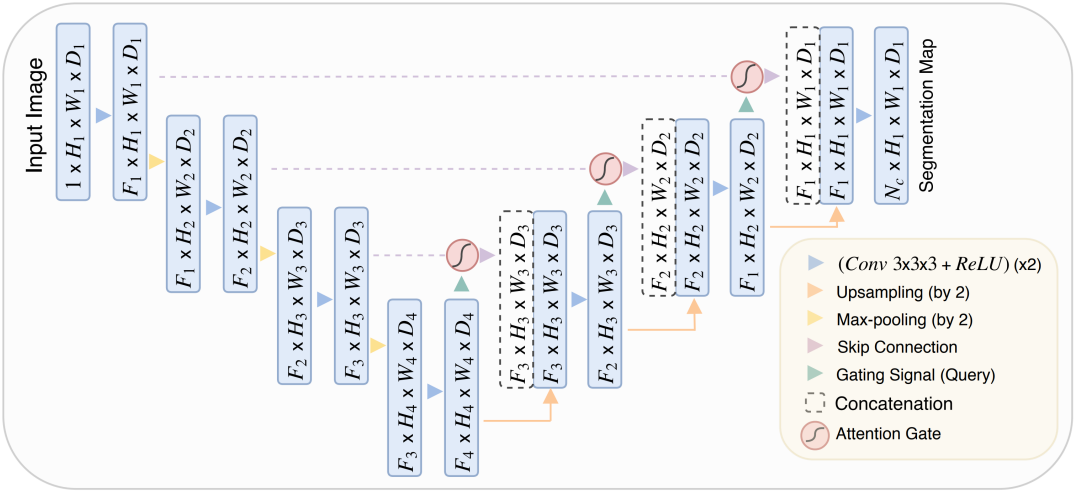


Figure 4.16: Architecture of attention U-Net. Attention gates filter the features propagated through the skip connections. Feature selection in the attention gates is achieved by the use of contextual information (gating) extracted in coarser scales. Credit: [131]

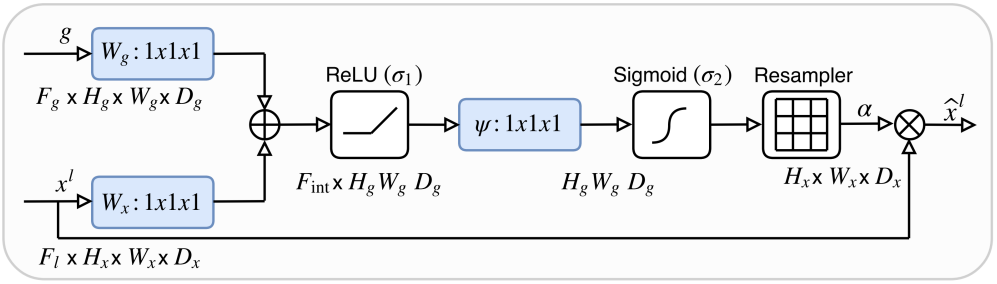


Figure 4.17: Schematic of the additive attention gate (AG). Input features (x^l) are scaled with attention coefficients (α) computed in the attention gate. Spatial regions are selected by analyzing both the activations and contextual information provided by the gating signal (g) which is collected from a coarser scale. Grid resampling of attention coefficients is done using trilinear interpolation. Credit: [131]

a channel descriptor with shape [1, 1, channel] by the global average pooling. Then to fully capture the aggregated channel-wise information, a simple gating mechanism was employed with linear transformations and non-linear activation functions:

$$\mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \rho(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \tag{4.9}$$

where $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are linear transformations, C and r are channel size and reduction rate, δ refers to ReLU and ρ refers to sigmoid activation. To finally emphasize differently the feature maps, the channel-wise multiplication was operated between the scaled squeeze-excitation scalar and the feature maps (Figure 4.18 a). The SE block can be combined with other convolution architectures since it aims at providing additional interdependencies between the feature maps obtained from convolution blocks.

The IRB proposed in MobileNetV2 [115] refers to ResNet. It consists in a series of 1×1 convolutions, depth-wise 3×3 convolutions and 1×1 convolutions, and an identity mapping. The IRB follows an inverse order of the feature map number compared to the original residual block [77]. In IRB the network is expanded by the first 1×1 convolutions and squeezed by the second 1×1 convolutions (Figure 4.18 b). This conception was initially intended for the lightweight networks for mobile applications thanks to the reduced number of parameters of the depth-wise convolution.

Selective Kernel (SK) [126] aims to adaptively adjust the receptive field sizes. To enable the automatic kernel size selection, three operators are used in SK: split, fuse, and select. The split operator creates two branches for the next operators where the first branch passes through conventional 3×3 convolutions and another is 3×3 dilated convolutions with a dilation size of 2. Then, in the fuse operator, a third branch is created to store the multi-kernel information. In this branch, the feature maps obtained by the split operator are first fused by element-wise summation and then embedded by global average pooling. A fully-connected layer compacts the fused features into a lower dimension. Finally, in the select operator, the compact feature guides the selection of different spatial scales of information for the feature maps of the first and second branches by soft attention across channels. The definitive output of the SK block is the sum of the first and second branches considering the attention weights achieved by the soft attention across channels (Figure 4.18 c).

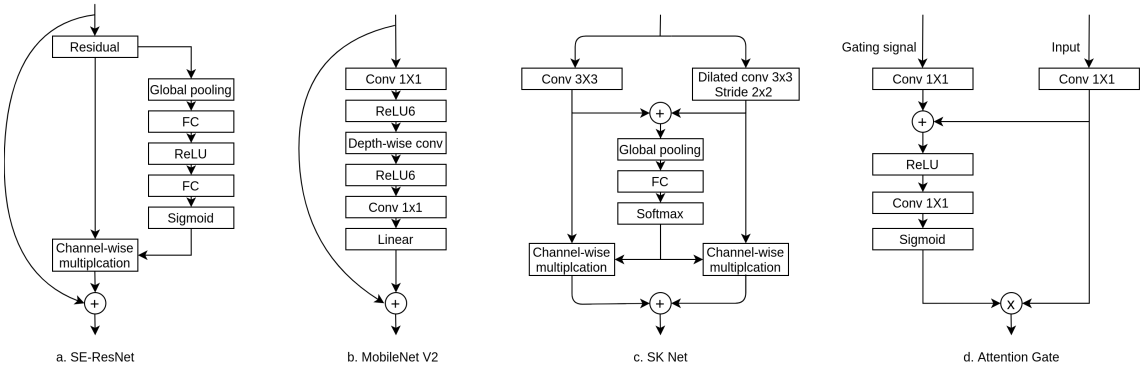


Figure 4.18: Building blocks referring to the attention mechanism for image interpretation. **a**. SE-ResNet: the residual SE block, **b**. MobileNet V2: the IRB from MobileNet V2, **c**. SK Net: the SK module can be deployed in the encoding or the decoding phases, **d**. Attention Gate: the attention gating from Attention U-Net should be deployed at the skip connection. The Gating signal comes from the encoding side and the input signal denotes the up-sampled features from the decoding side. The first two 1x1 convolution layers ensure the same number of channels for the two signals of the Attention Gate.

CE-Net CE-Net [125] outperforms multiple medical imaging data principally thanks to its Dense Atrous Convolution (DAC) and Residual Multi-kernel Pooling (RMP) blocks [59].

These additional manipulations on high-level semantic information aim at improving the interpretation of the size and form variations in medical images.

In CE-Net, the DAC block is inspired from both the dense block of Inception-ResNet-V2 [100] and the conventional convolution is replaced by the atrous convolution [86]. The atrous convolution is equivalent to the convolution with an upsampled filter. An atrous filter of rate r inserts $r-1$ zeros between two consecutive filter values on both dimensions. Rate r of 1 is effectively the standard convolution. Figure 4.19 shows the illustration of atrous convolution with different rates. The atrous convolution enlarges the receptive field without adding extra parameters and achieves better performance on different segmentation tasks. The DenseNet of Inception-ResNet-V2 as presented in Figure 4.13 is proved powerful at the cost of additional parameters. Thus CE-Net combines the atrous convolution and DenseNet only at the bottleneck to avoid excessive computation (Figure 4.20).

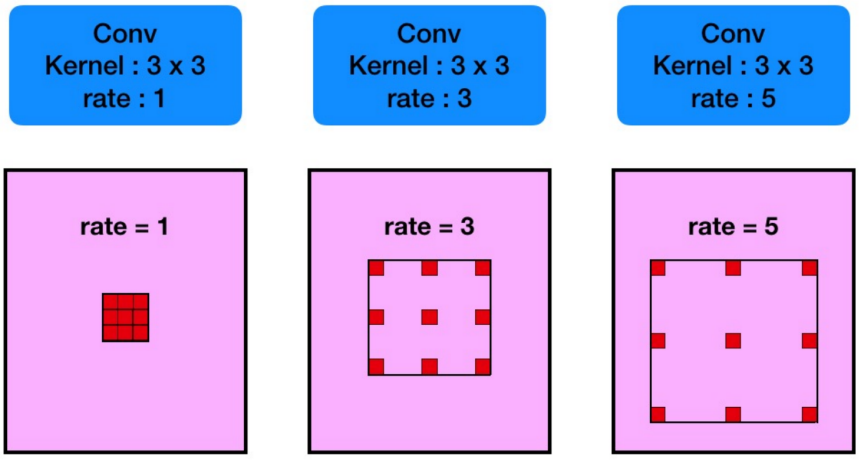


Figure 4.19: Illustration of atrous convolution with different rates. When the rate is 1, the atrous convolution is equivalent to the conventional convolution.

Feature maps convoluted by DAC block are then pooled by different pooling sizes and aggregated together as the output of the bottleneck (Figure 4.21). This block aims to extract context information of different scales. This conception responds to the challenge in medical images in which the objects may vary considerably. Nevertheless, in the experimental CE-Net, an input image is down-sampled 4 times as the baseline network, *i.e.* the 128 x 128 pixel^2 input will become 8 x 8 pixel^2 at the entry of the DAC block. Consequently, both DAC and RMP blocks may have little effect on such small feature maps. To take advantage of DAC and RMP blocks without changing the network's structure, input images are interpolated in order to increase their resolution by a factor of 5 using Nearest Neighbour Interpolation. This method duplicates the nearest neighboring pixels' values so that semantic information will not be modified.

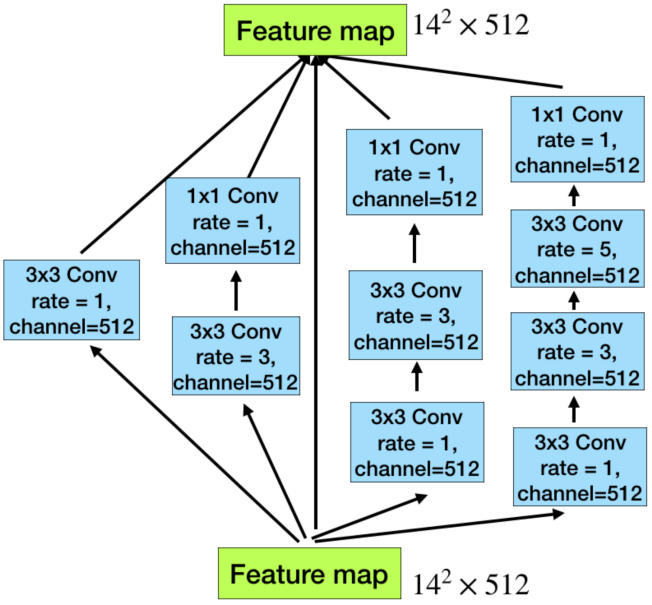


Figure 4.20: Illustration of Dense Atrous Convolution. A DAC block consists of four parallel atrous convolution branches. The number of convolutions between each branch gradually increases.

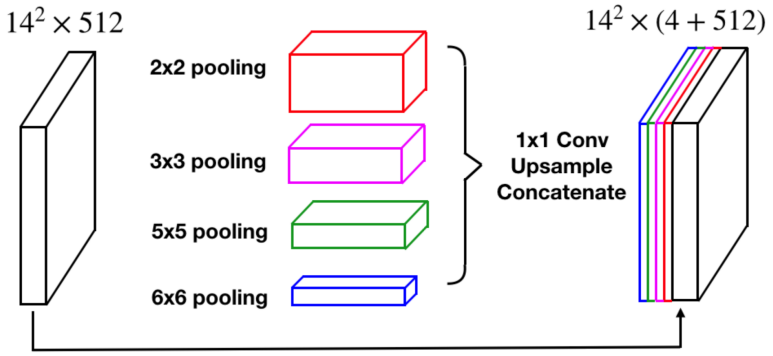


Figure 4.21: Illustration of Residual Multi-kernel Pooling. Feature maps passed from DAC are individually pooled by 4 different pooling and then upsampled and concatenated together.

4.3.4/ IN-LAYER NORMALIZATION

The in-layer normalization methods aim at fastening the velocity of convergence and mitigating some problems such as internal covariate shift or gradient vanishing [65]. The batch normalization is a widely employed in-layer normalization method. Batch normalization normalizes the samples in a batch that fixes the means and variances of each layer's inputs. Therefore, its formula can be referred to the image normalization methods, especially z-score normalization Eq. 5.4. The batch normalization is usually inserted after convolution, *i.e.,* the feature distribution of each mini-batch should be consistent.

Other usual in-layer normalization strategies are the instance normalization [83] and group normalization [117]. The main distinction between these in-layer normalization methods is the scope of their effect. For example, instance normalization is employed on every single instance instead of a mini-batch.

4.3.5/ STOCHASTIC OPTIMIZATION

Stochastic Gradient Descent (SGD) is a stochastic approximation of gradient descent optimization. Its objective is to minimize the distance between the output of a machine-learning model and the ground truth by updating the model's weight. The amount and the direction of each update are calculated according to the gradient of the loss function. More precisely, to achieve this goal, the following steps are executed by SGD:

1. A sample is randomly selected from the training dataset. The sample is fed to the model.
2. The gradient of all updateable parameters in the model are calculated according to the output of the loss function. The calculation follows the chain rule.
3. All the updateable parameters are updated. The update depends on the gradient and a hyper-parameter learning rate. The learning rate can be further optimized by some optimization strategies such as Momentum [7] and Adam [66].

The weight update of a parameter (weight) can be formulated as:

$$w := w - \eta \nabla Q_i(w) \tag{4.10}$$

where w is a parameter of the model, $Q_i(w)$ is the value of the loss function at i-th sample.

SGD is called "stochastic" due to the choice of the learning sample. SGD takes a randomly selected sample from the dataset for the gradient optimization in each iteration. Therefore, the gradient of SGD is an estimate thereof since it is calculated from one sample instead of the whole dataset. Moreover, if the gradient optimization is based on the gradient of all samples from a dataset at the same time, such optimization is called batch gradient descent optimization.

The mini-batch gradient descent optimization is a compromise between SGD and batch gradient descent. The mini-batch requires a more reasonable computation resource, especially the memory allocation compared to the batch gradient descent, and is more robust than SGD. In a big batch, the gradient of noisy samples has little impact on the weight update. Therefore, a bigger batch size can count more precisely the gradient over the dataset. Small batch size has a higher risk of misleading the update path. In most

CNN applications, the mini-batch gradient descent optimization has proved to be more suitable than the other two gradient optimizations [76].

4.3.6/ ACTIVATION FUNCTIONS

Like in most of the current deep learning models, the activation functions themselves are all nonlinear equations, their core functionality is to ensure that nontrivial problems can be fitted by deep neural networks. Sigmoid, rectifier (ReLU) [44] and its leaky variant leaky ReLU, exponential linear (ELU) [75], Swish [96], *etc.* are common activation functions for image-related tasks (Figure 4.22).

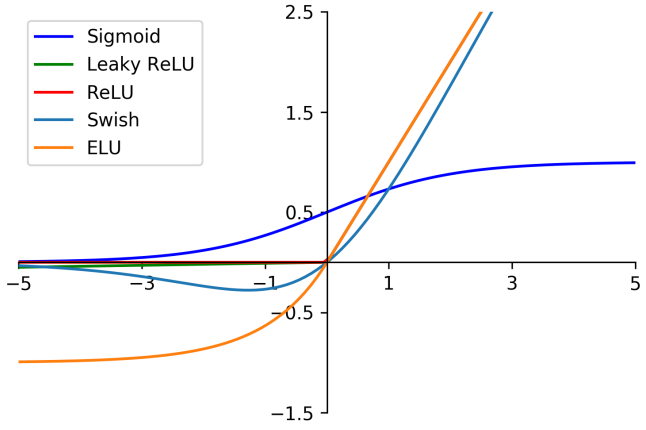


Figure 4.22: Common nonlinear activation functions for the image segmentation task.

Among the cited activation functions, sigmoid and ReLU are the ones that have been most applied. Sigmoid is suitable as the last activation function of a classification CNN since its output varies between 0 and 1. However, sigmoid risks of gradient vanishing because its gradient approaches to 0 when the input is at infinity. ReLU has no such problem [44] while it can not constrain its output. Moreover, ReLU is more computationally efficient compared to the exponential operations in sigmoid. In an evaluation of usual activation functions [113], results show that Swish's overall performance is better, although Swish may not always lead in every category, and Swish has limited advantages over other state-of-the-art activation functions.

4.3.7/ LOSS FUNCTIONS

Loss functions evaluate the similarity between the ground truth and the output. The loss function of a CNN should be differentiable so that the difference can be minimized by

backpropagation. The loss functions for the segmentation tasks can be classified according to the characteristics that are considered. The distribution, region and boundary are the three most common characteristics of the loss function conception. The compound of two or even more characteristics is also a frequently used penalization method. Figure 4.23 shows a summary of existent loss functions for image segmentation and their classification.

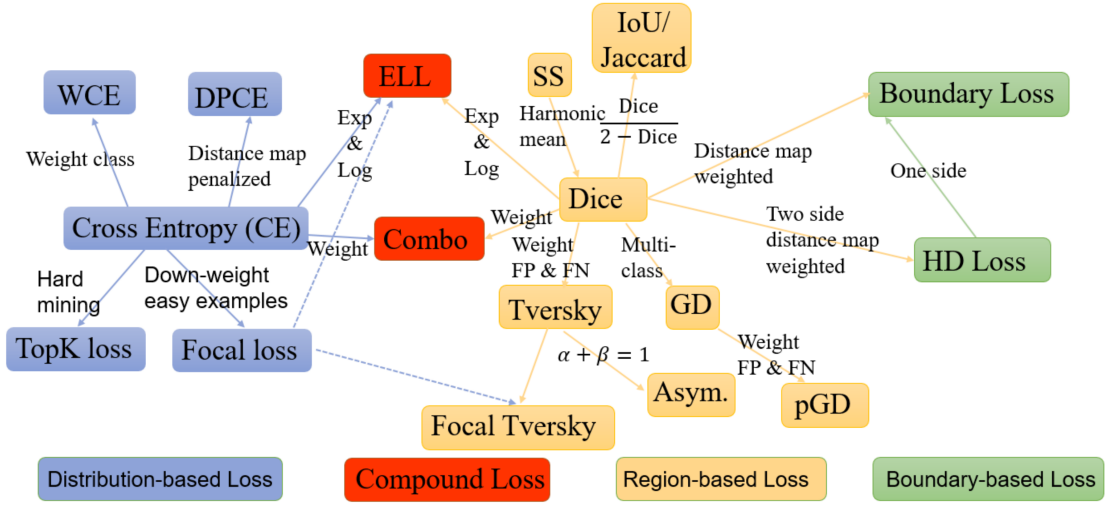


Figure 4.23: Overview and relationship among the existent loss functions. Credit: [156]

Distribution-based loss Cross Entropy (CE) has been widely employed in both medical imaging and semantic segmentation. CE derives from KL divergence. KL divergence is interpreted as the average difference of the number of bits required for encoding samples of a model using a code optimized for an approximation of the model rather than one optimized for the model. KL divergence of discrete probability distributions P(model) and Q(approximation) is formulated as:

$$\mathcal{D}_{KL}(P \parallel Q) = \sum_{x \subset \mathcal{X}} P(x) log \frac{P(x)}{Q(x)} \tag{4.11}$$

where $\mathcal{X}$ is the probability space of the two distributions. This formula is equal to:

$$\mathcal{D}_{KL}(P \parallel Q) = \sum_{x \subset \mathcal{X}} P(x) log P(x) - \sum_{x \subset \mathcal{X}} P(x) log Q(x) \tag{4.12}$$

$P(x)logP(x)$ is the entropy of the model P, which is a constant since P is invariant. Hence to describe the difference between the two distributions P and Q, the term $P(x)logP(x)$ can

be omitted so that the remaining part is the CE:

$$\mathcal{H}_{(P\|Q)} = -\sum_{x \subset \mathcal{X}} P(x) log Q(x) \tag{4.13}$$

For the multiclass segmentation task, the CE loss can thus be formulated as:

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{n}^{N}\sum_{c}^{C} g_{n,c} log(p_{n,c}) \tag{4.14}$$

where N denotes the number of samples (pixels or patches), C denotes the number of label classes, $g_{n,c}$ denotes the binary ground truth label of the sample n if it is of the class c, $p_{n,c}$ denotes the observation of a binary prediction if this sample is predicted as this label.

Region-based loss Region-based loss quantifies the region similarity between the prediction and the ground truth. The class imbalance is largely mitigated by the fact that the unit of calculation is an area rather than a pixel. For example, Dice loss is the widely applied loss for medical image segmentation tasks. The original Dice loss is formulated as:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{n}^{N} p_n g_n + \epsilon}{\sum_{n}^{N} p_n^2 + \sum_{n}^{N} g_n^2 + \epsilon} \tag{4.15}$$

Indeed, the Dice loss calculates the proportion of the common part compared to the sum of the individual surface, that is, $2\sum_{n}^{N} p_n g_n$ (p_n and g_n are binary bits) is the surface of the intersection region between the prediction and the ground truth, $\sum_{n}^{N} p_n^2$ and $\sum_{n}^{N} g_n^2$ are the surfaces of the prediction and of the ground truth respectively. ϵ is added to avoid the case where a non-target region is present in a sample.

The Dice loss aims at penalizing binary segmentation problems. For multiclass tasks, it can be transformed to prediction class by class after the softmax function. However, this intuitive implementation gives equal weight to each class, potentially resulting in the class imbalance issue. Despite Dice loss largely solving the pixel-level class imbalance issue, the instance imbalance should be considered. The instance imbalance issue refers to the inequality of regions between classes. For example, in a dataset consisting of multi-class images, the regions of one class may appear in few images. To address this shortcoming of the vanilla Dice loss, the weighted multi-class Dice loss has been proposed by Carole H *et al.* [99], termed Generalized Dice loss:

$$\mathcal{L}_{GDice} = 1 - 2\frac{\sum_{c}^{C} w_c \sum_{n} p_{cn} g_{cn}}{\sum_{c}^{C} w_c \sum_{n} p_{cn} + g_{cn}} \tag{4.16}$$

The weigh of a class w_c is calculated as $w_c = 1/\left(\sum_{n=1}^{N} g_{cn}\right)^2$, which is the inverse of the

square proportion of each class's pixels quantity in the training dataset.

Boundary-based loss Boundary-based losses are inspired by geometrical metrics that evaluate the segmentation mask such as the Hausdorff distance [38]. Hausdorff distance measures how far two subsets of a metric space are from each other. In the case of two segmentation masks in one image, the greatest of all the distances from a point in mask A to the closest point in mask B:

$$\mathcal{D}_H(A,B) = max\{sup_{a\subset A} d(a,B), sup_{b\subset B} d(b,A)\} \tag{4.17}$$

where sup represents the supremum, inf the infimum, $d(a,B) = inf_{b\subset B} d(a,b)$ quantifies the distance from a point a in A to the subset B. The loss function based on Hausdorff distance is then proposed [141] as:

$$\mathcal{L}_{HD} = \frac{1}{N}\sum_{n}^{N}[(p_n - g_n) * (d_{Pn}^2 + d_{Gn}^2)] \tag{4.18}$$

where d_P and d_G are distance transformations of the prediction and the ground truth masks.

Targeted variants for imbalance issue Although pixel imbalance problems have no effect on most region and boundary-based loss functions in binary segmentation tasks, the original implementations of these loss functions are insufficient for multi-class segmentation or instance imbalance problems, let alone distribution-based loss functions. A general solution for the above shortcomings is the weighting technique on the basis of the abundance of each class in a dataset.

The weighted Cross Entropy loss can be viewed as the baseline solution for both the challenge of multiclass and instance imbalance conditions. In weight CE loss, each class is weighted differently according to its abundance in the dataset. Its common implementation can be formulated as:

$$\mathcal{L}_{WC\mathcal{E}} = -\frac{1}{N}\sum_{n}^{N}\sum_{c}^{C} w_c g_{n,c} log(p_{n,c}) \tag{4.19}$$

where w_c denotes the weight of each class c. The weight can be calculated as the inverse of the class's pixel proportion. Since CE naturally calculates the similarity between two multiclass distributions, the additional weighting allows WCE loss to penalize most segmentation conditions.

The CE can be weighted in accordance with the sample's difficulty. The difficulty is defined as whether the prediction of the sample is high enough. Focal loss [93] introduces a

factor to down-weight the loss assigned to well-classified examples. To do so, the authors propose the following implementation of Focal CE for the binary classification problem:

$$FL = -\frac{1}{N}\sum_{n}^{N}(1-p_n)^{\gamma}log(p_n) \tag{4.20}$$

where $\gamma \subset (0, \infty)$ is a configurable factor. The Focal CE has such characteristics to ensure the down-weighting on easy samples and force to focus on hard samples:

- The additional factor term $(1-p_n)^{\gamma}$ controls the weighting amplitude. When γ is set to 0, the Focal loss and the vanilla binary CE are identical. The configuration of this value is empirical and the authors recommend a factor of 2.
- When the prediction on a sample is wrong, p_n should be very small and thus the term $(1-p_n)^{\gamma}$ is close to 1 regardless of the value of γ. In this case, the focal factor has little impact on hard cases since the difference between the focal CE and the original CE is negligible.
- When the prediction on a sample is satisfying, $(1-p_n)$ is close to 0 and $(1-p_n)^{\gamma}$ is smaller than 0. This produces a smaller loss value than the original CE in the case of a good (easy) prediction.

The above analyses explain the mechanism for reducing the weighting on easy samples. Moreover, the derivative of the introduced focal term $(1-p_n)^{\gamma}$ exists on $[0, 1]$, therefore the Focal CE is differentiable. To generalize the Focal CE for multiclass segmentation tasks, it can be implemented as:

$$\mathcal{L}_{FCE} = -\frac{1}{N}\sum_{n}^{N}\sum_{c}^{C}(1-p_{n,c})^{\gamma}g_{n,c}log(p_{n,c}) \tag{4.21}$$

TopK loss is another variant of the CE loss that focus more on difficult samples. Similar to Focal CE, during the training, only the samples whose prediction is not good enough will be taken into account. To realize such a filter according to the sample's difficulty, the formula is:

$$\mathcal{L}_{TopK} = -\frac{1}{\sum_{n}^{N}\sum_{c}^{C}\mathbf{1}\{g_{n,c}==1\&p_{n,c}<t\}}\sum_{n}^{N}\sum_{c}^{C}\mathbf{1}\{g_{n,c}==1\&p_{n,c}<t\}g_{n,c}log(p_{n,c}) \tag{4.22}$$

where $\mathbf{1}\{g_{n,c} == 1\&p_{n,c} < t\}$ is the binary indicator function that returns 1 if the inside conditions are met, $t \subset (0, 1]$ is the threshold value that evaluates if a sample is difficult. If the prediction value is higher than the threshold, the prediction on such sample is thought too easy, and thus the binary indicator function returns 0 so that the easy samples are omitted for the training of the current model.

Compound loss Combining two or even more loss functions is proven more robust or efficient in many applications. Some specific combinations have been proposed and evaluated such as the Combo loss and Exponential Logarithmic loss.

Combo loss combines the weighted CE and the Dice loss [133]. A coefficient is set as a hyperparameter that weighs between the WCE term and the Dice term. The formula for binary segmentation tasks is:

$$\mathcal{L}_{combo} = \alpha\left(-\frac{1}{N}\sum_{n}^{N}\beta(g_n log p_n) + (1-\beta)[(1-g_n)log(1-p_n)]\right) - (1-\alpha)\frac{2\sum_n^N p_n g_n + \epsilon}{\sum_n^N p_n^2 + \sum_n^N g_n^2 + \epsilon} \tag{4.23}$$

where in addition to the vanilla CE loss and the vanilla Dice loos, α controls the amount of Dice term contribution in the entire loss, and when β is set to smaller than 0.5, the false positive cases are more penalized than the false negative cases due to the heavier coefficient in $(1-g_n)log(1-p_n)$, and vice versa.

Furthermore, the equation can be extended to multiclass segmentation problems, however, the coefficient α that penalizes differently the false positive and false negative cases is no longer necessary:

$$\mathcal{L}_{combo} = -\alpha\frac{1}{N}\sum_{n}^{N}\sum_{c}^{C} w_c g_{n,c} log(p_{n,c}) - (1-\alpha)\frac{2\sum_c^C w_c \sum_n p_{cn} g_{cn}}{\sum_c^C w_c \sum_n p_{cn} + g_{cn}} \tag{4.24}$$

Besides the Combo loss, Exponential Logarithmic loss also combines CE and Dice losses with respectively additional exponential and exponential logarithmic:

$$\mathcal{L}_{exp} = -\alpha\frac{1}{N}\sum_{n}^{N}\beta(g_n log p_n)^{\gamma CE} - (1-\alpha)ln(\frac{2\sum_n^N p_n g_n + \epsilon}{\sum_n^N p_n^2 + \sum_n^N g_n^2 + \epsilon})^{\gamma Dice} \tag{4.25}$$

where γCE and $\gamma Dice$ adjust the nonlinearities of the loss. According to the authors' evaluations, with $\gamma > 1$, the loss focuses more on less accurate labels than the Focal loss while for the multiclass segmentation task, better results are obtained with $0 < \gamma < 1$.

4.3.8/ TRANSFER LEARNING

Transfer learning is a general machine learning technique that gains further knowledge from a related problem [41]. The idea of this technique is to learn the representation of a set of targeted characteristics with another similar set. The model is then considered to be able to transfer knowledge across tasks.

Medical image segmentation tasks are often faced with the limitation of available data volume. Transfer learning enlarges the quantity of knowledge the model can learn. To do so, a model can learn from a similar dataset that shares similar features as the target data.

For example, the training data can be acquired with different scanners or a kind of scanner with different sequences. Although the distribution between different data sources is different, transfer learning aims at selecting and then learning the shared features. For example, Huellebrand *et al.* [152] trained the myocardium delineation model with cine MRI from the ACDC dataset [105]. The target data is the DE-MRI that has the same short-axis orientation but different sequences as presented in Chapter 3.

Another common scenario of transfer learning in medical imaging assessment, not limited to the segmentation is the learning of the same data but for different tasks, such as the detection of multiple types of abnormalities, where the detection of each type of abnormality is a binary classification problem. More applications of transfer learning on medical image tasks have been discussed by Cheplygina *et al.* [121].

4.3.9/ Segmentation applications on cardiac MRI

The early segmentation work on DE-MRI mostly employed statistical models. Luciano *et al.* segmented the infarct on DE-MRI for canine MI. Mikami *et al.* compared the scar segmentation performances of FWHM, STRM with different SD and Otsu-auto-threshold on DE-RMI with hypertrophic cardiomyopathy [62]; Flett *et al.* studied the reproducibility of scar segmentation using FWHM and STRM on a dataset consisting of acute myocardial infarction, chronic myocardial infarction, and hypertrophic cardiomyopathy [43]. Lalande *et al.* segmented the challenging peri-infarct area of myocardial infarction with GMM [49].

Previous deep learning-based DE-MRI segmentation work mostly concentrates on the ventricle delineation. For example, Avendi *et al.* combined deep learning with a deformable model [73]; Yue *et al.* [135] combined a Shape Reconstruction Neural Network (SRNN) and a spatial constraint network. The number of the scar segmentation work is relatively limited. De La Rosa *et al.* [123] first applied morphology treatment for coarse infarction segmentation, then CNNs were employed for the false-positive elimination and segmentation refinement. Fahmy *et al.* [109] presented an initial proof-of-concept for the scar segmentation with U-Net. Moccia *et al.* [129] and Zabihollahy *et al.* [118] segmented the scar tissue on the given myocardium mask thanks to semi-automatic approaches. Recently organized challenges,

Some reviews and research compared different segmentation techniques. Based on the results of a MICCAI challenge in 2016, Karim *et al.* [78] summarized and investigated most of the rule-based approaches considering human and animal datasets. This review indicates that the myocardial infarction segmentation results of rule-based approaches do not completely satisfy the clinical requirements. Zabihollahy *et al.* justified the superiority of the CNN-based myocardial scar segmentation model compared to statistical models [136] on 3D MRI.

More recently, a MICCAI challenge in 2020 gathered challengers from all over the world in the contests of MI assessment [142]. One contest in the challenge was the myocardial tissue segmentation on LV. The segmentation contest winner proposed a concatenated two-stage segmentation model [159]. In the model, a 2D U-Net variant segments all the tissues as the preliminary result. The preliminary segmentation is then concatenated with 3D MRI as the input of a 3D U-Net variant for definitive segmentation. Other challengers employed or proposed one-stage or two-stage models with U-Net-based segmentation algorithms and specific image preprocessing and post-processing methods. The employed techniques and highlights are concluded in Table 4.1 and the segmentation results are shown in Table 4.2 for all the challengers of the segmentation contest in EMIDEC Challenge 2020 [146, 147, 150, 151, 152, 158, 160, 159].

Table 4.1: Principal concepts of the methods for the segmentation contest.

Challenger(s)	Framework	Methods	Highlights
Brahim *et al.*	Two-stages	Myocardium: 2D U-Net with Attention and IRB Infarct: 3D U-Net variant	3D Auto-encoder to perfect myocardium shape
Camarasa *et al.*	Two-stages	Myocardium: 3D U-Net variant Infarct: 3D U-Net variant	Uncertainty myocardial area generated by probabilistic auto-encoder for infarct segmentation
Feng *et al.*	One-stage	2D U-Net with dilated convolutions	Data augmentation with additional scar tissues
Girum *et al.*	Two-stages	Myocardium: 2D U-Net with SE block Infarct: 2D U-Net with SE block	Independent myocardium and infarct segmentation from non-cropped MRI
Huellebrand *et al.*	Two-stages	Myocardium: 2D U-Net variant Infarct: mixture model	Transfer learning with cine-MRI Post-processing with thresholding and morphological closing
Yang *et al.*	One-stage	2D U-Net with SE and SK blocks	RGB channel-like adjacent slices input Two decoder branches supervised by myocardium and infarct masks
Zhang	Two-stages	Preliminary: 2D U-Net variant Definitive: 3D U-Net variant	3D MRI with cascaded 2D segmentation as 3D input
Zhou *et al.*	One-stage	2D U-Net with Attention	Data augmentation with mix-up strategy Neighborhood penalty as neighboring loss

Table 4.2: Results of the segmentation contest. The metrics are given by target tissue (myocardium, infarct and PMO). The table is sorted by the general ranking of the contest, which is calculated from the nine subranks. Best results in bold.

Challenger(s)	Myocardium			Infarction			PMO				
	Dice	Vol. Diff. (cm^3)	Hausdorff (mm)	Dice	Vol. Diff. (cm^3)	Pct. Diff. (%)[1]	Dice	Vol. Diff. (cm^3)	Pct. Diff. (%)[1]	Acc. (case,%)[2]	Acc. (slice,%)[2]
Zhang	**0.879**±0.027	**9.26**±9.08	**13.01**±8.81	**0.712**±0.268	**3.12**±5.15	**2.38**±0.031	**0.785**±0.393	**0.63**±2.27	**0.38**±0.012	**84.00**	**94.97**
Feng *et al.*	0.836±0.124	15.19±16.41	33.77±111.63	0.547±0.340	3.97±8.36	2.89±0.045	0.722±0.432	0.88±3.41	0.53±0.017	80.00	90.78
Yang *et al.*	0.855±0.027	16.54±10.27	13.23±6.80	0.628±0.315	5.34±7.88	4.37±0.062	0.610±0.463	1.85±3.32	1.69±0.033	76.00	81.56
Huellebrand *et al.*	0.841±0.051	10.87±8.53	18.3±15.74	0.379±0.296	6.17±8.36	4.93±0.059	0.523±0.483	0.95±3.00	0.64±0.015	70.00	85.75
Camarasa *et al.*	0.757±0.111	17.11±15.45	25.44±21.71	0.308±0.280	4.87±8.49	3.64±0.047	0.605±0.485	0.87±3.27	0.52±0.016	74.00	84.36
Zhou *et al.*	0.825±0.057	13.29±11.34	83.42±158.97	0.378±0.309	6.10±9.45	4.71±0.06	0.520±0.487	0.88±3.38	0.54±0.017	64.00	86.87
Brahim *et al.*[3]	0.791±0.050	12.68±10.59	23.87±11.52	0.274±0.379	7.05±12.73	5.19±0.074	0.641±0.479	0.83±3.109	0.50±0.016	74.00	89.39
Girum *et al.*[3]	0.803±0.057	11.81±14.09	51.48±98.15	0.340±0.474	11.52±16.53	8.58±0.101	0.780±0.414	0.89±3.61	0.51±0.018	78.00	89.66

[1] Pct. Diff. : Difference between the percentage of the infarcted myocardium.
[2] Additional metrics. These metrics were not taken into account in the ranking.
[3] Co-author(s) come(s) from the challenge organization team. Do(es) not participate in rankings.

4.4/ CONCLUSIONS

Related works about automatic MI segmentation are presented in this chapter, including segmentation models and applications. Statistical models have been employed for cardiac segmentation tasks for a long time while U-Net-based deep learning models are showing more and more promising segmentation results on different medical images.

Popular deep learning-based segmentation models usually employ the backbone of U-Net in which the skip connection restores more details for the segmentation reconstruction. The current trends are the optimization of building blocks such as the residual block or the blocks referring to the attention mechanism, the employment of compound loss function that aims at solving the class imbalance issue, and diverse improvements including the data augmentation, data preprocessing, auxiliary networks, *etc.*

5

CONTRIBUTION 1: AUTOMATIC MI ASSESSMENT FROM MRI WITH 2D DL MODELS

5.1/ INTRODUCTION

As discussed in the previous chapters, the automatic assessment of MI from DE-MRI is challenging but of great clinical potential to improve the efficiency and accuracy of pathology diagnosis and therapeutic management. The current automatic assessment is based on image analysis techniques including probabilistic models and deep learning models.

According to image analysis techniques and anatomical and clinical prior knowledge, the assessment result can be in form of a delineation of myocardial tissues, or directly a comprehensive report of the presence/rate of infarct. In this chapter, only the proposed deep learning-based 2D segmentation approaches will be presented. The deep learning-based 2D segmentation approaches consist of the preprocessing methods, the design of networks, the segmentation frameworks and the post-processing methods.

The target of the segmentation is to delineate the contours of myocardial tissues on short-axis DE-MRI, incorporating the myocardium and the infarct. The PMO is viewed as a part of infarct and will not be independently assessed in this chapter. The CNN-based automatic assessment will be evaluated in comparison with the manual annotations and a popular probabilistic model. Proposed or evaluated segmentation approaches and their results on the mentioned private and public datasets will be presented.

The 2D deep learning-based myocardial tissues segmentation includes two applications: the segmentation on all-positive images and the segmentation on mixed datasets. The all-positive images consist of the DE-MRI involving at least one scar tissue area (with or without PMO). The mixed dataset consists of all the acquired DE-MRI exams, regardless

of whether the MI presents or not according to the manual annotations. The major difference between the two proposals is the class imbalance issue. With the mixed dataset, the image-wise false-positive or false-negative predictions should be specifically addressed.

This chapter is organized as follows:

- Preprocessing. The employment of preprocessing methods including data augmentation, data formatting and image normalization will be first stated.
- 2D segmentation models with deep learning. The conception details of our experimental 2D segmentation models will be presented. Particularly, the motivation for using a less common optimizer, stochastic gradient descent, will be discussed in length.
- Framework. Two frameworks for 2D segmentation will be illustrated.
- Post-processing. The proposed prior-based post-processing method incorporating three criteria will be introduced.
- Experimental results. The experiments are executed on all-pathological slices, then on all patients mixing normal and pathological slices.
- Discussion.
- Conclusions.

5.2/ PREPROCESSING

The data preprocessing is the first step to prepare a predictive pipeline. Data preprocessing has two objectives: ensuring a uniform and appropriate input data format, and enhancing the valuable features. The original DE-MRI images and corresponding manual annotations should be properly selected and processed to form a uniform format that can be correctly interpreted by CNN-based segmentation models. Some automatic transformations, such as the image normalization and denoising, can then be applied to DE-MRI to possibly enhance the feature interpretation.

5.2.1/ DATA AUGMENTATION

Although in the previous chapter, the relative data segmentation techniques have been presented and discussed, no data augmentation was employed for the preparation of this thesis since the generated data should follow the distribution of the original data, thus completely new features should not be produced. Otherwise, the training of the

segmentation model may be misguided due to the artificial characteristics. Moreover, the results of the EMIDEC challenge show that the proposals employing data segmentation techniques did not outperform others. For these reasons, data augmentation methods were not evaluated on experimental data in this work.

5.2.2/ DATA FORMATTING

The term formatting here refers to the uniformity of the segmentation model's input data format. The image formatting for 2D models naturally produces 2D images as training and test samples. The essential formatting includes the region selection and cropping.

The region selection depends on the dataset. The private dataset consists of the original acquisition of DE-MRI, therefore the position of the left ventricle is not necessary on the middle of the MRI. Indeed, the scanning plan is controlled by technicians and the left ventricle is just guaranteed to be fully displayed in the MRI. Furthermore, a CNN-based segmentation model normally requires an invariant input size, and ideally an aspect ratio of 1. On the basis of these characteristics of data and CNN models, a 128 x 128 image is cropped at the center of the initial DE-MRI of the private dataset. For the images whose width or length is less than 128, a zero-padding is added to fill them up. For the 11 images in which the left ventricle is too close to the image border, the position of the cropping box was manually shifted to ensure the presence of the entire myocardium in the networks' input. Although the result of interpolation is also a fixed-size image, this is not an optimal solution as the spatial information will be modified due to distortion.

For the models employing the public dataset, the cropping is slightly different. Since the DE-MRI cases in the public dataset have been manually aligned vertically and each slice has been shifted to the image center, the cropping box can be smaller to cover the left ventricle in all slices. After examining the minimum size of all the left ventricles, a center-cropping of 96 x 96 is applied on the public dataset to prepare its corresponding training and test data.

The side lengths of the two cropping boxes are deliberately set to 128 and 96. Such configurations assure consistent feature map sizes between the downsampling and upsampling. As introduced in Section 4.3.2, the feature maps are encoded and then decoded in U-Net-like segmentation models in which the downsampling and upsampling rates are usually set to 2 by (de)convolution with stride or pooling. At each level of paired downsampling-upsampling blocks, the skip connection requires an identical feature map size. However, when the feature maps to be encoded are not divisible by 2, zero-padding will be added. In the contrast, on the decoder side, the feature map size at all levels should be divisible by 2. The additional padding at the encoder side differentiates the feature map size if the same padding is not manually added to a level at the decoder

side. To avoid this additional manual parameter, the factorization of the side length of ideal input data should include the power of 2, and its exponent should not be smaller that the number of downsampling operations. Since the employed U-Net-based models have 5 downsampling blocks, 128 (2^7) or 96 ($2^6 \times 3$) are chosen as the optimal data sizes. Furthermore, the DE-MRI was acquired at 12-bit (intensity dynamics range: 0-4095). To keep the same intensity range and avoid the loss of original information, the cropped data are also registered as 12-bit .npz files. Figure 5.1 shows examples of cropping from the private and the EMIDEC dataset.

Apart from the image selection and cropping strategies of the DE-MRI, the mask drawing method is also different between the two datasets. The manual annotation of the private dataset is registered as JSON files, consisting of a set of points. The contour of each tissue area is enclosed by consecutive points. Each mask is thus enclosed by consecutive points. However, the manual annotation was drawn by experts on interpolated MRI to achieve a higher drawing precision, which makes the coordinates consist of floating point numbers. To convert the floating point coordinates to masks, the floating point numbers are rounded.

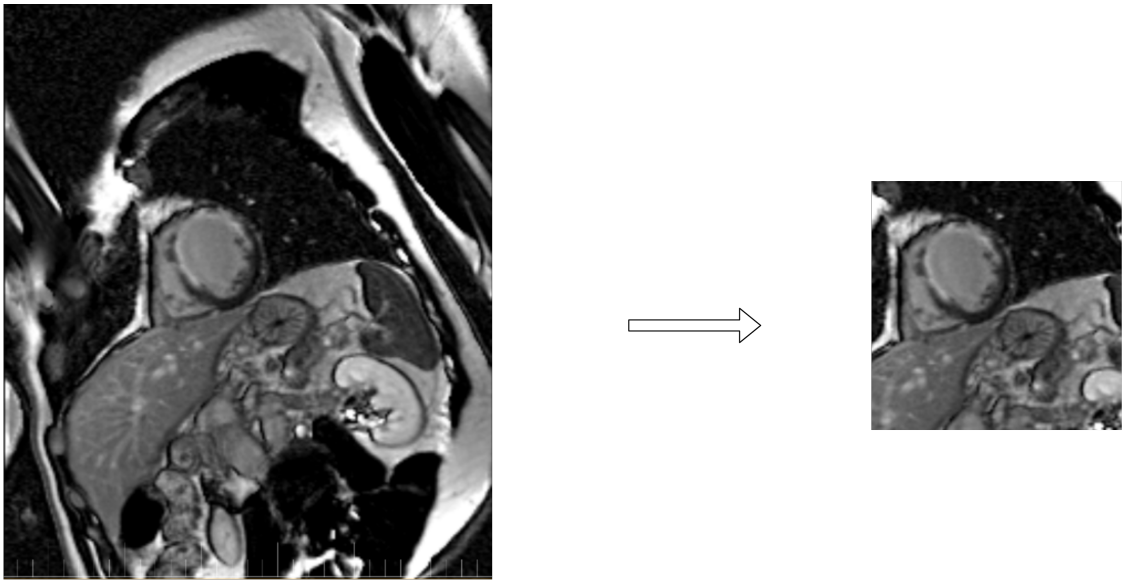
5.2.3/ IMAGE NORMALIZATION

The MRI signal was registered as a grayscale image after the acquisition. From a physician's point of view, the image normalization method could affect the judgment concerning the lesion area by modifying the contrast of the image. As a result of this observation, image normalization methods are proposed or evaluated on the input images as a preprocessing method.

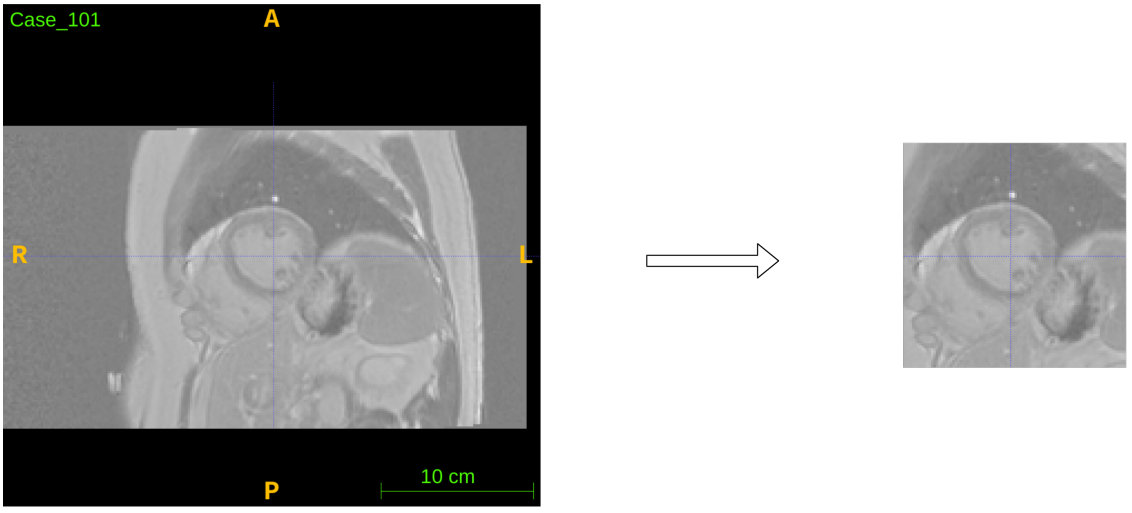
The image normalization aims at correcting the heterogeneous intensity between cases. Various normalization rules can be followed to correct the expected pixel distribution of images. The proposed normalization method is defined as follows:

$$\forall i \in I, i = \begin{cases} 0, & i \leq a \\ \frac{i-a}{b-a} \times 2^n - 1, & a < i < b \\ 2^n - 1, & i \geq b \end{cases} \tag{5.1}$$

where i is the intensity level of a pixel in image I. Contrary to the usual linear histogram normalization, top 5% and bottom 5% of intensity values on the histogram of the initial image are chosen as the lower and the upper borders (a and b) of the normalization window. An intensity level that is out of this window is then normalized as 0 or $2^n - 1$, where n is the depth of the normalized image (n in bits). All other intensities are linearly transformed to fill in the dynamic window. The contrast in the dynamic window is thus enhanced because of the shortened dynamic range. Nevertheless, the scope of the



a. A MRI slice from the private dataset



a. A MRI slice from the EMIDEC dataset

Figure 5.1: Examples of cropping from the private and the EMIDEC dataset. The cropped image from the private dataset is not aligned according to the center of the left ventricle, therefore its cropping box is larger than the one from the EMIDEC dataset to ensure the entire presence of the left myocardium.

normalization window can be adjusted, *i.e.* the top x and bottom y of intensity values on the histogram of the initial image. The higher the values of x and y, the smaller the scope of original dynamics and thus the higher the contrast in this window. Figure 5.2 visually compares three configurations of the scope of original dynamics on the same DE-MRI slice.

The second normalization method is the equalization which equalizes the intensity distri-

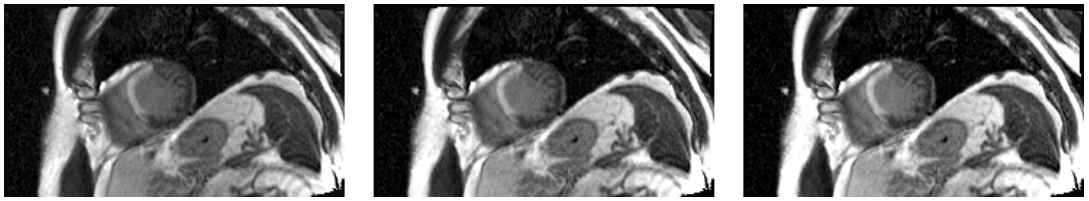


Figure 5.2: Effect of normalization dynamics scope on DE-MRI. From the left to the right, the three images are normalized according to Eq. 5.1 with the dynamics scope of 2%-98%, 5%-95%, 8%-92%, respectively.

bution on the histogram. The most frequent intensity values are spread out and vice versa [5]. The probability of an occurrence of a pixel of intensity level i can be represented as:

$$p_x(i) = p(x = i) = \frac{n_i}{n} \tag{5.2}$$

where $\{x\}$ is a grayscale image, n is the number of pixels whose intensity level is i, and n is the number of pixels in the image. Furthermore, the cumulative distribution function of an equalized $\{x\}$ according to i should be linear:

$$cdf_x(i) = \sum_{j=0}^{i} p_x(x = j) = iK \tag{5.3}$$

where K is constant.

In practice, since the grayscale image can be considered as a discret distribution, the intensity level i is usually replaced by a small range to facilitate the processing.

The last normalization method is Z-score normalization:

$$i = \frac{x - \mu}{\sigma} \tag{5.4}$$

where i is the pixel intensity after the Z-score normalization, μ the mean intensity at the level of the MR slice and σ the standard deviation of the slice intensity. Normalized images have a grey level distribution with zero mean and unit standard deviation so that the inter-case intensity distribution is uniform.

Figure 5.3 shows the normalized images processed with different methods and their original image and Figure 5.4 provides the corresponding histograms. The z-score is not shown in the figures since the obtained image is out of the dynamic range of a normal n-bit image.

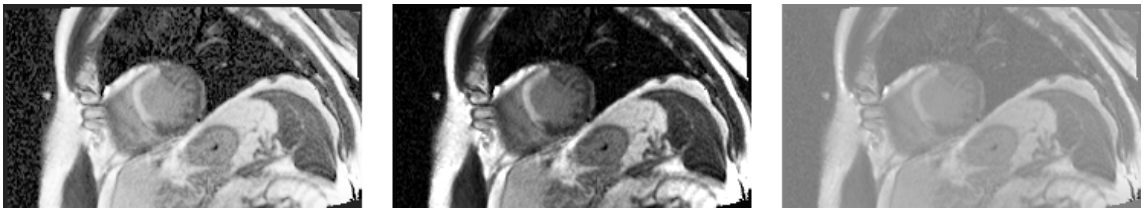


Figure 5.3: Effect of normalization methods on a DE-MRI slice. From the left to the right, the three images are equalization, normalization according to Eq. 5.1 with the dynamics scope 5%-95%, original image.

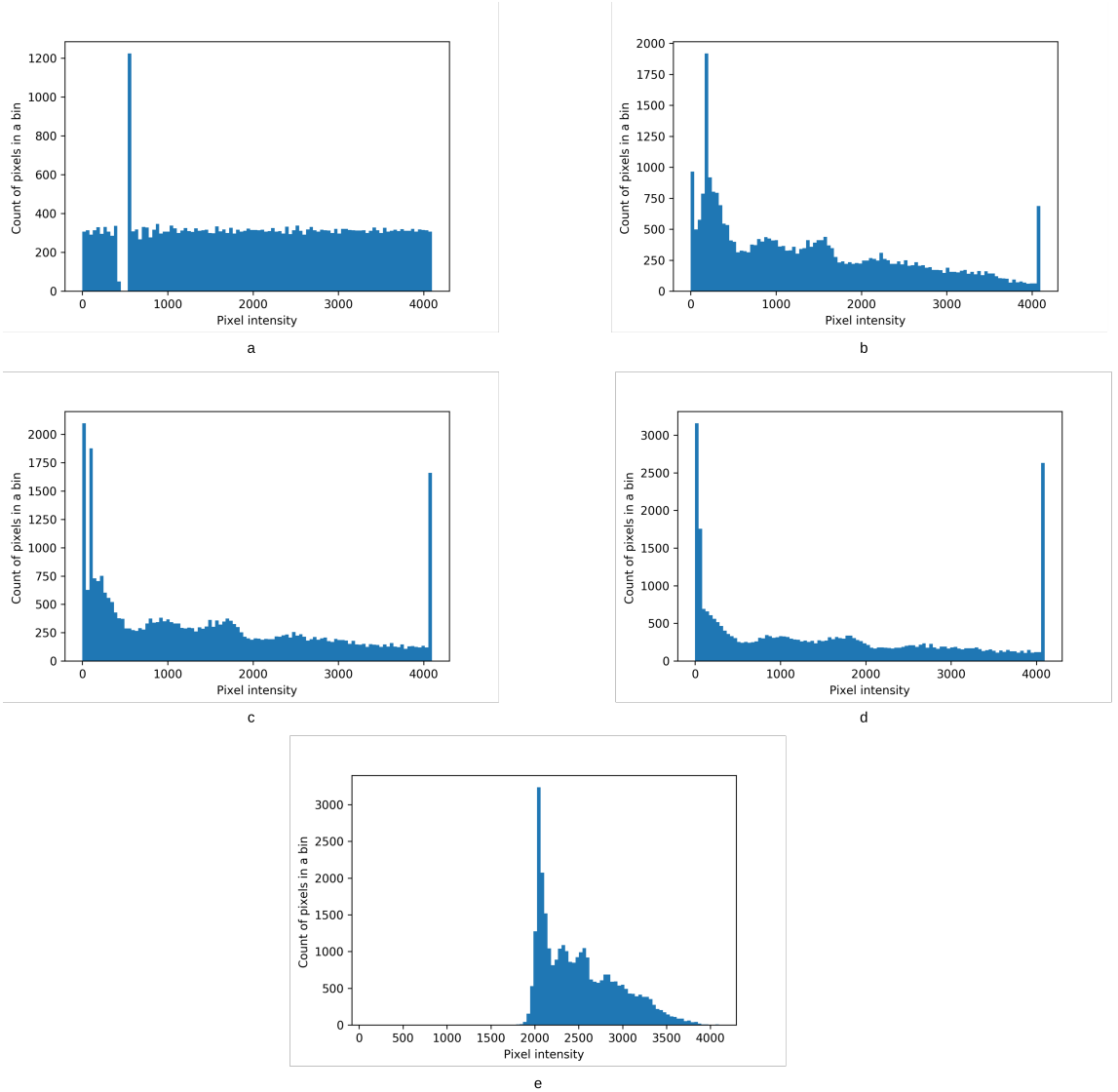


Figure 5.4: Histograms of different normalization methods on the same DE-MRI slice. **a**: equalization. **b**: normalization according to Eq. 5.1 with the dynamics scope 2%-98%. **c**: normalization according to Eq. 5.1 with the dynamics scope 5%-95%. **d**: normalization according to Eq. 5.1 with the dynamics scope 8%-92%. **e**: original image. The histogram of the equalization is not strictly balanced in every bin due to the discrete distribution.

5.3/ SEGMENTATION WITH 2D MODELS

The segmentation of MI and related myocardial tissues is more intuitive with 2D model. To obtain the segmentation of all the tissues in short-axis DE-MRI, a variety of techniques are proposed or studied, involving preprocessings, network conception details and post-processings. The proposed segmentation models have been performed on images in which the myocardium delineation is known or not, and on a dataset that includes only pathological slices or mixed images. The employed datasets are a private dataset (called private dataset) and a publically available dataset (called EMIDEC dataset) as introduced in Section 3.4, thus the methods may depend on the experimental dataset. Although the configuration of the target network, such as the framework and loss function, may vary depending on the conditions of the problem, most techniques can be shared for each specific task.

Therefore, in this part, the conception of the 2D predictive models will be presented with the motivation and implementation details. The methodology presentation will follow the stream of a predictive pipeline: the preprocessings of the training and test, then the predictive model itself including the building blocks, in-layer normalization methods, gradient optimization, loss function and framework, and finally the post-processing methods.

5.3.1/ BUILDING BLOCKS AND NETWORK

Baseline: U-Net with residual blocks The 2D segmentation networks are modified from the vanilla U-Net given its remarkable performance in many medical image segmentation tasks. In addition to the network conception, the residual block of ResNet has also been approved as revolutionary in various CNN-based applications. Therefore, U-Net with residual blocks is chosen as the baseline network.

After weighing network depth against input image size, the baseline network consists of 5 (de)convolution blocks. At the stage of the bottleneck block, the feature maps are downsampled to $1/16$ according to the size of the network's input, *i.e.*, feature maps of the size 8×8 if the input images are 128×128. Inspired by the implementation of Resnet50, The first convolution in the encoder employs a large kernel size of 7×7. This configuration allows to enlarge the receptive field on images of excessive scale. Despite the kernels of 7×7 bring more than 5 times the parameters than the kernels of 3×3, the extra computational overhead is limited as such large kernels are employed only at the first block. Furthermore, the next convolution blocks also refer to Resnet50. Each encoder block consists of stacks of residual blocks as shown in Figure 4.12 (right). The number of stacks is different as the authors of ResNet have proven by experiments after the trade-off.

On the decoder side, each deconvolution block consists of 1×1 convolution, 3×3 *deconvolution with stride* 2 and 1×1 convolution (Figure 5.5). The first 1×1 convolution fuses feature maps transmitted from the skip connection and the last upsampling, the last 1×1 convolution doubles the feature maps from the deconvolution. Each deconvolution block doubles the feature maps' size and halves the channel of feature maps. Figure 5.1 shows all the major parameters of 2D baseline segmentation model in the case of 128×128 input images.

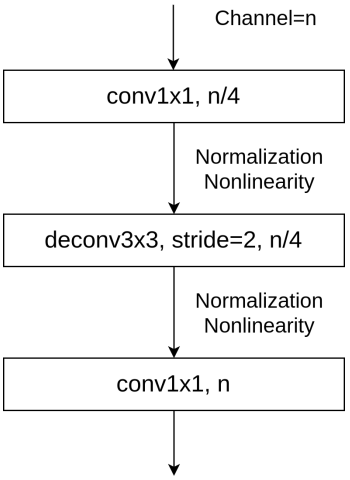
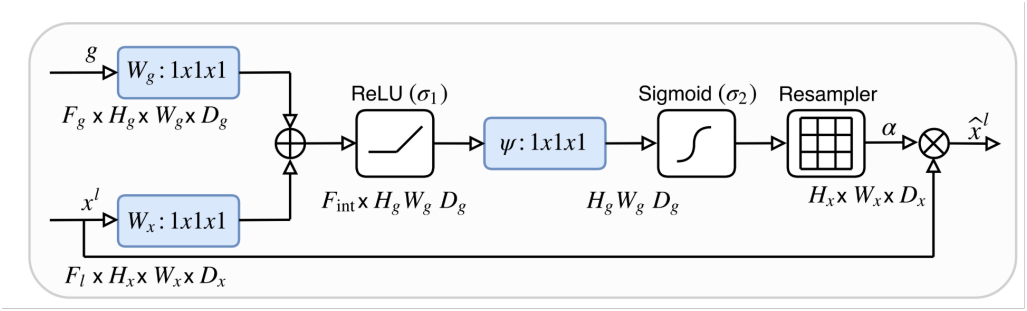


Figure 5.5: Illustration of the deconvolution block at the upsampling side. Each deconvolution block consists of a series of 1×1 convolution, 3×3 deconvolution and 1×1 convolution.

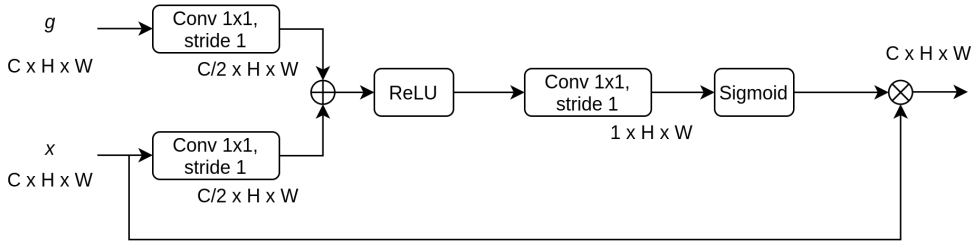
Table 5.1: Parameters of 2D baseline network. The deconv block consists in 1x1 convolution, transposed convolution and 1x1 convolution. The resconv denotes the residual block of ResNet. Skip connection, nonlinearity and in-layer normalization are not shown in the table.

Layer	Output size	Layers
Encoder 1	(16x)64x64	conv2d 7x7, 64 pool2d 3x3
Encoder 2	64x64	{resconv2d 3x3, 64)x3
Encoder 3	32x32	{resconv2d 3x3, 128)x4
Encoder 4	16x16	{resconv2d 3x3, 256)x6
Encoder 5	8x8	{resconv2d 3x3, 512)x3
Decoder 5	16x16	deconv2d 3x3, 512
Decoder 4	32x32	deconv2d 3x3, 256
Decoder 3	64x64	deconv2d 3x3, 128
Decoder 2	64x64	deconv2d 3x3, 64
Decoder 1	(16x)128x128	deconv2d 3x3, 32 conv2d 3x3, 32 conv2d 3x3, 5

Experimental networks Three experimental networks will be evaluated, consisting of the baseline residual U-Net, the baseline network with DAC and RMP blocks, and the baseline network with Attention Gate. The DAC and RMP blocks refer to CE-Net (Section 4.3.3). The Attention Gate refers to Attention U-Net (Section 4.3.3). Figure 5.6 shows the employed attention gate in the experimental network, which is slightly different to the original implementation in Figure 4.17. All the three networks share the same backbone while the specific blocks are different (Figure 5.7).



a. Original diagram of Attention Gate



b. Modified implementation for experiments

Figure 5.6: Diagrams of the original and the modified Attention Gate. The modified implementation removes the upsampling from Attention Gate as the gating signal and input features should have the dimension. Credit of diagram **a**: [131].

5.3.2/ IN-LAYER NORMALIZATION

As introduced in Section 4.3.4, batch normalization has become the routine in-layer normalization method. The uniformity among mini-batch in every layer intends for decreasing the distribution shift between feature maps. The such shift may be produced by the initialization of parameters, or the interval variance of dataset. Instance normalization has a different scope of the effect and the impact of the application scope may affect the distribution. The independence of a single instance may be diluted by batch normalization as its distribution will be shifted to the one of the mini-batch. To justify the impact of in-layer normalization scope, batch normalization and instance normalization will be trained on the proposed experimental models.

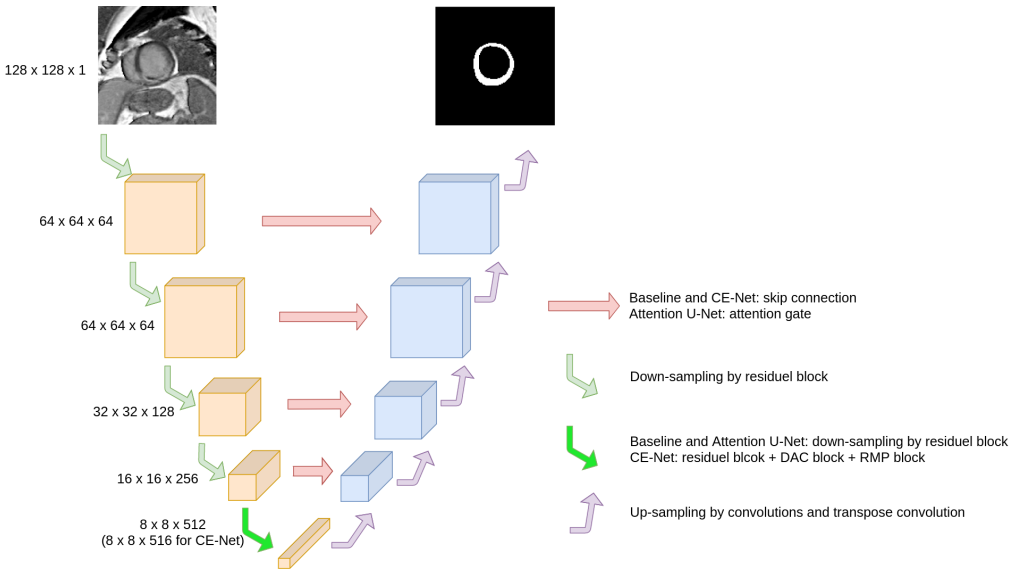


Figure 5.7: Architecture of the experimental networks. The three networks share a similar architecture except for the skip connection and the last down-sampling convolutions.

5.3.3/ STOCHASTIC OPTIMIZATION

Stochastic Gradient Descent (SGD) and mini-batch gradient descent optimization are introduced in Section 4.3.5. The major difference between the two stochastic optimizations is the amount of samples that participate in one calculation of gradient and update of parameters. Mini-batch gradient descent optimization takes the average gradient of a set of samples to update a model's parameters. Considering the important internal variability among the experimental images, mini-batch gradient descent optimization may omit features if few samples are characterized by these features. Therefore, both the SGD and the mini-batch will be tested together with the in-layer normalization methods.

5.3.4/ LOSS FUNCTION

The common loss functions for medical image segmentation such as the Dice loss, CE loss, weighted CE loss and their compounds are adopted as the baseline methods. The formulas of these loss functions have been presented in Section 4.3.7. In addition to the standard implementations, a variant based on Dice loss is proposed and then experimented with in response to the issue of the class imbalance.

Image-wise Dice loss For recall, the common Dice loss function is:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{n=1}^{N} p_n g_n + \epsilon}{\sum_{n=1}^{N} p_n^2 + \sum_{n=1}^{N} g_n^2 + \epsilon} \tag{5.5}$$

where N is the batch size, p_n and g_n stand for one prediction and its ground truth mask. $\sum$ sums the 3-D array of dimension [batch, height, width] and outputs a 1-D integer. ϵ is added on the numerator and denominator so that in case of true-negative, the loss will be correctly calculated as 0.

The vanilla Dice loss processes all images in a batch together as a single image. In contrast, an alternative Dice loss is proposed:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{n=1}^{N} \frac{p_n \hat{p_n} + \epsilon}{p_n^2 + \hat{p_n^2} + \epsilon}}{N} \tag{5.6}$$

This variant first calculates individually the loss of each sample $\frac{p_n \hat{p_n} + \epsilon}{p_n^2 + \hat{p_n^2} + \epsilon}$ in the batch, then the batch loss is counted as the average of all samples' losses in the batch of size N.

The difference between the common implementation and the proposed variant is the unit of the calculation. Both implementations will produce a similar loss value in most of cases. However, assuming that a batch consists of some negative images where no target class presents, while in the prediction a few false-positive pixels are predicted on these images, in such case the loss calculated according to Eq. 5.5 should be much lower than the one of Eq. 5.6 (inversely the Dice value).

5.4/ FRAMEWORK

To segment the myocardial infarction from the DE-MRI without the given myocardium contour, two frameworks are proposed employing 2D models. In Framework A, a CNN first segments the myocardium then another CNN finds the infarction area on the aforementioned segmented myocardium. Framework B independently segments the myocardium and the infarction areas from the whole image by two CNNs, then the in-common segmentation between the predicted myocardium and infarction is kept to remove the false-positive infarction area out of the myocardium. Figure 5.8 illustrates both frameworks.

5.5/ POST-PROCESSING

The proposed prior information-based filters [138] and the ensemble learning [97] are employed as post-processing methods. Other post-processing methods such as morphological transformations including erosion, dilation, opening and closing [162] are tested in preliminary experiments while the result is not satisfying, and therefore such methods will not be detailed.

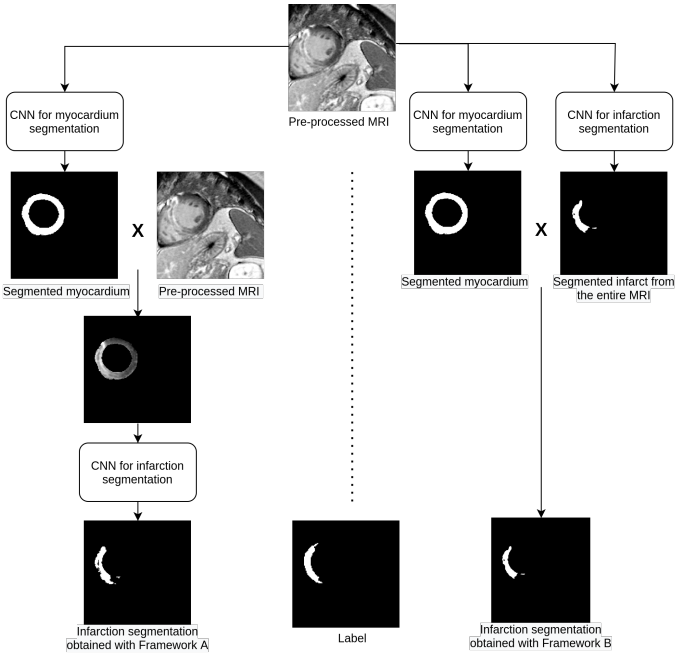


Figure 5.8: Illustration of both proposed frameworks. At the left, Framework A segments the myocardium with a first CNN, then another CNN segments the infarct on the obtained delimited myocardium. At the right, Framework B segments the myocardium and the infarct at the same step with two CNNs, then its final infarction segmentation is the intersection of the segmented myocardium and infarct.

Prior information-based filters The prior information-based post-processing refers to the rules that clinical experts employ for manual annotation drawing and inspection. Based on these rules and relative prior information in the data, criteria are constructed to eliminate false-positive contours in the CNN's segmentations. To design such criteria, we first investigated the image characteristics which can be considered as criteria, then we inspected the statistics of prior information in the training set to study the relevance of the proposed criteria. Since our data have significant variations, the criteria should be universal enough for the whole dataset. On the strength of this inspiration, three criteria were designed to improve the accuracy of the segmentation.

The first criterion focuses on the infarction size. In clinical practice, a validated myocardial infarction should have a minimum area. Otherwise, it will be considered a noisy region, e.g. an artifact. Therefore, this criterion compares the area of every single segmentation in CNN's outputs to a threshold value of the area. All segmentation contours which do not meet the condition:

$$Area(C) > \theta_{area} \tag{5.7}$$

are removed, where C is one contour belonging to a coarse segmentation and θ_{area} is the threshold area. Impacts of noises on the CNN's segmentations such as the partial volume of fat or cavity, artifacts of the high signal should be eliminated by this criterion.

The second and third criteria take advantage of the contrast agent's nature. The gadolinium contrast agent can enhance the signal intensity on our T1-weighted MR images and agent-enriched tissues show high signals. Blood transports the contrast agent in cardiac tissues. The absorption and the release rate of the agent depend on the type of tissue, which allows us to distinguish each tissue in DE-MRI. Therefore, around 10 minutes after the injection of the contrast agent, infarction tissues and cavity show high signals, while normal myocardium tissues appear dark. According to the above rules, the second criterion compares the median signal value of the cavity (signal of the blood) and the candidate area:

$$M(C) - M(cavity) > \theta_{signal} \tag{5.8}$$

where $M()$ means the median signal value in the region, C stands for the region of the candidate contour, and $cavity$ stands for the region of the cavity. If the signal in the region of the candidate contour is high enough compared to the cavity, this area will be kept. In this criterion, the intensity signal of the blood inside the cavity indicates the reference of low intensity. Indeed, the cavity is a reliable reference tissue because its contour is given in the dataset and the signal in the cavity is relatively homogeneous.

However, the absolute difference between the infarction tissue and cavity is not always consistent between examination cases as shown in the left part of Fig. 5.9. The delayed acquisition time (the images are not acquired exactly at the same delay time between examinations), the patient's physical state, characteristics of MRI equipment, etc. can affect the optimal threshold value of our criterion. To reduce the impact of such variations, the third criterion refers to both intensities of the cavities and non-pathological myocardium tissues, which constitute a scope of the signal. The candidate infarction area is then compared to this scope:

$$(M(C) - M(cav))/(M(cav) - M(myo)) > \theta_{\%} \tag{5.9}$$

where $M(myo)$ refers to the median signal of non-infarction myocardium. The threshold value $\theta_{\%}$ becomes a percentage of the scope rather than an absolute difference (θ_{signal}) as in 5.8. Fig. 5.9 calculated on the training set prior information proves that the index of scope on the right is more consistent that the absolute difference on the left. Nevertheless, the contours of healthy myocardium tissues are not given for the infarction segmentation task (only the whole myocardium contour annotation is provided). Reminding that our neural network ensures high recall, we assume that all the myocardium except the regions segmented by the neural network is healthy, which means that the non-selected myocardium is certainly non-pathological. The final segmentation results testify the relevance of this hypothesis.

According to the statistics of the prior information and experiments, our framework employs the first and third criteria to select true-positive segmentations. The final post-

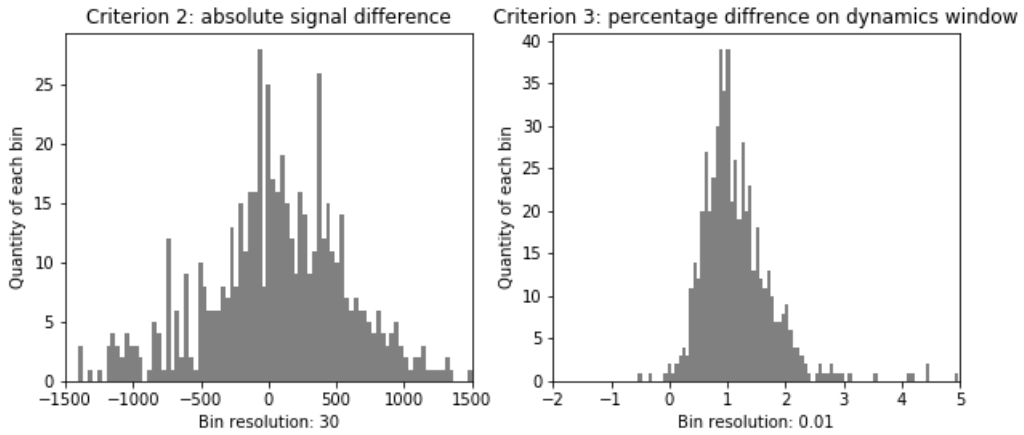


Figure 5.9: Prior statistics supporting Criterion (5.8) and Criterion (5.9). The absolute difference of signal (left) is less consistent than the proportional difference on scope (right).

processing is the summed-up in Algorithm 1. All threshold values are roughly determined on the training set's prior information 5.9, and then fine-tuned on the validation set to obtain the best performance on the test set and avoid data leakage.

Ensemble learning The best training epoch is normally determined by monitoring training and validation accuracy graphs (Dice metric for our case). Training should be stopped when validation accuracy does not increase anymore to avoid the overfitting effect. However, noise could temporally mislead the training path and a single model could not be the best for the segmentation of all pixels in a particular input image. In such a case the ensemble method is employed to strengthen the predictions' robustness.

The ensemble learning can be classified into homogeneous and heterogeneous ensembles. The heterogeneous ensemble learning, *i.e.* using different feature selection methods with the same training data, is adopted for further test. In our practice, during the training stage, several models are saved at different epochs prior to the early stopping point. Among the results, each model has the same network conception but different training epochs. During the test stage, all the trained models were used to predict a set of segmentation. The set of all predictions is then merged as a single segmentation for each input image. The merged segmentation looks like a heat map, therefore a threshold value that is determined on the validation set is used to transform the merged one into a binary segmentation mask.

Algorithm 1 Prior information-based post-processing

Result: removal of false-positive segmentation contour(s) of one image
Input : Set of infarction contours $\{S_i\}$
Input : Criterion (5.7) C_1; Criterion (5.9) C_3
Output: Final segmentation contour(s) S_{final}
$S_{ensemble} \leftarrow Avg(\{S_i\})$ $S_{ensemble} \leftarrow Binary(S_{ensemble})$
$S_{final} \leftarrow \{\}$
for $aContour \in S_{ensemble}$ **do**
  if $aContour$ fulfills conditions $\{C_1 \cup C_3\}$ **then**
    $S_{final} \leftarrow S_{final} + aContour$
  end
end
return S_{final}

5.6/ EXPERIMENTAL RESULTS

To justify the performance of the proposed myocardial segmentation approach on DE-MRI, experiments are conducted on the private dataset (Section 3.4). Data consisting of all-infarcted slices and consisting of mixed slices are separately trialled with different strategies, hence subsets are extracted from the private dataset to meet the conditions of each experiment, and therefore the value is not comparable between experiments.

For all the experiments, the neural networks are implemented with Python 3.6.9 and Pytorch (the version depends on the experiments). The training and test are executed on Tesla V100.

5.6.1/ EVALUATION METHODS AND METRICS

To evaluate the segmentation results of different tissues, clinical and geometrical metrics are employed. The clinical metrics are the most widely used in cardiac clinical practice, *i.e.* the average errors for the volume of the myocardium of the left ventricle, the MI (in cm^3), and the average errors of the percentages of MI (according to the myocardium). The average errors for the volume of a tissue is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|v_i - \hat{v}_i| \tag{5.10}$$

where MAE denotes Mean Absolute Error, v_i denotes the prediction volume of a target tissue on slice i, and $\hat{v}_i$ denotes the ground truth volume for the same tissue on the same slice. The average errors of the percentages of MI (or PMO) are then defined as:

$$Error_{PIM} = \frac{1}{n}\sum_{i=1}^{n}\frac{|v_i - \hat{v}_i|}{v\hat{m}_i} \tag{5.11}$$

where PIM denotes Percentage of Infarcted Myocardium (the infarcted myocardium can refer to the PMO if the PMO should be independently assessed), $\hat{vm}_i$ denotes the ground truth volume of the myocardium.

The geometrical metrics are the average Dice index for the different areas and the Hausdorff distance (in 3D in case of the segmentation with 3D model) for the myocardium. For MI, only the Dice is calculated due to the possible presence of multiple instances in an image. The formulas of Dice index and Hausdorff distance can refer to the corresponding loss functions.

5.6.2/ SEGMENTATION ON ALL PATHOLOGICAL SLICES

Experiments performed on all-pathological slices majorly examine all the conceptions of the segmentation models except the prior information-based post-processing. The impact of the image normalization and all the proposals concerning the CNN model are evaluated and compared. The post-processing referring to the pathological prior is not trialled on all-pathological data since the mechanism of this post-processing is the elimination of false-positive segmentation. The experiments do not distinguish between the MI and the PMO, therefore PMO is considered as a part of MI for both the segmentation and evaluation.

Since multiple methods are proposed, the choice of the image normalization method (Section 5.2.3) must be justified first as well as the gradient descent batch size (Section 4.3.5) and the loss function (Section 5.3.4). Secondly, the chosen methods are adopted on different CNNs (Section 5.3.1) considering both two frameworks A and B (Section 5.4). The ensemble method (Section 5.5) is then applied to all the trials to get a robust performance. To show the advantage of the learning-based models, the segmentation by GMM (Section 4.2) is also evaluated. Finally, the best practices proven by the above trials are integrated into our proposal and then compared with the inter- and intra-observer variation studies on different application conditions.

This trail focuses on images that include at least one infarction region. Hence, all the 904 slices with infarcted areas in the entire private dataset are selected for this experiment. These slices are then randomly assorted into 581, 168, and 155 images which, respectively, belong to the training, validation, and test sets. This splitting is based on the 60-20-20 rule, but as images coming from the same exam are assigned to the same set, the percentages are not strictly respected. The images of the test set are the same as those in the inter- and intra-observer study sets.

Image normalizations Table 5.2 reveals the effect of the image normalization method for the automatic infarction segmentation. The experiments are done on the same MRI

slices whose myocardium mask was given by an experienced physician. The experimental model takes the baseline CNN with batch size 1 and Dice loss and the only variable among the experiments is the image normalization method. The normalization according to Eq. 5.1 slightly outperforms the equalization for the segmentation task. The performance obtained using both above-mentioned normalizations largely exceeds the one with non-normalized images. Figure 5.10 provides examples of the infarction segmentation results with the three trials. Visually, Eq. 5.1 and the equalization enhance the contrast between the normal and the pathological tissues but the equalization markedly intensifies the noise.

Table 5.2: Impact of image normalizations on infarction segmentation

Model	Dice	Jaccard	PIM Difference(%)	Volume Difference(mL)
Normalization using Eq. 5.1	0.805	0.694	7.18	1.196
Equalization	0.804	0.693	7.58	1.310
No normalization	0.775	0.657	8.34	1.392

Batch size and loss functions The comparison of different batch sizes, layer normalization methods and loss function is completed on the normalized images whose myocardium mask is given by an experienced physician. The models' network is the baseline CNN. The results are shown in Table 5.3, the best Dice index obtained being 0.805 with SGD and Dice loss. The SGD converges more rapidly in terms of the training epochs, however, its total training time is much longer since the model updates more frequently its weights and the GPU computing resources can not be fully used.

Table 5.3: Impact of batch size, layer normalization methods and loss functions.

Batch size	Loss function	Layer normalization*	Dice	Jaccard	PIM Difference(%)	Volume Difference(mL)
1	Binary CE	Same effect	0.796	0.683	7.63	1.315
1	Weighted binary CE	Same effect	0.794	0.680	7.84	1.348
1	Dice	Same effect	**0.805**	**0.694**	**7.18**	**1.196**
64	Dice (per image)	Batch	0.757	0.627	8.73	1.485
64	Dice (per image)	Instance	0.748	0.617	8.55	1.473
64	Dice (per batch)	Batch	0.750	0.619	8.86	1.508
64	Dice (per batch)	Instance	0.755	0.628	8.30	1.452

*: When the batch size is one, the batch normalization and the instance normalization will have the same effect
In bold the best results.

Frameworks and neural networks In Table 5.4, the infarction segmentation performance is provided according to each neural network and framework. Contrary to the above trials, the myocardium mask is not provided for these evaluations. The image preprocessing and network configurations are adopted from the best practices in the above subsections, *i.e.* image normalization with Eq. 5.1, a small batch size of 1 and Dice

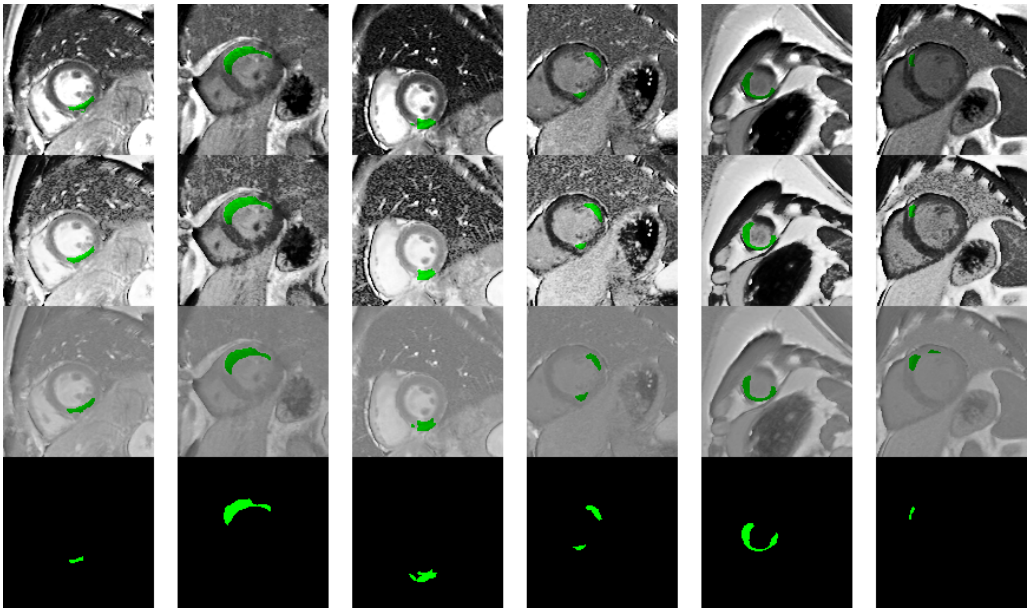


Figure 5.10: Visualization of image normalizations and their segmentation results. For each MRI slice (one per column), from top to bottom: automatic segmentation of infarction (in green) on images of normalization according to Eq. 5.1, equalization, and no normalization. The last line displays the ground truths of myocardial infarction.

index loss. To examine the impact of the receptive field size of convolution, the nearest neighbor interpolation is applied for the CNNs' input images. Results reveal that, globally, framework A gets better precision on Dice and Jaccard but framework B achieves better prediction of the PIM and volume, interpolated input significantly optimizes the segmentation and advanced convolution blocks have little impact on our data for the segmentation task.

Table 5.4: Performance of frameworks and neural networks for infarction segmentation.

Framework	Network	interpolation	Dice	Jaccard	PIM Difference(%)	Volume Difference(mL)
A	CE-Net	1x	0.595	0.450	11.06	1.927
A	CE-Net	5x	0.662	0.520	9.34	1.600
A	Attention	1x	0.551	0.410	12.55	2.175
A	Attention	5x	0.648	0.502	9.42	1.616
A	Baseline	1x	0.591	0.446	11.29	1.923
A	Baseline	5x	**0.668**	**0.524**	9.64	1.600
B	CE-Net	1x	0.546	0.403	9.93	1.728
B	CE-Net	5x	0.614	0.469	**8.95**	**1.487**
B	Attention	1x	0.521	0.381	10.59	1.899
B	Attention	5x	0.587	0.442	9.38	1.641
B	Baseline	1x	0.558	0.411	10.09	1.740
B	Baseline	5x	0.592	0.450	9.48	1.636

Evaluation results Table 5.5 lists the evaluation results from the proposal, the inter- and intra-observer variation studies as well as the results from GMM. The proposal is the model with the best result in the previous trials. Indeed, according to Table 5.4, both

the Framework A with baseline network and the Framework B with CE-Net achieve the best results based on different evaluation metrics. We choose the Dice as the principal analysis metric, therefore the proposal consists of the baseline network (residual U-Net) trained using SGD (batch size 1) with Dice loss and interpolated input under framework A. Both the conditions of the given and not given myocardium mask are evaluated. Since the base and the apex images are more difficult to segment, the results are calculated on the whole test set and only on the middle images without considering the most basal and apical slices of the test set.

The segmentation results of middle slices are more precise since basal and apical slices have more uncertainties. Without providing the myocardium mask, the infarction segmentation is much more challenging since the error of the myocardium segmentation is added to the infarction segmentation.

The dependent t-test for paired samples [50] is used to check if the difference between our proposal's results and other approaches is significant. A p-value inferior to 0.05 can be considered a significant difference between the results from both approaches. Results reveal that our proposal has a very similar performance to the inter-observer variation study. Compared to the intra-observer variation study, the gap is not always significant. The results from GMM are significantly inferior to those of our proposal.

Table 5.5: Proposal's results, inter-observer study and intra-observer study variations, and GMM's results.

Method	Myocardium Mask*	Image position	Dice	Jaccard	PIM Difference(%)	Volume Difference(mL)
Proposal	Unknown	All	0.668	0.524	9.64	1.600
Intra-observer study	Unknown	All	0.711**	0.572**	8.82	1.361**
Inter-observer study	Unknown	All	0.681	0.538	10.98	1.694
Proposal	Unknown	Middle	0.694	0.549	8.61	1.604
Intra-observer study	Unknown	Middle	0.734**	0.595**	6.81**	1.257**
Inter-observer study	Unknown	Middle	0.707	0.562	7.95	1.500
Proposal	Known	All	0.843	0.746	5.89	1.014
Intra-observer study	Known	All	0.845	0.751	4.63**	0.763**
GMM	Known	All	0.699**	0.566**	13.14**	2.118
Proposal	Known	Middle	0.845	0.748	5.38	1.000
Intra-observer study	Known	Middle	0.855	0.763	3.47**	0.636**
GMM	Known	Middle	0.693**	0.559**	12.90**	2.244**

*: Whether the manual delimitation of the myocardium is provided
**: p-value < 0.05 between our proposal and the compared approach under the same application conditions

Figure 5.11 shows the methods' segmentation results on typical images including apical images, images with important no-reflow tissue, images with artifacts, high contrast images and low contrast images.

5.6.3/ SEGMENTATION ON ALL PATIENTS

The experiment on all patients with 2D models majorly examines the impact of gradient optimisation methods, the 2.5D input (detailed in Section 6.2.1), CE-Net and the prior

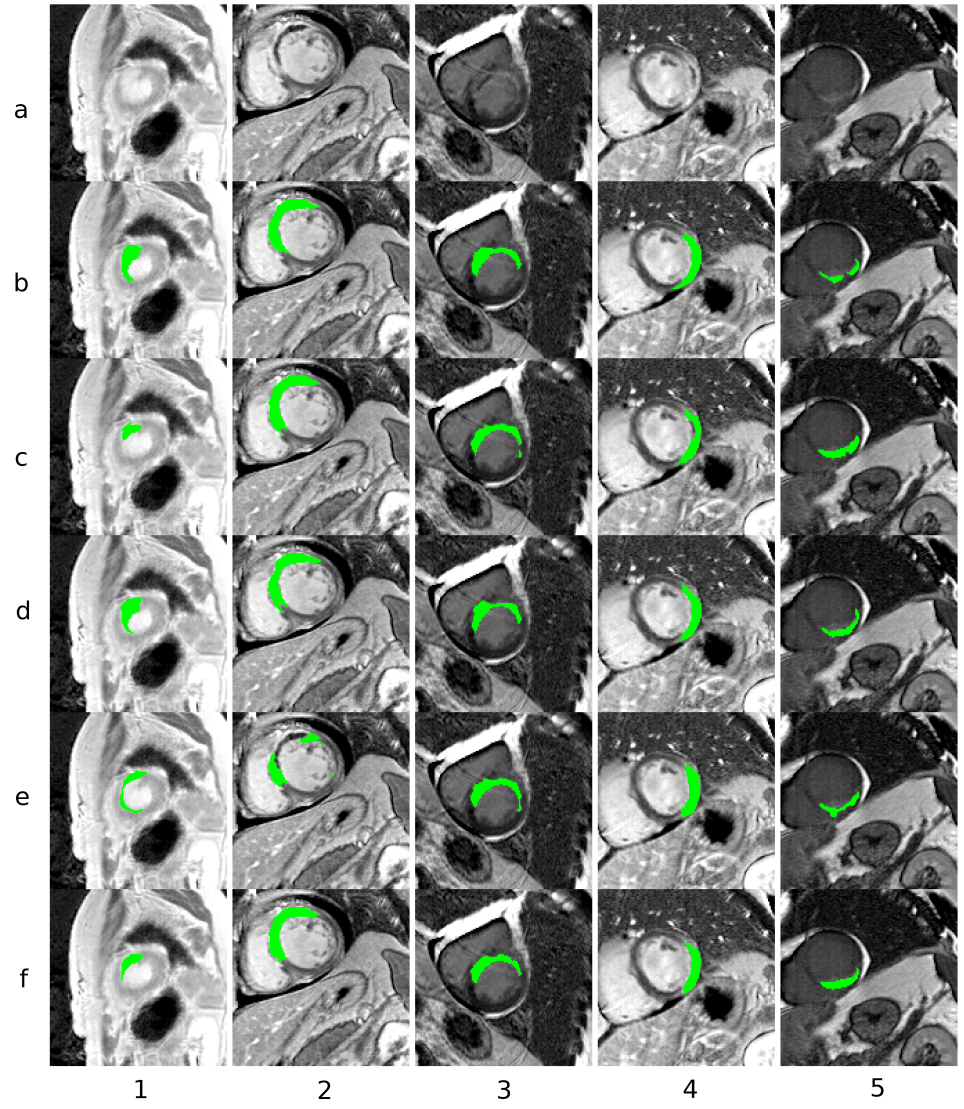


Figure 5.11: Myocardial infarction segmentation results on typical DE-MRI. For each row: **a** image to segment, **b** ground truth, **c** proposal with given myocardium mask, **d** intra-observer with given myocardium mask, **e** proposal without providing myocardium mask, **f** intra-observer without providing myocardium mask. For each column: **1** apical slices, **2** presence of important no-reflow tissue, **3** presence of artefact, **4** high contrast images, **5** low contrast images.

information-based post-processing. 888, 241, and 194 pathological images that were randomly selected from the private dataset are respectively used as the training, validation and test data. Intra- and inter-observer variation studies are performed by experienced cardiologists on the test set.

The mask of the myocardium is given in these experiments, therefore a simple one-stage framework is employed with different configurations. The vanilla Dice loss is the only loss function due to the single target class (PMO in considered as a part of MI and not be specifically segmented). The input is center-aligned according to the myocardium

and only the ROI (area of the myocardium) is kept in the input. Figure 5.12 shows the experimental framework of the tests on slices of all patients.

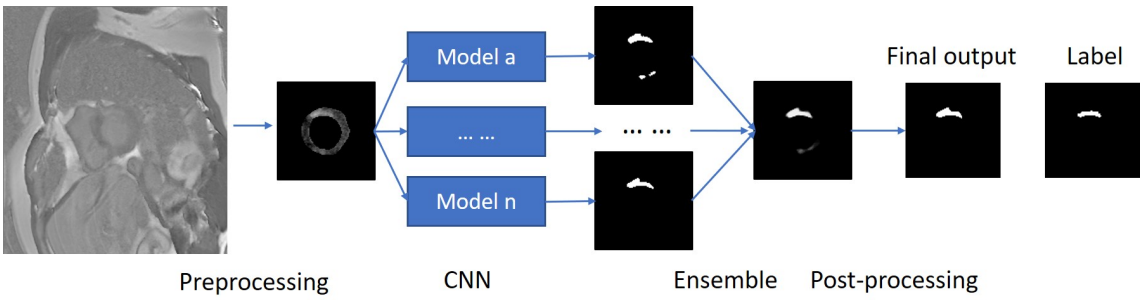


Figure 5.12: Illustration of the experimental framework with given mask of the myocardium on all patients. For a 2.5D input, three neighbouring images replace the single input image.

Table 5.6 shows the results between the baseline network or CE-Net, SGD or mini-batch SGD, 2D or 2.5D inputs, and with or without the prior information-based post-processing. In addition to the common metrics, recall (sensitivity) and accuracy are also calculated to emphasize the impact of the prior information-based post-processing. Values in brackets stand for the results before post-processing.

Table 5.6: Experiments results on all patients with given myocardium mask. 2D and 2.5D stand for input images format; BL and CE stand for BaseLine and CE-Net; B1 and B32 stand for the batch size of the training and inference.

Model[1]	Dice(%)	Dif. PIM(%)	Recall(%)	Accuracy(%)
Intra-observer	0.810	4.14	85.45	71.20
Inter-observer	0.774	8.66	90.26	67.50
2D BB B1	0.742 (0.663)	7.72 (8.29)	83.20 (85.62)	65.00 (56.28)
2D BB B32	0.718 (0.579)	8.87 (11.65)	77.41 (81.69)	61.32 (46.61)
2.5D BB B1	0.757 (0.690)	7.70 (7.83)	84.72 (86.80)	66.41 (59.15)
2.5D BB B32	0.711 (0.591)	8.09 (11.22)	75.94 (81.20)	60.29 (42.62)
2D BB 5x B1	0.780 (0.652)	5.33 (7.93)	88.93 (90.34)	68.39 (55.97)
2D BB 5x B32	0.731 (0.595)	7.25 (11.98)	79.40 (85.97)	63.34 (48.73)
2D CE 5x B1	0.779 (0.657)	5.47 (7.64)	89.13 (90.56)	68.07 (56.38)
2D CE 5x B32	0.732 (0.599)	7.39 (11.19)	78.86 (84.96)	63.57 (49.17)

The following conclusions can be obtained from the data in Table 5.6. First, no evident improvement is achieved from the 2.5D input. Then, CE-Net does not bring us a better result. The additional dense convolution and multi-kernel pooling on high-level semantic information do not provide useful information to the final segmentation. Third, interpolated images slightly improve the performance after post-processing. Indeed, the interpolation equivalently changes the size of the receptive field of convolution kernels.

Moreover, by observing between the high recall and relatively low accuracy, it can be inferred that most of the healthy images are wrongly segmented as false-positive by the CNN. The best CNN (according to Dice metric, before the post-processing) obtains

55.97% on accuracy and 90.34% on recall (2D interpolated inputs, baseline network, SGD), which means that most pathological images are segmented, whereas without the post-treatment many of the healthy test images are segmented as infarcts. The prior information-based post-processing obtains significant improvement in the segmentation results, especially on healthy images. This improvement majorly comes from the removal of false-positive contours from noisy areas as shown in the base image in Fig 5.13. Notice that the threshold values of the criteria can affect the post-treatment results: more strict thresholds decrease the rate of false-positive segmentation, but increase the false-negative and vice versa. The threshold values are fine-tuned on the validation set by inspecting the Dice values and the difference of proportion. Two metrics that can increase or decrease differently depending on the threshold values.

The results according to the categories of the slice position are also evaluated as shown in Tab. 5.7. Apex and base images result in more failed segmentation because of their higher uncertainty and less agreement between clinical experts. Focusing on middle images, the results are quite satisfying compared to the annotation's internal variations.

Table 5.7: Results On Different Image Positions

	Middle		**Apex + Base**	
	Dice(%)	**Dif. PIM(%)**	**Dice(%)**	**Difference(%)**
Proposal	81.37 (69.70)	4.29 (6.03)	65.63 (48.86)	9.21 (14.36)
Intra-observation	81.96	3.89	76.72	6.19

The qualitative results are provided as well. Fig.5.13 shows a segmentation result of an MRI exam. The result is globally satisfying in the middle images.

5.7/ DISCUSSION

Gradient optimisation method The results on both pathological-only and all-patients datasets with 2D models demonstrate that the SGD gets dramatically better results on our data. Indeed, since our dataset has considerable internal variability among images, the images' characteristics could not be so uniform. Indeed, the characteristics of basal and apical images could be very different from those of the images in the middle of the left ventricle. Moreover, the signal can be different among exams due to many factors such as renal clearance and contract media dose. As mini-batch gradient descent tends to ignore noisy features during the backpropagation, the internal variation of our data could be identified as noises. Some useful information is dropped out during training because of the mini-batch, which could explain why the recall of mini-batch gradient descent is lower. However, one drawback of SGD is its computation efficiency. A batch size of value 1 updates the model's weights more frequently than a big batch. Each update

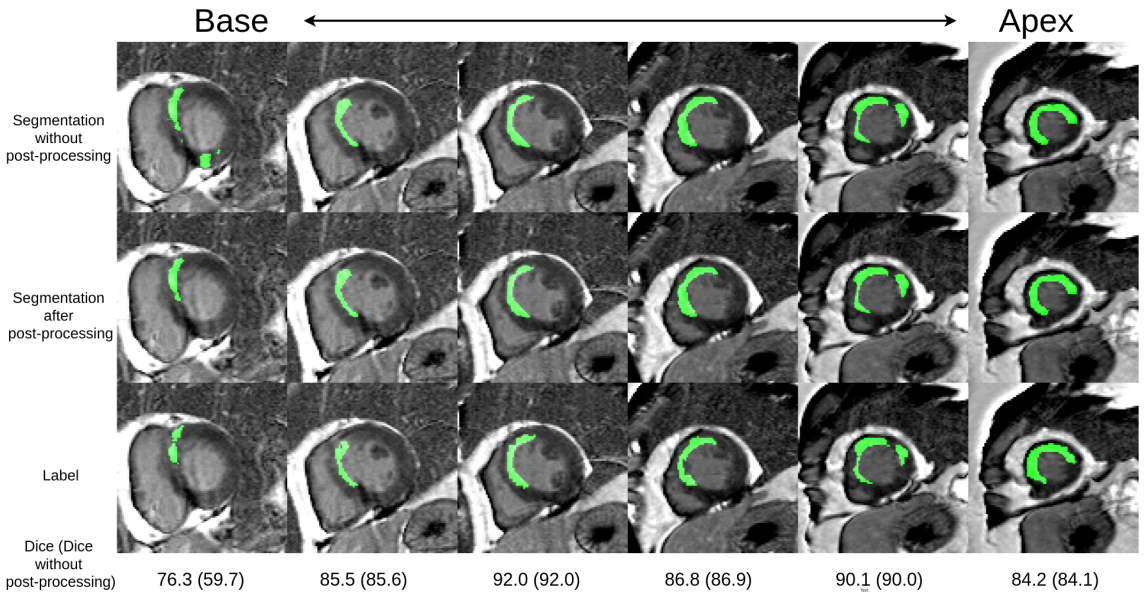


Figure 5.13: Segmentation result from the 2D baseline network. From left to right, a whole examination is segmented (base to apex). From top to bottom, the contours are respectively the merged segmentation (the set of predictions, without the post-processing), post-treated segmentation, and their labels. Dice values are given at the bottom of each slice (Dice value without post-processing in bracket). The artifact due to the MRI acquisition on the first image is wrongly segmented as an infarction by CNN, then the post-processing successfully removes this false-positive segmentation so that its Dice improves from 59.7% to 76.3%.

includes more noise so that the path of the weight updating is more rugged during the back-propagation. Although SGD learns both features of internal variation and noise and decelerates the training, SGD outperforms mini-batch gradient descent on our dataset.

Deep learning and statistical models for MI segmentation All the CNN-based models' results are significantly superior to the commonly used GMM. Compared to the state-of-the-art CNN models, the baseline network slightly outperforms them. This could be explained by the fact that the level of semantic information in our data is relatively low. The additional blocks that aim at improving the information interpretation could not dig further useful semantic information. As mentioned by Isensee *et al.* [153], a better adaptive model could be more efficient than a heavy and complex neural network for automatic medical image analysis.

Prior information-based post-processing The prior information-based post-processing provides a remarkable improvement, particularly on mixed healthy-infarcted data. A related segmentation strategy can be proposed where the model focuses on recall as the first metric. Then the post-processing eliminates the false-positive areas

from the high recall segmentation.

To generalize the prior information-based post-processing on other kinds of medical imaging, first, two (or more) labeled or region-confirmed tissues would be chosen to create a dynamics window. Second, the region signal in the candidate contour would be compared to a proportional threshold on this window. Only the conformed contours would be finally kept as the result of the segmentation. The representative value of the region could be the median of its histogram. The difficulty of this approach is the selection of characteristic tissues.

5.8/ CONCLUSIONS

The 2D segmentation approaches obtain very satisfying results on the automatic MI assessment. With the provided myocardium mask, the highest Dice metric of the segmented infarct is 0.843 on all-pathological slices, and 0.780 on slices of all patients. Without the provided myocardium mask, the highest Dice metric is 0.668 on all-pathological slices. Comparing the proposals and the variance of the manual annotations, the proposed 2D segmentation approach incorporating the proposed image normalization method, baseline network, SGD optimization, interpolated input images and proposed prior-based post-processing outperforms the inter-observer study on all patients (Table 5.6), and the same model conception without the post-processing achieves comparable result as the inter-observer study on all-pathological slices (Table 5.5).

The proposed image normalization and the prior-based post-processing methods are proved efficient (Table 5.6 and Table 5.10) without the requirement of additional heavy computation. The SGD is proved more convenient than the more popular mini-batch, which is counter-intuitive. Feeding interpolated images as networks' input increases the segmentation performance while the training and inference are slower due to the excessive operation on larger feature maps. However, the employment of more complex convolutional blocks has no convincing improvement on the DE-MRI segmentation.

6

CONTRIBUTION 2: AUTOMATIC MI ASSESSMENT FROM MRI WITH 3D DL MODELS

6.1/ INTRODUCTION

Since DE-MRI can be viewed as a tomographic imaging approach, the spatial information, *i.e.* the inter-slice relation can be potentially referred to enhance the image interpretation. To this end, 3D deep learning-based segmentation models are proposed to assess the MI in the left ventricle. The results are also compared to the 2D baseline method of the previous chapter to evaluate the effect of the introduced spatial information on the MI segmentation.

To take spatial information into account for the MI segmentation, three strategies are proposed. The first one feeds 3 adjacent MRI slices as the CNN's input where all the convolutions are 2D, termed 2.5D network. The second strategy takes the baseline 2D network (Section 5.4) as the backbone while all the convolutions at the encoding side are replaced by 3D convolutions, termed a fully 3D network. The last strategy is our original proposal, termed 3D network with Constrained Receptive Field (CRF). The CRF network integrates 3D and 2D convolutions at the encoding side to better control the impact of the inter-slice information.

To evaluate the proposed CRF network and other strategies, the tissues of interest including the LV myocardium, the LV infarct and LV PMO are segmented. The PMO will be independently segmented in this chapter, instead of being considered as part of the infarct as in the previous chapter. Therefore, loss functions for multi-class (more than two) should be considered. The segmentation results of the three tissues will be evaluated with the same employed metrics in the previous chapter.

This chapter is organized as follows:

- Segmentation with 3D models. Three variants of 3D models are proposed and a two-stage cascaded framework is introduced. Then a variant of generalized Dice loss is proposed.

- Experimental results. The multi-class segmentation loss functions are evaluated with the 2D baseline model. Then the best loss function is applied to the experiments comparing between proposed 3D models, 2D model and manual annotation variance.

- Discussion.

- Conclusions.

6.2/ SEGMENTATION WITH 3D MODELS

Although the DE-MRI is constituted of 2D slices, adjacent images are related to each other. Therefore, the inter-slice correlation can be potentially explored to optimize the segmentation result. With different input formats and network configurations, 2.5D and 3D models are proposed. Depending on the vertical alignment of the experimental dataset, a two-stage framework is also proposed where the first model segment the myocardium from 2D slices and aligns the 3D inputs according to the center of the obtained ROI.

6.2.1/ 2.5D MODEL

The 2.5D model inherits the same architecture of the 2D models (Section 5.3.1) except that the input layer receives 3 channels instead of 1. Three adjacent images are put into 3 channels as a single input and only 2D convolutions are performed in the CNN segmentation model. For the first and the last slice of a MRI case, the slice is copied to the former or later channel to form the 3-channel input. At the end of the model, only one segmentation mask is produced and compared to the image of the middle channel. Figure 6.1 illustrates the 2.5 input on an experimental network.

The model is called 2.5D because the inter-slice information can not be fully utilized. The 2D convolution interprets the correlation between channels without learning the characteristics of spatial information. In other words, exchanging the order of the three channels has no impact on the 2D convolution.

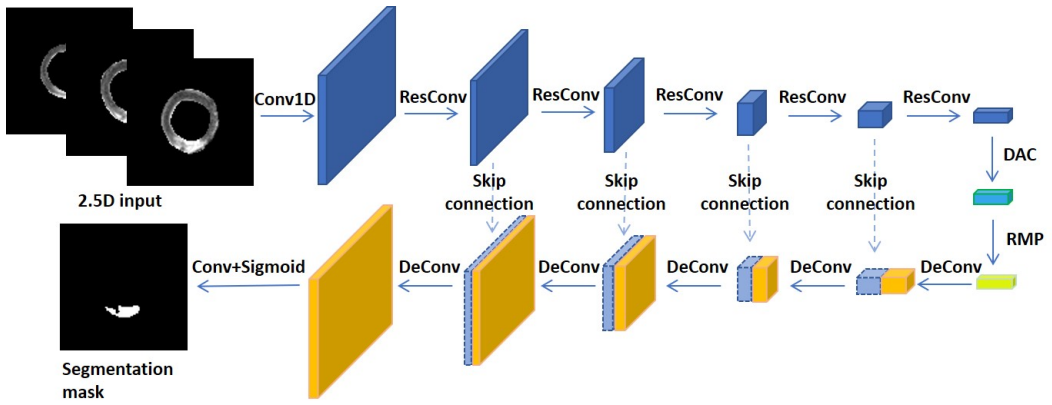


Figure 6.1: Illustration of experimented 2.5D CE-NET. ResConv doubles the number of feature maps and halves their resolution. Skip connection concatenates corresponding encoder feature maps and decoder feature maps. Relative to 2D input, the 1D convolution layer is additional to 2.5D input.

6.2.2/ 3D MODELS

The 3D models is fed by 3D inputs: each input of the size 16x128x128 consists of an entire MRI case. For cases having less than 16 slices, zero-padding is performed to ensure a uniform input size. Considering the impact of the size of the receptive field, two configurations are proposed with 3D convolutions.

The first variant is intuitively transformed from the baseline (Section 5.3.1) where all the 2D operations are replaced by 3D ones, including the kernel, stride and padding sizes, namely a fully 3D network. This network is very similar to V-Net [79] except that there is no down-sampling along the z-axis due to the inconsistency of the scale of the input data in the plane and z-axis. The second 3D variant is the proposed 3D Constrained Receptive Field network. As the name implies, the receptive field is limited to in-plane in the low semantic layers. In other words, in the first two down-sampling and the last two up-sampling layers, the convolution and pooling operations have the size of 1 at the z-axis for their kernel, stride, padding *etc.*. This conception aims at limiting the inter-slice interference on low semantic layers for a finer segmentation result and taking the advantage of 3D information on high semantic layers from the aligned 3D DE-MRI. Table 6.1 shows the parameters of the baseline, fully 3D and 3D CRF networks.

6.2.3/ LOSS FUNCTION FOR MULTI-CLASS SEGMENTATION

Since in this chapter, the target is to segment multi-class tissues (myocardium, infarct and PMO), loss functions for multi-class segmentation should be employed, such as the Cross-Entropy (CE) loss and the generalized Dice loss. Therefore, a variant of the generalized Dice loss is proposed to better address the class imbalance issue by optimizing

Table 6.1: Parameters of 2D and 3D networks. The deconv layer consists in 1x1 convolution, transposed convolution and 1x1 convolution. The resconv denotes the residual block of ResNet. Skip connection, LeakyReLU and batch normalization are not shown in the table. In bold the major conception difference between the Fully 3D and CRF networks.

Layer	Output size	baseline	Fully 3D	3D CRF
Encoder 1	(16x)64x64	conv2d 7x7, 64 pool2d 3x3	conv3d **3**x7x7, 64 pool3d **3**x3x3	conv3d **1**x7x7,64 pool3d **1**x3x3
Encoder 2	(16x)64x64	{resconv2d 3x3, 64}x3	{resconv3d **3**x3x3, 64}x3	{resconv3d **1**x3x3, 64}x3
Encoder 3	(16x)32x32	{resconv2d 3x3, 128}x4	{resconv3d 3x3x3, 128}x4	{resconv3d 3x3x3, 128}x4
Encoder 4	(16x)16x16	{resconv2d 3x3, 256}x6	{resconv3d 3x3x3, 256}x6	{resconv3d 3x3x3, 256}x6
Encoder 5	(16x)8x8	{resconv2d 3x3, 512}x3	{resconv3d 3x3x3, 512}x3	{resconv3d 3x3x3, 512}x3
Decoder 5	(16x)16x16	deconv2d 3x3, 512	deconv3d 3x3x3, 512	deconv3d 3x3x3, 512
Decoder 4	(16x)32x32	deconv2d 3x3, 256	deconv3d 3x3x3, 256	deconv3d 3x3x3, 256
Decoder 3	(16x)64x64	deconv2d 3x3, 128	deconv3d 3x3x3, 128	deconv3d 3x3x3, 128
Decoder 2	(16x)64x64	deconv2d 3x3, 64	deconv3d 3x3x3, 64	deconv3d 1x3x3, 64
Decoder 1	(16x)128x128	deconv2d 3x3, 32 conv2d 3x3, 32 conv2d 3x3, 5	deconv3d 3x3x3, 32 conv3d 3x3x3, 32 conv3d 3x3x3, 5	deconv3d 1x3x3, 32 conv3d 1x3x3, 32 conv3d 1x3x3, 5

the way of the class weighting.

Generalized Dice loss with different weighting The original generalized Dice loss [99] is formulated as:

$$1 - 2\frac{\sum_{l=1}^{L} w_l \sum_n g_{ln} p_{ln}}{\sum_{l=1}^{L} w_l \sum_n g_{ln} + p_{ln}} \tag{6.1}$$

where l denotes the class label, n denotes the pixel and w_l denotes the weight of the class label l that is calculated as $w_l = 1/\left(\sum_{n=1}^{N} g_{ln}\right)^2$, g and p are the ground truth and the prediction. With a different weighting method, we propose a variant of the generalized Dice loss:

$$1 - 2\sum_{l=1}^{L} w_l \frac{\sum_n r_{ln} p_{ln} + \epsilon}{\sum_n r_{ln} + p_{ln} + \epsilon} \tag{6.2}$$

In the proposed Eq. 6.2, the weight is $w_l = 1/\sum_{n=1}^{N} r_{ln}$, which is the same as in the cross-entropy loss. The proposal aims at reducing the difficulty of tuning the weighting.

6.2.4/ FRAMEWORK FOR NON-ALIGNED IMAGES

In a 3D volume, the connectivity between adjacent slices should be assured so that the inter-slice information can be correctly interpreted by the 3D network. For the experiments on non-aligned data, the 2D network is thus first employed to correct the potential misalignment of the heart between slices due to consecutive breath-hold at different positions.

In detail, in the two-stage framework, a 2D network segments the myocardium from the single MRI slices. The original MRI slices vary in resolution between exams, therefore all the slices were center-cropped at the size of 128x128 corresponding to the 2D network's

input. The gravity center of the segmented myocardium by the 2D network helps to relocate the slices of one MRI case so that the aligned MRI volume for each patient can be reconstructed. The cascaded 3D network then takes the volumes of the aligned MRI and the 2D segmentation mask as its 2-channel 3D input [159]. The volumes are made up of zero padding to have a fixed size of 16x128x128. The complementary input channel of the 2D segmentation mask provides pure in-plane information to the 3D network.

Figure 6.2 illustrates the proposed framework. The 3D network can either be the fully 3D network or the novel 3D CRF network, the two configurations of the proposed 3D network.

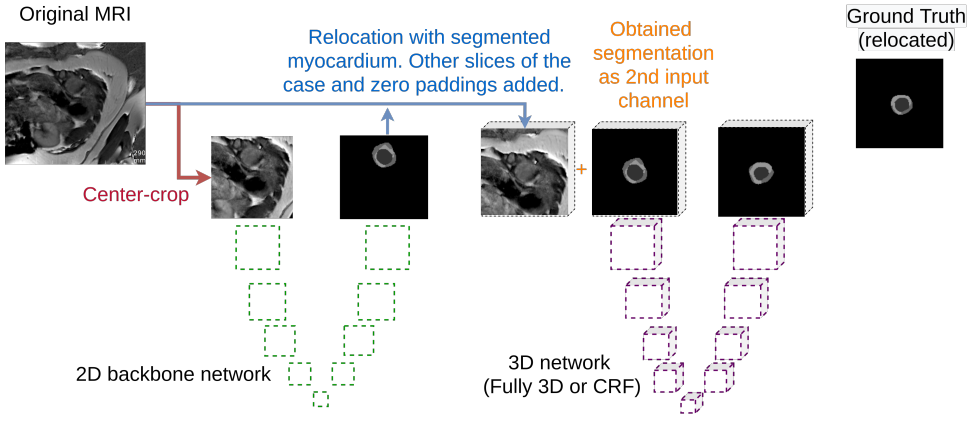


Figure 6.2: Framework of cascaded multidimensional networks for non-aligned DE-MRI.

6.3/ EXPERIMENTAL RESULTS

The evaluation metrics are the same as the proposed 2D segmentation methods (Section 5.6.1). One major difference compared to the evaluations in the previous chapter is the segmentation of the PMO.

The evaluation between 2D baseline (Section 5.3.1) and 3D models (Section 6.2.2), different binary and multi-class loss functions (Section 5.3.4), and configurations of the cascaded framework (Section 6.2.4) are performed on a subset of the private dataset, containing 100, 30and 30 cases respectively for the training, validation and test. With 3D models, the inputs are processed into a set of 16-layer volumes (Section 6.2). The batch size is set to 1 and the images are normalized by histogram normalization according to Eq. 5.1.

Different to the previous experiments, the evaluation with 3D models counts the PMO as an independent class, which brings more challenge due to its small size and low occurrence.

Loss functions The CE loss, Dice loss, both variants of generalized Dice loss (GD according to Eq.6.1 and Eq.6.2), as well as combinations of CE with Dice and GD losses (CE+Dice and CE+GD Eq.6.2) are trialled for a single-class and multiple-class segmentation tasks with the 2D baseline network. Tests are also performed on 5x interpolated MR images by nearest interpolation to demonstrate if the segmentation precision can be optimized. Table 6.2 shows that CE and CE+GD Eq.6.2 obtain better results than the GD Eq.6.2. The GD Eq. 6.1 loss does not pass the test because all the five classes of target tissues can not be segmented at the same time with the recommended weighting $w_l = 1/\left(\sum_{n=1}^{N} r_{ln}\right)^2$ of Eq. 6.1 or other tuned weighting.

Table 6.2: Results of single target and multiple target segmentation with different loss functions on 2D baseline network.

Loss	Interpolation	Myocardium			Infarction			PMO		
		Dif(mL)	HD(mm)	Dice	Dif(mL)	PIM(%)	Dice	Dif(mL)	PIM(%)	Dice
CE	1x	19.26	3.380	0.8114	19.26	5.89	**0.6058**	1.75	1.34	0.3774
GD Eq.6.1	1x					Failed				
GD Eq.6.2	1x	23.68	3.545	0.7954	23.68	7.28	0.6037	1.84	1.40	**0.4866**
CE+GD Eq.6.2	1x	16.60	3.071	0.8143	16.60	**5.16**	0.6001	**1.72**	**1.31**	0.2915
CE	5x	**10.17**	2.256	**0.8199**	**10.17**	5.41	0.5714	1.74	1.32	0.4498
GD Eq.6.2	5x	26.76	3.704	0.7849	26.76	7.95	0.5658	1.874	1.39	0.4453
CE+GD Eq.6.2	5x	11.98	**2.21**	0.8166	11.98	6.54	0.5943	1.86	1.43	0.4453
BinaryCE	5x	43.10	3.15	0.7915	-	-	-	-	-	-
Dice	5x	29.60	2.62	0.8100	-	-	-	-	-	-
BinaryCE+Dice	5x	42.25	3.37	0.7901	-	-	-	-	-	-

Framework and networks To configure the framework, different cascaded 2D segmentation masks are tested as the second input channel of the 3D network. Then the fully 3D network and the proposed CRF network are compared for single-class and multi-class segmentations under the same framework and loss functions. Results in Table 6.3 indicate that CRF outperformed the Fully 3D network on the primary metrics for the infarction and the PMO prediction. The 2D segmentation of all the tissues is a better choice for the 3D network's second input channel. The CRF with interpolated MR images achieves significant improvement in the volume prediction of the infarction. Compared to the 2D method and inter- and intra-observer variations, the proposed cascaded framework with CRF is more preferable to the 2D backbone, and even outperforms the inter-observer on the primary metrics. Figure 6.3 shows the segmentation result of different methods on an MRI case. By referring to the adjacent images, the 3D methods avoid the false positive PMO segmentation, which is generated by the 2D method. However, 3D methods extend the infarction segmentation to the suspected peri-infarction zone on the following slices, which may also explain the drop of Dice index compared to the results of the 2D network.

Table 6.3: Results of different frameworks and 3D networks. The second channel is the predicted mask from the 2D backbone network. 3D networks share the same CE+GD loss and the same weighting for multi-class segmentation, and Dice loss for the single-class (myocardium) segmentation. 2x and 5x stand for interpolation rate.

Method	2nd channel	Myocardium			Infarction			PMO		
		Dif(mL)	HD(mm)	Dice	Dif(mL)	PIM(%)	Dice	Dif(mL)	PIM(%)	Dice
Fully 3D	Myocardium	**11.92**	3.78	0.7867	6.65	5.42	0.5625	1.96	1.48	**0.6705**
CRF	Myocardium	15.45	3.76	0.7935	6.44	5.21	0.5792	1.84	1.35	0.6634
CRF	Non	15.60	3.84	0.7872	6.79	5.51	**0.6171**	1.95	1.46	0.5604
CRF	All tissues	14.56	3.91	0.7867	6.40	5.21	0.5874	**1.44**	**1.01**	0.6153
CRF 2x	All tissues	19.74	**3.46**	0.7838	**5.57**	**4.65**	0.5401	1.94	1.50	0.5081
Fully 3D	Myocardium	30.24	4.31	0.7785	-	-	-	-	-	-
CRF	Myocardium	21.22	3.89	**0.7948**	-	-	-	-	-	-
2D CE 5x	-	10.17	2.256	0.8199	10.17	5.41	0.5714	1.74	1.32	0.4498
Inter-observer	-	13.57	2.79	0.8220	6.34	5.29	0.6609	2.12	1.65	0.6014
Intra-observer	-	8.49	2.25	0.8443	5.16	4.35	0.7231	0.75	0.58	0.7214

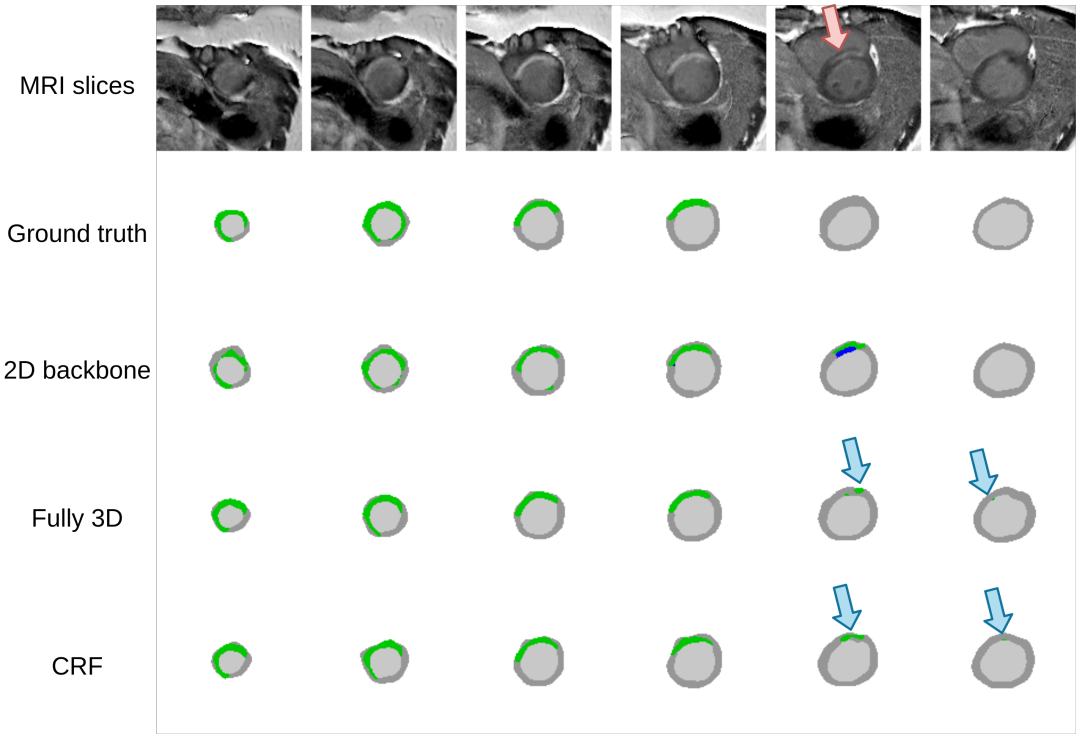


Figure 6.3: Segmentation results on one case. Green mask stands for the infarction and blue stands for the PMO. Red arrow points to a suspected peri-infarction zone; blue arrows point to the extended infarction segmentation by 3D networks.

6.4/ DISCUSSION

Input dimension and spatial information The 2.5D does not improve the segmentation results compared to 2D input under the same segmentation model. A crucial issue of 2.5D is the absence of the spatial information concerning the order of channels. Exchanging the channel order will have no effect on the feature extraction by the model, which is not logical. In such conditions, the network can not recognize between the essential im-

age (the middle one) and the supporting images that provide adjacent information. The final segmentation result on a 2.5D may be an average confusion of all the input channels.

The segmentation of 3D networks can be considered as a compromise between adjacent slices, which improves the volume prediction but sometimes weakens the Dice index. Under the cascaded framework, 3D CRF shows a better performance than the fully 3D network due to the constrained receptive field to intra-slice at low semantic layers. One possible reason for this is that although inter-slice information is meaningful, overly detailed inter-slice information can interfere with the reconstruction of segmentation details. Given the more coarse voxel size at the z-axis, this detailed inter-slice information introduces more bias relative to the gains.

Loss functions of multi-class segmentation The original generalized Dice loss (Eq. 6.1) fails the experiments due to its sensitivity of the class weighting ratio. Since the squaring operation is performed during the ratio calculation, the variance of the loss is enlarged when the class distribution is different between the samples in a batch and the whole training set. In case of important class imbalance, this formula is unstable, and as the result on our experimental data, no matter how this ratio is fine-tuned, there is always one class that is completely ignored in the segmentation mask. The proposed variant of generalized Dice loss (Eq. 6.2) refers to the weighting method of the cross entropy loss. The weighting is a more linear way and inaccuracy of the class weighting ratio will not result in a completely ignored class in the segmentation result.

6.5/ CONCLUSIONS

In this chapter, we presented the multiple class segmentation on DE-MRI with 3D CNN-based models. A new 3D segmentation model is proposed with a constrained receptive field (CRF network) and a variant of generalized Dice loss (Eq. 6.2) are proposed. The CRF network incorporates 2D convolutions at layers of large feature maps and 3D convolutions at layers of small feature maps. This conception forces the network only refers to inter-slice information for high-level (more abstract) semantic information. The CRF network obtains better scar tissue quantification since all slices are considered as a single entity, while the Dice metric is not improved due to the bias introduced from adjacent slices since the Dice is calculated on each slice. The proposed variant of generalized Dice loss obtains satisfying results when combined with cross-entropy loss. The original implementation of generalized Dice loss can not pass the experiments due to the non-linear way of the weighting ratio calculation.

7

Related works: automatic myocardial infarction assessment from multivariate diagnoses with machine learning

7.1/ Introduction

The automatic assessment of MI is not limited to the use of DE-MRI. Multiple medical diagnoses (Section 2.4) can be performed according to the progression of the condition. This chapter introduces the related works about the automatic MI assessment with different medical diagnoses.

Two types of automatic MI assessment works are investigated. The first works focus on the MI prediction from multivariate data. The difficulty for such works is to fuse the textual and image data: the image has higher dimension than the text hence their feature encoding should be operated differently. The second works aim at making a MI evaluation within feature patients other than imaging diagnoses. The application of these works are valuable in case of cardiac emergency when the imaging diagnosis has not been available.

This chapter is organized as follows:

- Prediction with multivariate data. The MI assessment works incorporating both textual and image data are presented. The machine learning-based prediction algorithms will be briefly described, then the data fusion and the prediction pipelines will be introduced.
- Classification and quantification of myocardial infarction from patient features. The MI classification or quantification works from patient features are presented.

7.2/ PREDICTION WITH MULTIVARIATE DATA

7.2.1/ BASIC DATA INTERPRETATION ALGORITHMS

Textual and image data should be processed to a similar dimension for further prediction as an ensemble. Provided with the MRI, a simple down-sampling CNN as AlexNet encodes the images to regression or classification outputs ([157, 155]), or optionally U-Net-based down-sampling up-sampling models yield the segmentation of different myocardial tissues so that the volume of each tissue can be quantified ([155, 151]).

To interpret the textual data of the clinical and physiological information, the choice of predictive models is more diverse. A crucial function of the common predictive models is their ability to solve non-linearly separable problems. For example, the MultiLayer Perceptron (MLP) ([30]) is a feedforward artificial neural network. Inputs are passed through multiple layers in which data are mapped with non-linear activation functions in the forward stage ([154, 157]). The decision tree ([9]) and the random forest ([13]) are flow-chart-like decision models that consist of nodes ([157, 154]). In the decision tree, each node examines an attribute and each branch represents the output of a test. Therefore, the branch where all the nodes are satisfied points to the prediction of the input. As its name suggests, the random forest is an optimization of the decision tree consisting of a stack of trees. The random forest corrects the overfitting habit of the decision trees by training uncorrelated trees and the final decision is made by individual trees. Boosting methods are the ensemble of sequentially connected weak learners ([14]). In the context of decision trees, the gradient boosting decision trees build a series of trees, which are the weak learners in this boosting method. Errors are passed between trees, with each tree attempting to reduce the errors passed from the previous tree ([21, 154]). Moreover, usual statistical models such as Support Vector Machine with non-linear kernel ([22]), k-Nearest Neighbors ([11]), the logistic regression ([26]) were investigated by Sharma *et al.*, Girum *et al.* and Ivantsits *et al.* ([157, 151, 154]).

7.2.2/ DATA FUSION AND DECISION OF THE PRESENCE OF MYOCARDIAL INFARCTION

The different format and dimensions between the images and the textual data constrain the decision with a single predictive model. Data fusion is therefore a challenging issue to achieve the maximum semantic information. Lourenço *et al.* [155], Girum *et al.* and [151] deployed the same strategy of predicting the volumes of different tissues as additional textual features alongside the 12 clinical and physiological features. Nevertheless, the volume estimation and the decision-making models are different among these approaches, *e.g.* Lourenço *et al.* and Girum *et al.* employed U-Net-based models to get

the segmentation. Apart from the surface regression methods, the concatenation of the surface information to other textual features was also variable. Lourenço *et al.* added the volumes of all myocardial tissues as four additional textual features. Girum *et al.* only considered if the case is pathological as one additional Boolean feature. Ivantsits *et al.* [154] tried to interpret the DE-MRI as textual information that the obtained textual information was radiomic features [106]. The radiomic features interpreted from the DE-MRI were intended to model the myocardial features such as the intensity, shape, and spatial characteristics. In practice, Ivantsits *et al.* investigated the shape and the Gray Level Co-occurrence Matrix (GLCM) that described the second-order joint probability function of an image region as the experimental radiomic features. Sharma *et al.* [157] proposed a stacked multi-modal approach without obtaining intermediate data such as the infarct volume or the radiomics features. The classifications were first achieved by a series of statistical models and a multi-modal CNN. Then the individual classifications were fed into an MLP to get the final decision. The application of the series of classification models could be thought as a boosting method and the models inside played the role of weak classifiers since their decisions would be judged together with the CNN's output by the MLP at the end of the proposal. Figure 7.1 concludes the above-mentioned classification pipelines that fuse both patient features and DE-MRI.

7.3/ CLASSIFICATION AND QUANTIFICATION OF MYOCARDIAL INFARCTION FROM PATIENT FEATURES

Research on the correlation between the physiological, clinical and paraclinical data and the symptoms of MI have been conducted for decades. More than 30 years ago, Goldman *et al.* [10] developed a computer protocol to diagnose related diseases when a patient received by an emergency department complained of chest pain unexplained by trauma or chest film abnormalities. Their decision protocol was based on a recursive partitioning approach [4]. About 50 potential predictive variables from the clinical history of the patient, physical examination, and ECG were incorporated into the decision protocol and the ultimate diagnosis of MI depended on three criteria namely the serum enzyme level, the comparison of the Q and R waves with the first ECG and the cardiac scintigraphy. In the case of a sudden unexplained death within 72 hours of receiving the patient, Than *et al.* [134] also predicted the likelihood of acute MI with decision trees afterward. Their prediction model incorporated the age, sex, and serial cardiac troponin 1 concentration and the ultimate diagnosis was adjudicated according to the Universal Definition of MI [116]. Similarly, Romero-Farina *et al.* [144] tried to predict the risk score for cardiac events. The gated SPECT metrics were considered with other clinical features for the prediction, which is the major highlight of the work.

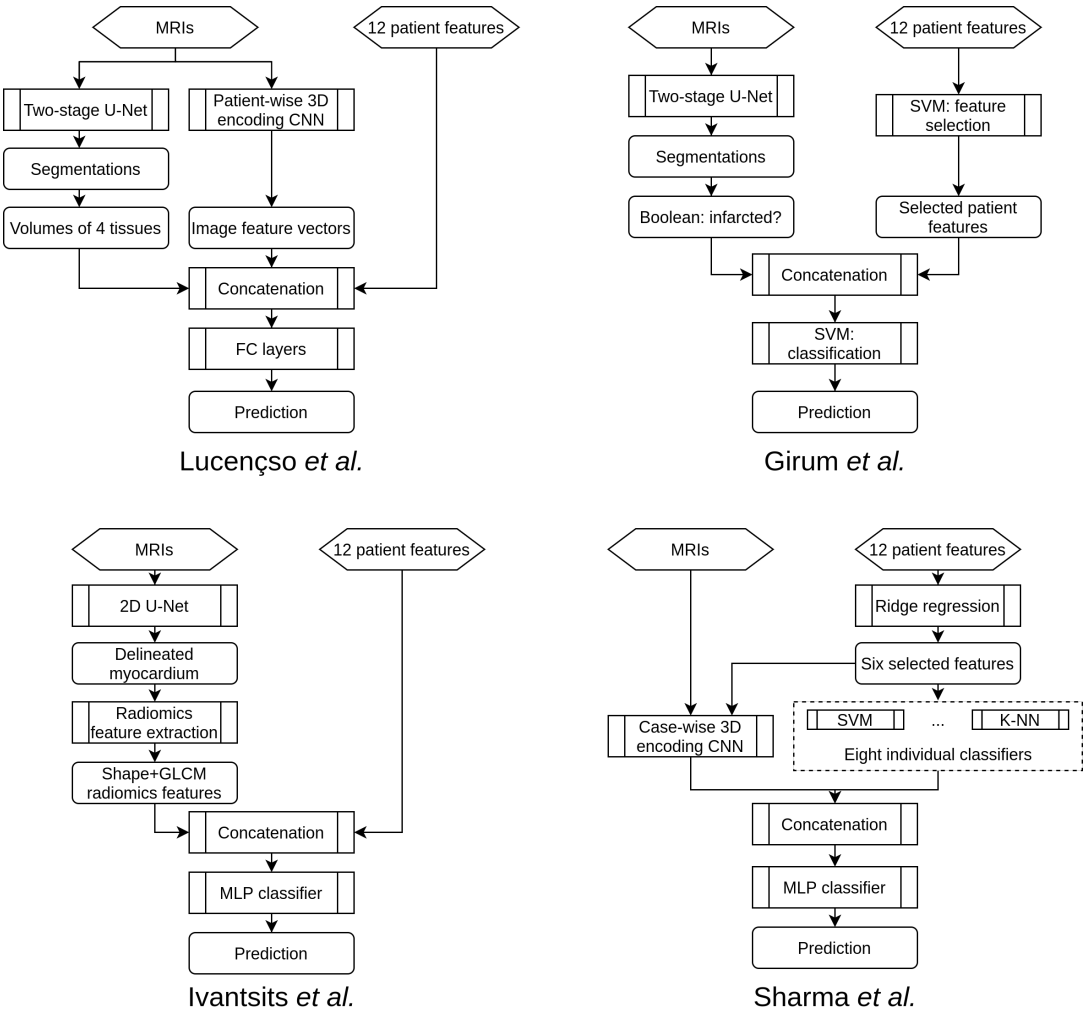


Figure 7.1: Multi-input classification pipelines for the classification.

Other works were dedicated to developing classification models of the presence of MI or other cardiac events. Al-Zaiti *et al.* [137] made the acute coronary syndrome prediction with temporal-spatial features of the 12-lead ECG. This work aimed at increasing the quality of pre-hospital ECG diagnosis. Faced with a huge number of features, Gárate-Escamila *et al.* [139] applied Principal Component Analysis (PCA) to reduce the dimensionality of the data, Daraei *et al.* [87] used Evolutionary Algorithms for the feature selection, then they classified the heart disease with Machine Learning models. Unlike the features studied in previous articles, Melillo *et al.* [68] proved that the heart rate variability-based classifier showed higher predictive values than the conventional echographic parameters for the cardiovascular event prediction.

7.4/ CONCLUSIONS

The works of MI assessment from multivariate data usually encoded the image data to a low dimension so that the final prediction was based on the fused textual and image data. The format of the encoded image is diverse, such as the radiomics features and the Boolean variable of the presence of MI.

Most of the works rely on patient features aimed at predicting the likelihood or classifying the presence of the MI and related cardiac events. Especially in pre-hospital or emergency cases, automatic predictive models have been shown to be of great clinical relevance. However, few researchers have attempted to develop predictive models of the MI based on clinical features in conjunction with medical imaging. Moreover, few works mentioned the automatic assessment of the PMO given that its clinical diagnosis mostly relies on invasive or imaging techniques [39]. The review of the literature reveals that no study has tried to train a predictive model through the quantitative data of MI provided by CMRI to obtain a prediction of its severity.

8

CONTRIBUTION 3: AUTOMATIC MI ASSESSMENT FROM MULTIVARIATE DIAGNOSES WITH MACHINE LEARNING

8.1/ INTRODUCTION

A variety of medical diagnosis methods can be proposed to detect or evaluate the extent of acute MI. However, the accuracy is often in conflict with the time required for the diagnosis, which remains a therapeutic emergency. For example, the DE-MRI is the gold standard for the diagnosis and the evaluation of the MI. The previous section also proposed and evaluated the automatic MI assessment approaches on DE-MRI that relied on Deep Learning and image processing methods. Nevertheless, the MI diagnosis with DE-MRI cannot be widely applied in the emergency department because of its required time for the acquisition and post-processing. In current practice, simple tools such as ECG, troponin assay and echocardiography are used to validate the emergency diagnosis of MI. ST segment analysis (Section 2.4.1) on ECG (especially in case of ST persistent elevation), intensity of troponin elevation, or LVEF assessment from TTE have been shown highly correlated with MI. Given these facts, when patients arrive in the emergency department complaining of chest pain, generally a series of indicators will be first listed with the help of the above-mentioned simple tools. If the examinations reveal the possibility of MI, the DE-MRI could be achieved in the next few days to have a more accurate evaluation of the myocardial impairment, after the acute phase and the early therapeutic management including revascularization and medications. Until the ultimate diagnosis based on the MRI exam is available, physicians mainly rely on the obtained physiological, clinical and paraclinical characteristics to determine the severity of a patient's condition and to give sound treatment advice.

In light of the above-mentioned facts, automatic prediction approaches are proposed to

precisely classify or quantify the severity of the acute MI only taking into account the physiological, clinical and paraclinical features. Standard machine learning algorithms involving linear models, Random Forest and Decision Trees, Support-Vector Machines (SVMs), Multilayer Perceptron and boosting models are employed for classification or quantification. Furthermore, the prediction from multivariate data involving both the patients features and DE-MRI are evaluated. An additional CNN branch encodes the DE-MRI and the encoded DE-MRI is concatenated to patient features for the MI classification by machine learning models. The fusion of data aims to investigate how different dimensions of data can be understood simultaneously and to quantify the gain in predictive accuracy that DE-MRI brings. The investigated data for the experiments come from the EMIDEC Challenge database, which consists of 150 cases of paired physiological, clinical and paraclinical features, and annotated DE-MRI (Section 3.4).

It is important to note that the proposal aims at providing early prediction to better orient patients in the emergency department. Therefore, the automatic prediction should only be used as an aid to clinical diagnosis and the associated risk of mispredictions should be taken into account by physicians.

This chapter is organized as follows:

- Classification of myocardial infarction from multivariate data. A classification model of MI from both DE-MRI and patient features is proposed. The data fusion strategy is proposed and the employed image encoding method and classification model are presented.

- Quantitative prediction of myocardial infarction from patient features. A quantitative and qualitative prediction model of the MI from patient features is proposed. The DE-MRI only participates in the training stage, interpreted as the ground truth of the MI volume.

- Experimental results. The experimental results of the above two proposals. The classification of the presence of MI is made from fused DE-MRI and patient features, and both the classification and the quantification of MI are predicted from patient features.

- Discussion.

- Conclusions.

8.2/ CLASSIFICATION OF MYOCARDIAL INFARCTION FROM MULTIVARIATE DATA

Classification of myocardial infarction from multivariate data The classification of MI from multivariate data refers to both patient features and DE-MRI with machine learning methods. In order to deal in an efficient way with the correlation between these data of different dimensions and semantic information, a mixed classification model is proposed. In the mixed model, a 3D CNN encodes the MRI as the surface of infarction then the surface is fed to Random Forest with other clinical characteristics to make the final decision. The proposal is designed for the application on the EMIDEC challenge dataset, where each exam consists of 12 pieces of 1D patient features, and a set of DE-MRI with manual annotations.

Each DE-MRI exam has multiple slices of 2D image but the clinical physiological information includes 12 pieces of 1D features. In order to handle these data that have different dimensions and semantic information for the classification of the myocardial infarction, the proposal contains two stages as shown in Figure 8.1: the encoding of DE-MRI then the classification on the fusion of encoded images and their paired clinical physiological features. The image encoding is realized by a 3D CNN and the classification of the myocardial infarction is done by Random Forest [13]. This conception aims at taking the advantage of the correlation between both types of data so that the classification result is more robust than on one single type of data.

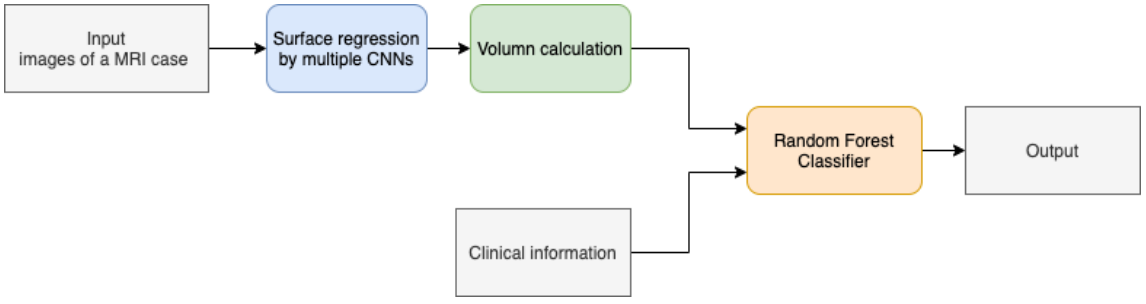


Figure 8.1: Overview of the proposed architecture of prediction combining CNNs and Random Forest. A set of three-slice images that come from the same MRI case are fed successively to the CNN. The output of the CNN is the regression of the predicted infarct surface of the middle slice in the input. Volume is calculated based on the predicted surfaces and the provided voxel spacing. Predicted volume and the clinical physiological information are concatenated as the input of the Random Forest classifier for the final classification.

8.2.1/ IMAGE PREPROCESSING

In order to fully catch semantic information from the DE-MRI, the image preprocessing is executed on each MRI case. To learn the spatial information between adjacent slices and ensure a fixed-size of CNN's inputs, three successive slices are taken as a single 3D input for the CNN. Assuming that an original MRI has N slices, the first and the last slice of the MRI are copied at the top and at the bottom side, hence N new 3D images are obtained and each 3D image is formed by three adjacent slices. Knowing that in the EMIDEC dataset the left ventricle myocardium is centered on the middle of each slice, to reduce the background's size, a center cropping of size (96,96) is performed on each slice. Therefore, each CNN's input has the same shape of (3,96,96). Figure 8.2 illustrates the way that the three-slice inputs are created. Despite a 3D input that may have more slices, three adjacent slices are sufficient to provide enough spatial information. With more slices, more bottom and top slices should be copied, which is not efficient for the surface regression.

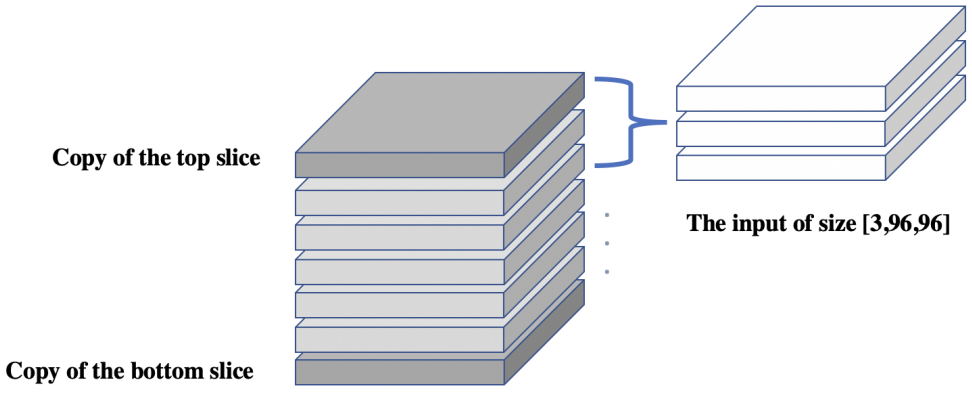


Figure 8.2: Preparation of three-slice input for the 3D CNN. For each center-cropped MRI case, firstly the top and the bottom slice are copied (dark gray). Secondly, every three consecutive slices are chosen to form a 3D input for the CNN.

8.2.2/ CNN WITH 3D MULTI-KERNEL CONVOLUTION BLOCK

The 3D convolution is added only at the first layer in the CNN since each input has only three slices. To expand the receptive field, multiple 3D convolution kernels of the size (3,3,3), (3,5,5) and (3,7,7) encoded the input image in parallel. This conception is inspired

by Inception structures [100] and the objective is to flexibly extract features for objects of various sizes. For the 3D convolutions, the zero padding is performed only at the width and length dimension in light of the thickness of 3 at the dimension Z. Hence, the feature maps generated by each 3D kernel are in the same size (96,96).

The 2D feature maps obtained by the 3D multi-kernel convolutional layer are then passed to residual modules which are similar to ResNet [77]. Each residual module performs 4 times of convolution + down-sampling + batch normalization + ReLU activation on feature maps. To reinforce the semantic information interpretation, the DAC block [125] is added at the last layer before the fully connected layers, motivated by the Inception-ResNet-V2 block [100] and atrous convolution. DAC has four cascade branches with a gradual increment of the number of atrous convolution, from 1 to 1, 3, and 5. Therefore, the network can extract high-level semantic information of different scales.

At the end of the CNN, the surface of the pathological tissue is predicted through the two fully connected layers as the CNN's final output. Smooth L1 loss is employed to penalize the error between the predicted surface and the ground truth. Figure 8.3 shows the structure of the proposed encoding CNN for the surface regression.

8.2.3/ VOLUME CALCULATION

The output of the CNN is the predicted surface of the infarct in one MRI slice. In order to calculate the predicted volume of the infarct in one MRI case, the sum of the surfaces multiplying the pixel spacing (provided as the MRI metadata) is calculated. The predicted volume is used as an additional feature of the subsequent Random Forest model for the definitive classification of the myocardial infarction. The volume calculation is illustrated in Figure 8.4.

8.2.4/ RANDOM FOREST CLASSIFIER

Random forest, developed by Breiman [13], is a classification algorithm that uses the ensemble of classification trees. Each of the classification trees is built using a bootstrap sample of the data, and at each split, the candidate set of variables is a random subset of the variables. Thus, random forest uses both bagging (Bootstrap Aggregation), a successful approach for combining unstable learners, and random variable selection for tree building.

As the predicted volume of infarct from the CNN is obtained during the first stage, at the second stage, the predicted volume is concatenated to the 12 clinical and physiological features. The Random Forest is trained on these 13 features and the output is binary that indicates if the case is pathological or not.

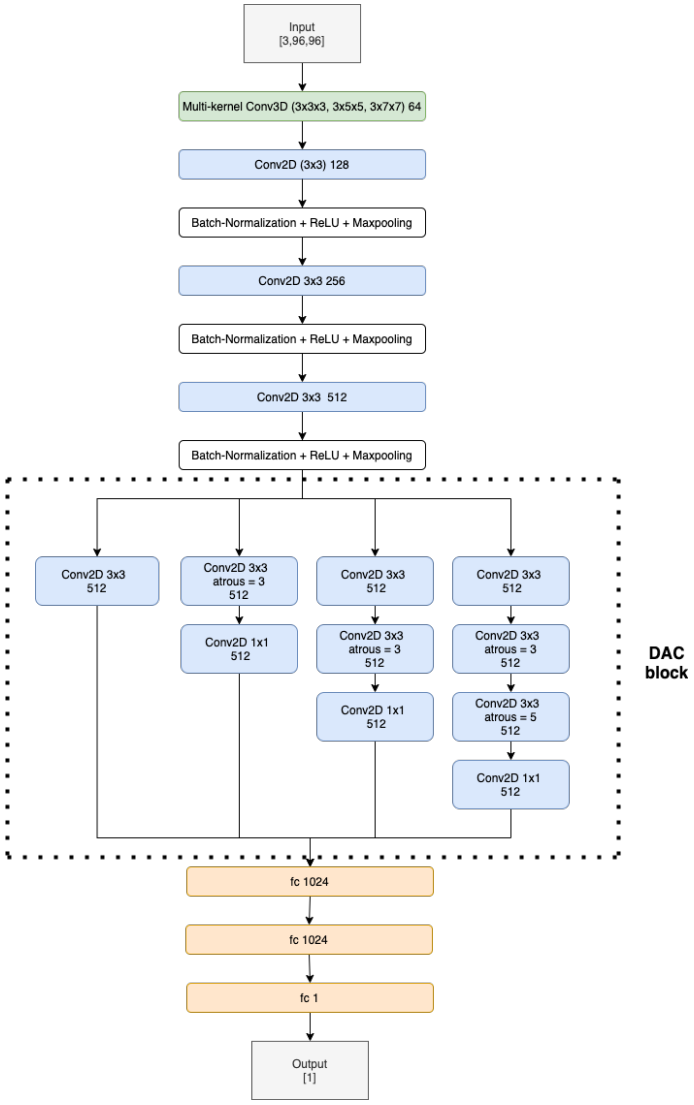


Figure 8.3: The structure of the neural network. The 3D convolution is adopted only at the first layer. Before the fully connected layers, the DAC block enhances the cognition for both large and small areas.

8.3/ QUANTITATIVE PREDICTION OF MYOCARDIAL INFARCTION FROM PATIENT FEATURES

An automatic prediction approach is proposed to precisely classify and quantify the severity of the acute MI only taking into account the physiological, clinical and paraclinical features. Furthermore, experimental attempts are also made to estimate the PMO individually with the same approach. The predictions are based on standard machine learning algorithms involving linear models, Random Forest and Decision Trees, SVMs, Multilayer Perceptron and boosting models. For each patient, the annotated DE-MRI provides the quantitative ground truth of the infarction and the selected patient features are thought as the input data. For the training stage, the features are the inputs of an appropriate

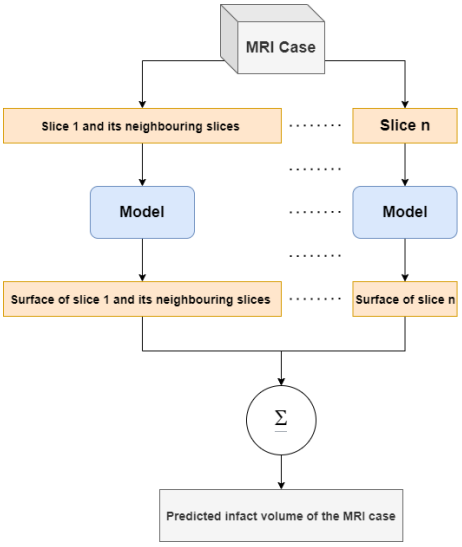


Figure 8.4: The algorithm of volume calculation. The CNN predicts the surface of each slice. The predicted volume of an MRI case is calculated based on the predicted surfaces and the voxel spacing provided by the MRI metadata.

machine learning model. Depending on the classification or the quantification task, the state of the myocardium or the PIM calculated from the DE-MRI are the target output of the model. For the inference stage, once the model is well-trained with the paired patient features and DE-MRI, it can predict the severity of the patient's infarction only according to the patient features. The investigated data for the experiments come from the EMIDEC Challenge database, which consists of 150 cases of paired physiological, clinical and paraclinical features, and annotated DE-MRI.

More precisely, the proposal targets two prediction problems: the quantification of PIM and the classification of the state of the myocardium. Predictions are carried out with classification and regression algorithms. Each training case incorporates the 12 patient features shown in Table 3.2, and the PIM or the state of the myocardium is evaluated from the DE-MRI and its annotations. The machine learning models are first trained with paired input data *i.e.* patient features and the ground truth. During the inference stage, only the clinical features are fed to a trained predictive model and the model's output is the predicted PIM or the state of the myocardium. Figure 8.5 shows the workflow of the proposal during the inference stage.

8.3.1/ DATA PREPROCESSINGS

Physiological, clinical and paraclinical data should be first preprocessed so that the machine learning models can manage the features correctly. The data format of model's input should be numerical, therefore, the categorical features, *i.e.* Tobacco and Killip max are converted to one-hot encoding, and Boolean features are encoded as 0 or 1.

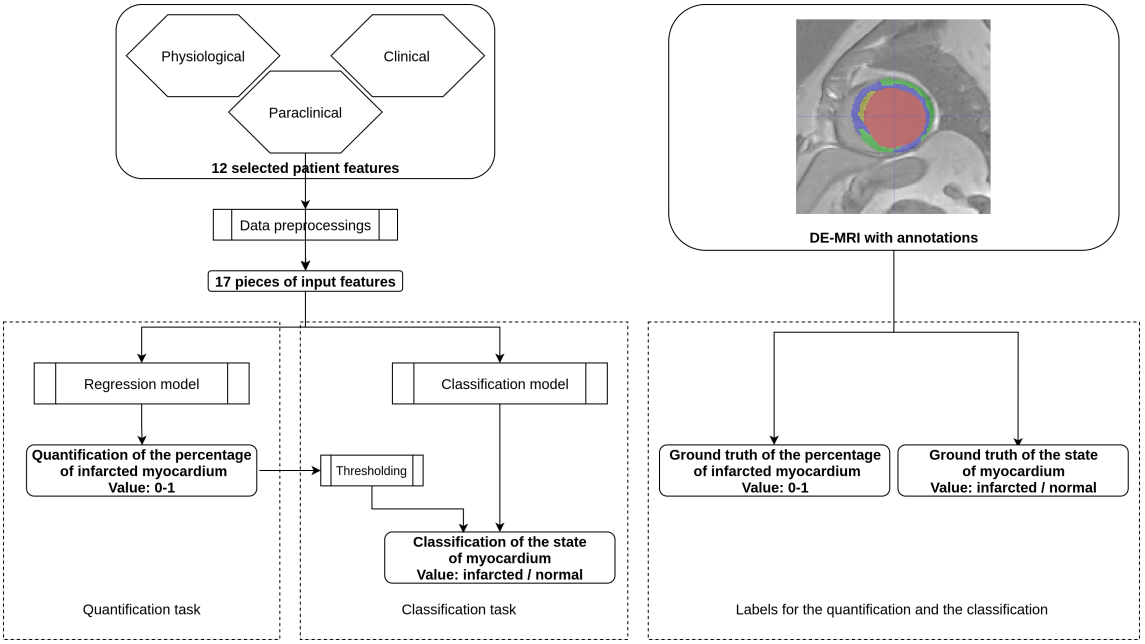


Figure 8.5: Workflow of the proposal. This figure presents the inference stage for the quantification task and the classification task of the automatic MI prediction. On the left part, selected patient features are first preprocessed to 17 pieces of numerical or Boolean features. For the quantification task, the features are incorporated through a regression model so that the obtained value is the PIM ranging from 0 to 1. In the case of the classification task, the prediction can be obtained by either a regression model followed by a thresholding or a classification model. During the training stage, the regression model is supervised by the ground truth PIM, and the classification model is supervised by the ground truth state of the myocardium. Both ground truths are defined from the DE-MRI and manual annotations.

Gradient descent-based and distance-based algorithms are sensitive to the feature scaling. Therefore, normalizing the features to a universal interval may improve the predictive performance of such models [145]. To address this fact, in addition to the numerical encoding, normalization is also applied on the features of age, troponin, LVEF and NT-ProBNP.

Being the ground truth of the machine learning models for the quantification task, the PIM is calculated from the manual annotations of myocardium and scar contours. Given that the voxel size is constant for each DE-MRI case, the calculation of the volumes of myocardium and infarcted areas only relies on the manually drawn contours of all the myocardial tissues. Thus, the PIM of one patient is calculated as

$$PIM = Volume(scartissue)/Volume(myocardium) \tag{8.1}$$

where the volume refers to the voxel quantity of a scar tissue in the DE-MRI case. The scar tissue can be either MI or PMO. For the MI evaluations, the PMO is considered as a part of MI area. The PIM value, which is theoretically ranged from 0 to 1, describes the

severity of MI. For the classification task, the state of myocardium denotes the presence of the infarction. Therefore, the data format of the state of myocardium is Boolean: the MI or the PMO is detected or not from the DE-MRI.

8.3.2/ MACHINE LEARNING ALGORITHMS

The employed predictive models include regression models and classification models. Regression models estimate the relationships between the dependent variable of numerical values, *i.e.* the predicted PIM and the input patient features, while for the classification models, their dependent variable is the binary state of the myocardium. For the quantification task, regression models should be applied since the expected output is the PIM. Therefore, the target label during the training stage for the regression model is the calculated PIM according to the DE-MRI. For the classification task, both regression and classification models are feasible. Indeed, a regression model followed by a thresholding can also provide a binary classification result of the state of the myocardium.

For the regression task, the linear model is first studied. Linear regression tries to establish a linear function that links the input features and the regression label [8, 2]. Although the linear model can predict the PIM with the scalar patient features, non-linear models may be of more interest since the PIM regression task may be a non-linearly separable problem due to the complexity of the input features. Therefore, more learning algorithms using non-linear models including Support-Vector Machines (SVMs) with non-linear kernel function, Random Forest and Decision Trees, Multilayer Perceptron and boosting models were investigated. SVM is not necessarily a non-linear model, however, with a non-linear kernel function that maps the data to a higher dimension, the SVM can solve the non-linearly separable problem [22]. Decision Trees algorithm has a flowchart-like structure that consists of nodes [9]. Optimized from Decision Trees, Random Forest is trained on uncorrelated Decision Trees as its name suggests, and the inference is made by the individual trees. Random Forest is by definition more robust to overfitting so it generally outperforms Decision Trees [13]. Boosting methods are the ensemble of sequentially connected weak learners [14]. For example, Gradient Boosting Decision Trees consist of a series of trees, which are the weak learners in this boosting method. Errors are passed between trees, with each tree attempting to reduce the errors passed from the previous tree [21]. Gradient Boosting Decision Trees algorithm tends to outperform Random Forest in practice, however, the sequential structure results in its longer computation time than the parallel structure in Random Forest. Multilayer Perceptron is a kind of feedforward artificial neural network. Inputs are passed through multiple layers in which data are mapped with nonlinear activation functions in the forward stage [30]. Moreover, knowing that a regression model and a classification model could share the same learning algorithm but different optimization functions, the widely used SVM (with

linear kernel function) and Random Forest were selected as the learning algorithm for the classification models.

In addition to the predictions obtained from an individual model, the ensemble method is also investigated. The ensemble method is used to improve the predictive reliability by combining the predictions from all individual models into a single set of predictions [18]. In this work, for each prediction target, the ensemble method's prediction is derived by adding up and then averaging all the models' predictions.

8.3.3/ Statistical analysis

Primary analysis

The quantitative prediction of the severity of the MI, *i.e.* the quantification of the PIM including MI (PMO inclusive) and PMO (MI exclusive), is the main objective of the work. The absolute scar tissue volume, which is another possible severity indicator, is not adopted since its severity evaluation can be biased by patient's physiological conditions. The proposal's performance can be described as the absolute quantification error between the inference result of machine learning models and the PIM calculated from DE-MRI for each scar tissue. Multiple machine learning-based regression models were compared with the ground truth PIM. The comparison results are presented as the mean absolute difference. In order to make the best use of the available data, given the limitations of available data cross-validation is used for the experiment evaluations.

Secondary analysis

In addition to the quantitative analysis, a qualitative analysis is also performed as the secondary analysis, *i.e.* the classification of the state of the myocardium. As in the primary analysis, MI and PMO are the two assessed scar tissues. The classification can be carried out by both classification and regression models. However, the data formats of the training labels and the predicted values are different when both the models are applied to the classification task. The classification carried with regression models consists of the regression models and a thresholding. The regression models are the same as in the primary analysis: the training is supervised with the PIM therefore the model's output is the PIM. A discrimination threshold differentiates if the patient is pathological from the prediction of the regression model. Differently, when the classification is carried out with classification models, the classification models are trained with Boolean target labels which are annotated if the case is normal or pathological. Therefore the predictive value of the classification models is also Boolean. Both the classification methods can be refereed in Figure 8.5.

For classification models, the classification performance could adequately be evaluated by the sensitivity (or recall), specificity (or selectivity), precision and accuracy metrics. However, the discrimination threshold used in the regression model could impact the confusion matrix, thus, the Receiver Operating Characteristic curve (ROC) was also adopted to plot the true positive rate (TPR) against the false positive rate (FPR) to intuitively compare between classification and regression models by considering the Area Under Curve (AUC) [36].

ADDITIONAL ANALYSES

Additional analyses are investigated under the contests of the quantification of the PIM and the classification of the state of the myocardium. Training database size is crucial for machine learning models. Therefore, the quantification performance with different training data volumes is compared. Then the quantification error is studied according to the severity of MI with the help of Bland–Altman plot [6]. Moreover, the importance of individual features for different predictive models is extracted. The quantification of scar tissues with limited selected features is also investigated. Finally, cases who has important inconsistencies between the prediction and the ground truth are listed to undertake further medical interpretations.

8.4/ EXPERIMENTAL RESULTS

8.4.1/ CLASSIFICATION OF MYOCARDIAL INFARCTION FROM MULTIVARIATE DATA

The CNN and random forest classifiers are implemented by Pytorch and Scikit-learn. The CNN has been trained for 500 epochs and the predicted volume is the ensemble of multiple models' prediction of different epochs, i.e. using the ensemble method. To show the advantage of the classification on merged MRI and clinical physiological information, the classifications made only on the CNN and only on Random Forest are also performed. The comparative tests use the same method from stage one or stage two and the data repartition is identical to that of the above two-stage method, which can be considered as the baseline approaches.

The experiments are executed on two different subsets from the EMIDEC dataset. The first experiment employs five-fold cross-validation on the training set of the EMIDEC dataset. Table 8.1 shows the results using 80 cases of training data. The two-stage method achieves $95\,\% \pm 3\,\%$ accuracy, which is respectively $4\,\%$ and $8\,\%$ superior to the CNN only and Random Forest only approaches.

The second experiment takes the entire training set of 100 cases for training and the

Table 8.1: Classification accuracy of three approaches on the training set of EMIDEC dataset. The five-fold cross-validation is employed.

	Random Forest	CNN	Random Forest + CNN
Accuracy (%)	87	91	95
Standard deviation (%)	3	2	3

prediction is made on the test set of 50 cases. This experiment result is released by the classification contest of EMIDEC challenge. Table 8.2 shows the proposal's result with other participants. The proposal obtains the best result in EMIDEC challenge and only 4 cases among the 50 cases of the test set are wrongly classified.

Table 8.2: Results of the classification contest. Best results in bold. Results in brackets are obtained from predictions based solely on textual data.

Challengers	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
Lourenço *et al.*	87.88	70.59	85.29	82 (70)
Ivantsits *et al.*	72.73	82.35	88.89	76
Sharma *et al.*	72.73	41.18	70.59	62
Girum *et al.*	78.79	88.24	92.86	82 (78)
Proposal	**90.91**	**94.12**	**96.77**	**92** (74)

8.4.2/ QUANTITATIVE PREDICTION OF MYOCARDIAL INFARCTION FROM PATIENT FEATURES

The machine learning models are implemented with scikit-learn [46], XGBoost [74] and lightGBM [91] libraries, and Python 3.6.9. The training and the inference are only conducted with CPU and since the operation time is in the order of seconds, the computational time is not specifically listed. The experiments are conducted on the EMIDEC Challenge (2020) dataset. Except for the tests of the training set volume in Section 8.4.2, all other experiments employ 10-fold cross-validation, *i.e.* the training set volume of each split is 135 patients.

REGRESSION MODELS FOR THE PIM QUANTIFICATION

The performance of different regression models is presented in Table 8.3. Linear regression model of the ordinary least squares and other non-linear regression models are evaluated. The ensemble of all the models' predictions is also examined. Using the mean PIM calculated from the ground truth as the predicted PIM, the assumed quantification is achieved as the baseline.

Multilayer Perceptron and SVM (with non-linear kernel function) respectively obtain the

lowest mean PIM difference for the MI and PMO prediction, and Random Forest achieves relatively low mean quantification error with a small variance. Results also reveal that the ensemble of all the prediction outperforms each single regression model for the MI quantification. The satisfying regression performance of SVM and Random Forest and their much shorter processing time compared to Multilayer Perceptron justify the choice of experimental learning algorithms for the classification task. Random Forest will be performed in all following experiments concerning the classification, SVM will be employed only for the classification in the analysis of the importance of patient features.

Table 8.3: Prediction error of regression models for the PIM quantification. In bold the best result of a single model. The ensemble is the average prediction of all the regression models.

Regression model	**Predicted PIM error**	
	MI[1]	**PMO**
Linear Regression	0.0639±0.0677	0.0152±0.0214
Support Vector Regression	0.0579±0.0632	**0.0116±0.0238**
Decision Tree Regressor	0.0679±0.0741	0.0162±0.0293
Random Forest	0.0587±0.0597	0.0149±0.0227
Multilayer Perceptron	**0.0578±0.0609**	0.0179±0.0229
Gradient Boosting Regressor	0.0602±0.0584	0.0152±0.0228
XGBoost	0.0646±0.0572	0.0172±0.0199
Light Gradient Boosting	0.0590±0.0616	0.0161±0.0227
Ensemble	0.0555±0.0594	0.0141±0.0210
Mean predicted PIM [2]	0.1070±0.0693	0.0162±0.0206

[1] PMO inclusive
[2] Calculated from the ground truth

CLASSIFICATION OF THE STATE OF THE MYOCARDIUM

The classification of the state of the myocardium is performed on both the presence of MI and PMO in two ways. Figure 8.6 shows the classification results of the presence of the infarction and the PMO. The results shown in Table 8.4 present the statistical metrics of the same methods as in Figure 8.6. For the thresholding, the best threshold value is obtained by iterating from 0 to 1 with a step of 0.001 and observing the best accuracy. Table 8.4 reveals that with the regression model and thresholding, the infarction classification error mostly results from the false-negative predictions according to the relatively low recall. Moreover, classifying on the ground truth of MI and PMO with the retained threshold values, the sensitivity of the infarction classification is correct (87.00%) while many cases suffering from PMO (sensitivity=56.86%) may be omitted. It also implies that with the classification report, physicians should pay particular attention to the missed suspected patients in case of negative prediction.

The obtained results reveal that for the classification of the presence of a particular target tissue, the regression model significantly outperforms the classification model that shares the same learning algorithm. A relatively high threshold value (PIM below 0.064) obtains the best accuracy for the infarction classification when the classification is done with the regressor followed by thresholding.

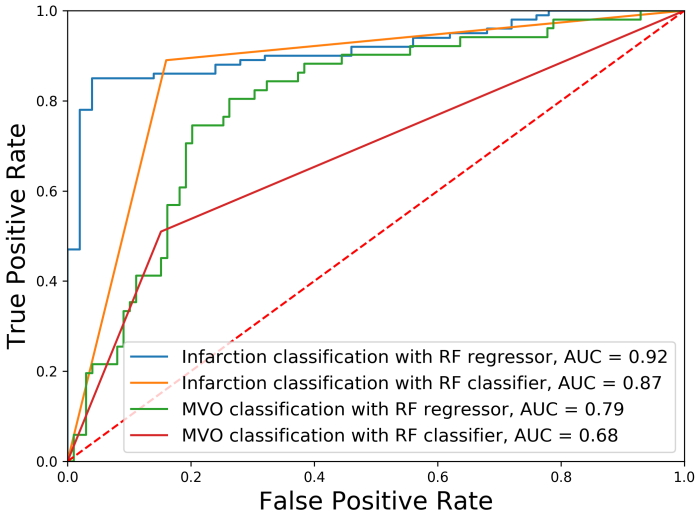


Figure 8.6: Receiver operating characteristic curves of classification results. The classification of different tissues is realized with Random Forest (RF) algorithms with different optimization functions.

Table 8.4: Classification results under different metrics. Threshold value θ is derived when the best classification accuracy is achieved from the RF regressor. RF: Random Forest

Prediction model	Target tissue	Sensitivity	Specificity	Precision	Accuracy
RF Regressor	Infarction, $\theta = 0.064$	85.00%	96.00%	97.70%	88.67%
	PMO, $\theta = 0.013$	70.59%	80.81%	65.45%	77.33%
RF Classifier	Infarction	89.00%	84.00%	91.75%	87.33%
	PMO	50.98%	84.85%	63.41%	73.33%
GT with thresholding	Infarction, $\theta = 0.064$	87.00%	100.00%	100.00%	91.33%
	PMO, $\theta = 0.013$	56.86%	100.00%	100.00%	85.33%

IMPACT OF TRAINING SET VOLUME

Supervised machine learning models are sensitive to the volume of training data. To justify if the quantity of cases in the dataset is the bottleneck for the proposal, and to estimate the potential of the predictive models if more training data could be available, a Random Forest regression model of 5000 estimators is trained several times feeding different quantities of training cases into the model each time. To ensure that the results are comparable, the cross-validation of different folds is applied to control the difference in the training data quantity.

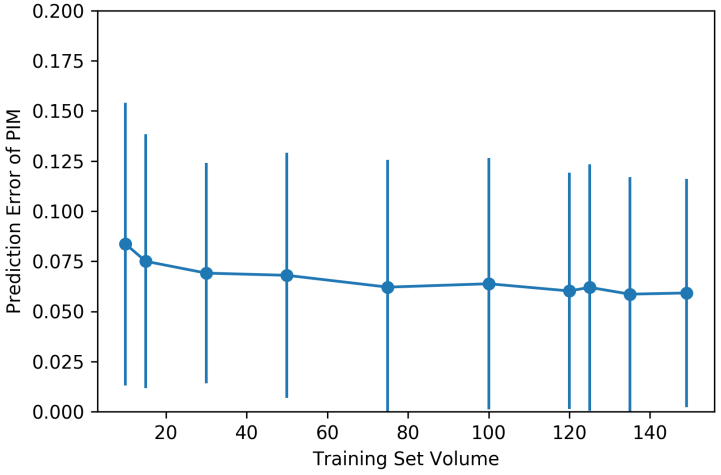


Figure 8.7: Impact of the training set volume on the mean and the standard deviation of quantification error. The dataset of 150 cases is randomly split into multiple folds to have different amounts of training data. The training set volume ranged from 10 to 149 for each cross-validation.

Figure 8.7 shows the improvement of the PIM prediction as the training set gets larger. Both the mean error and standard deviation of the PIM quantification decrease along with the increasing evolution of the training set volume. However, as the amount of data increases, the performance improvement becomes less and less obvious.

Performance on cases of different severity of the MI

To show the accuracy of the PIM quantification on the cases of different severity of MI, in Figure 8.8 the prediction error between the ensemble model and the PIM calculated from the MRI is presented as a Bland-Altman plot only considering the cases with visible MI on DE-MRI. Prediction error rises gradually with the increasing PIM, *i.e.* the prediction error is larger in the more severe cases.

Importance of patient features

The importance of physiological features for the prediction result can be visualized on some regression and classification algorithms. The feature importance for linear or non-linear, regression or classification models, for both the infarction and PMO predictions are presented in Table 8.5. Random Forest is selected as the non-linear regression and classification models, the ordinary least squares Linear Regressor is the compared linear regression model, and the SVM classifier with the linear kernel is the selected linear classifier. The regression models are trained with PIM as the target, while the classification

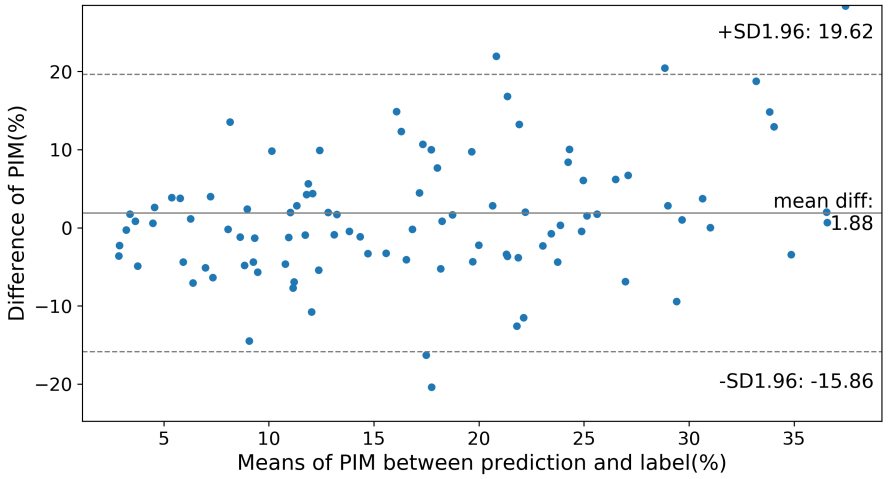


Figure 8.8: Prediction error on cases of different severity according to the infarction. Normal cases were not considered in this figure.

models are trained with the presence of infarction.

The importance of features in the case of linear models is signed, indicating the direction of influence (positive or negative correlation) of the feature on the prediction. To simplify the results, the absolute normalized importance is kept for the linear models. The mean importance for all the models of each feature is also calculated at the bottom of the table.

Table 8.5: Feature importance to linear and non-linear models for classification and quantification tasks. Categorical features are converted to one-hot encoding. Classification models are trained with the Boolean values, quantification models are trained with PIM. The importance of each model and task is normalized therefore the sum of importance is 100%. A higher importance shows a closer relationship between the feature and the disease. In bold, the four most important features for each model.

Target	Task	Model[1]	Feature importance(%)[2]															
			Sex	Age	SK	N-SK	F-SK	OW	HT	DB	HD	ST+	Troponin	KL1	KL2	KL3	KL4	LVEF	NTp
Infarction	Quantification	RFR	2.91	**5.55**	1.04	0.67	0.73	1.06	1.73	0.55	0.61	1.70	**63.01**	0.45	0.26	0.28	0.10	**10.80**	**8.55**
		LR	6.50	3.64	0.00	0.00	0.00	1.17	5.68	1.77	**10.21**	**8.42**	**46.80**	0.00	0.00	0.00	0.00	**11.72**	4.08
	Classification	RFC	**10.30**	**11.90**	2.44	1.65	2.18	2.02	3.21	1.37	2.70	**9.94**	**30.60**	1.18	1.22	0.17	0.10	9.29	9.71
		SVCL	**15.84**	0.93	4.57	5.71	1.14	2.13	1.14	5.22	**21.86**	**14.53**	**21.99**	0.87	0.72	0.50	0.35	0.99	1.50
PMO	Quantification	RFR	1.48	**20.09**	2.58	0.70	1.92	2.31	1.67	0.18	0.17	0.43	**37.38**	1.24	0.46	0.81	0.67	**17.82**	**10.11**
		LR	1.30	**9.40**	0.00	0.00	0.00	2.30	**12.95**	2.02	2.61	10.63	**40.70**	0.00	0.00	0.00	0.00	3.55	**14.54**
	Classification	RFC	4.04	**13.23**	2.57	1.79	2.37	3.42	2.73	1.64	0.88	6.14	**28.88**	1.21	0.99	0.57	0.47	**12.83**	**16.24**
		SVCL	4.90	5.88	1.17	2.27	3.44	0.99	2.13	**9.24**	3.81	**11.71**	**21.75**	5.71	7.55	**8.08**	5.19	5.16	1.02
Mean			5.91	**8.83**	1.80	1.60	1.47	1.93	3.90	2.75	5.36	7.94	**38.38**	1.33	1.40	1.30	0.86	**9.02**	**8.22**

[1] RFR: Random Forest Regression; RFC: Random Forest Classifier; LR: Ordinary least squares Linear Regression; SVCL: Support Vector Classifier with Linear Kernel

[2] SK: Smoker; N-SK: Non-smoker; F-SK: Former Smoker; OW: Overweight; HT: Arterial hypertension; DB: Diabetes; HD: History of coronary artery disease; KL1: Killip max=1; KL2: Killip max=2; KL3: Killip max=3; KL4: Killip max=4; NTp: NTProBNP

The results reveal that for different algorithms and predicted tissues, the feature importance can be slightly inconsistent. Combining all the results, the most important features in order of significance are troponin, sex, history of coronary artery disease and ST+ MI for the infarction, and troponin, age, NTProBNP and ST+ MI for the PMO. It should be

noted that the ground truth the prediction models learned between the regression and classification models is different, which explains the difference of the feature importance between Random Forest Classifier and Random Forest Regressor.

Furthermore, using Random Forest Regressor as the experimental model, the PIM prediction results are trialled with only selected important features according to Table 8.5. Results in Table 8.6 show that the impact of the feature quantity is inconsistent between the MI and the PMO. The PIM of the MI is generally more reliable with complete features according to its relatively low mean error and lowest standard deviation, while the PMO achieves the opposite results. Nevertheless, the high p-values (> 0.05) between the prediction relying on selected features and the prediction of all 12 features proves that even with the sole feature Troponin, the prediction is already reasonable.

Table 8.6: PIM prediction error with selected important features using Random Forest Regressor

Selected features	PIM	
	MI	PMO
Troponin	0.0634±0.0615 *	0.0128±0.0226 *
Troponin, LVEF	0.0585±0.0620 *	0.0122±0.02121 *
Troponin, LVEF, NTP, Age	0.0645± 0.0598 *	0.0145±0.0230 *
All 12 features	0.0587±0.0597	0.0149±0.0227

*: If the t-test between the selected features model and the all 12 features model obtains p-value>0.05: the difference is not significant.

DISCORDANT CASES

Even though the results are in general very encouraging, the proposal still provides inaccurate predictions on a few specific cases (Table 8.7). Incorrect predictions are divided into two categories: the predictions with an important difference in the quantification of the PIM and the wrong classifications. Several clinical reasons can explain these mismatches. Firstly, the automatic quantification of the PIM is sometimes underestimated when the troponin was relatively high and associated with ST+ MI but with normal ejection fraction (cases 22, 69 and 119). In these cases, discordance between the level of troponin and the LVEF can be observed, certainly due to an overestimation of the ejection fraction acquired at the acute phase. Indeed, a decrease of this value between the acute phase and a measurement carried out at an early moment after the revascularization can be produced, and then the lowest ejection fraction may increase the PIM value. Sometimes the proposal underestimates the PIM despite a relatively high troponin level and low ejection fraction (cases 1, 19 and 105) and it can be considered as a limit of the method. In particular, for the case 105, previous cardiovascular events can explain a high PIM. Sometimes, the results are incomprehensible and maybe the ground truth obtained from DE-MRI underestimates the PIM (as for the cases 7 and 110 where with a high

troponin level and altered ejection fraction, a PIM higher than 25% seems to be correct) or provides a value higher than expected, maybe due to a pre-existing necrosis (such as for the case 94). Counterexamples can also be found in the classification task, suggesting that the patient may suffer from another cardiovascular disease without reflection on DE-MRI and with preserved cardiac function (such as cases 16, 65, 68 or 117) or without preserved cardiac function (such as cases 34, 114 and 145, suggesting for these cases the presence of a chronic disease).

Table 8.7: Cases with incorrect prediction. Incorrect predictions are presented as the important quantification error and the wrong classification (false positive and false negative). The ground truth (GT) and the prediction values are the PIM. Classification results in the table are obtained by regression model and thresholding.

Case[2]	Feature values[1]												GT	Prediction
	Sex	Age	TB	OW	HT	DB	HD	ST+ MI	Troponin	KL	LVEF	NTp		
Cases with an important PIM Quantification error														
1	0	32	0	0	0	0	0	1	130	1	35	447	51.64%	23.25%
7	0	66	0	0	0	0	0	0	200	1	45	532	9.33%	25.64%
19	0	52	0	1	0	0	0	0	87	3	20	7139	48.04%	14.84%
22	0	53	0	0	0	0	0	1	170	1	60	43	42.56%	23.81%
69	1	45	0	0	0	0	0	1	120	1	55	649	39.06%	18.62%
94	0	61	0	0	1	1	0	1	3.9	1	46	5810	29.75%	12.94%
105	0	54	2	1	0	0	1	1	25	1	21	4153	46.41%	16.52%
110	0	49	0	1	0	0	0	1	200	1	45	29	7.54%	27.95%
119	0	66	2	1	0	0	0	1	73	1	70	159	31.79%	9.86%
Wrongly Classified Cases														
16	0	76	0	1	0	0	0	1	14	1	60	192	0.00%	9.32%
34	1	78	2	1	0	0	0	1	1.8	1	35	22577	0.00%	11.20%
65	0	57	1	1	0	0	0	1	19	1	60	71	0.00%	8.06%
68	0	39	0	1	0	0	0	1	9	1	60	23	0.00%	10.40%
114	1	54	0	1	1	0	1	0	1	1	45	68	0.00%	7.98%
117	0	53	1	1	0	0	0	1	83	1	60	94	0.00%	18.83%
145	1	66	2	1	0	0	0	0	2.5	2	45	6209	14.93%	1.38%

[1] TB: Tobacco; OW: Overweight; HT: Arterial hypertension; DB: Diabetes; HD: History of coronary artery disease; KL Killip max; NTp: NTProBNP. For Boolean features, 0 stands for negative (man for Sex) and 1 stands for positive (woman for Sex)

[2] Case number of the EMIDEC dataset. Complete data (patient features and MRI) can be accessed on the official website.

8.5/ DISCUSSION

Classification of MI from multivariate data Despite the classification accuracy is correct with sole textual data or MRI, the fusion of both types of the data obtains a more satisfying result. These results proves the fact that the DE-MRI is the gold standard of the MI assessment. The proposal can be further applied on other classification task with both image and textual diagnosis data, in which the image should be first encoded as the information of volume.

Quantitative prediction of MI from patient features The prediction results are very satisfying for both the quantification task and the classification task. For the quantification task, the ensemble method achieves the best predicted PIM, which shows only 0.056 of error for the PIM comparing with the DE-MRI ground truth. For the classification task, with the Random Forest regressor followed by a thresholding, 133 and 112 of the 150 cases are correctly predicted for the presence of infarction and PMO respectively, representing accuracies of 88.67% and 75.33%. As the data are collected from daily clinical practice and not specifically selected, the prediction accuracy is encouraging given the effect of data collection inaccuracies. Thanks to the statistical analyses, the machine learning brings a more comprehensive interpretation of multiple scalar indicators. It is not straightforward for a standard statistical model to construct such well-fitted non-linear model considering multiple features in a comprehensive manner, and meanwhile to analyse the importance of each feature.

For the automatic PIM quantification, all the regression models involved in the tests obtain satisfying results. Contrary to the findings in the literature studies, Boosting models do not significantly outperform the Random Forest, even though the Boosting models are computationally intensive. This fact reveals that the predictive model's complexity is not the key issue in improving the prediction performance. The training set volume does have an obvious impact on the prediction error. Indeed, once the volume of data reaches approximately 100 cases, the predictive accuracy of the model struggles to increase further as the volume of data increases. Comparing the regression results with few training samples (Figure 8.7) and the prediction error when the mean ground truth PIM is used to predict all cases (Table 8.3), the smaller prediction error in Figure 8.7 justifies the efficient use of very small sample data. Besides the data volume, the analysis of the discordant cases may reveal the biggest bottleneck in this method for improving performance: ambiguity that originates at the time of data collection. These inaccuracies both increase the bias of the model training and reduce the reliability of the label data during the performance evaluations.

To classify the presence of MI or PMO, the regression models trained with the PIM label followed by a thresholding slightly outperform the classification models trained with the Boolean label (Tab. 8.4). This fact demonstrates that even for the classification task, the inference results may benefit from richer information provided by the DE-MRI and annotations in the ground truth label. It also justifies the advantage of this work, namely a DE-MRI-guided physiological, clinical and paraclinical feature learning system.

When the regression results are taken for the classification task, a relatively high threshold value (PIM of 0.064) for the infarction is observed. This observation reveals that cases close to the critical hyperplane have higher predictive instability. For cases where the regression prediction is around this threshold value, additional complementary clinical

exams should be conducted to achieve greater certainty.

In terms of the importance of features for predictive models, the troponin is proven to be the dominant factor in the automatic prediction. This finding echoes the recommendations of The American College of Cardiology/American Heart Association (ACC/AHA) and the European Society of Cardiology (ESC) guidelines that cardiac troponin is the only biomarker for the diagnosis of acute MI due to its superior sensitivity and accuracy [57, 55, 81, 90]. Age, ST elevation on ECG, LEVF from cardiac TTE and NTProBNP level also demonstrated their obvious contributions. Between the infarction and the PMO, models rely more on the age and NTProBNP for the PMO prediction. This fact indicates a higher relationship between these factors and the presence of the PMO. The history of coronary artery disease only has a noticeable effect on linear models for the infarction prediction. This exception could be explained as the drawback of the linear models: linear models attempt to find a linear combination of the clinical features to distinguish the problem. However, the ideal critical hyperplane for the tasks is apparently not linear, which produces exceptional importance to some features. The observation on the prediction performance with selected features may suggest that the evaluation of the infarction and the available features are well linked, thus the PIM for MI increases when more features participate in the prediction. The opposite results on the PMO may indicate a weaker link between it and the available features.

Techniques employing artificial intelligence can become essential to improve cardiologists' work and performances in all aspects of cardiovascular diseases. In clinical practice, the prediction of the presence or not of a MI and the quantification of myocardial necrosis have a certain interest, first and foremost to confirm or invalidate a diagnosis, and therefore to provide information to guide treatment. In the case of an important extent of necrosis with a reduction in LVEF, treatment adapted to heart failure can immediately be introduced, a LifeVest wearable defibrillator can be proposed and the doses of diuretics can be better adjusted. Saving time is also important for physicians, and reducing delays with the help of artificial intelligence can allow more patients to benefit from high-performance exams and increase the global quality of care.

The major limitation of the proposal is the reliability of the data. Erroneous predictions may be produced because of the measurement inaccuracies, or the interference from other diseases like myocarditis, coronary spasms or the Tako Tsubo syndrome. In future work, to improve the proposal's performance with the given data, the confidence level can be estimated while making the predictions. Since most of the discordant cases can be explained as suspected feature acquisition error, the confidence level can be predicted by analyzing the correlation between features that have strong consistency. Then, in clinical practices, doctors should review the patient reports more thoroughly when the proposed confidence level is relatively low.

8.6/ CONCLUSIONS

The proposal that fuses DE-MRI and patient features achieves the first place in the classification contest of EMIDEC challenge. The classification accuracy in the challenge is 92%, which shows an important improvement over the classification based only on patient features (accuracy of 78%). This proposal proves the benefit of combining various data types for a more robust MI classification.

The proposal incorporates basic physiological, clinical and paraclinical features to provide a rapid and accurate physiological prediction of the severity of acute MI with the help of machine learning approaches. In clinical applications, an automatic assessment of the state of the myocardium and the PIM quantification can be obtained with just these minimal tests including the blood test, ECG and echocardiography. The MI classification and quantification results are promising, and the PMO prediction also shows the value of reference since the PMO evaluation was very difficult without cardiac imaging diagnosis. The study of the feature importance shows the highest correlation between the MI and the troponin, while age, ST elevation on ECG, LEVF from cardiac TTE and NTProBNP level are also proven to be associated with the MI. This proposal can thus potentially speed up the disease diagnosis of the acute MI in the emergency cardiology and can also indicate a rethinking of each patient feature's importance for the diagnosis of the disease.

9

CONCLUSIONS AND PERSPECTIVES

9.1/ CONCLUSIONS

This thesis concentrates on the automatic assessment of MI from multivariate medical diagnoses employing machine learning algorithms. The diagnosis relying on DE-MRI with deep learning methods intends to provide a more robust prediction (Chapters 5 and 6), while the diagnosis relying on patient features aims at achieving a quick and accurate analysis of the pathology with the minimum medical information in case of emergency (Section 8.3). Moreover, the classification of MI from both DE-MRI and patient features demonstrates the benefit of data fusion for a more reliable prediction (Section 8.2).

MI segmentation with 2D and 3D models The results of the DE-MRI segmentation task are globally satisfying compared to the human variance of the inter- and the intra-observer studies. The segmentation of myocardium and infarct in the majority of high-quality slices is comparable to that of cardiological specialists. The 2D segmentation model incorporating the proposed image normalization method, baseline network, SGD optimization, interpolated input images and proposed prior-based post-processing achieves comparable results (Dice metric of 0.843 on all-pathological slices, and 0.780 on slices of all patients for the infarct segmentation when the myocardium delineation is known, and 0.668 on all patients when the myocardium mask is not known) to the inter-observer study (0.845, 0.774 and 0.681, respectively). Compared with statistical models, the proposal significantly outperforms Gaussian Mixture Model on infarct segmentation task, showing an obvious advantage of deep learning-based segmentation approaches.

The 3D network with constrained receptive field obtains lower MI volume prediction difference relative to 2D models. Compared with 2D approaches, the volume prediction of scar tissues is significantly improved due to the reference from the adjacent slices, while the Dice metric is not improved due to the introduced bias on local segmentation.

However, the recognition of the PMO is still challenging regardless the proposal or other

state-of-the-art work, and human and computer assessments disagree even more on fuzzy images. Nevertheless, the present results of the segmentation appear to be ready for assisting physicians, providing a baseline for the contours of each tissue. The proposal will be soon integrated to the cardiac MRI delineation software QIR (CASIS Company, Quetigny, France) and hopefully will share the workload of manual annotation by physicians.

MI assessment from multivariate data The results of the proposals incorporating both DE-MRI and patient features for MI classification and quantification are also proved promising. The proposal that identifies the presence of MI from both DE-MRI and patient features won the EMIDEC challenge and a great improvement is obtained compared to the classification with sole patient features (classification accuracy of 92% compared to 78%). This proposal confirms the advantage of the fusion of multivariate data, and proposes a generalized fusion technique for medical imaging and clinical data.

The other proposal based on the patient features provides a subject and comprehensive MI assessment from all the available characteristics at the first stage of the pathology management. Besides the classification of MI, a quantification of the infarct volume can be also estimated from patient features, which was not possible without imaging diagnoses. This proposal achieves 0.0578 the error of PIM quantification and 88.67% the accuracy of classification for the MI prediction. The prediction of the PMO is also of value for reference since its evaluation was difficult without imaging diagnosis. Moreover, the study of the discordant cases has indicated that the prediction errors majorly result from the inaccuracy during the acquisition of clinical characteristics, while the comprehensive way of taking into account all the available features mitigates the error. For further clinical applications, the qualitative and quantitative predictions relying on patients features provide cardiologists with a subject index for the subsequent management before the gold standard DE-MRI is performed.

9.2/ PERSPECTIVES

PMO segmentation The PMO takes up only a small part of area in myocardium and its density is often similar to the normal myocardium tissue. According to this study and the results of the EMIDEC challenge, the improvement of the MI segmentation on DE-MRI originates from more adaptive configurations instead of blindly applying more complex networks. Therefore, one of the next work will focus on the optimization of the loss function to better respond to the class imbalance issue since the segmentation of PMO is still challenging.

Comprehensive study of 3D models Although the proposed 3D CRF model obtains an improvement on the scar volume prediction, the local segmentation result evaluated by Dice metric is not satisfying due to the bias from adjacent slices. Therefore, a more efficient way should be designed to control the fusion of the inter-slice information so that both the Dice metric and volume prediction is increased.

MI segmentation with alternative deep learning approaches During the preparation of the thesis, all segmentation work was based on fully convolutional U-Net networks. Meanwhile, the Generative Adversarial Network (GAN) [107] is showing promising results on medical image segmentation tasks. The GAN can be employed as the segmentation network [98, 143], or as the data augmentation generator [130]. Furthermore, the Transformer [101], initially proposed the attention mechanism for Natural Language Processing (NLP), has been expanded to the domain of computer vision. Recent works have integrated the Transformer into U-Net [149] for medical image segmentation tasks. Therefore, the GAN and the Transformer U-Net that replaces the vanilla U-Net should be further studied for the MI segmentation on DE-MRI.

MI segmentation with image denoising techniques The image denoising techniques [124] have been a popular computer vision topic. The objective is to remove the noise from the image and retain the essential features. Image denoising techniques have been studied by researchers on medical images to improve the visual evaluation by physicians. However, many related works evaluated the denoising approaches from the ratio of noise-to-signal, which limits their translational value. In our further work, state-of-the-art denoising approaches will be optimized and compared according to the automatic segmentation results, providing an end-to-end evaluation of the denoising techniques.

Expanded applications of the segmentation proposals Despite the proposed segmentation approaches intend for the assessment of MI on cardiac MRI, some of these techniques can be expanded to other tissues on similar medical images. For example, the stochastic gradient descent optimization and the image normalization method can be trialled on most medical images. Then the proposed variants of Dice loss and Generalized Dice loss can be employed if the quantity of instances of each tissue is significantly unequal. The prior-based post-processing can be also fine-tuned if the density distribution of the target tissue is consistent. And finally, the 3D models can be employed on tomographic images if the voxel size is inconsistent between the x-y plane and the z-axis. Therefore, new applications can be further evaluated using relevant proposed techniques on images like MRI and computerized tomographic imaging (CT), as well as microscopic images.

Confidence estimation of MI prediction from patient features For the assessment of patient features, in order to give physicians a more complete result for their reference, confidence estimates should be obtained in complement to the severity prediction. Since the study of discordant cases proves that an important amount of MI prediction error comes from the inaccuracy of the patient features acquisition, the confidence level can be obtained by examining the correlation between patient features. For clinical applications, physicians should pay more attention if the automatic prediction indicates a low confidence level.

PUBLICATIONS

JOURNAL PAPERS

- Alain Lalande, Zhihao Chen, Thomas Decourselle, Abdul Qayyum, Thibaut Pommier, Luc Lorgis, Ezequiel de la Rosa, Alexandre Cochet, Yves Cottin, Dominique Ginhac, Michel Salomon, Raphaël Couturier, and Fabrice Meriaudeau *"Emidec: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI"*. In **Data**, 5(4), 89 (2020). DOI: 10.3390/data5040089.

CONFERENCE PAPERS

- Zhihao Chen, Alain Lalande, Michel Salomon, Thomas Decourselle, Thibaut Pommier, Gilles Perrot, and Raphaël Couturier. *"Myocardial Infarction Segmentation From Late Gadolinium Enhancement MRI By Neural Networks and Prior Information"*. In **2020 International Joint Conference on Neural Networks (IJCNN)**, 1-8 (2020). DOI: 10.1109/IJCNN48605.2020.9206759.

- Jixi Shi, Zhihao Chen, and Raphaël Couturier. *"Classification of Pathological Cases of Myocardial Infarction Using Convolutional Neural Network and Random Forest"*. In **International Workshop on Statistical Atlases and Computational Models of the Heart (STACOM) 2020**, 406-413 (2020). DOI: 10.1007/978-3-030-68107-4_43.

- Ke Du, Jiaxing Yan, Zhou Hang, Zhihao Chen, and Lulu Wu. *"An LSTM Based Deep Learning Method for Airline Ticket Price Prediction"*. In **International Conference on Neural Information Processing (ICONIP) 2020**, 762-769 (2020). DOI: 10.1007/978-3-030-63823-8_86.

SUBMITTED PAPERS

- Alain Lalande, Zhihao Chen, Thibaut Pommier, Thomas Decourselle, Abdul Qayyum, Michel Salomon, Dominique Ginhac, Youssef Skandarani, Arnaud Boucher, Khawla Brahim, Marleen de Bruijne, Robin Camarasa, Teresa M. Correia, Xue Feng, Kibrom B. Girum, Anja Hennemuth, Markus Huellebrand, Raabid

Hussain, Matthias Ivantsits, Jun Ma, Craig Meyer, Rishabh Sharma, Jixi Shi, Nikolaos V. Tsekos, Marta Varela, Xiyue Wang, Sen Yang, Hannu Zhang, Yichi Zhang, Yuncheng Zhou, Xiahai Zhuang, Raphaël Couturier, and Fabrice Meriaudeau. *"Deep learning methods for automatic evaluation of delayed enhancement-MRI. The results of the EMIDEC challenge."*. In **Medical Image Analysis**. Submitted in July 2021.

- Zhihao Chen, Jixi Shi, Thibaut Pommier, Michel Salomon, Thomas Decourselle, Alain Lalande, and Raphaël Couturier. *"Prediction of Myocardial Infarction from Patient Features with Machine Learning"*. In **Frontiers in Cardiovascular Medicine**. Submitted in July 2021.

- Zhihao Chen, Alain Lalande, Michel Salomon, Thomas Decourselle, Thibaut Pommier, Abdul Qayyum, Jixi Shi, Gilles Perrot, and Raphaël Couturier. *"Automatic Deep Learning-based Myocardial Infarction Segmentation From Delayed Enhancement MRI"*. In **Computerized Medical Imaging and Graphics**. Submitted in January 2021.

BIBLIOGRAPHY

[1] KILLIP III, T., AND KIMBALL, J. T. **Treatment of myocardial infarction in a coronary care unit: a two year experience with 250 patients**. *The American journal of cardiology 20*, 4 (1967), 457–464.

[2] HOERL, A. E., AND KENNARD, R. W. **Ridge regression: Biased estimation for nonorthogonal problems**. *Technometrics 12*, 1 (1970), 55–67.

[3] DEMPSTER, A. P., LAIRD, N. M., AND RUBIN, D. B. **Maximum likelihood from incomplete data via the em algorithm**. *Journal of the Royal Statistical Society: Series B (Methodological) 39*, 1 (1977), 1–22.

[4] GOLDMAN, L., WEINBERG, M., WEISBERG, M., OLSHEN, R., COOK, E. F., SARGENT, R. K., LAMAS, G. A., DENNIS, C., WILSON, C., DECKELBAUM, L., FINEBERG, H., AND STIRATELLI, R. **A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain**. *N Engl J Med 307*, 10 (1982), 588–596.

[5] BOCKSTEIN, I. M. **Color equalization method and its application to color image processing**. *JOSA A 3*, 5 (1986), 735–737.

[6] MARTIN BLAND, J., AND ALTMAN, D. **Statistical methods for assessing agreement between two methods of clinical measurement**. *Lancet 327*, 8476 (1986), 307–310.

[7] RUMELHART, D. E., HINTON, G. E., AND WILLIAMS, R. J. **Learning representations by back-propagating errors**. *nature 323*, 6088 (1986), 533–536.

[8] SANTOSA, F., AND SYMES, W. W. **Linear inversion of band-limited reflection seismograms**. *SIAM J Sci Statist Comput 7*, 4 (1986), 1307–1330.

[9] QUINLAN, J. **Simplifying decision trees**. *Int J Man Mach Stud 27*, 3 (1987), 221 – 234.

[10] GOLDMAN, L., COOK, E. F., BRAND, D. A., LEE, T. H., ROUAN, G. W., WEISBERG, M. C., ACAMPORA, D., STASIULEWICZ, C., WALSHON, J., TERRANOVA, G., GOTTLIEB, L., KOBERNICK, M., GOLDSTEIN-WAYNE, B., COPEN, D., DALEY, K.,

BRANDT, A. A., JONES, D., MELLORS, J., AND JAKUBOWSKI, R. **A computer protocol to predict myocardial infarction in emergency department patients with chest pain**. *N Engl J Med 318*, 13 (1988), 797–803.

[11] FIX, E., AND HODGES, J. L. **Discriminatory analysis. nonparametric discrimination: Consistency properties**. *International Statistical Review / Revue Internationale de Statistique 57*, 3 (1989), 238–247.

[12] GUTGESELL, H. P., AND REMBOLD, C. M. **Growth of the human heart relative to body surface area**. *The American Journal of Cardiology 65*, 9 (1990), 662–668.

[13] HO, T. K. **Random decision forests**. In *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1* (1995), p. 278.

[14] BREIMAN, L. **Bias, variance, and arcing classifiers**. Tech. rep., 1996.

[15] HOULT, D. I., AND BHAKAR, B. **Nmr signal reception: Virtual photons and coherent spontaneous emission**. *Concepts in Magnetic Resonance: An Educational Journal 9*, 5 (1997), 277–297.

[16] HOCHREITER, S. **The vanishing gradient problem during learning recurrent neural nets and problem solutions**. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 6*, 02 (1998), 107–116.

[17] KIM, R. J., FIENO, D. S., PARRISH, T. B., HARRIS, K., CHEN, E.-L., SIMONETTI, O., BUNDY, J., FINN, J. P., KLOCKE, F. J., AND JUDD, R. M. **Relationship of mri delayed contrast enhancement to irreversible injury, infarct age, and contractile function**. *Circulation 100*, 19 (1999), 1992–2002.

[18] OPITZ, D., AND MACLIN, R. **Popular ensemble methods: An empirical study**. *J Artif Intell Res 11* (1999), 169–198.

[19] ALPERT, J., THYGESEN, K., ANTMAN, E., AND BASSAND, J. **for the joint european society of cardiology/american college of cardiology committee. myocardial infarction redefined–a consensus document of the joint european society of cardiology/american college of cardiology committee for the redefinition of myocardial infarction**. *J Am Coll Cardiol 36*, 3 (2000), 959–69.

[20] CARR, J. C., SIMONETTI, O., BUNDY, J., LI, D., PERELES, S., AND FINN, J. P. **Cine mr angiography of the heart with segmented true fast imaging with steady-state precession**. *Radiology 219*, 3 (2001), 828–834.

[21] FRIEDMAN, J. H. **Greedy function approximation: A gradient boosting machine.** *Ann Statist 29*, 5 (2001), 1189–1232.

[22] SCHOLKOPF, B. **The kernel trick for distances**. *Adv Neural Inf Process Syst* (2001), 301–307.

[23] WEIK, M. H. **full-width at half-maximum**. Springer US, Boston, MA, 2001, pp. 661–661.

[24] KELLMAN, P., ARAI, A. E., MCVEIGH, E. R., AND ALETRAS, A. H. **Phase-sensitive inversion recovery for detecting myocardial infarction using gadolinium-delayed hyperenhancement**. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 47*, 2 (2002), 372–383.

[25] MILDENBERGER, P., EICHELBERG, M., AND MARTIN, E. **Introduction to the dicom standard**. *European radiology 12*, 4 (2002), 920–927.

[26] PENG, C.-Y. J., LEE, K. L., AND INGERSOLL, G. M. **An introduction to logistic regression analysis and reporting**. *The journal of educational research 96*, 1 (2002), 3–14.

[27] ATISHA, D., BHALLA, M. A., MORRISON, L. K., FELICIO, L., CLOPTON, P., GARDETTO, N., KAZANEGRA, R., CHIU, A., AND MAISEL, A. S. **A prospective study in search of an optimal b-natriuretic peptide level to screen patients for cardiac dysfunction**. *American heart journal 148*, 3 (2004), 518–523.

[28] LARSON, A. C., WHITE, R. D., LAUB, G., MCVEIGH, E. R., LI, D., AND SIMONETTI, O. P. **Self-gated cardiac cine mri**. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 51*, 1 (2004), 93–102.

[29] SELVANAYAGAM, J. B., KARDOS, A., FRANCIS, J. M., WIESMANN, F., PETERSEN, S. E., TAGGART, D. P., AND NEUBAUER, S. **Value of delayed-enhancement cardiovascular magnetic resonance imaging in predicting myocardial viability after surgical revascularization**. *Circulation 110*, 12 (2004), 1535–1541.

[30] HINTON, G. E., OSINDERO, S., AND TEH, Y.-W. **A fast learning algorithm for deep belief nets**. *Neural comput 18*, 7 (2006), 1527–1554.

[31] HINTON, G. E., AND SALAKHUTDINOV, R. R. **Reducing the dimensionality of data with neural networks**. *science 313*, 5786 (2006), 504–507.

[32] MACEIRA, A., PRASAD, S., KHAN, M., AND PENNELL, D. **Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance**. *Journal of Cardiovascular Magnetic Resonance 8*, 3 (2006), 417–426.

[33] TORIO, C. M., AND MOORE, B. J. **National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013: Statistical Brief #204**. Agency for Healthcare Research and Quality (US), Rockville (MD), 2006.

[34] VOGEL-CLAUSSEN, J., ROCHITTE, C. E., WU, K. C., KAMEL, I. R., FOO, T. K., LIMA, J. A., AND BLUEMKE, D. A. **Delayed enhancement mr imaging: utility in myocardial assessment**. *Radiographics 26*, 3 (2006), 795–810.

[35] SCHINKEL, A. F., POLDERMANS, D., ELHENDY, A., AND BAX, J. J. **Assessment of myocardial viability in patients with heart failure**. *Journal of Nuclear Medicine 48*, 7 (2007), 1135–1146.

[36] ZOU, K. H., O'MALLEY, A. J., AND MAURI, L. **Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models**. *Circulation 115*, 5 (2007), 654–657.

[37] DINOV, I. D. **Expectation maximization and mixture modeling tutorial**. 2008.

[38] ROCKAFELLAR, R. T., AND WETS, R. J.-B. **Variational analysis**, vol. 317. Springer Science & Business Media, 2009.

[39] BEKKERS, S. C., YAZDANI, S. K., VIRMANI, R., AND WALTENBERGER, J. **Microvascular obstruction: Underlying pathophysiology and clinical diagnosis**. *Journal of the American College of Cardiology 55*, 16 (2010), 1649–1660.

[40] NACIF, M. S., MELLO, R. A., LACERDA JUNIOR, O. O., SIBLEY, C. T., MACHADO, R. A., AND MARCHIORI, E. **Double-chambered left ventricle in an adult: diagnosis by cmri**. *Clinics 65*, 12 (2010), 1393–1395.

[41] PAN, S. J., AND YANG, Q. **A survey on transfer learning**. *IEEE Transactions on Knowledge and Data Engineering 22*, 10 (2010), 1345–1359.

[42] VINCENT, P., LAROCHELLE, H., LAJOIE, I., BENGIO, Y., MANZAGOL, P.-A., AND BOTTOU, L. **Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.** *Journal of machine learning research 11*, 12 (2010).

[43] FLETT, A. S., HASLETON, J., COOK, C., HAUSENLOY, D., QUARTA, G., ARITI, C., MUTHURANGU, V., AND MOON, J. C. **Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance**. *JACC: cardiovascular imaging 4*, 2 (2011), 150–156.

[44] GLOROT, X., BORDES, A., AND BENGIO, Y. **Deep sparse rectifier neural networks**. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (Fort Lauderdale, FL, USA, 11–13 Apr 2011), vol. 15 of *Proceedings of Machine Learning Research*, PMLR, pp. 315–323.

[45] NG, A., AND OTHERS. **Sparse autoencoder**. *CS294A Lecture notes 72*, 2011 (2011), 1–19.

[46] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. **Scikit-learn: Machine learning in Python**. *J Mach Learn Res 12* (2011), 2825–2830.

[47] TANINDI, A., AND CEMRI, M. **Troponin elevation in conditions other than acute coronary syndromes**. *Vascular health and risk management 7* (2011), 597.

[48] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. **Imagenet classification with deep convolutional neural networks**. *Advances in neural information processing systems 25* (2012), 1097–1105.

[49] LALANDE, A., VALINDRIA, V., ANGUE, M., VIGNON, N., COCHET, A., AND BRUNOTTE, F. **Automatic evaluation of the peri-infarct area of myocardial infarction from delayed enhancement mri**. In *European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), Lisbonne, Portugal* (10 2012), p. 123.

[50] RONALD E. WALPOLE, RAYMOND H. MYERS, S. L. M., AND YE, K. E. In *Probability & statistics for engineers & scientists*, 9th ed. Pearson, 2012.

[51] STEG, P. G., JAMES, S. K., ATAR, D., BADANO, L. P., BLOMSTROM-LUNDQVIST, C., BORGER, M. A., DI MARIO, C., DICKSTEIN, K., DUCROCQ, G., FERNANDEZ-AVILES, F., GERSHLICK, A. H., GIANNUZZI, P., HALVORSEN, S., HUBER, K., JUNI, P., KASTRATI, A., KNUUTI, J., LENZEN, M. J., MAHAFFEY, K. W., VALGIMIGLI, M., VAN 'T HOF, A., WIDIMSKY, P., ZAHGER, D., BAX, J. J., BAUMGARTNER, H., CECONI, C., DEAN, V., DEATON, C., FAGARD, R., FUNCK-BRENTANO, C., HASDAI, D., HOES, A., KIRCHHOF, P., KNUUTI, J., KOLH, P., MCDONAGH, T., MOULIN, C., POPESCU, B. A., REINER, Z., SECHTEM, U., SIRNES, P. A., TENDERA, M., TORBICKI, A., VAHANIAN, A., WINDECKER, S., HASDAI, D., ASTIN, F., ASTROM-OLSSON, K., BUDAJ, A., CLEMMENSEN, P., COLLET, J. P., FOX, K. A., FUAT, A., GUSTIENE, O., HAMM, C. W., KALA, P., LANCELLOTTI, P., MAGGIONI, A. P., MERKELY, B., NEUMANN, F. J., PIEPOLI, M. F., VAN DE WERF, F., VERHEUGT, F., AND WALLENTIN, L. **ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation**. *Eur Heart J 33*, 20 (2012), 2569–2619.

[52] FARABET, C., COUPRIE, C., NAJMAN, L., AND LECUN, Y. **Learning hierarchical features for scene labeling**. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 8 (2013), 1915–1929.

[53] GEIGER, A., LENZ, P., STILLER, C., AND URTASUN, R. **Vision meets robotics: The kitti dataset**. *International Journal of Robotics Research (IJRR)* (2013).

[54] MAKHZANI, A., AND FREY, B. **K-sparse autoencoders**. *arXiv preprint arXiv:1312.5663* (2013).

[55] O'GARA, P. T., KUSHNER, F. G., ASCHEIM, D. D., CASEY, D. E., CHUNG, M. K., DE LEMOS, J. A., ETTINGER, S. M., FANG, J. C., FESMIRE, F. M., FRANKLIN, B. A., GRANGER, C. B., KRUMHOLZ, H. M., LINDERBAUM, J. A., MORROW, D. A., NEWBY, L. K., ORNATO, J. P., OU, N., RADFORD, M. J., TAMIS-HOLLAND, J. E., TOMMASO, C. L., TRACY, C. M., WOO, Y. J., AND ZHAO, D. X. **2013 accf/aha guideline for the management of st-elevation myocardial infarction**. *Circulation 127*, 4 (2013), e362–e425.

[56] SCHULZ-MENGER, J., BLUEMKE, D., BREMERICH, J., FLAMM, S., FOGEL, M., FRIEDRICH, M., KIM, R., VON KNOBELSDORFF, F., KRAMER, C., PENNELL, D., PLEIN, S., AND NAGEL, E. **Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for cardiovascular magnetic resonance (scmr) board of trustees task force on standardized post processing**. *J Cardiovasc Magn Reson 15* (2013), 35.

[57] AMSTERDAM, E. A., WENGER, N. K., BRINDIS, R. G., CASEY, D. E., GANIATS, T. G., HOLMES, D. R., JAFFE, A. S., JNEID, H., KELLY, R. F., KONTOS, M. C., LEVINE, G. N., LIEBSON, P. R., MUKHERJEE, D., PETERSON, E. D., SABATINE, M. S., SMALLING, R. W., AND ZIEMAN, S. J. **2014 aha/acc guideline for the management of patients with non–st-elevation acute coronary syndromes**. *Circulation 130*, 25 (2014), e344–e426.

[58] HAMIRANI, Y. S., WONG, A., KRAMER, C. M., AND SALERNO, M. **Effect of microvascular obstruction and intramyocardial hemorrhage by cmr on lv remodeling and outcomes after myocardial infarction: A systematic review and meta-analysis**. *JACC: Cardiovasc Imaging 7*, 9 (2014), 940 – 952.

[59] HE, K., ZHANG, X., REN, S., AND SUN, J. **Spatial pyramid pooling in deep convolutional networks for visual recognition**. In *ECCV 2014* (2014), D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., pp. 346–361.

[60] KUMAR, V., ABBAS, A. K., FAUSTO, N., AND ASTER, J. C. **Robbins and Cotran pathologic basis of disease, professional edition e-book**. Elsevier health sciences, 2014.

[61] LIN, T.-Y., MAIRE, M., BELONGIE, S., HAYS, J., PERONA, P., RAMANAN, D., DOLLÁR, P., AND ZITNICK, C. L. **Microsoft coco: Common objects in context**.

In *Computer Vision – ECCV 2014* (Cham, 2014), D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Springer International Publishing, pp. 740–755.

[62] MIKAMI, Y., KOLMAN, L., JONCAS, S. X., STIRRAT, J., SCHOLL, D., RAJCHL, M., LYDELL, C. P., WEEKS, S. G., HOWARTH, A. G., AND WHITE, J. A. **Accuracy and reproducibility of semi-automated late gadolinium enhancement quantification techniques in patients with hypertrophic cardiomyopathy**. *Journal of Cardiovascular Magnetic Resonance 16*, 1 (2014), 1–9.

[63] PREIM, B., AND BOTHA, C. **Chapter e16 - visual exploration and analysis of perfusion data**. In *Visual Computing for Medicine (Second Edition)*, B. Preim and C. Botha, Eds., second edition ed. Morgan Kaufmann, Boston, 2014, pp. e19–e60.

[64] DÉGANO, I. R., SALOMAA, V., VERONESI, G., FERRIÉRES, J., KIRCHBERGER, I., LAKS, T., HAVULINNA, A. S., RUIDAVETS, J.-B., FERRARIO, M. M., MEISINGER, C., ELOSUA, R., AND MARRUGAT, J. **Twenty-five-year trends in myocardial infarction attack and mortality rates, and case-fatality, in six european populations**. *Heart 101*, 17 (2015), 1413–1421.

[65] IOFFE, S., AND SZEGEDY, C. **Batch normalization: Accelerating deep network training by reducing internal covariate shift**. In *International conference on machine learning* (2015), PMLR, pp. 448–456.

[66] KINGMA, D. P., AND BA, J. L. **Adam: A method for stochastic gradient descent**. In *ICLR: International Conference on Learning Representations* (2015), pp. 1–15.

[67] LONG, J., SHELHAMER, E., AND DARRELL, T. **Fully convolutional networks for semantic segmentation**. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.

[68] MELILLO, P., IZZO, R., ORRICO, A., SCALA, P., ATTANASIO, M., MIRRA, M., DE LUCA, N., AND PECCHIA, L. **Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis**. *PLOS ONE 10*, 3 (2015), 1–14.

[69] RONNEBERGER, O., FISCHER, P., AND BROX, T. **U-net: Convolutional networks for biomedical image segmentation**. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (10 2015), pp. 234–241.

[70] SIMONYAN, K., AND ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition**. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015), Y. Bengio and Y. LeCun, Eds.

[71] SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V., AND RABINOVICH, A. **Going deeper with convolutions**. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1–9.

[72] SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V., AND RABINOVICH, A. **Going deeper with convolutions**. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1–9.

[73] AVENDI, M., KHERADVAR, A., AND JAFARKHANI, H. **A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri**. *Medical Image Analysis 30* (2016), 108 – 119.

[74] CHEN, T., AND GUESTRIN, C. **Xgboost: A scalable tree boosting system**. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), p. 785–794.

[75] CLEVERT, D., UNTERTHINER, T., AND HOCHREITER, S. **Fast and accurate deep network learning by exponential linear units (elus)**. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings* (2016), Y. Bengio and Y. LeCun, Eds.

[76] GOODFELLOW, I., BENGIO, Y., AND COURVILLE, A. **Deep Learning**. The MIT Press, 2016.

[77] HE, K., ZHANG, X., REN, S., AND SUN, J. **Deep residual learning for image recognition**. pp. 770–778.

[78] KARIM, R., BHAGIRATH, P., CLAUS, P., HOUSDEN, R. J., CHEN, Z., KARIMAGHALOO, Z., SOHN, H.-M., RODRÍGUEZ, L. L., VERA, S., ALBÀ, X., HENNEMUTH, A., PEITGEN, H.-O., ARBEL, T., BALLESTER, M. A. G., FRANGI, A. F., GÖTTE, M., RAZAVI, R., SCHAEFFTER, T., AND RHODE, K. **Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement mr images**. *Medical Image Analysis 30* (2016), 95 – 107.

[79] MILLETARI, F., NAVAB, N., AND AHMADI, S. **V-net: Fully convolutional neural networks for volumetric medical image segmentation**. In *2016 Fourth International Conference on 3D Vision (3DV)* (2016), pp. 565–571.

[80] QUAN, T. M., HILDEBRAND, D. G. C., AND JEONG, W. **Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics** .

[81] ROFFI, M., PATRONO, C., COLLET, J.-P., MUELLER, C., VALGIMIGLI, M., ANDREOTTI, F., BAX, J. J., BORGER, M. A., BROTONS, C., CHEW, D. P., GENCER, B., HASENFUSS, G., KJELDSEN, K., LANCELLOTTI, P., LANDMESSER, U., MEHILLI, J., MUKHERJEE, D., STOREY, R. F., WINDECKER, S., AND GROUP, E. S. D. **2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC)**. *European Heart Journal 37*, 3 (01 2016), 267–315.

[82] SARKISIAN, L., SAABY, L., POULSEN, T. S., GERKE, O., JANGAARD, N., HOSBOND, S., DIEDERICHSEN, A. C., THYGESEN, K., AND MICKLEY, H. **Clinical characteristics and outcomes of patients with myocardial infarction, myocardial injury, and nonelevated troponins**. *The American Journal of Medicine 129*, 4 (2016), 446.e5–446.e21.

[83] ULYANOV, D., VEDALDI, A., AND LEMPITSKY, V. **Instance normalization: The missing ingredient for fast stylization**. *arXiv preprint arXiv:1607.08022* (2016).

[84] VOS, T., ALLEN, C., ARORA, M., BARBER, R. M., BHUTTA, Z. A., BROWN, A., CARTER, A., CASEY, D. C., CHARLSON, F. J., CHEN, A. Z., AND OTHERS. **Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015**. *The lancet 388*, 10053 (2016), 1545–1602.

[85] ZHUANG, X. **Multivariate mixture model for cardiac segmentation from multi-sequence mri**. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016), Springer, pp. 581–588.

[86] CHEN, L.-C., PAPANDREOU, G., KOKKINOS, I., MURPHY, K., AND YUILLE, A. L. **Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs**. *IEEE transactions on pattern analysis and machine intelligence 40*, 4 (2017), 834–848.

[87] DARAEI, A., AND HAMIDI, H. **An efficient predictive model for myocardial infarction using cost-sensitive j48 model**. *Iran J Public Health 46* (2017), 682–692.

[88] FOURURE, D., EMONET, R., FROMONT, E., MUSELET, D., TREMEAU, A., AND WOLF, C. **Residual Conv-Deconv Grid Network for Semantic Segmentation**. In *BMVC 2017* (Londre, United Kingdom, Sept. 2017).

[89] HUANG, G., LIU, Z., VAN DER MAATEN, L., AND WEINBERGER, K. Q. **Densely connected convolutional networks**. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4700–4708.

[90] IBANEZ, B., JAMES, S., AGEWALL, S., ANTUNES, M. J., BUCCIARELLI-DUCCI, C., BUENO, H., CAFORIO, A. L. P., CREA, F., GOUDEVENOS, J. A., HALVORSEN, S., HINDRICKS, G., KASTRATI, A., LENZEN, M. J., PRESCOTT, E., ROFFI, M., VALGIMIGLI, M., VARENHORST, C., VRANCKX, P., WIDIMSKÝ, P., AND GROUP, E. S. D. **2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC)**. *European Heart Journal 39*, 2 (08 2017), 119–177.

[91] KE, G., MENG, Q., FINLEY, T., WANG, T., CHEN, W., MA, W., YE, Q., AND LIU, T.-Y. **Lightgbm: A highly efficient gradient boosting decision tree**. In *Advances in Neural Information Processing Systems* (2017), vol. 30, pp. 3146–3154.

[92] LIN, T.-Y., DOLLÁR, P., GIRSHICK, R., HE, K., HARIHARAN, B., AND BELONGIE, S. **Feature pyramid networks for object detection**. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 936–944.

[93] LIN, T.-Y., GOYAL, P., GIRSHICK, R., HE, K., AND DOLLAR, P. **Focal loss for dense object detection**. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (Oct 2017).

[94] MEHTA, R., AND SIVASWAMY, J. **M-net: A convolutional neural network for deep brain structure segmentation**. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (2017), pp. 437–440.

[95] POHLEN, T., HERMANS, A., MATHIAS, M., AND LEIBE, B. **Full-resolution residual networks for semantic segmentation in street scenes**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 4151–4160.

[96] RAMACHANDRAN, P., ZOPH, B., AND LE, Q. V. **Searching for activation functions**. *CoRR abs/1710.05941* (2017).

[97] SEIJO-PARDO, B., PORTO-DÍAZ, I., BOLÓN-CANEDO, V., AND ALONSO-BETANZOS, A. **Ensemble feature selection: Homogeneous and heterogeneous approaches**. *Knowledge-Based Systems 118* (2017), 124 – 139.

[98] SOULY, N., SPAMPINATO, C., AND SHAH, M. **Semi supervised semantic segmentation using generative adversarial network**. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 5688–5696.

[99] SUDRE, C. H., LI, W., VERCAUTEREN, T., OURSELIN, S., AND CARDOSO, M. J. **Generalised dice overlap as a deep learning loss function for highly unbal-**

anced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support - Third International Workshop, DLMIA* (2017), pp. 240–248.

[100] SZEGEDY, C., IOFFE, S., VANHOUCKE, V., AND ALEMI, A. A. **Inception-v4, inception-resnet and the impact of residual connections on learning**. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (2017), AAAI'17, p. 4278–4284.

[101] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł., AND POLOSUKHIN, I. **Attention is all you need**. In *Advances in neural information processing systems* (2017), pp. 5998–6008.

[102] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L. U., AND POLOSUKHIN, I. **Attention is all you need**. In *Advances in Neural Information Processing Systems* (2017), I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30, Curran Associates, Inc.

[103] XIE, S., GIRSHICK, R., DOLLAR, P., TU, Z., AND HE, K. **Aggregated residual transformations for deep neural networks**. pp. 5987–5995.

[104] ALAOUR, B., LIEW, F., AND KAIER, T. E. **Cardiac Troponin - diagnostic problems and impact on cardiovascular disease**. *Ann Med 50*, 8 (2018), 655–665.

[105] BERNARD, O., LALANDE, A., ZOTTI, C., CERVENANSKY, F., YANG, X., HENG, P.-A., CETIN, I., LEKADIR, K., CAMARA, O., GONZALEZ BALLESTER, M. A., SANROMA, G., NAPEL, S., PETERSEN, S., TZIRITAS, G., GRINIAS, E., KHENED, M., KOLLERATHU, V. A., KRISHNAMURTHI, G., ROHÉ, M.-M., PENNEC, X., SERMESANT, M., ISENSEE, F., JÄGER, P., MAIER-HEIN, K. H., FULL, P. M., WOLF, I., ENGELHARDT, S., BAUMGARTNER, C. F., KOCH, L. M., WOLTERINK, J. M., IŠGUM, I., JANG, Y., HONG, Y., PATRAVALI, J., JAIN, S., HUMBERT, O., AND JODOIN, P.-M. **Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?** *IEEE Transactions on Medical Imaging 37*, 11 (2018), 2514–2525.

[106] CETIN, I., SANROMA, G., PETERSEN, S. E., NAPEL, S., CAMARA, O., BALLESTER, M.-A. G., AND LEKADIR, K. **A radiomics approach to computer-aided diagnosis with cardiac cine-mri**. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges* (Cham, 2018), M. Pop, M. Sermesant, P.-M. Jodoin, A. Lalande, X. Zhuang, G. Yang, A. Young, and O. Bernard, Eds., Springer International Publishing, pp. 82–90.

[107] CRESWELL, A., WHITE, T., DUMOULIN, V., ARULKUMARAN, K., SENGUPTA, B., AND BHARATH, A. A. **Generative adversarial networks: An overview**. *IEEE Signal Processing Magazine 35*, 1 (2018), 53–65.

[108] DESAIX, P., BETTS, J. G., JOHNSON, E., JOHNSON, J. E., KOROL, O., KRUSE, D. H., POE, B., WISE, J. A., YOUNG, K. A., AND OTHERS. **Anatomy & Physiology: OpenStax**. 2018.

[109] FAHMY, A., RAUSCH, J., NEISIUS, U., CHAN, R., MARON, M., APPELBAUM, E., MENZE, B., AND NEZAFAT, R. **Automated cardiac mr scar quantification in hypertrophic cardiomyopathy using deep convolutional neural networks**. *JACC: Cardiovascular Imaging 11*, 12 (08 2018), 1917–1918.

[110] HU, J., SHEN, L., AND SUN, G. **Squeeze-and-excitation networks**. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 7132–7141.

[111] LOUKADAKIS, M., CANO, J., AND O'BOYLE, M. **Accelerating deep neural networks on low power heterogeneous architectures**. In *Eleventh International Workshop on Programmability and Architectures for Heterogeneous Multicores (MULTIPROG-2018)* (2018).

[112] MORENO-BAREA, F. J., STRAZZERA, F., JEREZ, J. M., URDA, D., AND FRANCO, L. **Forward noise adjustment scheme for data augmentation**. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (2018), pp. 728–734.

[113] RAMACHANDRAN, P., ZOPH, B., AND LE, Q. V. **Searching for activation functions**. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings* (2018), OpenReview.net.

[114] RICE, J. **Medical Terminology for Health Care Professionals**, 3 ed. Pearson, 2018, ch. 9.

[115] SANDLER, M., HOWARD, A., ZHU, M., ZHMOGINOV, A., AND CHEN, L.-C. **Mobilenetv2: Inverted residuals and linear bottlenecks**. pp. 4510–4520.

[116] THYGESEN, K., ALPERT, J. S., JAFFE, A. S., CHAITMAN, B. R., BAX, J. J., MORROW, D. A., AND WHITE, H. D. **Fourth universal definition of myocardial infarction (2018)**. *Journal of the American College of Cardiology 72*, 18 (2018), 2231–2264.

[117] WU, Y., AND HE, K. **Group normalization**. In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 3–19.

[118] ZABIHOLLAHY, F., WHITE, J. A., AND UKWATTA, E. **Myocardial scar segmentation from magnetic resonance images using convolutional neural network**. In *Medical Imaging 2018: Computer-Aided Diagnosis* (2018), N. Petrick and K. Mori, Eds., vol. 10575, International Society for Optics and Photonics, SPIE, pp. 663 – 670.

[119] ZHANG, H., CISSÉ, M., DAUPHIN, Y. N., AND LOPEZ-PAZ, D. **mixup: Beyond empirical risk minimization**. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings* (2018).

[120] ZHUANG, X. **Multivariate mixture model for myocardial segmentation combining multi-source images**. *IEEE transactions on pattern analysis and machine intelligence 41*, 12 (2018), 2933–2946.

[121] CHEPLYGINA, V., DE BRUIJNE, M., AND PLUIM, J. P. **Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis**. *Medical Image Analysis 54* (2019), 280–296.

[122] CUBUK, E. D., ZOPH, B., MANÉ, D., VASUDEVAN, V., AND LE, Q. V. **Autoaugment: Learning augmentation strategies from data**. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 113–123.

[123] DE LA ROSA, E., SIDIBÉ, D., DECOURSELLE, T., LECLERCQ, T., COCHET, A., AND LALANDE, A. **Myocardial infarction quantification from late gadolinium enhancement mri using top-hat transforms and neural networks**. unpublished.

[124] FAN, L., ZHANG, F., FAN, H., AND ZHANG, C. **Brief review of image denoising techniques**. *Visual Computing for Industry, Biomedicine, and Art 2*, 1 (2019), 1–12.

[125] GU, Z., CHENG, J., FU, H., ZHOU, K., HAO, H., ZHAO, Y., ZHANG, T., GAO, S., AND LIU, J. **Ce-net: Context encoder network for 2d medical image segmentation**. *IEEE Transactions on Medical Imaging 38*, 10 (Oct 2019), 2281–2292.

[126] LI, X., WANG, W., HU, X., AND YANG, J. **Selective kernel networks**. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019* (2019), Computer Vision Foundation / IEEE, pp. 510–519.

[127] MAZIAR ZAFARI, A. **Myocardial infarction**. https://emedicine.medscape.com/article/155919-overview, May 2019. Accessed: 2021-06-03.

[128] MENCHÓN-LARA, R.-M., SIMMROSS-WATTENBERG, F., CASASECA-DE-LA HIGUERA, P., MARTÍN-FERNÁNDEZ, M., AND ALBEROLA-LÓPEZ, C. **Reconstruction techniques for cardiac cine mri**. *Insights into imaging 10*, 1 (2019), 1–16.

[129] MOCCIA, S., BANALI, R., MARTINI, C., MUSCOGIURI, G., PONTONE, G., PEPI, M., AND CAIANI, E. **Development and testing of a deep learning-based strategy for scar segmentation on cmr-lge images**. *Magnetic Resonance Materials in Physics, Biology and Medicine 32* (2019), 187–195.

[130] SANDFORT, V., YAN, K., PICKHARDT, P. J., AND SUMMERS, R. M. **Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks**. *Scientific reports 9*, 1 (2019), 1–9.

[131] SCHLEMPER, J., OKTAY, O., SCHAAP, M., HEINRICH, M., KAINZ, B., GLOCKER, B., AND RUECKERT, D. **Attention gated networks: Learning to leverage salient regions in medical images**. *Medical Image Analysis 53* (2019), 197 – 207.

[132] SHORTEN, C., AND KHOSHGOFTAAR, T. M. **A survey on image data augmentation for deep learning**. *Journal of Big Data 6*, 1 (2019), 1–48.

[133] TAGHANAKI, S., ZHENG, Y., ZHOU, S. K., GEORGESCU, B., SHARMA, P., XU, D., COMANICIU, D., AND HAMARNEH, G. **Combo loss: Handling input and output imbalance in multi-organ segmentation**. *Computerized Medical Imaging and Graphics 75* (05 2019), 24–33.

[134] THAN, M., PICKERING, J., SANDOVAL, Y., SHAH, A., TSANAS, A., APPLE, F., BLANKENBERG, S., CULLEN, L., NEUMANN, J., TWERENBOLD, R., WESTERMANN, D., BESHIRI, A., MILLS, N., AND CHAPMAN, A. **Machine learning to predict the likelihood of acute myocardial infarction**. *Circulation 140* (2019), 899–909.

[135] YUE, Q., LUO, X., YE, Q., XU, L., AND ZHUANG, X. **Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors**. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2019, Proceedings, Part II* (2019), vol. 11765 of *Lecture Notes in Computer Science*, pp. 559–567.

[136] ZABIHOLLAHY, F., WHITE, J. A., AND UKWATTA, E. **Convolutional neural network-based approach for segmentation of left ventricle myocardial scar from 3d late gadolinium enhancement mr images**. *Medical physics 46*, 4 (2019), 1740–1751.

[137] AL-ZAITI, S., BESOMI, L., BOUZID, Z., FARAMAND, Z., FRISCH, S., MARTIN-GILL, C., GREGG, R., SABA, S., CALLAWAY, C., AND SEJDIC, E. **Machine learning-**

based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nat Commun 11* (2020).

[138] CHEN, Z., LALANDE, A., SALOMON, M., DECOURSELLE, T., POMMIER, T., PERROT, G., AND COUTURIER, R. **Myocardial infarction segmentation from late gadolinium enhancement mri by neural networks and prior information**. In *2020 International Joint Conference on Neural Networks (IJCNN)* (2020), IEEE, pp. 1–8.

[139] GÁRATE-ESCAMILA, A. K., HAJJAM EL HASSANI, A., AND ANDRÈS, E. **Classification models for heart disease prediction using feature selection and pca**. *Inform Med Unlocked 19* (2020), 100330.

[140] ITO, R., YAMASHITA, J., CHIKAMORI, T., KONDO, S., MITSUHASHI, Y., IWATA, H., SAJI, M., ASANO, T., WAKABAYASHI, K., YAHAGI, K., SHINKE, T., MASE, T., ABE, K., MIYACHI, H., HIGUCHI, S., KISHI, M., TANAKA, H., YAMASAKI, M., MIYAUCHI, K., YAMAMOTO, T., NAGAO, K., AND TAKAYAMA, M. **Clinical Differences of Recent Myocardial Infarction Compared With Acute Myocardial Infarction - Insights From the Tokyo CCU Network Multicenter Registry**. *Circ J 84*, 9 (2020), 1511–1518.

[141] KARIMI, D., AND SALCUDEAN, S. E. **Reducing the hausdorff distance in medical image segmentation with convolutional neural networks**. *IEEE Transactions on Medical Imaging 39*, 2 (2020), 499–513.

[142] LALANDE, A., CHEN, Z., DECOURSELLE, T., QAYYUM, A., POMMIER, T., LORGIS, L., DE LA ROSA, E., COCHET, A., COTTIN, Y., GINHAC, D., SALOMON, M., COUTURIER, R., AND MERIAUDEAU, F. **Emidec: A database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac mri**. *Data 5*, 4 (2020).

[143] LEI, B., XIA, Z., JIANG, F., JIANG, X., GE, Z., XU, Y., QIN, J., CHEN, S., WANG, T., AND WANG, S. **Skin lesion segmentation via generative adversarial networks with dual discriminators**. *Medical Image Analysis 64* (2020), 101716.

[144] ROMERO-FARINA, G., CANDELL-RIERA, J., AGUADÉ-BRUIX, S., AND DORADO, D. **A novel clinical risk prediction model for myocardial infarction, coronary revascularization, and cardiac death according to clinical, exercise, and gated spect variables (vh-rs)**. *Eur Heart J: Cardiovasc Imaging 21* (2020), 210–221.

[145] SINGH, D., AND SINGH, B. **Investigating the impact of data normalization on classification performance**. *Appl Soft Comput 97* (2020), 105524.

[146] BRAHIM, K., QAYYUM, A., LALANDE, A., BOUCHER, A., SAKLY, A., AND MERIAUDEAU, F. **Efficient 3d deep learning for myocardial diseases segmentation**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 359–368.

[147] CAMARASA, R., FAURE, A., CROZIER, T., BOS, D., AND DE BRUIJNE, M. **Uncertainty-based segmentation of myocardial infarction areas on cardiac mr images**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 385–391.

[148] CASELLA, G., AND BERGER, R. L. **Statistical inference**. Cengage Learning, 2021.

[149] CHEN, J., LU, Y., YU, Q., LUO, X., ADELI, E., WANG, Y., LU, L., YUILLE, A. L., AND ZHOU, Y. **Transunet: Transformers make strong encoders for medical image segmentation**. *arXiv preprint arXiv:2102.04306* (2021).

[150] FENG, X., KRAMER, C. M., SALERNO, M., AND MEYER, C. H. **Automatic scar segmentation from de-mri using 2d dilated unet with rotation-based augmentation**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 400–405.

[151] GIRUM, K. B., SKANDARANI, Y., HUSSAIN, R., GRAYELI, A. B., CRÉHANGE, G., AND LALANDE, A. **Automatic myocardial infarction evaluation from delayed-enhancement cardiac mri using deep convolutional networks**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 378–384.

[152] HUELLEBRAND, M., IVANTSITS, M., ZHANG, H., KOHLMANN, P., KUHNIGK, J.-M., KUEHNE, T., SCHÖNBERG, S., AND HENNEMUTH, A. **Comparison of a hybrid mixture model and a cnn for the segmentation of myocardial pathologies in delayed enhancement mri**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 319–327.

[153] ISENSEE, F., JAEGER, P., KOHL, S., PETERSEN, J., AND MAIER-HEIN, K. **nnu-net: a self-configuring method for deep learning-based biomedical image segmentation**. *Nature Methods 18* (02 2021), 1–9.

[154] IVANTSITS, M., HUELLEBRAND, M., KELLE, S., SCHÖNBERG, S. O., KUEHNE, T., AND HENNEMUTH, A. **Deep-learning-based myocardial pathology detection**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 369–377.

[155] LOURENÇO, A., KERFOOT, E., GRIGORESCU, I., SCANNELL, C. M., VARELA, M., AND CORREIA, T. M. **Automatic myocardial disease prediction from delayed-enhancement cardiac mri and clinical information**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 334–341.

[156] MA, J., CHEN, J., NG, M., HUANG, R., LI, Y., LI, C., YANG, X., AND MARTEL, A. **Loss odyssey in medical image segmentation**. *Medical Image Analysis 71* (03 2021), 102035.

[157] SHARMA, R., EICK, C. F., AND TSEKOS, N. V. **Sm2n2: A stacked architecture for multimodal data and its application to myocardial infarction detection**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 342–350.

[158] YANG, S., AND WANG, X. **A hybrid network for automatic myocardial infarction segmentation in delayed enhancement-mri**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 351–358.

[159] ZHANG, Y. **Cascaded convolutional neural network for automatic myocardial infarction segmentation from delayed-enhancement cardiac mri**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021), E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 328–333.

[160] ZHOU, Y., ZHANG, K., LUO, X., WANG, S., AND ZHUANG, X. **Anatomy prior based u-net for pathology segmentation with attention**. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges* (2021),

E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young, Eds., pp. 392–399.

[161] JONES J, G. F. **Mri sequences (overview)**. https://radiopaedia.org/articles/mri-sequences-overview. Accessed: 2021-06-04.

[162] SERRA, J. **Courses on mathematical morphology**. https://https://people.cmm.minesparis.psl.eu/users/serra/cours/index.htm. Accessed: 2021-07-01.

List of Figures

List of Tables

Document generated with LaTeX and:
the LaTeX style for PhD Thesis created by S. Galland — http://www.multiagent.fr/ThesisStyle
the tex-upmethodology package suite — http://www.arakhne.org/tex-upmethodology/

école doctorale **sciences pour l'ingénieur et microtechniques**

Title: Deep learning for automatic detection and quantification of the disease areas of the myocardium from DE-MRI after myocardial infarction

Keywords: Myocardial infarction, DE-MRI, Image segmentation, Automatic assessment, Deep learning, Machine learning, Classification, Regression

Abstract:

Myocardial Infarction (MI) has become one of the most common cardiovascular diseases. The MI occurs when the blood flow decreases or stops in a part of the heart so that the myocardium gets damaged due to ischaemia. The MI can lead to persistent microvascular obstruction (PMO) where the myocardium dysfunctions because of the no-reflow phenomenon after the reperfusion therapy of an occluded coronary artery in a proportion of patients. To assess the MI and the extended PMO, Delayed Enhancement MRI (DE-MRI) has become the gold standard. However, to precisely quantify the pathology, physicians have to manually draw the delineations of different myocardial tissues on the short-axis MRI slices then the volume of normal and pathological tissues will be calculated to obtain the severity evaluation. The manual annotation procedures are highly time-consuming and subjective, which shows a great potential of an automatic evaluation approach for the MI assessment. Current automatic delineation methods on DE-MRI are mostly probabilistic approaches such as the Mixture Models etc. while more and more applications with deep learning achieve state-of-the-art performance compared to the probabilistic approaches. Therefore, in this thesis, the different aspects of deep learning-based approaches have been investigated for the automatic MI evaluation from DE-MRI.
This dissertation mainly focuses on the automatic MI assessment using the MRI and the clinical and physiological features. With the MRI, three major aspects consisting in the pre-processing, the deep learning models and the post-processing have been investigated or proposed. The pre-processing aims at preparing more consistent and clear inputs for deep learnings models using image normalizations, cropping etc. To design well-adaptive deep learning models, many parts (building blocks, loss functions, segmentation models, etc.) have been proposed or compared. Moreover, the post-processing methods including morphological treatments and prior information-based filters have been applied to the coarse segmentations to eliminate false positive contours. Furthermore, merging the MRI and textual patient features, a machine learning-based approach was proposed to obtain a quantitative estimate of MI incorporating only the clinical and pathological features for emergency cardiological assessment.

Titre : Deep learning for automatic detection and quantification of the disease areas of the myocardium from DE-MRI after myocardial infarction

Mots-clés : Imagerie par résonance magnétique), Caractéristiques cliniques, Détection automatique, Segmentation de l'image, Infarctus du myocarde

Résumé :

L'infarctus du myocarde (IM) est devenu l'une des maladies cardiovasculaires les plus courantes. L'infarctus se produit lorsque le flux sanguin diminue ou s'arrête dans une partie du cœur, ce qui peut endommager le myocarde en raison de l'ischémie. L'infarctus peut entraîner une obstruction microvasculaire persistante (PMO) même après le traitement de reperfusion d'une artère coronaire occluse, ce qui touche une proportion importante de patients. Pour évaluer l'infarctus et l'obstruction microvasculaire persistante, l'IRM avec rehaussement tardif (IRM-RT) est devenue la référence. Cependant, pour quantifier précisément la pathologie, les médecins doivent d'abord tracer manuellement les délimitations des différents tissus myocardiques sur les tranches d'IRM à petit axe ; ensuite, le volume des tissus normaux et pathologiques sera calculé pour évaluer la gravité de l'attaque. Les procédures d'annotation manuelle sont très chronophages et subjectives, ce qui renforce l'intérêt potentiel d'une approche d'évaluation automatique pour l'évaluation de l'IM. Les méthodes actuelles de contourage automatique en IRM-RT sont principalement des approches probabilistes telles que les modèles de mélange de distributions Gaussiennes, alors que de plus en plus d'applications d'apprentissage profond atteignent des performances de pointe par rapport aux approches probabilistes. Par conséquent, dans cette thèse, les différents aspects des approches basées sur l'apprentissage profond ont été étudiés pour l'évaluation automatique du MI à partir de l'IRM-RT.
Pour développer le pipeline d'évaluation de l'IM basé sur l'apprentissage profond, trois aspects majeurs, à savoir le prétraitement, les modèles d'apprentissage profond et le post-traitement, ont été étudiés ou proposés. Le prétraitement vise à préparer des entrées plus cohérentes et plus claires pour les modèles d'apprentissage profond en utilisant des normalisations d'images, des recadrages, etc. Pour concevoir des modèles d'apprentissage profond bien adaptés, de nombreux éléments (blocs de construction, fonctions de perte, modèles de segmentation, etc.) ont été proposés ou comparés. En outre, les méthodes de post-traitement, y compris les traitements morphologiques et les filtres basés sur l'information préalable, ont été appliquées aux segmentations grossières afin d'éliminer les contours faussement positifs. En plus de l'évaluation automatique par segmentation, en fusionnant l'IRM et les caractéristiques physiologiques du patient, une approche basée sur l'apprentissage automatique a été proposée pour obtenir des estimations quantitatives de l'IM en utilisant uniquement les caractéristiques cliniques et pathologiques pour une évaluation cardiologique rapide dans le service des urgences cardiaques.

UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besançon, France