Cardio-Respiratory Motion Compensation for Radiation Dose Reduction in X-Ray Guided Cardiac Interventions

by

Fariba AZIZMOHAMMADI

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, NOVEMBER 22, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Luc Duong, Thesis supervisor
Department of Software and Information Technology Engineering, École de technologie supérieure

Ms. Rita Noumeir, Chair, Board of Examiners
Department of Electrical Engineering, École de technologie supérieure

Mr. Stéphane Coulombe, Member of the Jury
Department of Software and Information Technology Engineering, École de technologie supérieure

Mr. Sylvain Deschênes, External Examiner
Department of Radiology, Radiation Oncology and Nuclear Medicine, Université de Montréal

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON OCTOBER 7, 2022

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

The achievements presented in this thesis have been possible thanks to many people. Professor Luc Duong has mentored me with patience and perseverance; he gave me the confidence to thrive in my research training. His support has been key to my personal and professional development. Under his supervision, professor Luc Duong provided me with the opportunity to come to Canada to pursue my doctoral studies.

This thesis was inspired and developed thanks to the insights provided by Dr. Joaquim Miró and Dr. Iñaki Navarro Castellanos. They shared their medical knowledge and clinical experience in understanding heart complexities. Thanks to them, I had the opportunity to attend several cardiac interventions for pediatric patients. In addition, Francisco, and Victor, cardiology residents, supported me with data collection. Nathalie also helped me with administrative duties and coordination with the cathlab. I would like to thank Denise, Julie, and Hélène from Sainte-Justine's cathlab.

Thank you, Professor Stéphane Coulombe, Professor Rita Noumeir, and Dr. Sylvain Deschênes, for accepting the invitation to be part of the board of examiners of this thesis.

I would also like to show gratitude to Dr. Paul Segars and Ehsan Samei from Duke University Medical Center and the Center for Virtual Imaging Trials (CVIT) for their collaboration in my research and for providing me access to the XCAT simulator.

I want to give special thanks to Samuel Docquier, Olivier Rufiange technicians, and Patrice Dion, analyst from the département de génie logiciel et des TI for their tremendous support and help during my study.

As a LIVE/LiNCS lab member, I would like to thank all lab members for their friendships and support. I have special thanks to Rémi as my best friend who has always been a great support during my Ph.D. We have been through many things together and discovered new sports and challenges aside from our scientific discussions that fascinated our research topics. You have helped me through difficult periods and given me helpful advice and ideas for my research. I also would like to thank Gerardo, Houda, Neda, and Ahmed, from who I have benefited from their experiences and friendships. The fact that we came from different countries made our interaction an enriching experience. To all of you, thank you very much.

My motivations for pursuing doctoral studies in Canada included enriching my academic experience and improving my language skills. But perhaps the primary motivation was broadening my view of the world. Nevertheless, it implied departing and staying away from my relatives, to whom I ask forgiveness for being absent for too long. I am deeply grateful for the constant support I received from my mother, Azam, my sister Parisa, my brother-in-law Mehdi and my niece Parmis. I also have had great emotional support from my family here in Canada, including my brother Kourosh, who has been the greatest support for me in the most difficult periods during my study, my aunt and uncle Fati and Mostafa, my cousins Mohammad, Pouyan, and Ali. I hope to be able to give back all the love that I received from all of you. Thank you very much.

Last but not least, thanks to the financial support provided by NSERC Discovery grant and Impulsion program. The GPU used in this work was provided by the NVIDIA Applied Research Accelerator Program.

Compensation des mouvements cardio-respiratoires et réduction de la dose de rayonnement pour les interventions cardiaques guidées par rayons X

Fariba AZIZMOHAMMADI

RÉSUMÉ

Les maladies des artères coronaires (MAC) affectent un grand nombre d'enfants chaque année et constituent le type de malformations congénitales le plus courant dans le monde. Les MAC sont dus à des anomalies du cœur avant la naissance. Il est impératif d'améliorer les techniques de traitement pour qu'elles soient le moins invasives possible. Les procédures interventionnelles, guidées par imagerie par rayons X, ont connu un essor considérable au cours des deux dernières décennies. Elles constituent, à ce jour, l'un des traitements les plus populaires des MAC.

Le guidage de la navigation pendant les interventions cardiaques, telles que l'angioplastie et principalement la pose d'endoprothèses, est généralement effectué sous fluoroscopie par rayons X. Les patients atteints de MAC sont exposés à des quantités importantes de rayonnements ionisants lors des procédures de diagnostic et de traitement. Ces dernières années, le nombre d'interventions cardiaques pédiatriques complexes et de longue durée a considérablement augmenté.

Pendant une intervention cardiaque, plusieurs organes, y compris les artères, sont en mouvement en raison des battements du cœur, des mouvements respiratoires et parfois des mouvements du patient. Ces mouvements dégradent l'acquisition d'images et rendent le processus de navigation plus difficile. L'objectif principal de cette recherche était de trouver des techniques moins invasives à appliquer pour les interventions cardiaques pédiatriques. Nous avons pour cela proposé la réduction de la dose de radiation reçue par le patient et le personnel, ainsi que de la quantité d'agents de contraste utilisés, en mettant en place des techniques de prédiction de mouvement. Dans le cadre de cette recherche, nous avons développé et validé nos approches en utilisant des ensembles des données issues de l'hôpital Sainte-Justine mêlées à des simulations cardio-respiratoires. Nous avons d'abord étudié des séquences cardio-respiratoires simulées de rayons X générées en utilisant le simulateur XCAT. Nous avons simulé un total de 56 patients différents (32 hommes et 24 femmes) totalisant 112 séquences (2 séquences par patient, montrant l'artère coronaire gauche ou droit). Toutes les séquences générées avaient une longueur de 75 images et une fréquence de15 images par seconde.

Notre ensemble de données médicales est composé de 52 patients. Chaque patient présente un nombre différent de séquences, avec des longueurs variables. Il y a un nombre total de 340 séquences, avec une longueur minimale et maximale de 15 et 70 images respectivement. Toutes les données ont été acquises à une fréquence de 15 images par seconde.

Dans le premier objectif, une approche basée sur l'apprentissage automatique a été proposée pour prédire les images d'angiographie à rayons X afin de réduire l'exposition aux radiations des patients pédiatriques et du personnel pendant les interventions cardiaques tout en préservant la qualité de l'image.

Dans le deuxième objectif, nous nous sommes d'abord concentrés sur l'extraction des caractéristiques 2D de mouvement à partir des séquences de rayons X, puis sur la construction d'un modèle prédictif de mouvement pour la navigation des interventions. Notre approche est capable de prédire le signal cyclique du mouvement cardio-respiratoire bruité par les mouvements soudains causés par les patients ou autres irrégularités. Cette approche a été développée et validée sur les ensembles de données simulées et celles des patients.

Pour le troisième objectif, en utilisant l'ensemble de données simulées, nous avons étudié les différents patrons de mouvement entre les hommes et les femmes. La recherche offre une solution générale qui ne nécessite pas de modalité d'imagerie supplémentaire tout en fournissant une prédiction, une estimation et une compensation précises.

Les méthodologies pour atteindre ces objectifs sont basées sur des algorithmes d'apprentissage profond (réseaux neuronaux récurrents) en extrayant d'abord les caractéristiques de mouvement des images et en suivant ces caractéristiques.

Nous pensons que les approches basées sur l'apprentissage peuvent ouvrir la voie à une meilleure évaluation des maladies cardiovasculaires en proposant une réduction de la dose de radiation pour les patients et le personnel. Dans cette thèse nous appliquons des méthodes d'apprentissage profond pour les interventions cardiaques minimalement invasives.

Mots-clés: Angiographie à rayons X, mouvement cardio-respiratoire, réduction de la dose, prédiction du mouvement

Cardio-Respiratory Motion Compensation for Radiation Dose Reduction in X-Ray Guided Cardiac Interventions

Fariba AZIZMOHAMMADI

ABSTRACT

Cardiac diseases affect a large population and especially children, every year. Congenital Heart Disease (CHD) is the most common type of birth malformation worldwide. CHD is caused by abnormalities in heart structure before birth. It is imperative to advance the treatment techniques to be as less invasive as possible. One popular treatment for CHD is X-ray image-guided interventional procedures that have gained popularity over the past two decades.

Navigation guidance during cardiac interventions, such as balloon angioplasty and stent placement, is generally performed under X-ray fluoroscopy. Patients with CHD are exposed to substantial amounts of ionizing radiation from diagnostic and treatment procedures. In recent years, the number of complex, long-duration pediatric cardiac interventions has risen significantly.

During the cardiac intervention, several organs, including the arteries, are moving, given the heart beating, respiratory movement, and sometimes the patient's movements. These movements degrade image acquisition and make the navigation process more difficult.

This research's main goal was to develop less invasive techniques to apply for pediatric cardiac interventions. We pursued this goal by minimizing the radiation dose the patient and staff received and compensating for the induced motions by estimating and predicting the targets' (arteries) movements. Moreover, while the targets' movements are tracked in the images, the need to inject the contrast agent to visualize the vessels will be reduced.

In the context of this research, we developed and validated our approaches using both simulated and patient X-ray angiography datasets from Sainte-Justine Hospital. Simulated X-ray sequences generated from realistic XCAT computational phantoms with cardio-respiratory motion were first investigated. The simulated motion included the beating heart and respiratory motion. We simulated 56 different patients (32 male and 24 female) and 112 sequences (2 sequences per patient, showing either the left or the right coronary artery). All the generated sequences had a length of 75 frames and were generated at 15 frames per second (fps). The patient X-ray angiography dataset comprises 52 different patients with contrasted coronary arteries. Each patient presents a different number of sequences with varying lengths. There is a total number of 340 sequences, with a minimum and maximum length of 15 and 70 frames, respectively. All the data were acquired at 15 fps.

In the first objective, a generative learning-based approach was proposed to predict X-ray angiography frames to reduce the amount of radiation exposure to pediatric patients and the staff during cardiac interventions while preserving the image quality.

In the second objective, we focused on extracting 2D motion features from the X-ray sequences first and then building up a "predict-ahead" motion model for navigating the interventions. Our model-free cardio-respiratory motion estimation approach can predict cyclic cardio-respiratory motion signals artifacted by sudden motions caused by the patients or some irregularities. This approach was developed and validated with both simulated and patient datasets.

For the third objective, using the simulated dataset and based on our experiment on the second objective, we investigated the different motion patterns between males and females. The research will offer a general solution that does not require additional imaging modality while providing an accurate motion prediction and estimation. The methodologies to achieve these objectives are based on deep learning algorithms by first extracting motion features from the images and tracking these features.

We believe that learning-based approaches can pave the road for better assessment of cardiovascular motion and radiation dose reduction for the patients and staff. Thus, in this thesis, we have applied deep learning methods to facilitate the desired less-invasive cardiac interventions.

Keywords: X-ray angiography, cardio-respiratory motion, dose reduction, motion prediction

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AC	Attenuation Correction
ALARA	As Low As Reasonably Achievable
ARIMA	Autoregressive Integrated Moving Average
AT	Affine Transformation
CAD	Coronary Arteries Disease
CFF	Critical Flicker Frequency
CHD	Congenital Heart Disease
CNN	Convolutional Neural Network
CPD	Coherent Point Drift
CT	Computed Tomography
DAP	Dose Area Product
ECG	Electrocardiogram
ED	End diastole
EKF	Extended Kalman Filter
EMBC	Engineering in Medicine and Biology Society
ES	End Systole
FH	Feet–Head
fps	frame per second
FR	Frame Rate

GANs	Generative adversarial networks
GMM	Gaussian Mixture Model
IHD	Ischaemic Heart Disease
LCA	Left Coronary Artery
LCM	Local Circular Motion
LFH	Last Fluoroscopy Hold
LLE	Locally Linear Embedding
LR	Linear Regression
LSTM	Long Short-Term Memory
KAP	Kerma At the reference Point
KDE	Kernel Density Estimation
MAE	Mean Absolute Error
MC-LDDM	Multichannel Large Deformation Diffeomorphic Metric Mapping
MICs	Minimally Invasive Surgeries
MRI	Magnetic Resonance Imaging
MTS	Magnetic Tracking System
MULIN	Multi-step Linear Method
NLMS	Normalized Least Mean Square
NN	Neural Network
PCI	Percutaneous Coronary Interventions

PR	Pulse Rate
PSNR	Peak Signal-to-Noise Ratio
RCA	Right Coronary Artery
RDR	Required Dose Rate
RNN	Recurrent Neural Networks
SGMA	Simultaneous Group-wise Manifold Alignment
SSIM	Structural SIMilarity
SVM	Support Vector Machine
SVR	Support Vector Regression
TP	Transformation Parameter
WMA	Wavelet-based Multi-scale Auto-regression

INTRODUCTION

Coronary Arteries Disease (CAD) is a leading cause of mortality in Western countries. It is the most common cause of death in the adult population. CAD happens when the arteries are narrowed or occluded by plaque build-up or deposited cholesterol on arteries' inner wall (Sun, 2010). CAD mainly involves the progressive narrowing or occlusion of the arteries that nourish the heart, which means an interruption or insufficiency of the blood circulation in the heart muscle, thus limiting the supply of nutrients, and oxygen. It might cause chest pain or, in more severe cases, can lead to heart failure.

In the pediatric population, Congenital Heart Disease (CHD) affects 1% of the population and is the most common type of birth malformation worldwide. In sever cases, interventions are required for treatment. In recent years, the number of complex, long-duration pediatric cardiac interventions has risen significantly (Liu *et al.*, 2019a). Image-guided interventions for children are usually more challenging and have technical or procedural complexities due to small size, high heart rates, and motion artifacts from cardiac pulsation, respiration, and the patients themselves (Goo, 2015).

Minimally invasive surgeries (MICs) are image-guided interventional procedures that have emerged over the past two decades to treat CADs and CHDs. MIC techniques are famous for two reasons. First, the success rate of MICs is comparable to traditional interventions. Second, the risk of complications is lower in the case of MICs. Percutaneous coronary angioplasty is a frequent MIC that aims to treat stenosis. X-ray angiography is one the most used imaging modality in cardiac interventions providing adequate spatial and temporal image resolution.

0.1 Problem statement and motivations

We proposed a novel radiation dose reduction approach in our first contribution by predicting dynamic X-ray angiography sequences using a generative model. A video frame prediction model was developed to predict new X-ray angiography frames. We introduced a new loss function to predict the temporal and spatial information of the arteries in angiography sequences. Then, a predictive RNN-based motion model was trained to estimate the motion and content of single and/or multiple future frame(s) based on previously acquired frames in an end-to-end system. A detailed description of the proposed method and results was published in Medical Physics (Azizmohammadi *et al.*, 2022)

During the cardiac intervention, several organs are moving, given the heartbeat, respiratory movement, and sometimes the patient's movements (Shechter, Ozturk, Resar & McVeigh, 2004). These movements cause artifacts in image acquisition and make navigation and guidance more challenging. The cardiac motion, being distinct for each patient and hardly perceptible on the angiographies, generates additional complexity for the cardiologists. With a precise recovery of a patient's cardiac motion, it becomes possible to offer sophisticated tools to help with minimally invasive surgeries. Moreover, respiratory movement degrades the detection and quantification capabilities of imaging modalities. The mismatches between some imaging techniques due to respiratory motion results in additional Attenuation Correction (AC) artifacts and inaccurate localization. Given the increasing demand for accurate image-guided treatments in CADs and CHDs, a robust method to account for reducing the effects of respiratory, cardiac, and even unexpected motions during cardiology interventions is needed.

0.2 Research objectives and contributions

The main goal of this research was to investigate learning-based predictive cardio-respiratory motion models. Predictive models can facilitate less invasive cardiac intervention processes. To pursue this goal, we proposed new techniques to compensate for the cardio-respiratory motion artifacts and minimize the radiation used in X-ray angiography sequences. In this research, we defined three objectives, and the described final results were our contributions.

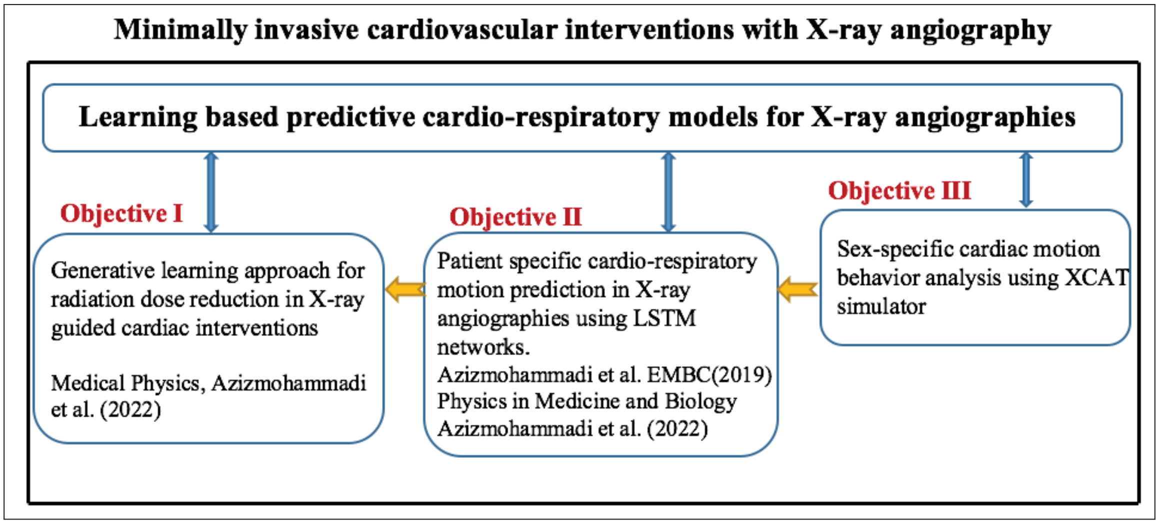

Figure 0.1 Research objectives to provide less invasive navigation process for cardiac interventions

In our first contribution, we proposed a novel radiation dose reduction approach by predicting dynamic X-ray angiography sequences using a generative model. A video frame prediction model was developed to predict new X-ray angiography frames. We introduced a new loss function to predict the temporal and spatial information of the arteries in angiography sequences. Then, a predictive RNN-based motion model was trained to estimate the motion and content of single and/or multiple future frame(s) based on previously acquired frames in an end-to-end system. A detailed description of the proposed method and results was published in Medical Physics (Azizmohammadi *et al.*, 2022).

The second contribution targeted cardio-respiratory motion compensation. We presented a patient-specific model-free approach for cardiorespiratory motion prediction from X-ray angiography time series based on Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN). Cardiorespiratory motion prediction was defined as a problem of estimating the future displacement of the coronary arteries in the next image frames in an X-ray angiography sequence. The displacements of the arteries were represented as a sequence of 2D affine transformation matrices allowing successive registration in a sequence. The new displacement parameters from a sequence of transformation matrices are predicted using an LSTM model. This motion prediction model was fully adaptive to any unanticipated changes or sudden movement. The need to inject contrast iodine agents can also be reduced since the geometry of the arteries can be predicted even if they are not contrasted. This contribution has been developed in two steps. The preliminary results were published in IEEE Engineering in Medicine and Biology Society (EMBC) (Azizmohammadi, Martin, Miro & Duong, 2019). The model was improved and validated using a larger dataset. This work was submitted to Physics in Medicine and Biology.

The third contribution was to analyze and compare the cardiac motion behavior between men and women using simulated X-ray angiographies. To develop a general sex-specific and non-patient-specific cardiac motion pattern that can be used for diagnostic or treatment planning by cardiologists, we investigated the differences between cardiac motions for males and females.

0.3 Thesis outline

This document is organized as follows. Chapter 1 presents the clinical description and treatments of patients with CHD. In addition, it provides a general description of motion compensation prediction methods and radiation dose reduction approaches in minimally invasive interventions. Chapter 2 describes our first contribution, a generative learning approach for radiation dose reduction in X-ray-guided cardiac interventions. Our second contribution, cardio-respiratory motion tracking in pediatric X-ray angiography sequences using a simple LSTM network, is presented in Chapter 3. Chapter 4 presents our third contribution, sex-specific cardiac motion analysis for treatment and planning the cardiac interventions.

CHAPTER 1

LITERATURE REVIEW

1.1 Minimally invasive treatments for CADs and CHDs

Image-guided interventions are currently minimally invasive treatments using different imaging modalities. X-ray angiography is one of the most popular modalities for cardiac interventions since it provides adequate spatial and temporal image resolution for cardiologists. This modality is used in catheterization, balloon angioplasty, atherectomy, laser treatment, stent placement, and other procedures. Fluoroscopy and fluorography are the two main fluoroscopically guided intervention modes in X-ray imaging. In fluoroscopy mode, the X-ray images are generated instantaneously and continuously to observe moving objects by capturing the motion. The images in this mode are used to navigate the medical devices to specific locations within the patient in real-time. Fluorography, the mode requires a higher radiation exposure to generate and record high-resolution images for interpretation after the termination of the exposure (Dauer, 2011).

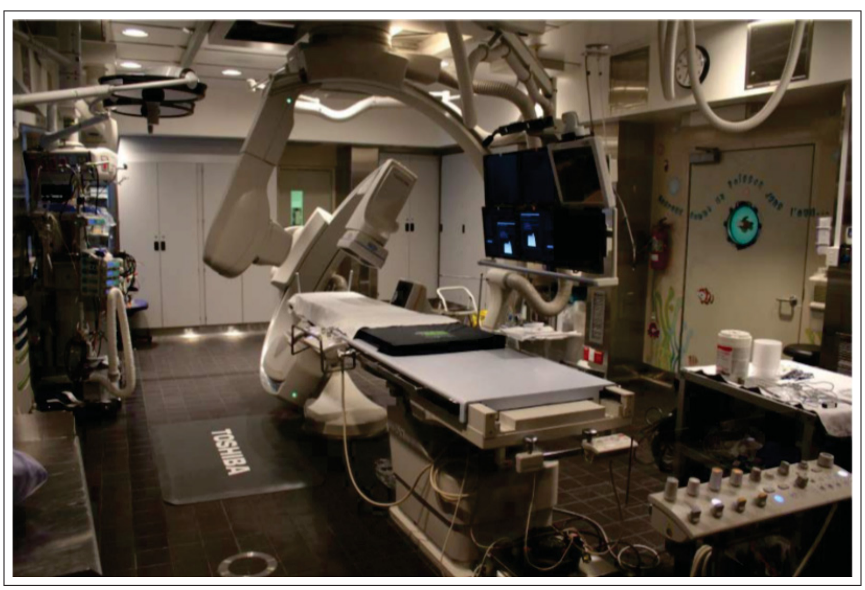

Figure 1.1 X-ray angiography system at CHU Sainte-Justine

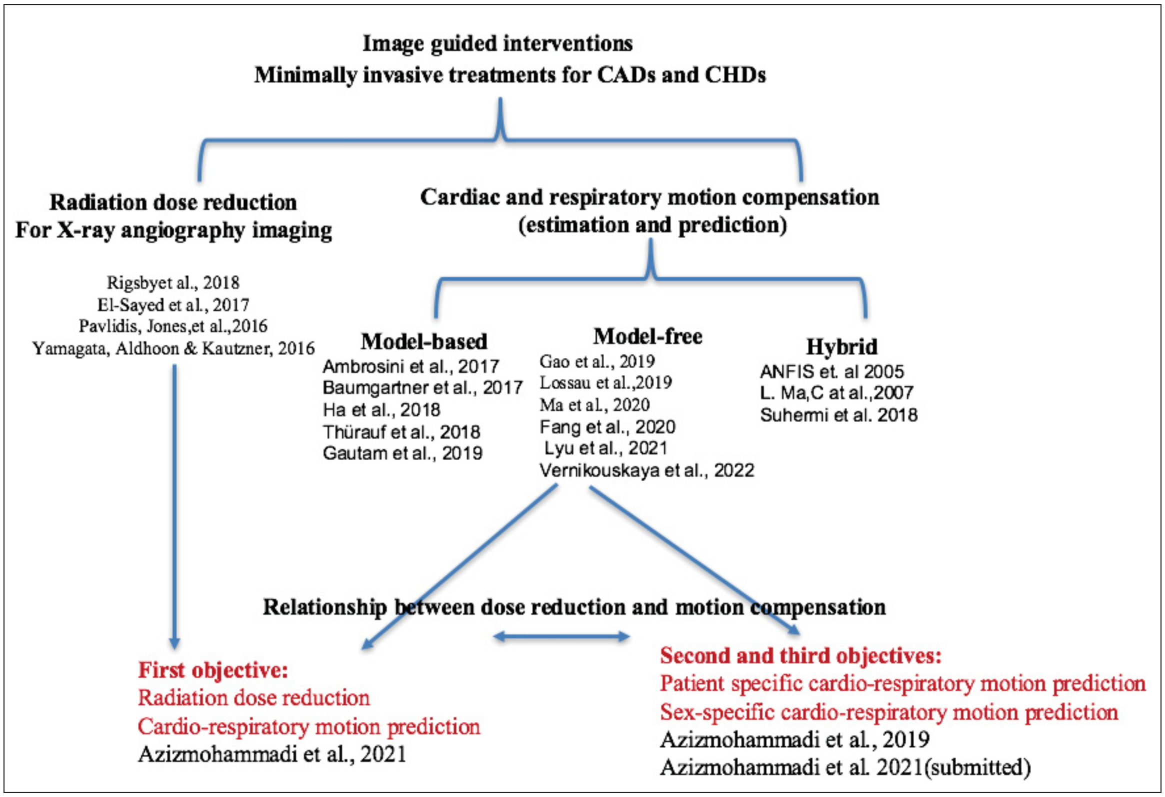

Figure 1.2 A summary of the related works to this study and the link of our objectives to the literature review

1.1.1 Radiation dose reduction in X-ray angiography

In recent years, the number of complex, long-duration pediatric cardiac interventions has increased significantly. Although most long-term sequelae of medical radiation exposure can be attributed to the exposure, the elevated cancer risk is the most significant concern. Depending on the radiation dose, patients are subjected to stochastic effects, such as an increased risk of radiation-induced cancer (El-Sayed *et al.*, 2017) (Pavlidis, Jones, Sirker, Mathur & Smith, 2016). Minimizing radiation exposure in pediatric cardiology is essential since children are more vulnerable to the potentially harmful effects of ionizing radiation than adults (Rigsby *et al.*, 2018). Moreover, the same exposure can be more critical for smaller patients because of their reduced tissue debilitation since body weight plays an essential role in determining the required dose (Rigsby *et al.*, 2018). In addition to pediatric patients' radio sensitivity, any complex congenital pediatric heart disorders they suffer must be catheterized repeatedly, thereby increasing their risk of radiation-induced cancer (Chida *et al.*, 2010). Hence, radiation exposure is a concern for the pediatric population, and determining the optimal dose is a highly relevant research topic in pediatric cardiology.

X-ray angiography is widely accepted for minimally invasive interventions and provides adequate spatial and temporal image resolution. The radiation dose is a function of the required image quality, the patient's size, and the time required to perform the procedure. The Fluoroscopy Time (FT) in minutes is one of the most fundamental factors in choosing the patient's dosage and comprises all the time spent using fluoroscopy for image acquisition (Yamagata, Aldhoon & Kautzner, 2016).

Analog X-ray devices are used to deliver X-ray energy in a continuous dose. The most common approach used to mitigate the radiation dose involves minimizing the beam-on time for both the fluoroscopy and cine acquisition modes (Hirshfeld *et al.*, 2004). Recent X-ray systems are equipped to deliver energy in pulses that can be adjusted to 1, 2, 3, 7.5, 10, 15, and 30 frames per second (fps). In pulsed fluoroscopic imaging, the X-ray beam is switched on and off for every fluoroscopic image. Thus the pulse width, or time duration of each image, is lower than the time required in continuous fluoroscopy imaging. The FT is reduced by replacing the continuous exposure with a pulsed beam delivery. Accordingly, the dose rate is determined by the required dose per frame and acquisition Frame Rate (FR). Hence, for a specific FR, the amount of radiation exposure is reduced, given a specific Pulse Rate (PR). A sequence of pulsed images, including moving objects, appears more continuous and less flickering at high pulse rates or frequencies based on Critical Flicker Frequency (CFF). At low FRs, one way to avoid flicker and minimize the unsharpness of the moving targets is to use Gap Filling. Gap Filling consists of replicating each acquired frame multiple times. This conventional method could be applied to keep the low FR and reduce the dose rate accordingly while having a low image quality. Therefore, the amount of noise perception by human eyes can be reduced (Balter, 2014).

1.1.2 Relationship between motion estimation and the dose reduction for cardiac interventions

To keep the radiation dose as low as possible during diagnostic and interventional procedures, motion compensation techniques are required to reduce potential misinterpretations caused by motion while preserving the image quality. Cardio-respiratory motion prediction has always been preferred in cardiac applications as it facilitates more accurate navigation procedures.

Deep learning architectures such as Recurrent Neural Network (RNN) models are popular in predicting the cardio-respiratory motion in diagnostic and interventional imaging processes for different imaging modalities(Azizmohammadi *et al.*, 2019), (Fang *et al.*, 2020), (Lyu *et al.*, 2021). Yet, there are not many learning based methods for X-ray angiography.

In these approaches, motion features (temporal and spatial) are extracted from image frames and memorized by the RNN model to predict upcoming images. However, predicting and generating realistic images and motion in an end-to-end system continues to present issues using existing models. Generative adversarial networks (GANs) are the tools used for learning deep representations. They can be used for both supervised and semisupervised learning by implicitly modeling high-dimensional data distribution. The main structure of GANs is based on training a pair of networks competing against each other. These two networks are generators and discriminators. The generator is like an art forger and produces realistic synthetic samples like images using a distribution. The discriminator acts as an art expert to distinguish the real sample from the synthetic generated one. These two networks are trained simultaneously, allowing them to improve their respective abilities until the discriminator is unable to tell the real and synthetic samples apart (Creswell *et al.*, 2018). Recently, GANs have been used for video frame prediction (Hu & Wang, 2019). The Prediction quality has been improved considerably using GANs, and the combination with RNNs has made it possible to predict multiple frames as well.

Reducing the PR during complex invasive cardiovascular procedures results in a considerable reduction of the mean energy required for X-ray imaging (fluoroscopy and cine). Pyne et al. demonstrated that reducing the PR from 15 pulses per second (pps) to 10 pps reduces the required dose by 34% (Pyne *et al.*, 2014). The average required dose rate scales as the square root of FR with an equal noise perception for the operator's eyes in pulsed fluoroscopy imaging. Hence, if the FR is reduced from 15 fps to 7.5 fps, the required dose rate is reduced by 30%, while doubling the FR from 15 fps to 30 fps increases the required dose rate by about 40% (Balter, 2014), (Aufrichtig, Xue, Thomas, Gilmore & Wilson, 1994). One recent approach for reducing the FT for diagnostic cine acquisition involves the Last Fluoroscopy Hold (LFH) technique. The cumulative DAP, FT, and contrast used in this approach are reduced as compared to the cine mode by dynamically caching the last current frame of the fluoroscopy sequence. Thus, the last acquired image is saved and displayed during percutaneous coronary interventions (Olcay *et al.*, 2015). Radiation dose reduction is managed in different applications. For instance, X-ray beam filtering and anti-scatter grid removal at a given contrast-to-noise ratio were applied for endovascular interventions (Rana, Singh, Jain, Bednarek & Rudin, 2015), and alternative exposure is applied in mammography using a tungsten/silver (W/Ag) target-filter spectrum (Freer *et al.*, 2017), and dose flattening approaches are used in tumor delivery treatments (Lohse *et al.*, 2011). Despite these advances, current methods still carry certain limitations and costs.

1.2 Cardio-respiratory motion compensation

In recent years progression in imaging techniques has opened up numerous potential applications for medical images for diagnosis, treatment planning, and image-guided interventions. Albeit, given the moving organs in the abdomen and thorax, there are still some problems, such as having artifacts in the acquired images and misalignment between the static guidance information and the moving anatomy in image-guided interventions. These problematic restrictions are caused by respiratory and cardiac motions as well as unexpected motions such as patient's movements (McClelland, Hawkes, Schaeffter & King, 2013).

One approach to overcome the problems caused by moving organs during image acquisitions and image-guided interventions is called breath holding, in which the respiratory motion is suppressed (Oshinski *et al.*, 1996). However, given several restrictions, such as scan time constraints and poor steady breath-holding, free-breathing techniques needed to be improved (Nehrke, Bornert, Manke & Bock, 2001) (Taylor *et al.*, 1997).

Motion tracking is another approach to compensate for the effects of cardio-respiratory motion during the image-guided interventions (Schweikard, Glosser, Bodduluri, Murphy & Adler, 2000) (Kesner & Howe, 2010). Several methods have been developed to segment and track the visible medical devices under the X-ray images in real-time. Deep convolutional neural networks are applied as state-of-the-art approaches for real-time tracking of the segmented catheter and guidewires in 2D X-ray fluoroscopic sequences. For the input, the network takes the current image and the three previous ones and segments the catheter and guidewire in the current image (Ambrosini, Ruijters, Niessen, Moelker & Walsum, 2017). Nagata *et al.* (2009) implemented markers into the region of interest to track the motion in a sequence of images based on these markers. However, the implantation can be invasive for the patients. The motion information is also limited to the marker points rather than the entire region of interest.

The previous motion compensation approaches are discussed in different categories in the following sections. First, the motion prediction strategies are categorized into three groups: model-based, model-free, and hybrid.

1.2.1 Model-based motion prediction algorithms

In model-based methods, the motion is represented using a special mathematical model like linear prediction, Bayesian filtering (Kalman filter, Extended Kalman filter, and particle filtering), sinusoidal model, finite state model, autoregressive moving average model, Support Vector Machine (SVM) and Hidden Markov Model (HMM) (Gautam, Kaur & Sharma, 2019).

Motion models are usually patient-specific and take surrogate data as input and come up with a motion estimation as output. Depending on the applications, motion models can differ in the data used to form the models, the type of model employed, how this model is computed, the type of surrogate data used as input to the model, and what form this output should take. A motion model can be predictive, estimative, and a combination of both depending on the future prediction or current signal values.

A motion model consists of four components. The first component is the input signal(s) to the model or surrogate data (single signal, respiratory phase, single signal, gradient, etc.). The second component is the internal representation of the model, such as rigid/affine transformation coefficients and control points, and the third component is the correspondence of motion representation and surrogate signals (Polynomial, B-spline, Neural networks, Fourier series, etc.). The fourth one is the proper method to match the correspondence model to training data (Linear least squares, rigid registration, principal component regression, etc.).

In a predictive motion model, the future values for the motion signal depend on the past and current motion signal information. Motion estimation models estimate the current motion signal values based on some other simple available signals (surrogate signals).

Many model-based approaches are available for predicting and tracking moving arteries and compensating the cardio-respiratory motion in X-ray angiography (King, Rhode, Razavi & Schaeffter, 2009), (Baka, Lelieveldt, Schultz, Niessen & van Walsum, 2015b), (M'hiri *et al.*, 2016).

Extended Kalman filters (EKF) based on the Local Circular Motion (LCM) Model were applied in (Jung, Kim & Hong, 2013) for predicting respiratory motion. Based on this approach, the first and second-order Kalman filters were implemented based on LCM. Based on LCM, the target location is predicted by evaluating the local dynamic model equations at the required prediction length. Kalman filter estimates the unknown variables by measuring statistical noise over time for predicting respiratory motion, and the LMC characterizes the local circular respiratory motion in an augmented plane.

An auto-adaptive motion model was also introduced in (Baumgartner, Kolbitsch, McClelland, Rueckert & King, 2017) for estimating respiratory motion. This method is based on an extension of the Simultaneous Group-wise Manifold Alignment (SGMA) technique. The adaptive motion model is generated from multiple 2D motion fields derived from sagittal 2D MR slices acquired at different anatomical positions. Then to estimate the 3D motion, the 2D motion fields are combined. To extend the SGMA, they (1) applied the manifold alignment on motion fields. (2) a combination of coronal and sagittal slices was used. (3) The missing sampled respiratory motion 2D points are estimated using K-nearest neighbors as an interpolation scheme on the manifold. This method is automatically adapted to breathing patterns.

In the auto-adaptive motion model, the correspondence is between data acquired from different anatomical positions slide-by-slide of motion fields but with similar respiratory phases. This approach has some limitations in estimating the motion in the direction orthogonal to slides since they assumed the Left-Right movement is negligible compared to Sagittal-Lateral and Anterior-Posterior motion direction. The manifold learning used in this approach is based on Locally Linear Embedding (LLE) for dimensionality reduction. In LLE approaches, the k-nearest neighbor graph of data (motion fields) is based on the shortest distance between data points. K-nearest neighbor algorithm is considered an instance-learning method where the function is only approximated locally, and all computation is deferred until classification. Hence, in this work, each point is reconstructed as a linear combination of its nearest neighbors. Thus, the non-linear regression in 3D for motion fields is reconstructed by considering having linear relocation of any point in 2D slides. This assumption could reduce the accuracy of the work since R-L motion is neglected.

Ha et al. introduced a novel approach that targets real-time imaging modalities by employing contrast-invariant feature descriptors. This work combined an image-based real-time, sparsely distributed feature point tracking with a dense patient-specific motion model. In this combination, a unified optimization framework was introduced based on sparse-to-dense interpolation. This model-based sparse-to-dense image registration method achieved a 2 mm prediction error for respiratory motion estimation (Ha, Wilms, Handels & Heinrich, 2018).

Another novel model-based approach was developed for calibrating a robotic C-arm system for X-ray imaging. According to this method, it can be possible to calibrate a mechanical C-arm system using X-ray images. By reorientating the C-arm system, a static model of the system is required to achieve the desired accuracy (1.5 mm) for 2D/3D overlays. Thus, apart from 2D X-ray images, C-arm systems can perform 2D/3D overlays to provide additional information to the clinician (Thürauf *et al.*, 2018).

1.2.2 Model-free motion prediction algorithms

Model-free methods are heuristic learning-based algorithms that find a pattern for the cardio-respiratory motion. Linear adaptive filters and neural network variables are examples of model-free heuristic learning methods to compensate for the impaired breathing signal with various breathing patterns. The following section explains the adaptive filters and artificial neural networks in more detail. Adaptive filters can update their coefficients and adjust themselves based on the prediction errors with their optimizer over time. Usually, the motion prediction using adaptive filters is calculated using the combination of previous respiratory motion multiplied by its coefficient values. It has been proved by (Vedam *et al.*, 2004) that the prediction accuracy of adaptive filtering can be less than 2mm, which is robust enough to outperform some model-based methods, such as sinusoidal models. However, there are some restrictions for adaptive filters in one-dimensional predictions. Thus, they are extended to multidimensional adaptive filters and adjusted to update the weights of neural networks. Artificial neural networks are mathematical function structures in three sections (input, hidden layers, and outputs). The sections are interconnected with some weights (M.J. Murphy, 2009).

Gao *et al.* (2019) intruded a learning-based approach to compensate for the effects of the patient movement during imaging for digital subtraction angiography using a single live image. Based on the obtained results, the digital subtraction images can be done with fewer artifacts and have more quality for diagnosis. A supervised generative adversarial network strategy achieved better vesselness details.

Lossau *et al.* (2019) also presented a motion estimation and correction method for CT angiography imaging using a CNN model. This work showed supervised learning for coronary motion estimation by patch analysis in CT data. The data for supervised learning was simulated by the coronary motion forward artifact model for CT data, including the simulated motion to 19 artifact-free clinical CT cases with a step-and-shoot acquisition protocol. Then, CNN models were trained to estimate underlying 2D motion vectors from 2.5D image patches based on the coronary artifact appearance. The motion direction and the magnitude with average test accuracies of 13.37 degrees and 0.77 mm, respectively.

One other recent model-free approach introduced a novel real-time dynamic coronary roadmapping method for X-ray fluoroscopy that provides dynamic vessel visualization without using a contrast agent. An accurate catheter trip tracking based on deep learning-based Bayesian filtering was applied to compensate for the respiratory motion. In this method, the posterior of the catheter tip was tracked via a particle filter, for which a likelihood probability map was computed for updating the particle weights using CNN model (Ma, Smal, Daemen & van Walsum, 2020).

Deep Recurrent Neural Network (RNN) models have demonstrated outstanding potential in cardiac imaging and in predicting the cardio-respiratory motion in diagnostic and interventional imaging processes (Azizmohammadi *et al.*, 2019), (Fang *et al.*, 2020), (Lyu *et al.*, 2021). Based on these approaches, motion features (temporal and spatial motion features) are extracted from image frames and memorized by the RNN model to predict the future dynamic behavior of the structures. Nevertheless, it should be noted that predicting and generating realistic future images and motion in an end-to-end system continues to present issues using existing models.

Vernikouskaya et al. represented a novel learning-based method to extract motion features from X-ray images using CNN. The motion features extraction was applied to update the overlay of a static model concerning respiratory and cardiac motion. The automatic motion compensation presented in this work during the fusion of 3D anatomic models with XR fluoroscopy appears to improve robust augmentation during catheter interventions Vernikouskaya, Bertsche, Rottbauer & Rasche (2022).

1.2.3 Hybrid motion prediction algorithms

Hybrid motion prediction algorithms are a combination of model-based and model-free methods while benefiting from both algorithms and could outperform the others. This approach is also categorized in adaptive neuro-fuzzy interference systems, a hybrid model with adaptive filter and nonlinear model and interacting multiple model filter ((2005), ANFIS) (L. Ma,C. Herrmann, 2007).

In Suhermi, Prastyo, Ali *et al.* (2018), a deep artificial neural network model with multiple hidden layers was combined with the Autoregressive Integrated Moving Average (ARIMA) as a hybrid model to capture the nonlinear and linear patterns for the prediction of roll motion. Based on their experiments, the results of this hybrid model prediction outperformed every other available forecast of the roll motion.

1.3 Cardiovascular motion estimation and its effects on dose reduction for cardiac interventions

During fluoroscopy imaging, many factors may influence image quality. These include the patient size, the clinical task, the imaging geometry, the contrast, the signal-to-noise ratio, and the anatomical target movements. Cardiologists must choose an optimized fluoroscopic dose rate that counterbalances these factors (Balter, 2014). If the image quality is not high enough and the image is not sufficiently detailed, the cardiologist's clinical confidence might be impacted. Nevertheless, acquiring better quality images could still increase the radiation rate without necessarily improving clinical outcomes (Balter, 2014).

To keep the radiation dose as low as possible during long-term interventions and short-term acquisitions, motion compensation, and prediction techniques are required to reduce potential misinterpretations caused by motion while keeping decent image quality.

CHAPTER 2

CONTRIBUTION #1: GENERATIVE LEARNING APPROACH FOR RADIATION DOSE REDUCTION IN X-RAY GUIDED CARDIAC INTERVENTIONS

2.1 Proposed contribution

The goal of this study is to predict dynamic X-ray angiography sequences using a generative model. A video frame prediction model is introduced to predict new X-ray angiography frames. We introduced a new loss function to predict the temporal and spatial information of the arteries in angiography sequences. To minimize the vesselness structure differences between the predicted and ground truth images, a multi-scale Hessian-based loss term is added to the loss function presented by Mathieu et al. (Mathieu, Couprie & LeCun, 2016). Then, a predictive RNN-based motion model is trained to estimate the motion and content of single and multiple future frame(s) based on previously acquired frames in an end-to-end system. This contribution was published in Medical Physics (Azizmohammadi *et al.*, 2022)

2.2 Materials and methods

2.2.1 Data description

We developed and validated our method using simulated and patient X-ray angiography datasets from Sainte-Justine Hospital. Simulated X-ray sequences generated from realistic XCAT computational phantoms with cardio-respiratory motion (Segars, Sturgeon, Mendonca, Grimes & Tsui, 2010) were first investigated. The simulated motion included the beating heart and respiratory motions. The simulated dataset included 56 patients (32 male and 24 female) and 112 sequences (2 sequences per patient, showing either the left or right coronary artery). All the generated sequences had a length of 75 frames capturing multiple heart and respiratory cycles, and were acquired at 15 fps. The patient database comprises 52 patients with contrasted coronary arteries. This study was reviewed and approved by the Institutional Review Board of Sainte-Justine Hospital. Each patient presents a different number of sequences with varying lengths. There are

340 sequences, respectively, with a minimum and maximum length of 15 and 70 frames. All the data was acquired at 15 fps. Fig 2.1 shows samples of simulated and patient data for coronary artery branches.

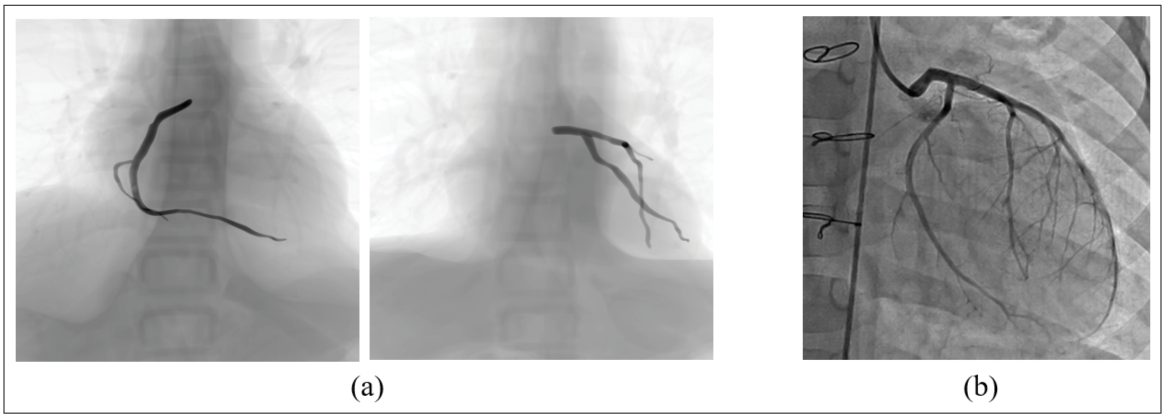

Figure 2.1 (a): Simulated RCA (left) and LCA (right) coronary artery branches. (b): Contrasted LCA branch for patient data.

2.2.2 X-ray angiography frame predictions

In this section, the effects of frame predictions on dose reduction are assessed regarding the required dose rate and the total fluoroscopy time. The quantitative results of this assessment illustrate that reducing the total fluoroscopy time can considerably impact cumulative radiation exposure reduction.

2.2.2.1 Assessment of the impact of pulse rate reduction on the total radiation dose reduction

In our approach, we assumed that for any specific frame rate (7, 15, 30, 60 fps), the number of pulses required could be reduced during an X-ray imaging process such that the predicted frames can replace the real X-ray frames. Depending on the X-ray manufacturers, the dose for a given exposure duration is directly related to the pulse rate (Kobayashi & Hirshfeld Jr, 2017) (Mahesh, 2001), or it can scale as the square root of the frame rate for uniform noise perceived by the operator's eyes (Balter, 2014) (Aufrichtig *et al.*, 1994). In this work, we considered the square root model.

According to this approach, for the same frame rate, a smaller pulse rate (i.e., dose rate) is required since T frames are predicted (Fig. 2.2 (a)). Considering K as the number of previously generated and visited frames and T as the number of predicted frames at each prediction mode, for every $K+T$ frames, T frames are predicted. Thus, the number of pulses required at every second can be reduced by $FR \times (\frac{T}{K+T})$. Hence, the Required Dose Rate (RDR) scales proportionally as:

$$RDR \propto \sqrt{FR \times \frac{K}{K+T}} \tag{2.1}$$

where the FR is the selected frame rate for the intervention or acquisition (7, 15, 30, 60 fps).

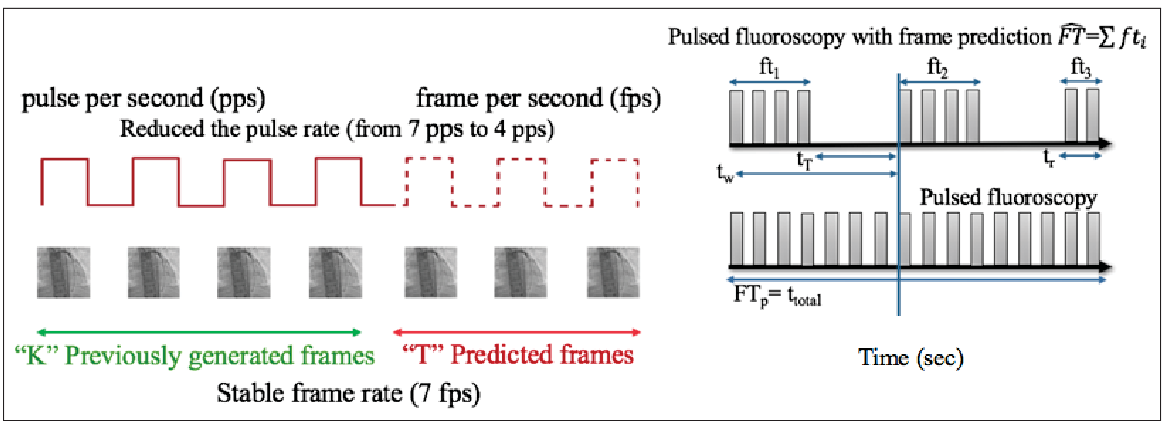

Figure 2.2 An example of different fluoroscopy techniques. Less fluoroscopy time is required for pulsed discrete fluoroscopy by pausing the radiation beam after K acquired images for a prediction time t_T in each time window t_w compared to other methods ($\hat{FT} < FT_p$)

Given the parameter K, which is the number of previously generated and visited frames contributing to the prediction of the new frame/s, the X-ray exposure can pause at each predicting mode and resume in acquisition mode. Assuming t_T as the required time for T frames prediction, t_w as the required time window for $K+T$ acquisitions, and FT_p as the entire required fluoroscopy time (in seconds), the $\hat{FT}$ is the reduced fluoroscopy time:

$$\hat{FT} = FT_p - (\lfloor \frac{FT_p}{t_w} \rfloor \times t_T) + t_r \tag{2.2}$$

In any time window (t_w), the exposure time is reduced by the amount of time that is required to acquire T frames (t_T). The t_r is the remaining time in the X-ray angiography sequence.

Fig. 2.2 (b) shows the difference between conventional continuous fluoroscopy, pulsed fluoroscopy, and our method in terms of fluoroscopy time. For the pulsed fluoroscopy with frame prediction the $\hat{FT} = \Sigma(t_w - t_T) + t_r = \Sigma ft_i$. In pulsed fluoroscopy, less energy is exposed as compared to continuous fluoroscopy. In our approach, the X-ray device is supposed to pause at each prediction mode and resume in each acquisition mode. Thus, the total fluoroscopy required in an X-ray imaging process is reduced.

2.2.2.2 Cardio-respiratory motion and content estimation in X-ray sequences

The prediction of upcoming frames of a video sequence requires two components, namely, the visual content and pixel displacement through time or motion. Thus, the proposed network learns the internal representation of image evolution through the sequence based on its content and motion. The model in this work consists of two different encoders, one for the visual content and a second one for the motion of the image sequence. These two key components must be decomposed among the images and predicted separately. The motion features are extracted by an RNN-based encoder with Long Short Term Memory (LSTM) and CNN, while the visible content features are only extracted from the last visited image with a CNN-based model. Deep learning methods have been applied successfully for video frame prediction in the literature (Villegas, Yang, Hong, Lin & Lee, 2017), (Hsieh, Liu, Huang, Fei-Fei & Niebles, 2018), (Tulyakov, Liu, Yang & Kautz, 2018).

2.2.3 Model architecture

A generative model is built on an Encoder-Decoder framework. To extract the motion and content features of the images in sequences, a CNN model is used in combination with an LSTM network. The LSTM cells memorize the complex non-periodic cardio-respiratory motion in the angiography sequences. According to our previous work (Azizmohammadi *et al.*, 2019), the LSTM structure is robust enough to deal with different motion patterns in the cardio-respiratory motion signals during prediction. Therefore, an LSTM-CNN combination is used for a general motion estimator. To effectively handle the complex evolution of pixels in sequences, the motion and content are predicted independently using two encoders. Thus while the motion encoder captures the temporal dynamics of a sequence, the content encoder extracts the spatial information from the most recent image. Hence, the X-ray images' spatial and temporal dynamic features are extracted and encoded separately. The model architecture also includes a concatenating section that combines the outputs of these encoders and a multi-scale residual used to avoid information loss before pooling in the network. The last part is the decoder, which reconstructs the predicted images. Fig. 2.3 shows the complete structure of the model.

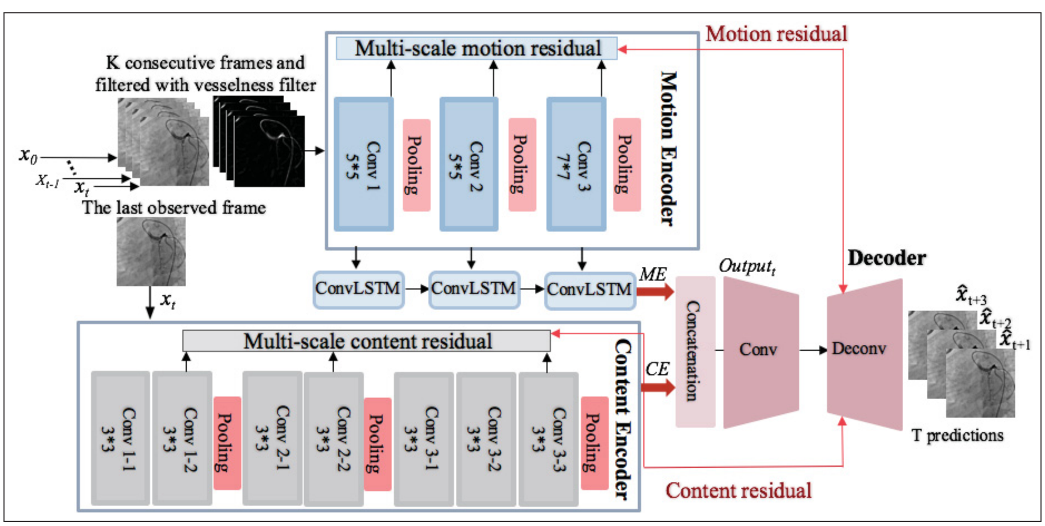

Figure 2.3 The motion-content model structure. Two encoders extract the motion and content features separately (ME and CE). The input for the motion encoder is a sub-sequence of the previous filtered frames filtered. The input for the content encoder is the last seen frame. The motion and content residuals are added to avoid information loss.

2.2.3.1 Motion encoder

A Convolutional LSTM (ConvLSTM) extracts the dynamic features in X-ray sequences. While the pixel-level features are extracted by a Convolutional Neural Network (CNN), the sequential information is provided by the LSTM cells in the motion encoder. The motion encoder captures the local motions from one frame to the next in X-ray sequences. The cardio-respiratory movements of the objects (arteries, devices, catheters, wires, stents, etc.) are predicted directly (without using a surrogate object) and independently in the sequences. The original presented motion encoder in (Villegas *et al.*, 2017) takes the element-wise image subtraction between (x_t and x_{t+1}) as an input. Since there are background movements in angiography images, the subtraction of original frames includes a lot of artifacts. In our approach, we filtered the input images by vesselness filter first and then subtracted the filtered input images to overcome the artifact caused by the background movement. Thus, the motion encoder tracks only the contrasted arteries' movement to encode the temporal dynamics of transformed images through the sequence (d_t). Considering $d_t = x_{v(t)} - x_{v(t-1)}$ as the element-wise subtraction between frames at time t and $t-1$ that were filtered with vesselness filter, and d_{t-1} as the feature tensor encoding the motion across the observed difference filtered image inputs and c_{t-1} as the memory cell retains information of the dynamics observed through time, then the $f^{motion}(.)$ is a fully convolutional function that allows the model to identify the local dynamics of consequent frames. The output of the motion encoder is as follow:

$$ME = [d_t, c_t] = f^{motion}(x_{v(t)} - x_{v(t-1)}, d_{t-1}, c_{t-1}) \tag{2.3}$$

2.2.3.2 Content encoder

The content encoder extracts the essential spatial features from the visible contents, such as contrasted moving objects (arteries) and the background (ribs, bones, and devices) in the images. It takes the last observed frame x_t as input and encodes the spatial information in the image (CE_t) using a CNN network. The last observed frame has the most recent and essential information required for predicting the future frame(s). Thus, CE is the feature encoding the spatial content of the last observed image, and $f^{content}(.)$ is a CNN model that extracts content features from one single image x_t. The output of the content encoder is as follow:

$$CE = f^{content}(x_t) \tag{2.4}$$

2.2.3.3 Final prediction using the content and motion encoders' outputs

A multi-scale encoder residual is used to compute the residual Res_t at each scale or layer just before the pooling layers of both motion and content encoders. The outputs of both encoders are concatenated and combined with the residual outputs (ME_t, CE_t, Res_t) to perform pixel-level predictions in the decoder. These predictions can represent one or more frames in the future. The output of the model (Villegas *et al.*, 2017) is as follows:

$$Res_t^h = f^{residual}([ME^h, CE^h]) \tag{2.5}$$

$$Output_t = f^{combination}([ME, CE]) \tag{2.6}$$

$$\hat{x}_{t+1} = f^{decoder}(Output_t, Res_t) \tag{2.7}$$

Where Res^h is the residual link at layer h being used to avoid information loss after pooling for each layer, $f^{residual}(.)$ is the residual function implemented as consecutive convolution layers, and rectification with a final linear layer. The $Output_t$ represents the combination layer that concatenates the outputs of both motion and content encoders (ME and CE). The new frame is generated as the output of the decoder going through a $tanh(.)$ activation function.

2.2.3.4 Loss function

A combination of terms (image space and generator loss terms) is minimized in this approach. We adjusted this loss function to predict the cardiac angiography sequences, given that the targets to track and predict are contrasted arteries. The total loss function combines the image loss

(L_{IM}) and generator loss (L_{GAN}) in adversarial training. Considering the α and β as constant weights:

$$L_{Total} = \alpha L_{IM} + \beta L_{GAN} \tag{2.8}$$

$$L_{IM} = \alpha L_{gdl} + \beta L_P + \gamma L_{Vss} \tag{2.9}$$

Where L_{IM} represents the image space loss as a combination of terms that match the average pixel intensities with L_P, gradient difference to sharpen the predictions, and the newly added sub-loss called vesselness difference L_{Vss}.

We penalized the difference between the second derivative of the Gaussian filter applied on the predicted and ground truth images with six different scales (vesselness σ range: 0.5 - 3 with step size: 0.5). The output of the vesselness filter on the images is the vesselness response image. The second derivatives encode the shape information, and the eigenvector corresponding to the smallest eigenvalue is the direction of the blood vessel locally. Hence, the L_{Vss} is applied to minimize the local differences between the predicted and ground truth images, which refer to the shape of the arteries.

The gradient difference term L_{gdl} (Mathieu *et al.*, 2016) (Villegas *et al.*, 2017) is applied to sharpen the generated images. This term directly assesses the gradient discrepancy between the ground truth and the predictions. The gradient difference between the ground truth image Y and the prediction $\hat{Y}$ is given by:

$$L_{gdl}(\hat{Y}, Y) = \Sigma_{i,j}(||Y_{i,j} - Y_{i-1,j}| - |\hat{Y}_{i,j} - \hat{Y}_{i-1,j}||^{\lambda} + ||Y_{i,j-1} - Y_{i,j}| - |\hat{Y}_{i,j-1} - \hat{Y}_{i,j}||^{\lambda}) \tag{2.10}$$

Where λ is an integer greater or equal to 1 (here, the $\lambda = 1$) and $|.|$ is the absolute value function (Mathieu *et al.*, 2016). The new vesselness difference term L_{Vss} matches the vesselness responses of the predicted and ground truth images. The vesselness difference between the ground truth image Y and the prediction $\hat{Y}$ is given by:

$$L_{Vss}(\hat{Y}, Y) = \Sigma_{i,j}|I(Y) - I(\hat{Y})| \tag{2.11}$$

Where $I(Y)$ represents the ground truth image filtered by vesselness, $I(\hat{Y})$ is the predicted image filtered by vesselness, and $I(Y) - I(\hat{Y})$ is the element-wise subtraction of filtered ground truth and prediction frames. We only considered pixel intensity differences for the vesselness filter, not direction differences. To generate images correctly and avoid blurred images, the generator loss in adversarial training L_{GAN} is added to solve the blurriness problem. It induces realism in the image sequences and sharpens the images (Mathieu *et al.*, 2016).

$$L_{GAN} = -\log D([x_{1:t}, G(x_{1:t})]) \tag{2.12}$$

while $D(.)$ represents the discriminator in adversarial training, and $x_{1:t}$ is the input image concatenation. The adversarial discriminator loss (L_d) is defined by:

$$L_d = -\log D([x_{1:t}, x_{t+1:t+T})]) - \log(1 - D[x_{1:t}, G(x_{1:t})]) \tag{2.13}$$

the concatenation of future ground truth images and all of the predictions are represented as $x_{t+1:t+T}$ and $G(x_{1:t}) = \hat{x}_{t+1:t+T}$ respectively (Mathieu *et al.*, 2016) (Villegas *et al.*, 2017).

2.3 Results and validations

The parameters for the X-ray angiography sequences were optimized for both the simulated and patient datasets. The number of iterations was evaluated between 1000 to 1500 for the simulated dataset and between 2000 and 2500 for the patient datasets. We divided the dataset into two parts: 80% of the dataset for training and 20% for testing. The model was evaluated on each dataset separately. Each sequence was divided into time slots or time windows of minimum $(K + T)$ frames. A single frame was repeatedly predicted at a time, and the prediction was included through the time slot while the previous predicted frame(s) contributed to new predictions. The number of previously generated and visited frames (K = 7 to 10) contributing to predicting the future frame(s) for the motion encoder was set based on capturing a complete heart cycle in time (0.8s to 1s) and on the length of the shortest sequences in our dataset. All the parameters and hyperparameters were selected based on different experiments. The hyper-parameters controlling the effect of each sub loss function represented as α, β, and γ were set to 1, 0.02, and 0.01, respectively, based on the experiments. The quality of the predicted images was reduced by increasing the number of predictions. The visual quality of the predicted images using our method (the vesselness-based MCnet) was assessed as compared to the original MCnet in terms of certain similarity measurement metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) (Table 2.1 and Table 2.2). We used PSNT and SSIM as two baselines to compare our method with the original MCnet and then evaluated the quality results in terms of motion and content prediction based on other approaches as well. In our experiments, we predicted up to 3 frames with over 0.6 SSIM for both the original and vesselness-based MCnets (Table 2.1 and Table 2.2).

Table 2.1 Average similarity measurements of the predicted images over the testing data on three predicted images for the simulated dataset

Frame	Vss PSNR	Original MCnet PSNR	Vss SSIM	Original MCnet SSIM
Simulated data K=7, T=1,2,3				
Frame 1	28.28	27.98	0.94	0.89
Frame 2	25.47	24.85	0.92	0.85
Frame 3	23.90	23.01	0.88	0.82
Simulated data K=10, T=1,2,3				
Frame 1	29.13	28.82	0.97	0.86
Frame 2	27.65	25.10	0.93	0.83
Frame 3	24.14	23.12	0.87	0.81

According to the experiments, the quality of the predicted images is reduced by increasing the number of predictions. With the simulated data, the first three frames were well-predicted with 0.87 to 0.97 SSIM (Table 2.1). For the patient dataset, the best results refer to K = 10, in which the first three predicted frames reach between 0.68 and 0.82 SSIM (Table 2.2). Our experiments show that the parameter K must be equal to or greater than the frames required to

Table 2.2 Average similarity measurements of the predicted images over the testing data on three predicted images for patient dataset

Frame	**Vss PSNR**	**Original MCnet PSNR**	**Vss SSIM**	**Original MCnet SSIM**
Patient data K=7, T=1,2,3				
Frame 1	27.10	26.75	0.79	0.80
Frame 2	24.42	23.59	0.68	0.70
Frame 3	23.10	21.54	0.61	0.61
Patient data K=10, T=1,2,3				
Frame 1	27.97	26.80	0.82	0.78
Frame 2	25.65	24.62	0.74	0.69
Frame 3	24.14	23.32	0.68	0.63

cover a cardio-respiratory cycle. Moreover, the values for the parameter K in our experiments depend on the length of the shortest sequences in our patient dataset such that the $K+T$ must be equal to or less than the length of the shortest sequence in our dataset (13 frames). Based on the overall experiments with patient and simulated datasets (Table 2.1 and 2.2), the first three predicted frames have over 0.6 SSIM, and the vesselness structure is visible. Thus, at each second during the X-ray imaging process, the patients can be exposed to 3 fewer pulses while keeping the same frame rate (15 fps). The required frame acquisition (i.e., pulses) for a 15 fps sequence can drop by 23% to 30% (for K = 10 and K = 7 respectively), and according to (2.1), the average required dose rate for 15 fps imaging on every second can be reduced by 0.63 to 0.47, as compared to real acquisition. Fig. 2.4 (a) and (b) show the samples of prediction with K = 7 and 10, respectively and Fig. 2.6 shows the overlay of the manually segmented ground truth arteries (in green) and the predictions.

Table 2.3 Euclidean distance between the centrelines of arteries in the predicted frames and ground truth for the frame prediction and Gap-Filling

Frames	**Euclidean distance (mm)**					
	Frame prediction			Gap-filling		
	Mean	Max	SD	Mean	Max	SD
Frame 1	0.28 mm	0.76 mm	(+/-) 0.19 mm	0.33 mm	0.79 mm	(+/-) 0.22 mm
Frame 2	0.30 mm	0.78 mm	(+/-) 0.20 mm	0.39 mm	0.85 mm	(+/-) 0.31 mm
Frame 3	0.32 mm	0.84 mm	(+/-) 0.21 mm	0.51 mm	0.93 mm	(+/-) 0.35 mm

To evaluate the motion prediction, we applied optical flow to estimate the motion between consecutive predicted and ground truth frames. Optical flow is one common approach for detecting moving objects' motion in an image sequence. It is defined as the distribution of visible velocities of moving objects in an image.

Fig. 2.5 shows the estimated movements between the four consecutive frames with optical flow. The first row extracts the motion arrows from the ground truth sequence. In the second and third rows, the motion arrows are extracted from the predicted images using the vesselness MCnet and the original MCnet, respectively. The optical flow fields between each moving frame and the previous (source) frame are overlaid by moving frames (F = 7,8,9). The motion vectors in the frames predicted using the vesselness MCnet have the same directions and intensities in the region of interest (arteries) as the ground truth in all frames. In contrast, the intensities and directions of the detected motion vectors are different in the predicted frames using the original MCnet.

We randomly selected 30% of the sequences from the test dataset to evaluate the predicted content of the generated images with K = 7 (visited frames) and T = 3 (predicted frames). Coronary arteries were segmented in three consecutive frames of each selected sequence in both groups (ground truth and predictions) by a trained operator. We computed the DICE coefficients and Euclidean distances between the ground truth and the predicted images from the resultant masks.

Euclidean distance was calculated between the extracted centrelines of the segmented masks. Additionally, we reported the results of a conventional Gap-Filling method (baseline) in the selected dataset (Table 2.3).

The Gap-Filling method copied the last visited frame multiple times instead of being predicted. The Euclidean distance and DICE coefficients of the predicted images in our method and the ground truth were computed and compared with the Euclidean distance and DICE coefficients of the ground truth and the copied frames used as the Gap-Filling.

The average computed DICE coefficients (between the ground truth and predicted images) over three predictions using our method was 0.78 ± 0.07 while this value for the conventional Gap-Filling was 0.63 ± 0.05. Table 2.3 shows the comparison of our approach and Gap-Filling in terms of the computed Euclidean distances between the centerlines of the ground truth and the predictions. Based on these evaluations, for three consecutive frames, the results of the frame prediction with our approach outperform the baseline method (Gap-Filling).

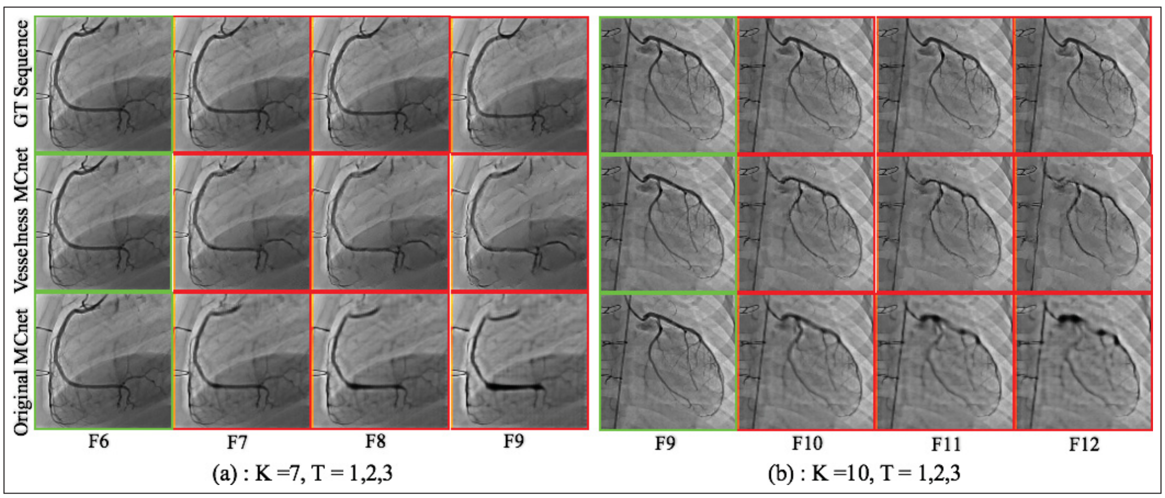

Figure 2.4 The ground truth sequence (first row). The results of vesselness-MCnet and original MCnet (second and third rows). The predicted images in red outline, and the last visited frame is in green outline

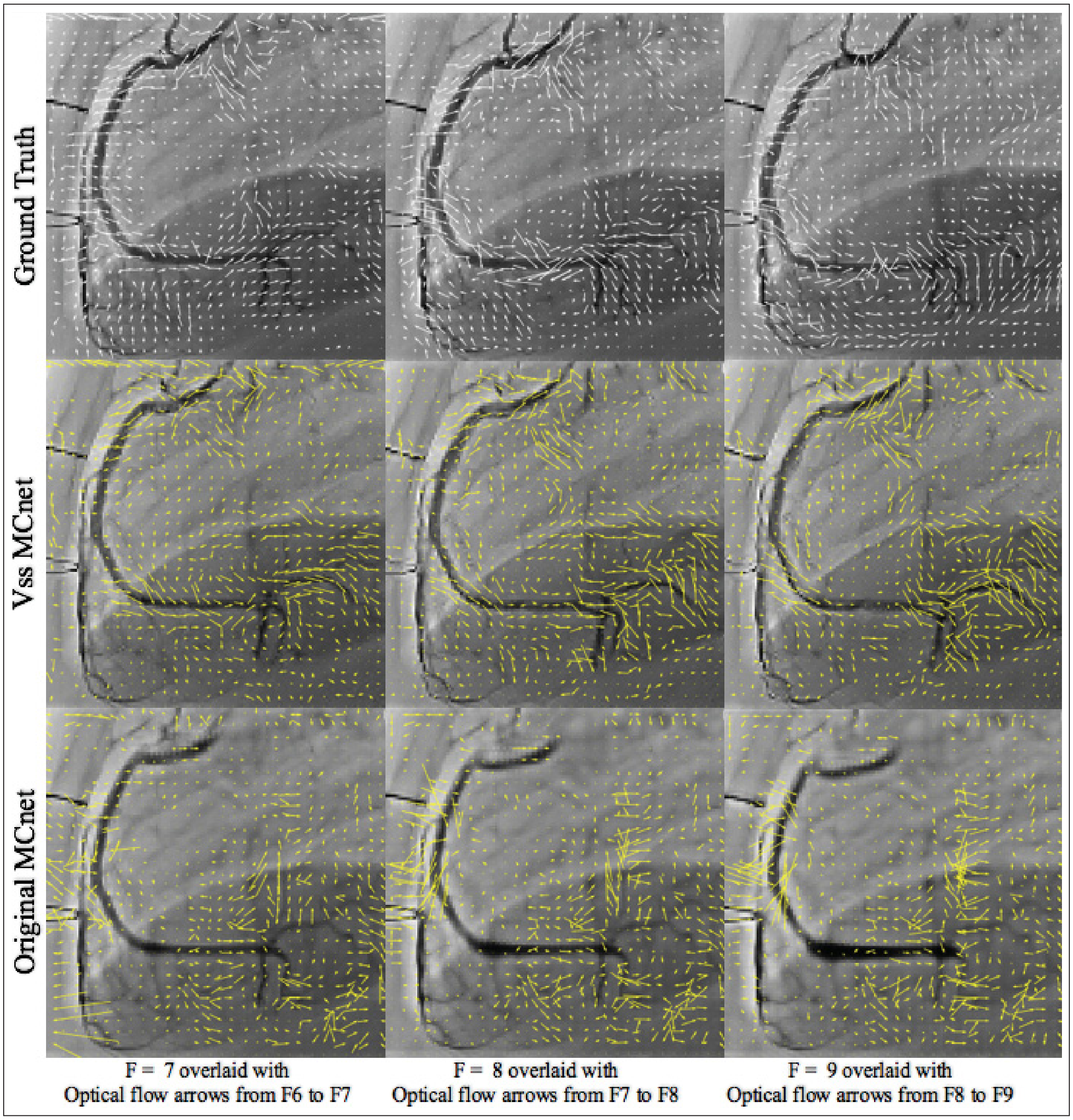

Figure 2.5 Optical flow estimated motion fields of the ground truth sequence are shown in white arrows overlaid with consecutive ground truth frames (F7, F8, and F9) on the top row. The generated frames with vesselness-based MCnet on the second row and the original MCnet on the third row are shown with yellow arrows. The optical flow fields are overlaid to the predicted frames F7, F8, F9

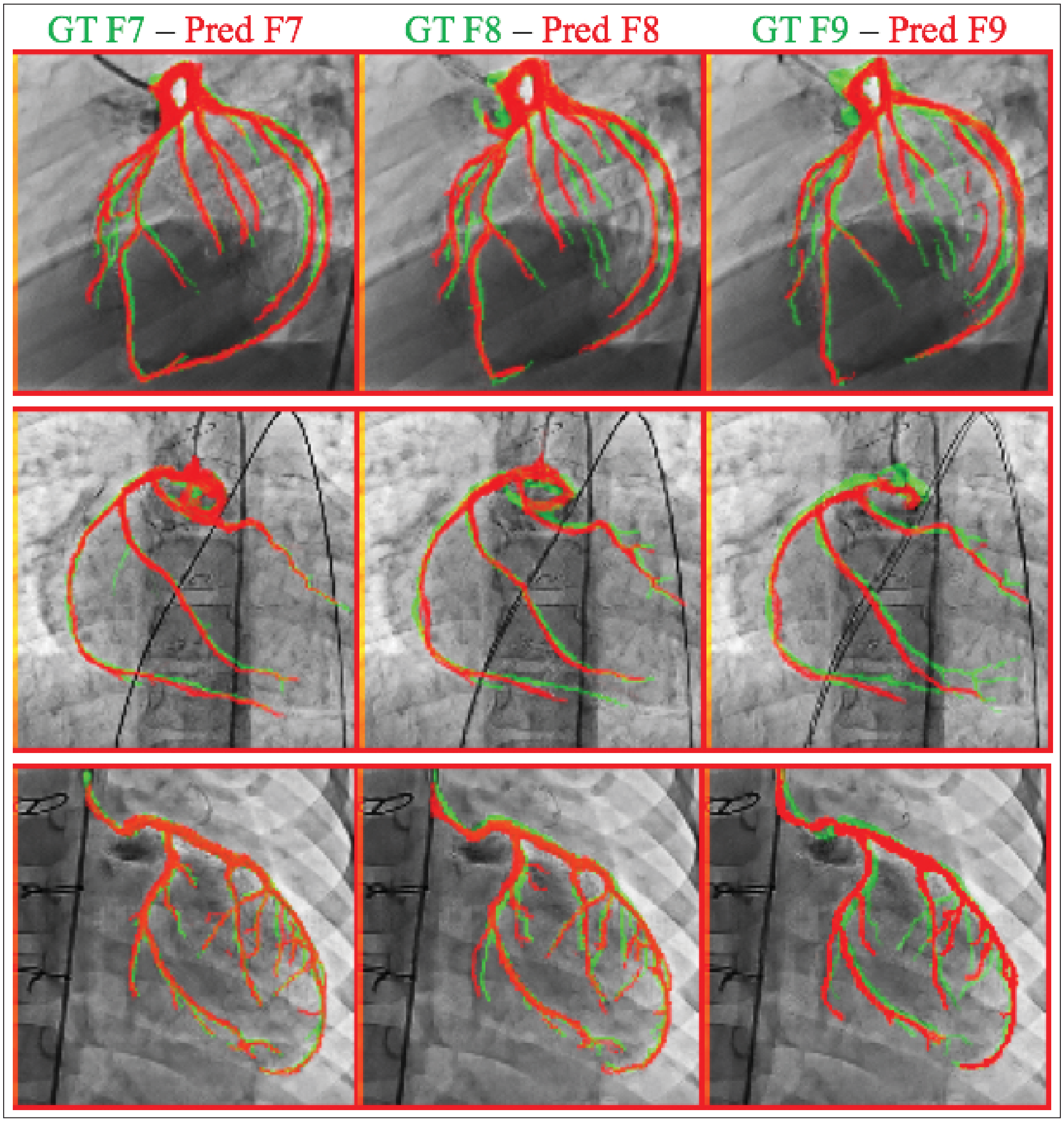

Figure 2.6 An overlay of the manual segmentation masks for the ground truth in green and predicted sequences in red

2.4 Discussion

This work presents a novel radiation dose management approach for pediatric interventional cardiology using a generative learning-based video frame prediction approach. This study can also facilitate the navigation of X-ray-guided interventions, given its intrinsic motion compensation strategy in the frame predictions. Our approach has introduced a predictive model rather than an interpolation approach since interpolation methods require both future and former information.

In frame prediction using this model, the idea is to extract the cyclic cardio-respiratory motion features from the previous frames and combine them with the visual content of the last visited frame. The correlations between spatial and temporal features extracted from the previous frames allow self-supervision of the prediction of single or multiple frame(s) in an end-to-end system. This model can be transferable to adult patients by training on clinical data from adults. The presented model can fully adapt to different patients with distinct respiratory and cardiac motion patterns.

Compared to other video frame applications, X-ray sequences have less inherent uncertainty and variety when estimating upcoming frames since their grayscale images include limited objects for tracking, and the cardio-respiratory motion is periodic. However, the main challenge with X-ray sequence prediction compared to natural video prediction lies in the moving background, which makes motion prediction more complex in the former. In this work, we applied a new loss function and changed the input of the motion encoder using a vesselness filter to overcome the artifacts caused by the moving background. Obtaining a minimum required image quality in X-ray angiography is challenging since different types of interventions may require different image qualities. Our results show the potential of our method for reducing the fluoroscopy time for pediatric cardiac interventions.

This work only focused on the pulse rate and fluoroscopy time reduction since our dataset was retrospective. Other dose indicators, such as cumulative air kerma, should be considered along with fluoroscopy time in our future work. Significant efforts have been invested in improving the new generation of X-ray devices, given the importance of radiation dose reduction not only for pediatric patients with high potential risks of cancer but also for adult patients, cardiologists, and medical staff (Gislason-Lee *et al.*, 2016) (Gislason-Lee *et al.*, 2017) (McNeice *et al.*, 2018).

This study can thus pave the way for the next generation of X-ray imaging devices, it allows for optimization of the induced radiation dose for patients and staff.

Future work will consider incorporating the heart cycle information using the ECG signal for more accurate motion estimation. Other model-based or hybrid approaches can be investigated to improve motion prediction accuracy. Additionally, video super-resolution methods can be included in the content predictor to improve the image quality of predictions.

CHAPTER 3

CONTRIBUTION #2: PATIENT-SPECIFIC CARDIO-RESPIRATORY MOTION PREDICTION IN X-RAY ANGIOGRAPHY USING LSTM NETWORKS

3.1 Proposed contribution

We present a novel patient-specific cardio-respiratory motion prediction approach using a simple supervised LSTM network for angiography sequences. The 2D displacements of the moving objects in an angiography sequence are extracted from the images and represented through 2D affine transformation matrices. Then, a many-to-many LSTM model is trained for every sequence to predict the next geometrical transformation of the arteries in the future sequence frames from previous ones.

3.2 Materials and methods

2D transformation parameters (translation, rotation, scaling, and shearing) are extracted from the X-ray angiography sequences representing the motion features in matrix form. Thus, a sequence of affine transformation matrices resulting from frame-by-frame image registrations of the original X-ray sequence is fed into the sequential LSTM model for training. Then, the future displacements of the moving targets (contrasted arteries and medical devices) in the upcoming frames are predicted.

3.2.1 Data description

This approach was developed with a simulated X-ray angiography dataset using a realistic XCAT computational phantom simulatorSegars *et al.* (2013); Segars *et al.* (2010). It was validated with real patient X-ray angiography sequences from Sainte-Justine Hospital.

Computerized phantoms play a major role in medical imaging research today. They are very helpful in providing a practical means for evaluating and improving imaging techniques and devices.

In this study, we employed realistic XCAT computational phantoms with the cardio-respiratory motion for adult and pediatric patients. Initially, our experiments were done on three different adult patients Azizmohammadi *et al.* (2022). The length of the sequences used to capture at least two to five heart and/or respiratory cycles varied between 75 and 150 frames. For each patient, three different types of motions (only cardiac, only respiratory, and both motions) were generated. Respiratory motion is not gated with cardiac motion, and the misalignment between the motions makes the motion estimation more complicated. Fig.3.1 shows the mismatching of the ECG signal and the respiratory motion signal. The corresponding acquisition cardiac phase is expressed using a percentage interval between two consecutive R-wave (R-R interval). Therefore, in this method, we simulated different motion modes to assess the predictions based on the motion complexity.

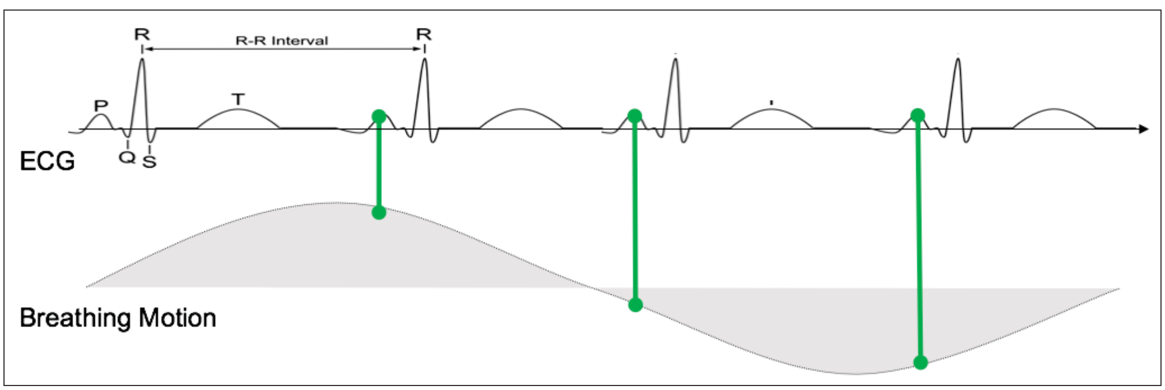

Figure 3.1 Gating the ECG signal and breathing motion. The end of each heart cycle (R-R interval) is synchronized with different breathing phases.

Simulations were also carried out for different circumstances, with the patient having normal and abnormal respiratory and heartbeat cycles. A normal heartbeat cycle is 1 a second long, while the respiratory cycle is 5 seconds. These values can vary between patients and change if the patient is under stress or not breathing normally. We then simulated 64 pediatric and adult patients (36 male and 28 female) falling within the 8-month newborn to 85-year age range, including the left and right coronary arteries. The pediatric simulated dataset included 112 sequences (2 sequences per patient, showing either the left or right coronary artery), while our adult simulated dataset included 12 sequences.

For the real patient dataset, we selected 10 different sequences with lengths ranging between 84 and 166 frames and having visible moving objects (contrasted arteries and/or medical devices such as catheters and guide wires).

3.2.2 Segmentation and centerline extraction of the simulated and patient X-rays sequences

To track the motion signal in an X-ray sequence, pre-processing steps were applied to segment and extract the centerlines of the moving targets. Using the Frangi filter and the connected components, we segmented the arteries and skeletonized the segmentation to extract the centerlines. The 2D motion features were extracted by registering the X-ray images frame-by-frame in the sequence.

For the simulated data, the vascular structures were extracted from the image frames for each sequence and segmented by applying image processing filters and connected components to remove the small objects. Fig. 2.2 and Fig.2.3 show the vessel structures extracted from the original X-rays (simulated and patient data, respectively). The Frangi filter parameters were set based on the diameter of the coronary arteries. For sigmas, an interval of $[1, 6]$ was considered with a step size of 0.1. We applied the same segmentation steps for the patient data, although the segmentation was noisy in the background as compared to the simulated images. Moreover, since the arteries' structures were not continuously contrasted in all the image frames of a given sequence, we segmented medical devices as moving objects, where the arteries were not visible or faded in some frames. We changed the Frangi filter parameters as a function of the visible objects in the patient sequences.

3.2.2.1 Coherent Point Drift (CPD) registration

The point set registration algorithm is widely used in computer vision problems such as image registration. The registration can be rigid or non-rigid. Given two point sets (centerlines of two consequent frames), we applied CPD as a non-rigid registration to map one point set to the other, yielding a non-rigid transformation. Non-rigid transformations include affine transformations such as scaling and shear, as well as translation and rotation.

The CPD algorithm is a Gaussian Mixture Model (GMM) based algorithm that assigns correspondence points between two sets of point clouds. It retrieves the transformations for mapping each point cloud to the other using a specified registration Myronenko & Song (2010). The alignment of the two point clouds is a probability density estimation problem. The first point set is centered on the second one by fitting the GMM algorithm and maximizing the likelihood. The GMM moves coherently and retains the topological structure of the point

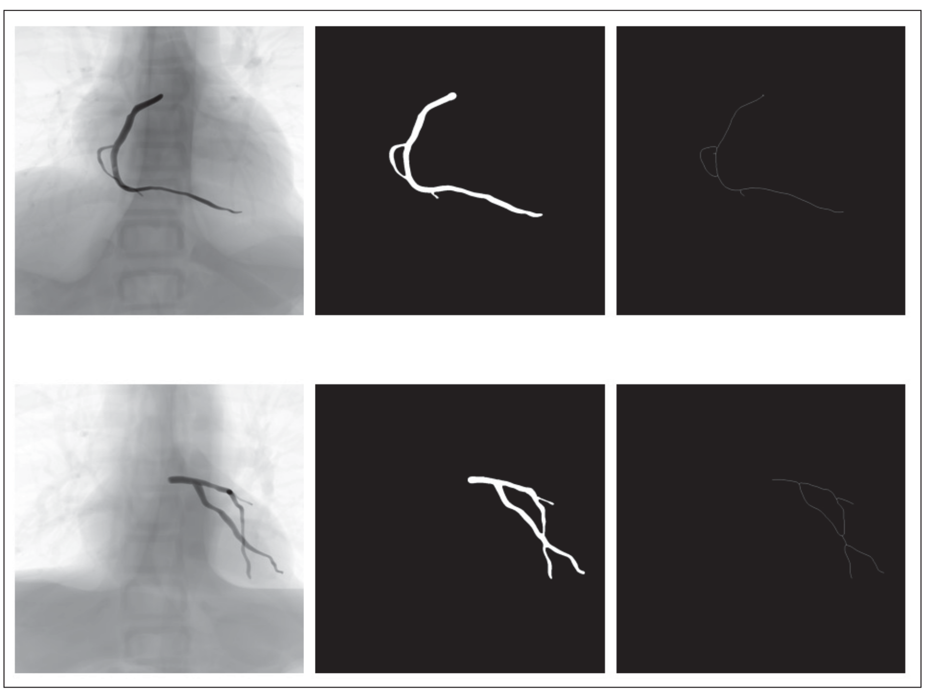

Figure 3.2 Pre-processing steps (segmentation, denoising, and centerline extraction) on simulated data. The first row is a Right Coronary Artery (RCA) branch and the second row shows the Left Coronary Artery (LCA).

Figure 3.3 Pre-processing steps on patient data (segmentation, denoising, and centerline extraction)

sets Myronenko & Song (2010). A coherence constraint is imposed for affine registration by re-parametrizing the GMM centroid locations with affine transformation parameters (translation, rotation, shearing, scaling). These parameters are concatenated to build the Affine Transformation (AT) matrix as follows:

$$AT = \begin{bmatrix} s_x cos(\theta) & s_y sin(\theta) & x - c_x s_x cos(\theta) - c_y s_y sin(\theta) \\ -s_x sin(\theta) & s_y cos(\theta) & y + c_x s_x sin(\theta) - c_y s_y cos(\theta) \\ 0 & 0 & 1 \end{bmatrix} \tag{3.1}$$

$$ATM = \begin{bmatrix} A00 & A01 & Tx \\ A10 & A11 & Ty \\ 0 & 0 & 1 \end{bmatrix} \tag{3.2}$$

While $A_{00} = s_x cos(\theta)$, $A_{01} = s_y sin(\theta)$, $A_{10} = -s_x sin(\theta)$, $A_{11} = s_y cos(\theta)$, $T_x = x - c_x s_x cos(\theta) - c_y s_y sin(\theta)$ and $T_y = y + c_x s_x sin(\theta) - c_y s_y cos(\theta)$. We used notations A_{00}, A_{01}, A_{10}, A_{11}, T_x, T_y for the predicted parameters. The extracted centerlines of the arteries are considered bright point sets. Every centerline point set in each frame is registered to the previous frame in a sequence using the CPD algorithm.

3.2.3 LSTM models for sequence prediction

Long Short-Term Memory (LSTM) is a Recurrent Neural Network (RNN)-based architecture with optimized memory, which can solve the vanishing and exploding gradient problem in conventional RNNs. LSTM structures have memory blocks which include memory cells that can store the temporal information of sequential data as well as specific multiplicative units, called gates, to control the flow of information. Each memory block contains an input gate to control the flow of input activations into the memory cell, an output gate to control the output flow of cell activations into the rest of the network, and a forgetting gate Azzouni & Pujolle (2017). Therefore, an LSTM network can keep only the necessary information from the past and forget the rest, thus optimizing its memory.

Initially, the model could predict the transformation matrix for a single future frame given the previous frames using a many-to-one LSTM structure such that given a sequence of frames as an input, we were expecting one single frame as output. The new proposed model can forecast multiple values in the future after receiving multiple inputs to improve time complexity. Fig.3.4 and 3.5 show the structures of many-to-one and many-to-many frame prediction, respectively.

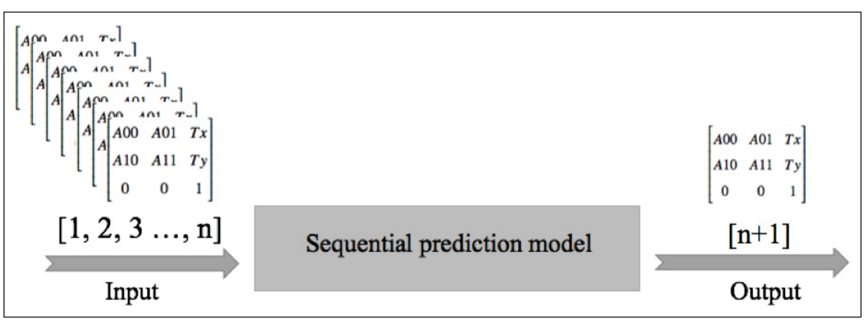

Figure 3.4 The many-to-one structure of the sequential prediction

In both models, the input for the model is a sequence of transformation matrices extracted from the X-ray images by registering the consequent frames in pairs, and the LSTM network is trained to predict the arteries' transformation in the next frame/s from the previous ones.

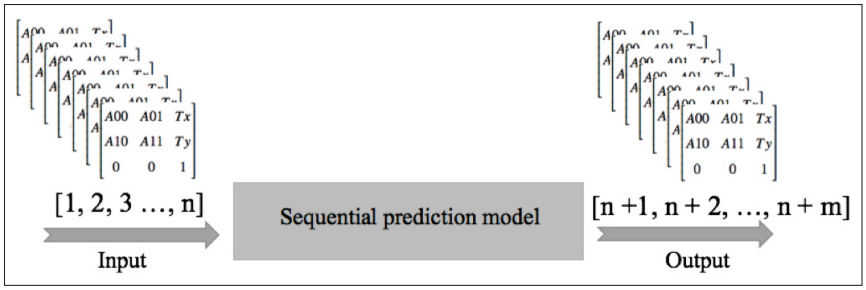

Figure 3.5 The many-to-many structure of the sequential prediction

Considering $N = 6$ as the number of 2D affine transformation parameters representing translation, rotation, shearing, and scaling $(Tx, Ty, A00, A01, A10, A11)$, and T as the number of transformation matrices. To effectively feed the models, we sort the parameters in a vector X^t of size $N * T$. This vector is called the transformation parameter vector (TP). Then, the values in the TP vector are normalized to be fed into the network. The normalization was required since the range of values for some parameters is so small or big, and in that case, the network can not learn or converges slowly. Then, at the end of the prediction, they can be de-normalized to have real values. Now, the prediction problem is defined as solving the predictor of X^t (denoted by $\hat{X}^t$) via a series of previously measured TP vectors.

Compared to a many-to-one model, a many-to-many model was developed to predict all of the $N = 6$ transformation parameters at the same time for multiple matrices in the future. To feed the input matrix sequences to the model, a fixed number of previous frames (matrices) was considered as a time window to learn from to predict the new TPs in the future (Fig.3.6).

Fig.3.7 shows the deep LSTM model structure. First, the input image sequence is pre-processed, the TP vectors are extracted, and then the deep LSTM model is fed by multiple TP vectors. ,A linear regression layer passes the output of the N-layer LSTM model and the model's output is multiple future TP vectors.

3.2.3.1 Assessing metric

We used Mean Absolute Error (MAE) between the ground truth and predicted values for transformation parameters to assess the quality of the predictions because this metric is not overly sensitive to outliers and can simply evaluate the overall error. Given that the segmentation is not perfect, some parts of vessel centerlines in the frames may be lost or extended by noise. Hence, we had to choose a metric to be less sensitive to this problem while we compare the predicted values to the ground truth resulting in CPD registrations. Additionally, we first calculated the distance transform of the original centerlines image to evaluate the overall error of predicted transformed centerlines. For each pixel of the background, we obtained its distance to the closest centerline point. The distance transform or distance field for each white pixel on the extracted centerline assigns a number that is the distance between that pixel and the nearest nonzero pixel of the vessels. Thus, to calculate the final distance, we projected the predicted transformed centerline on the distance transform matrix and averaged the obtained values as an overall prediction error.

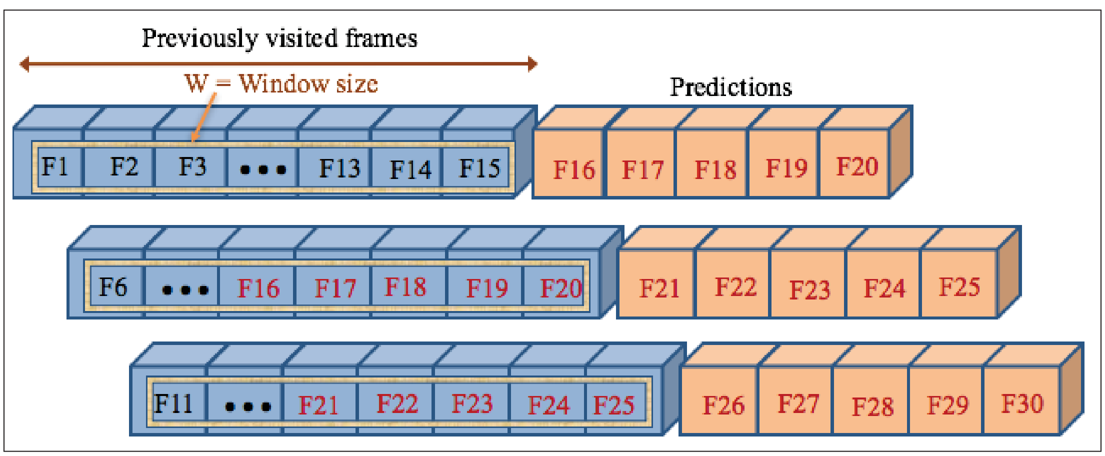

Figure 3.6 The many-to-many LSTM model. All TP elements are predicted at the same time. $W = 15$ is the window size and the number of outputs or predictions is $P = 5$

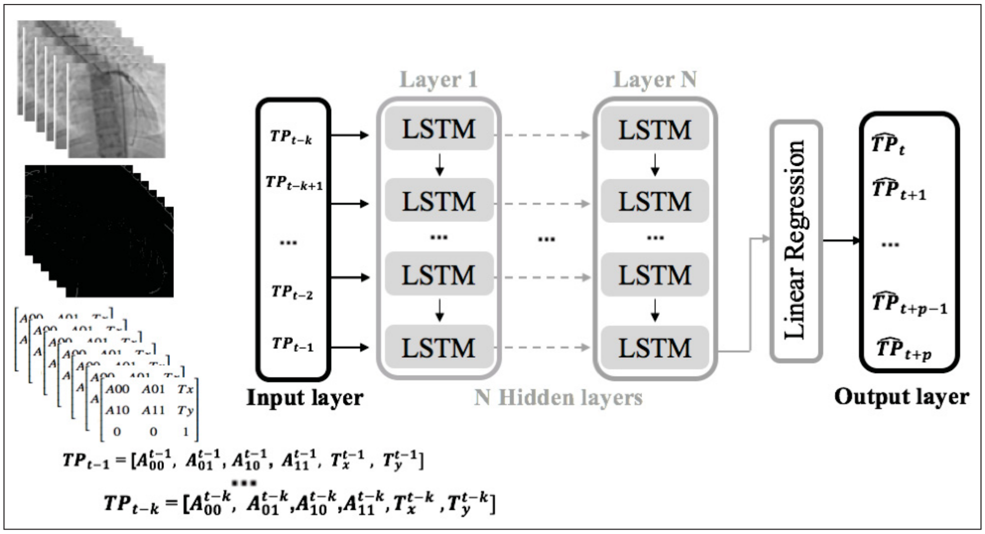

Figure 3.7 The deep LSTM model structure for multiple TP predictions

3.3 Experiments and results

We used 64 patients (pediatric and adult) simulated in normal and abnormal modes for different motions (cardiac only, respiratory only, and both motions). The sequences simulated with both motions were 75 frames in length, while other sequences having only cardiac or only respiratory motion were generated using a 150 frame length. Additionally, 10 sequences from the patient dataset with between 84 and 166 frames in length were selected based on the visibility of contrasted arteries or medical devices through the sequences. The motion features extracted from the centerlines of the segmented arteries were represented as 6 affine transformation parameters collectively called the TP vector (translations Tx, Ty, and rotation, shearing, scaling A00, A01, A10, A11). For each sequence, the LSTM model was trained separately with a sequence of TP vectors, while 80 percent of each sequence was considered as the training set and 20 percent as the testing set. TP vectors were normalized by dividing by the maximum value in each TP vector. Since the prediction is considered a regression problem, we used a linear activation function for our model and the RMSProp as an optimizer for compilation. Based on the experiments, the optimal number of epochs within the 100 to 1000 range was 500 for the simulated data with a length of 150 frames and 200 for sequences with 75 frames. We trained the patient data sets with 500 epochs as well.

Keras library was used to build and train the model. The accuracy of the method was evaluated by computing the Mean Absolute Error (MAE) between the predicted values using our model and the results of the CPD registration as the ground truth (Table. 3.1, Table. 3.2). Fig.3.9 and Fig.3.8 depict the predictions of TP for a simulated data sample, and Fig.3.10 and Fig.3.12 show a patient data TP prediction sample. To evaluate the overall error of predicted transformed centerlines, we first calculated the distance transform of the original centerline images. For each background pixel, we obtained its distance to the closest centerline point. The distance transform or distance field for each white pixel on the extracted centerline assigns a number, the distance between that pixel, and the nearest nonzero pixel of the vessels. Thus, to calculate the final distance, we projected the predicted transformed centerline on the distance transform matrix and averaged the obtained values as the overall prediction error. We applied the predicted parameters in matrix form to transform the arteries' centerlines and overlaid the transformed centerlines with predicted parameters on the distance transform of the extracted centerlines from the original images. Fig.3.13, Fig.3.14, and Fig.3.15 show the overlay of the transformed segmented vessels with predicted transformation parameters and the original transformed images for simulated and patient data, respectively. Therefore, we first evaluated our results by computing the Mean Absolute Error between the predicted TP values and the ground truth TP (resulting from CPD registration) (Tables 3.1 and 3.2), and then we compared our results to the original images by applying for the CPD registration on the originally extracted and predicted centerlines (Table 3.3). Apart from the simulated data, we validated our results by applying the method on 10 real patients, with a more realistic and noisy segmentation of the moving objects. As shown in the results presented in Table 3.3, we obtained a low accumulated error for the prediction of the transformation matrix using the distance transform of the original segmented arteries for both simulated data with different motions as well as the patient data samples.

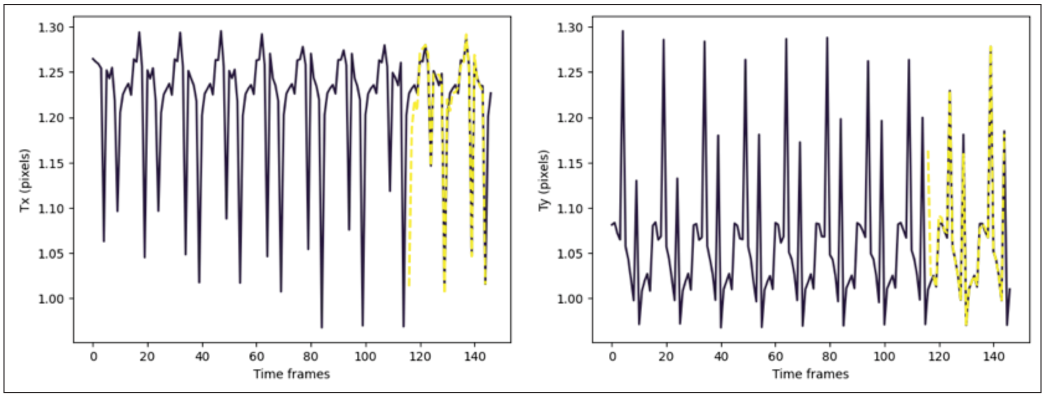

Figure 3.8 Prediction of 2D translation parameters for only cardiac motion (simulated). The ground truth values are shown in purple, while the yellow dashed lines show the predictions.

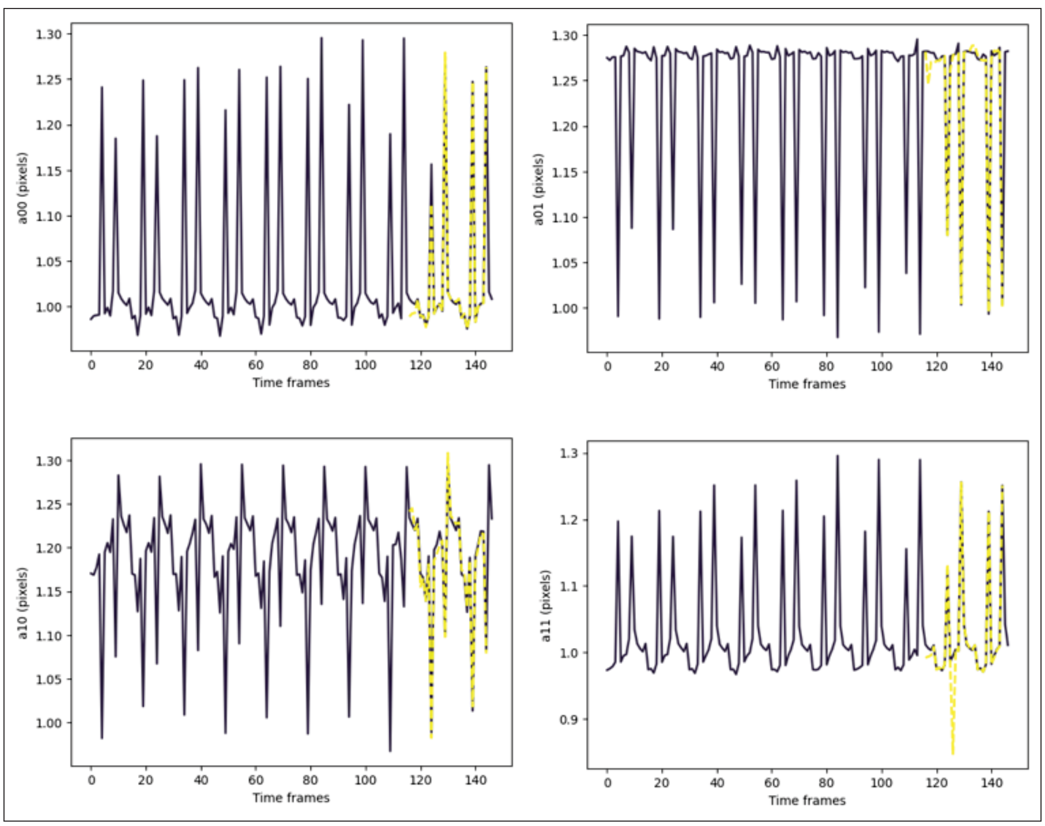

Figure 3.9 Transformation parameters predictions (simulated only cardiac motion). The ground truth values are in purple, and the predicted values are in yellow dashed lines.

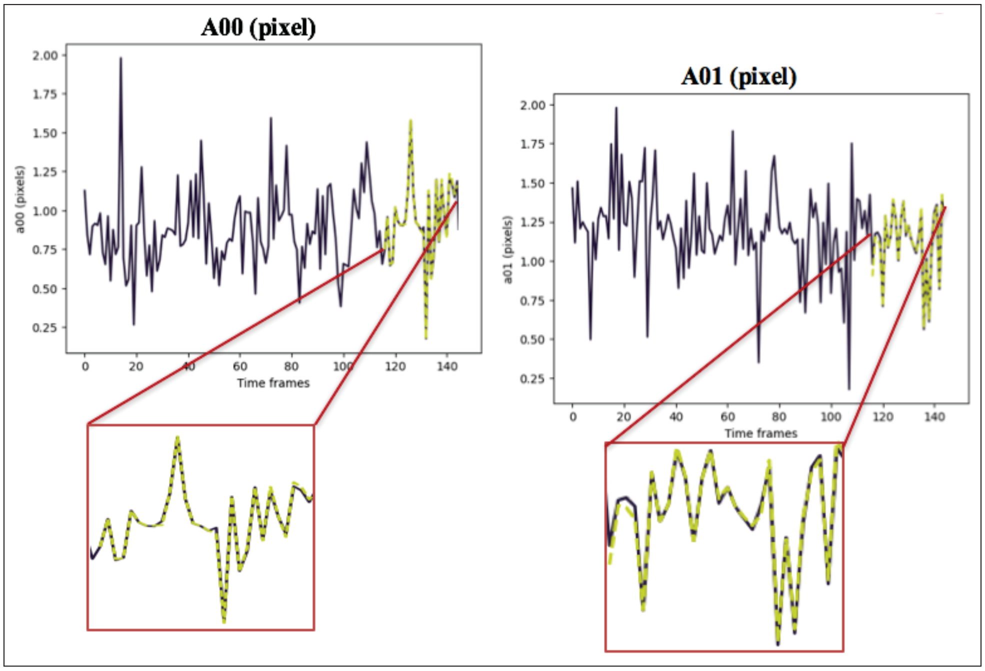

Figure 3.10 Transformation parameters predictions (patient data). The ground truth values are in purple, and the predicted values are in yellow dashed lines.

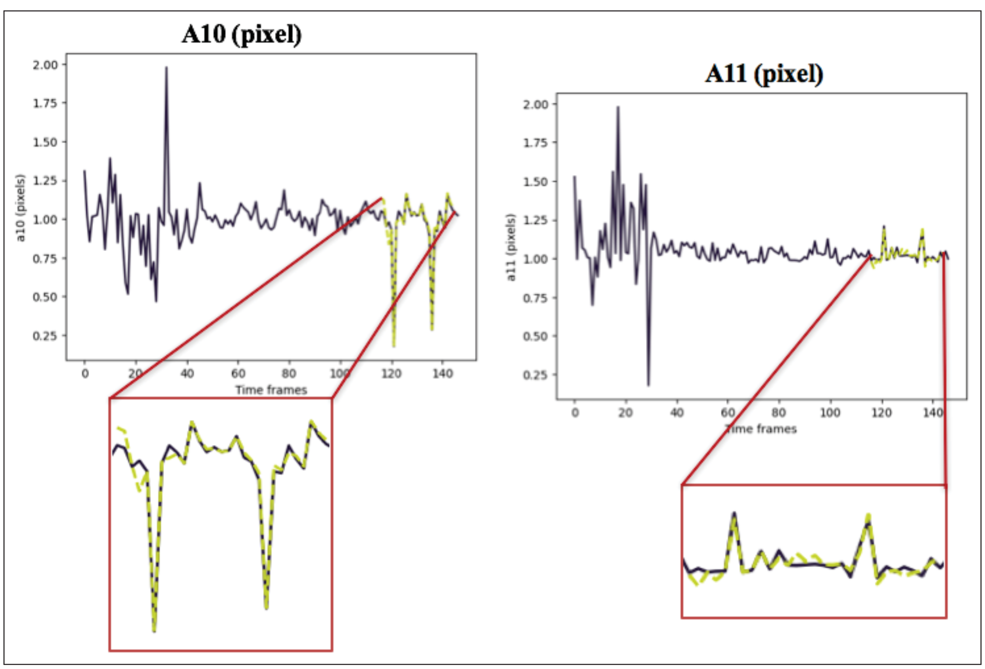

Figure 3.11 Transformation parameters predictions (patient data). The ground truth values are in purple, and the predicted values are in yellow dashed lines.

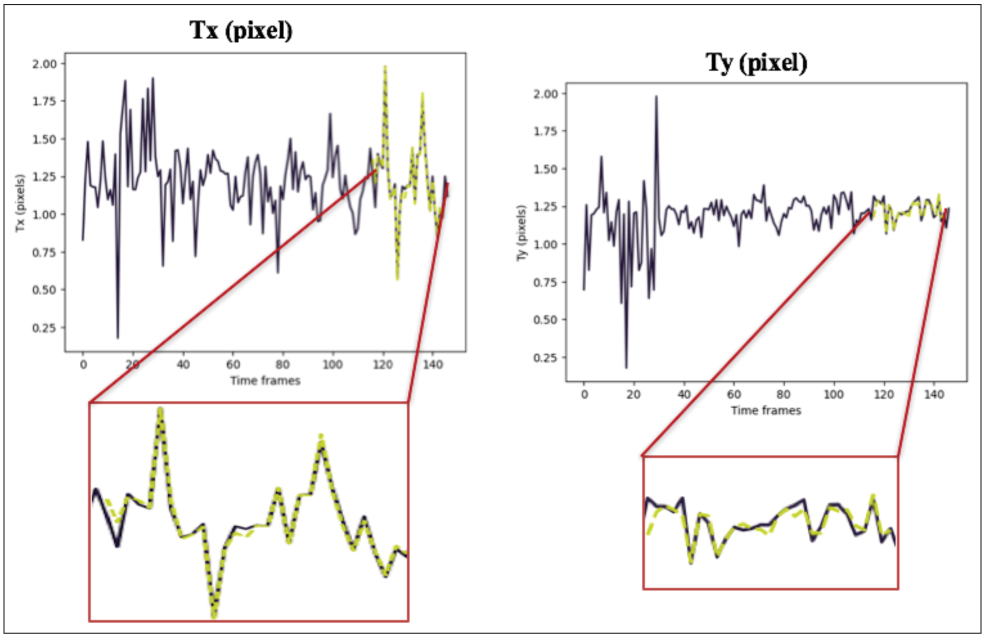

Figure 3.12 Transformation parameters predictions (patient data). The ground truth values are in purple, and the predicted values are in yellow dashed lines.

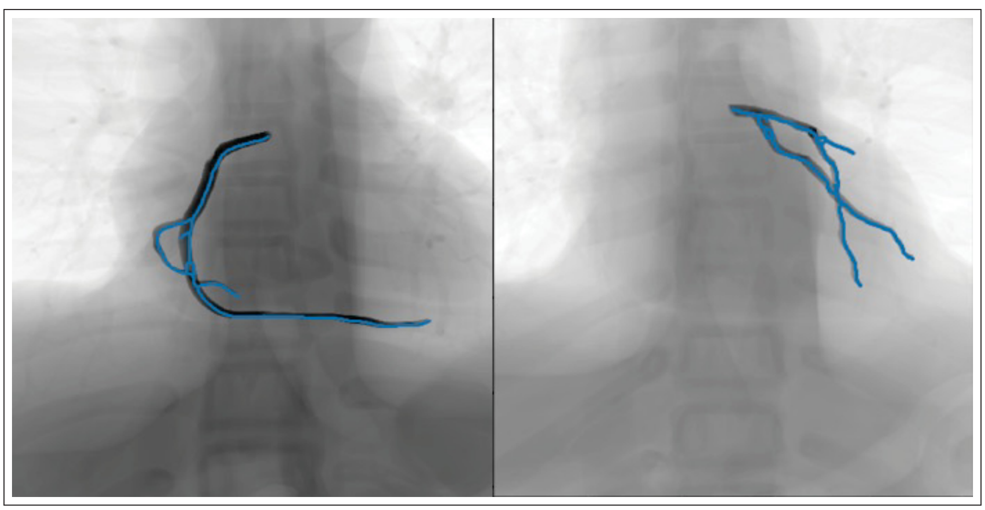

Figure 3.13 Simulated data example for overlaying the transformed LCA and RCA arteries with predicted transformation parameters (blue colored center-lines) with the original transformed images

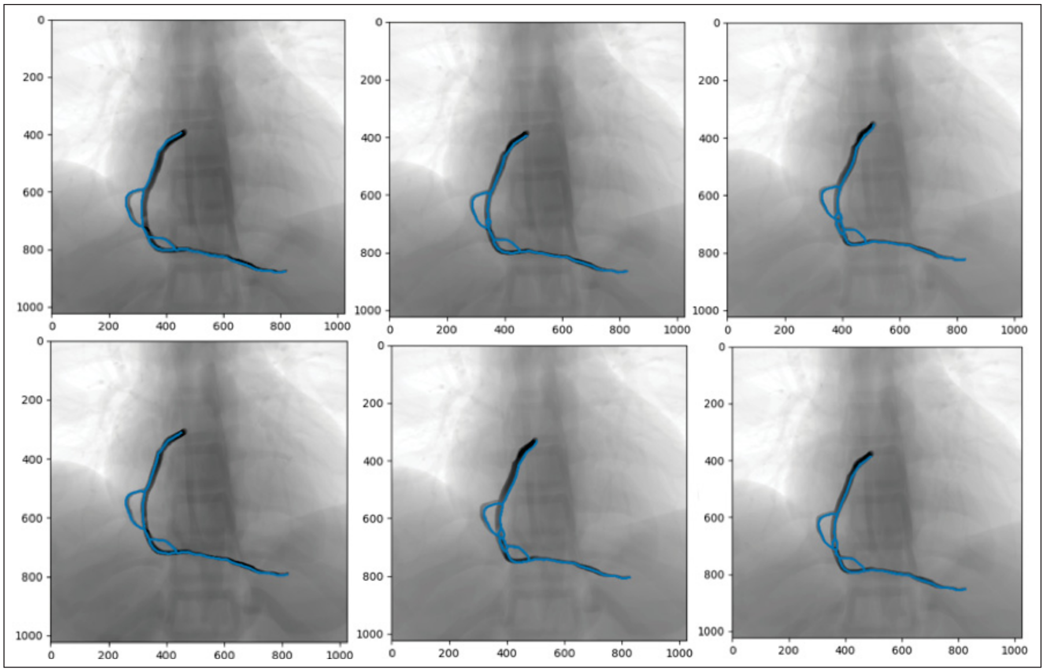

Figure 3.14 Simulated data example for overlaying the transformed arteries in 6 consecutive frames with predicted transformation parameters (blue colored center-lines) and the original transformed images

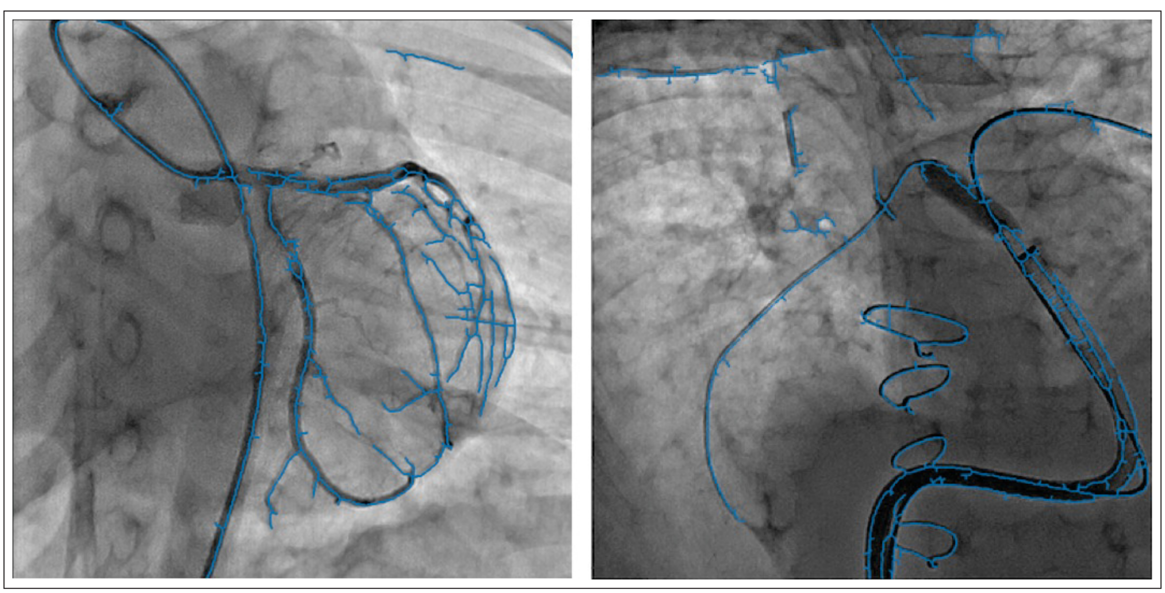

Figure 3.15 Two different Patient data samples for overlaying the transformed vessels with predicted transformation parameters (in blue) with the original transformed images

Table 3.1 The average overall simulated data MAE between the ground truth and predicted values

Both motions						
MAE (pixel)	**Tx**	**Ty**	**a00**	**a01**	**a10**	**a11**
Mean	0.24	0.21	0.32	0.35	0.32	0.32
Max	0.41	0.42	0.40	0.51	0.52	0.51
Min	0.0012	0.0084	0.06	0.074	0.029	0.064
Only cardiac motion						
Mean	0.13	0.11	0.14	0.12	0.13	0.13
Max	0.35	0.43	0.42	0.42	0.44	0.45
Min	0.0002	0.004	0.003	0.06	0.001	0.05
Only Respiratory motion						
Mean	0.10	0.13	0.11	0.14	0.11	0.15
Max	0.21	0.43	0.37	0.50	0.41	0.45
Min	0.0001	0.0001	0.002	0.0012	0.005	0.004

Table 3.2 The average overall patient data MAE between the ground truth and predicted values.

Patient data						
MAE (pixel)	**Tx**	**Ty**	**a00**	**a01**	**a10**	**a11**
Mean	0.19	0.20	0.20	0.20	0.18	0.17
Max	0.55	0.45	0.54	0.55	0.39	0.47
Min	0.025	0.03	0.01	0.00	0.02	0.01

3.4 Discussion

We have presented a learning-based patient-specific cardio-respiratory motion prediction method using a simple LSTM model. This model can predict the temporal and spatial changes for moving objects (contrasted arteries and/or medical devices) in a sequence. The model is computationally efficient, simple to implement, and fast to train.

Table 3.3 The average over all samples distance transform error of the original centerline image to the predicted transformed one in mm.

Mean DT Error(mm)	Cardaic	Respiratory	Both	Patient data (Both)
	0.33 mm	**0.47 mm**	**0.39 mm**	**0.58 mm**

The proposed motion prediction method was validated on a wide range of simulated cardiac and respiratory rates for pediatric and adult phantoms and a patient dataset. According to our results, the transformation parameters for moving objects, including medical devices in a sequence, are predicted well, with a very low amount of error (mostly less than a few pixels in the images). Medical devices such as guide wires or catheters were considered moving objects in a sequence for the patient data since they are continuously visible and contrasted under the X-ray images. Hence, the visible devices can represent the movement of arteries, while the contrasted arteries gradually become fade through a sequence as they lose their contrast. Motion prediction is thus not completely dependent on the visibility of the contrasted arteries with the contrast agent. As a robust prediction algorithm, the motion tracking approach can be applied to estimate future heart trajectories in robotic-assisted operations on a free-beating heart. This method is less sensitive to low X-ray dosage and image quality than other motion-tracking methods that track the heart edges.

Moreover, this approach requires no additional imaging modality and preoperative information for motion tracking. Not only this model-free cardio-respiratory motion prediction approach can facilitate the navigation process of cardiac interventions but also potentially aids in reducing the required amount of contrast agent and radiation dose for cardiac interventions.

The accuracy of the prediction indirectly depends on the accuracy of the segmentation and registration algorithms in the pre-processing steps. Future work will include investigating a non-parametric and accurate method for automatizing the segmentation in our pre-processing steps. Moreover, we are planning to integrate this motion prediction approach into an End-to-End system for X-ray image prediction.

CHAPTER 4

CONTRIBUTION #3: SEX-SPECIFIC CARDIAC MOTION BEHAVIOR ANALYSIS BASED ON XCAT SIMULATED ANGIOGRAPHY

4.1 Proposed contribution

A general comprehension of the differences between males and females for cardiovascular diseases is still missing. It has been proved that sex-specific cardiac and vascular aging patterns are significant issues that need to be considered for diagnosing and treating cardiovascular diseases (Merz & Cheng, 2016). In general, the female heart is smaller than the male heart (usually one-fourth of the men's heart). Accordingly, the number of heartbeats per minute (bpm) for adult men is between 70 to 72 bpm, while for adult women is between 78 to 82 bpm. Hence, the heart cycle is shorter for women (average 0.75 sec) compared to men (average 0.85 sec). Apart from size differences, there are geometric differences such as heart mass, left and right ventricular mass, and wall thickness (Pierre, Peirlinck & Kuhl, 2022).

Ischaemic Heart Disease (IHD) is one of the most common cardiovascular diseases and is the cause of death in men and women nowadays. Based on many pieces of research, the risk for IHD is significantly lower for premenopausal women compared to their counterparts men. However, the risk of IHD for women increases after menopause. Moreover, it has been experimentally confirmed that there are sex-specific differences in cardiac tolerance to ischemia (Ostadal & Ostadal, 2014). In the past 20 years, clinical and experimental research on investigating sex-related differences has grown significantly. This incremental interest reflects the significance of this topic and the extreme necessity for comprehending the sex determinants of outcomes and minimizing the bias in the management and treatment of IHD in women (Ostadal & Ostadal, 2014).

Compared to men, there is a greater prevalence of heart failure with preserved ejection fraction for female patients, and this is why the differences between males and females are important for age-related cardiac remodeling. Moreover, how coronary artery disease appears and develops typically or atypically over the life course mainly depends on sex variation. Hormonal and non-hormonal factors underlay the sex variance for cardiovascular aging (Merz & Cheng, 2016).

Different factors affect the risk for cardiovascular diseases, such as aging, gender, frailty, obesity, and diabetes. Among these factors, age and gender are the most fundamental factors to be considered. The elderly population is usually at more risk of being exposed to cardiac disease. Based on pieces of evidence, older females are more exposed to cardiovascular problems than age-matched males, and for both males and females, the risks associated with cardiovascular disease enhance with age. This fact corresponds to an overall decrease in sex hormones, primarily estrogen and testosterone (Rodgers *et al.*, 2019).

This research investigated the general male and female heart motion differences represented by transformation parameters in simulated X-ray sequences. This study's motivation is to develop a general sex-specific motion prediction model. Previously we introduced a patient-specific motion perdition approach for facilitating the navigation process of the cardiac intervention. This research was focused on considering the sex diversity in cardiac motion behaviors and was submitted to IJCARS 2022.

4.2 Methodology

4.3 Sex specific population simulation using 4D XCAT phantom

The 4D XCAT phantom was developed to provide virtual patients for medical imaging research. It provides detailed whole-body anatomies for males and females using a nonuniform rational B-spline. Visible Male and Female anatomical datasets were segmented and used to generate subdivisions of highly flexible surfaces. The visible human anatomies were matched to the body measurements and organ volumes of adult males and females. Thousands of anatomical structures and cardiac and respiratory motion models can be simulated in a series of computerized phantoms for adult and pediatric patients. For each phantom, there is a parameterized model and anatomy. Pediatric and adult reference models were used to generate the pediatric and adult phantoms, including females and males of various ages (1 to 78 years old), heights, body masses, and ethnicity.

High-resolution Computed Tomography (CT) positrons were selected and validated by an experienced radiologist regarding anatomic regularity. The target model using nonuniform rational B-spline surfaces was formed by segmenting the CT portions. The cardiac and respiratory motions were defined for the targets based on the calculated transformation for matching the pediatric XCAT reference phantom and the patient target using a Multichannel Large Deformation Diffeomorphic Metric Mapping (MC-LDDM) algorithm. The complete phantoms were then manually inspected for anatomical precision. The mass for each main tissue was computed and evaluated for different ages. The generated pediatric phantoms were as detailed as the original XCAT reference phantoms, and the cardiac and respiratory motions were simulated with parameterized models. Male and female anatomies differed by including both sets of reproductive organs for the phantoms that were 10 yr old and younger (cut-off age). The phantom models include wide anatomical variation in organ shape and size for models of the same age and sex. One limitation of the pediatric phantoms generated by XCAT is that not every structure is defined by segmenting the patient data. The MC-LDDMM algorithm estimated the structures that were not visible in the patient data by deforming a pediatric template to match the segmented patient framework (Segars *et al.*, 2015) (Segars *et al.*, 2010) (Segars *et al.*, 2013).

4.3.1 Data simulation with XCAT

Using the XCAT simulator, we generated two sets of datasets. First, we have simulated X-ray angiography sequences for 33 patients (17 females and 16 males) older than the cut-off age, which is 10 yr and ranges between 11 to 45 yr, including the contrasted left and right coronary arteries. This dataset was generated based on the default parameter setting for the heart cycle. The default parameter setting for a normal length of beating heart cycle in males and females and all ranges of ages is equal to 1 second. A general blood pressure heart curve is defined by two heart geometries (End Diastole (ED) and End Systole (ES)). Starting with the ED phase and four other parameters, including ES (Fig 4.1), the heart curve is generated for motion simulation in all phantoms while the time duration to complete the heart curve is considered as 1 second. The second simulated dataset included only 20 patients (10 females and 10 males) over 50 years old with different heart curves for males and females. All of the generated sequences had a length of 150 frames. Then we investigated the cardiac motion behavior represented by 2D affine transformation matrices resulting from the frame-by-frame sequence registration. The respiratory motion was removed, and only the cardiac motion was simulated in X-ray sequences. The arteries were segmented and registered in pairs using point cloud registration for all of the frames in each sequence. The registration matrices, including the affine transformation parameters, were saved for each sequence.

time (0 to 1)	heart_curve	1-heart_curve		
0	0	1	ED Phase	hrt_v1
0.038461538	0.055	0.945		
0.076923077	0.2094	0.7906		
0.115384615	0.3798	0.6202		
0.153846154	0.4866	0.5134		
0.192307692	0.5747	0.4253		
0.230769231	0.6542	0.3458		hrt_t1
0.269230769	0.748	0.252		
0.307692308	0.8303	0.1697		
0.346153846	0.888	0.112		
0.384615385	0.9422	0.0578		
0.423076923	0.993	0.007		
0.461538462	1	0	ES Phase	hrt_v2
0.5	1	0	ES Phase	hrt_v2
0.538461538	0.978	0.022		
0.576923077	0.93	0.07		hrt_t2
0.615384615	0.792	0.208		
0.653846154	0.582	0.418		
0.692307692	0.376	0.624		hrt_v3
0.730769231	0.281	0.719		hrt_t3
0.769230769	0.226	0.774		
0.807692308	0.174	0.826		hrt_v4
0.846153846	0.136	0.864		
0.884615385	0.112	0.888		hrt_t4
0.923076923	0.0975	0.9025		
0.961538462	0.0614	0.9386		
1	0	1	ED Phase	hrt_v5

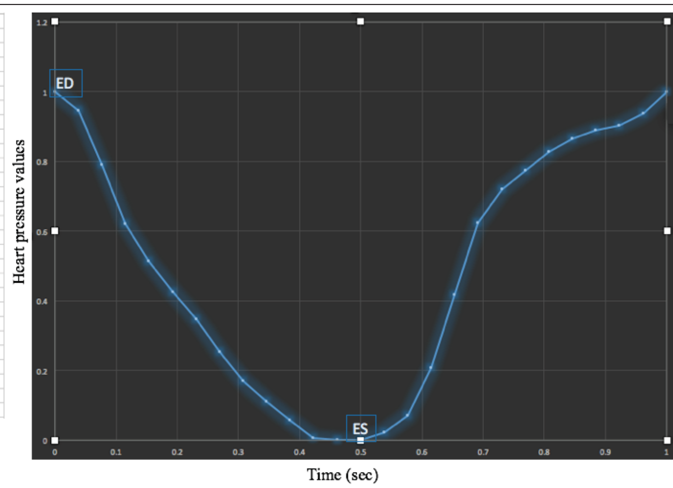

Figure 4.1 The default XCAT heart curve (ventricle (LV) volume). ED and ES are the two main parameters (heart geometries) for defining the heart curve.

We generate sex-specific heart curves for adult males and females based on the differences in heart rates between men and women. The average heart rate for adult men is 75 bpm and for adult women is 80 bpm; accordingly, the amount of time to complete a heart curve is 0.84 sec and 0.75 sec for males and females, respectively. Fig. 4.2 shows the difference between the default heart curve for XCAT and new sex-specific heart curves.

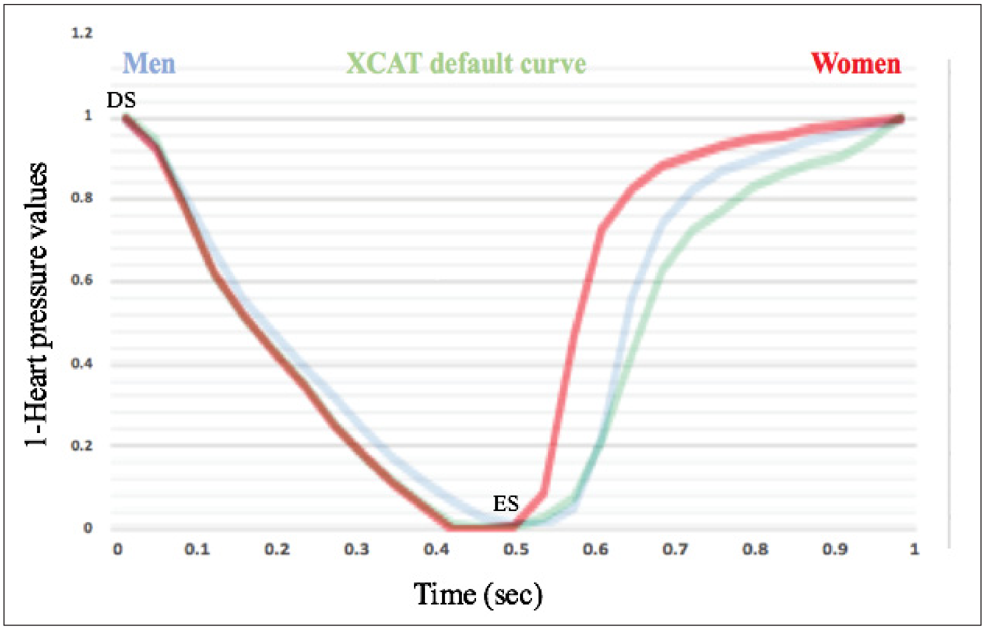

Figure 4.2 The default XCAT parameters setting for the heart curve is in blue. The women's heart curve is in red, and the males' heart curve is in green.

4.3.2 Data analysis and results

For the first dataset, according to our previous work, the 2D affine transformation parameters of visible arteries in the X-ray images through a sequence represent cardiovascular motion. To obtain the arteries displacements from frame to frame in a sequence, we first segmented the arteries, extracted the center lines for each frame, and applied Coherent Point Drift (CPD) algorithm to register the center-lines frame by frame through the sequence. The results of the registration with CPD are sequences of 2D affine transformation matrices. The transformation metrics contain information about translation in 2D (Tx and Ty), rotation, scaling and shearing ($A00$, $A01$, $A10$, $A11$). We compared the heart motion behavior of male and female patients using the 2D transformation matrices. In Tables (4.1) and (4.2), we averaged the transformation parameters' values over patients for males and females and computed the average range of displacements in 2D for both groups.

The range of motion for the male group is approximately 20 pixels in x direction and 68 pixels in y direction, while for the female group, the range of motion in x direction is about 54 pixels and 60 pixels in y direction (Table (4.1) and Table (4.2)). Thus, according to these results, the range of motion for female patients in x is more than for male patients. For the other transformation parameters ($A00$, $A01$, $A10$, $A11$), the values for males and females are slightly different (approximately a pixel).

Fig.4.3, Fig.4.4, and Fig.4.5 show the comparisons between the average values for transformation parameters over the male and female patients. Although there are outliers in both females and males, each category has a specific pattern associated with each transformation parameter. Moreover, it can be seen that for female patients, the irregularity for the motion parameters is seen more than for men patients (A00 and A10), and for females, the range of values is bigger than for males. Regarding translation parameter differences between males and females, there is more visible diversity between different sexes in the x direction compared to the y direction.

Table 4.1 The average range of values for 2D translation of the arteries in a sequence for both males and females

	Mean Tx pixel		**Mean Ty pixel**	
	Max	**Min**	**Max**	**Min**
Males	6.918	-13.725	26.300	-41.937
Range	**20 pixels**		**68 pixels**	
Females	9.400	-44.580	26.600	-33.915
Range	**53.98 pixels**		**60.515 pixes**	

Table 4.2 The average range of values for 2D transformations (A00, A01, A10, A11) of the arteries in a sequence for both males and females

	Mean A00 pixel		**Mean A01 pixel**		**Mean A10 pixel**		**Mean A11 pixel**	
	Max	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**
Males	1.020	0.984	0.017	-0.016	1.012	-0.005	0.029	0.983
Range	**0.036 pixels**		**0.033 pixels**		**1.017 pixels**		**1.063 pixels**	
Females	1.070	0.991	0.009	-0.005	0.014	-0.013	1.068	1.002
Range	**2.061 pixels**		**1.05 pixels**		**0.027 pixels**		**2.07 pixels**	

In our first experiments, we included male and female patients of different ranges of ages (12 yr to 45 yr). We considered all patients older than the cuff of age (10 years old). Thus for each group of men and women, there exist different ages, patient body weights, heights, and ethnicity. There are outliers for both categories with a bigger range of values for the motion. One reason for having these outliers can be the effects of other mentioned parameters (age, weight, height, ethnicity). The other reason for having these values for each group can be the registration failure using the CPD algorithm.

Complementary to our previous work on predicting the 2D affine transformation of coronary arteries in the X-ray angiography sequences, we investigated the differences between females and males in our predictions. The parameter predictions for affine transformation with our LSTM model worked well for patient-specific motion prediction. To develop a more general (sex-specific) model, we compared the parameters prediction errors in different groups (male and female). Based on the presented results in Table. 4.3, the transformation parameter prediction errors for females were slightly bigger than the prediction error for men.

Table 4.3 Comparison of transformation parameters prediction error (MAE) between the average of 17 females and 16 males with a range of 12 to 45 yr ages

Females						
MAE (pixel)	**Tx**	**Ty**	**a00**	**a01**	**a10**	**a11**
Mean	0.17	0.26	0.26	0.16	0.19	0.15
Max	0.57	0.53	0.54	0.58	0.71	0.55
Min	0.005	0.023	0.021	0.012	0.003	0.004
Males						
Mean	0.13	0.12	0.10	0.12	0.11	0.11
Max	0.49	0.47	0.51	0.51	0.46	0.44
Min	0.002	0.004	0.005	0.007	0.006	0.005

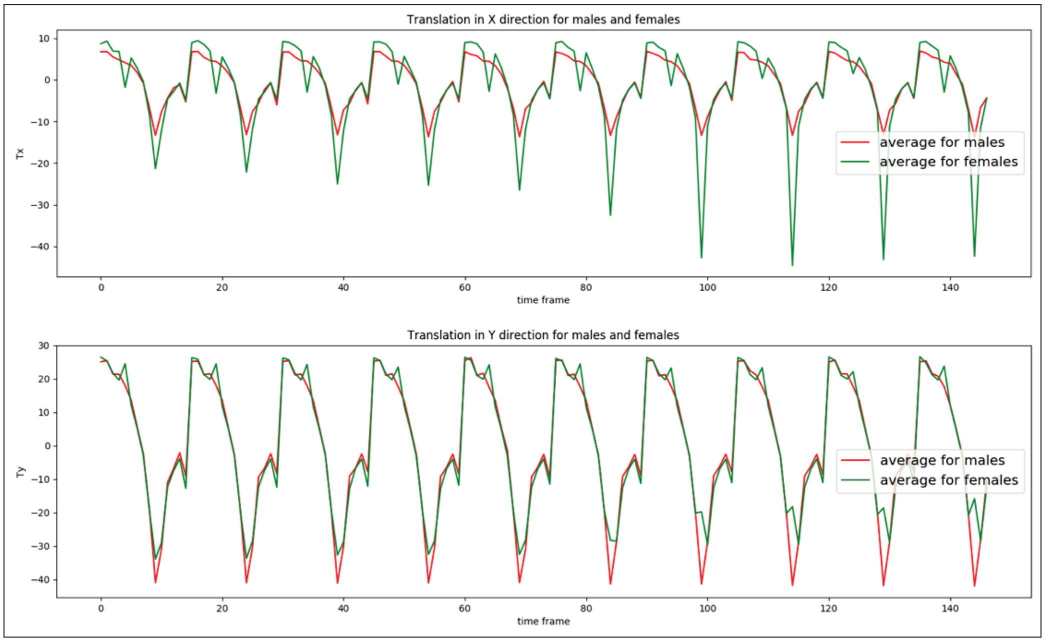

Figure 4.3 Average overall females and males for 2D translation parameters in pixel

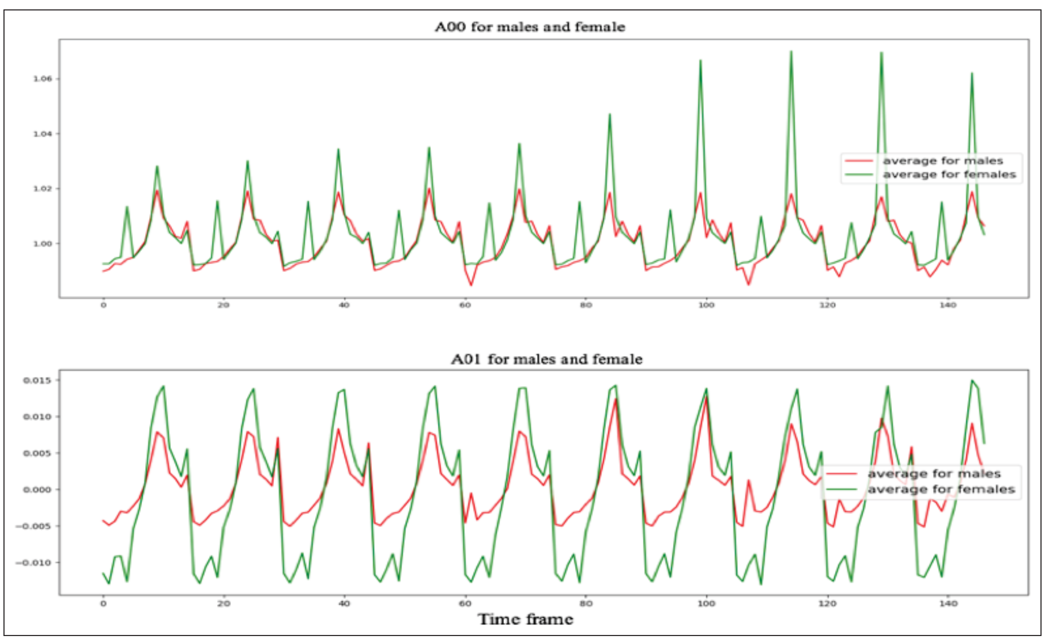

Figure 4.4 Average overall females and males for 2D transformation parameters (A00 and A01) in pixel

In the second experiment with the second simulated data, we differentiated the heart curves for males and females for 50 yr. Figs. 4.6 and 4.7 show the average overall females and males over

50 yr for 2D transformation parameters in pixels. Our results show that for the population over 50 yr, there is not much difference between males and females for transformation parameters in a sequence. However, there is a shift between male and female transformation parameter values because of different heart curves. Based on Figs. 4.6 and 4.7, the pattern for the motion features for males and females can be almost the same while having a faster heartbeat for females compared to males.

4.4 Discussion and conclusion

In this chapter, we investigated the differences between males and females in cardiac motion behavior based on two datasets. Cardiac motion in a sequence was represented as the 2D registration matrices, including translation, rotation, scaling, and shearing information. Our results show that not only are these parameters different between men and women, but the prediction error for predicting the future frames of angiography sequences is slightly higher for women than men. This fact can be considered for developing a general sex-specific cardiac motion model.

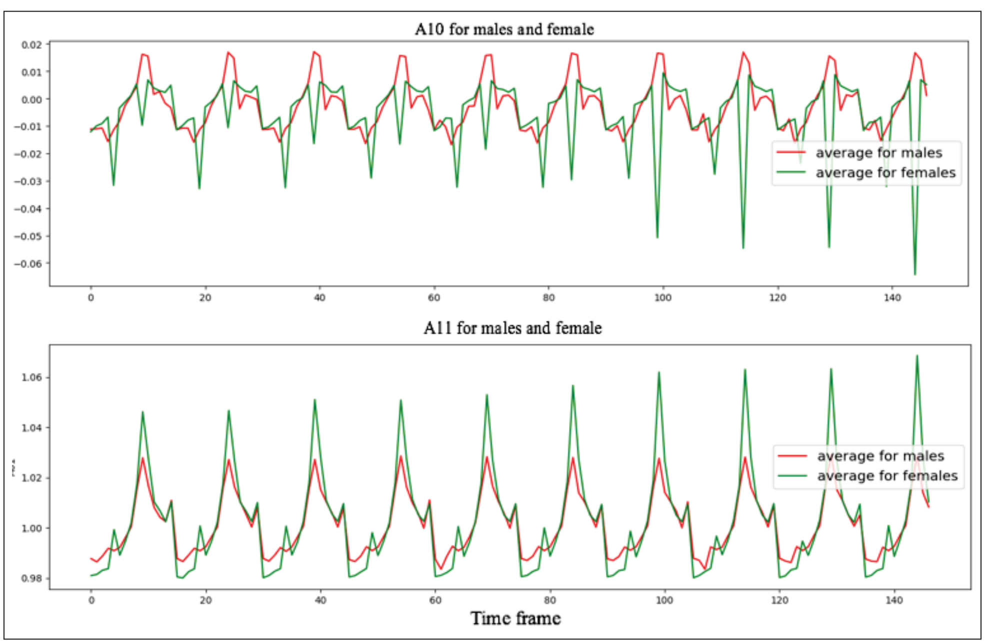

Figure 4.5 Average overall females and males for 2D transformation parameters (A10 and A11) in pixel

The sequences of 2D registration matrices resulted from the CPD algorithm. There were pre-processing steps to remove the background noises, segment the arteries, extract the center lines of the arteries, and finally register the consequent frames of a sequence in pairs. Therefore, other pre-processing steps might include errors in the resulting transformation parameters apart from the registration error.

In the first experiments, we simulated our X-ray angiographies based on the XCAT default parameters for the heart curve. Except for the different geometries (ES and ED) for all of the patients of any age and sex, the same interpolation method was used to create the heart curve, and the time to complete the heart cycle was considered as 1 second. We compared different genders, and based on the results, for males and females, there are differences in the cardiac motion behavior represented by transformation parameters. However, given the fundamental differences between males' and females' hearts, including the size of the heart, the heart cycle, etc., the differences can be more significant if considering the different heart curves for males and females in our simulations.

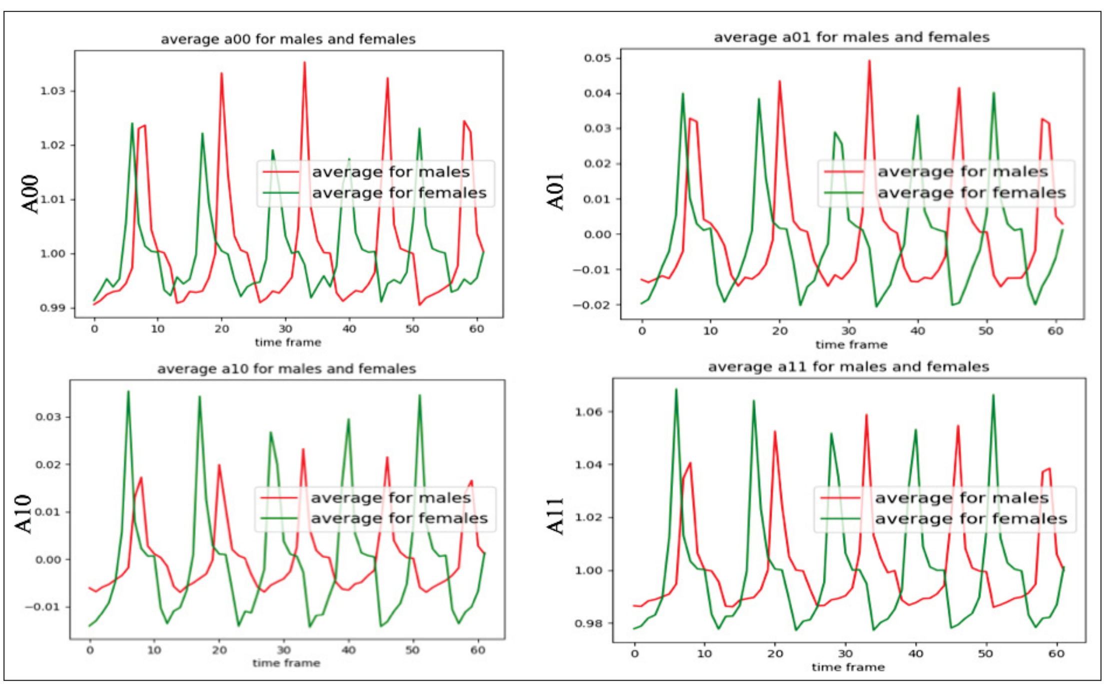

Figure 4.6 Average overall females and males over 50 yr for 2D transformation parameters (A00, A01, A10 and A11) in pixel

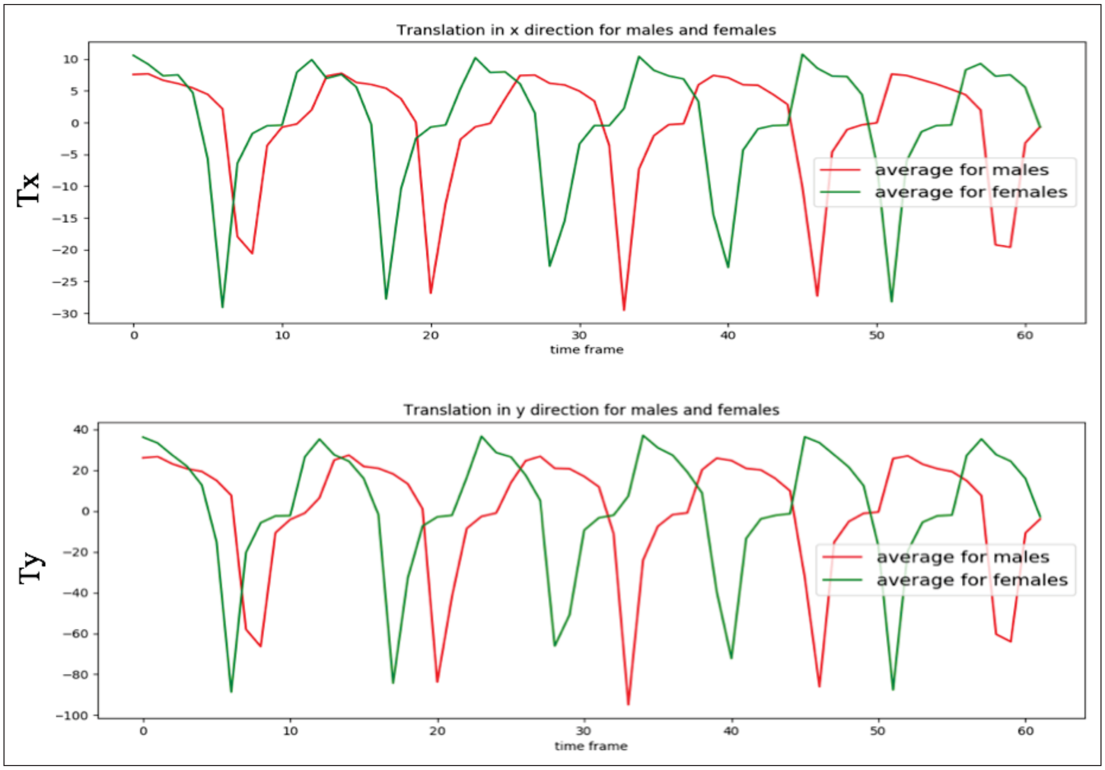

Figure 4.7 Average overall females and males over 50 yr for 2D translation parameters (Tx and Ty) in pixel

The second experiment included patients over 50 yr and with different heart curves for males and females. Based on the results, the heart motion pattern can differ between populations based on different sexes and age groups. Therefore, a more general motion prediction model can be developed considering some parameters such as sex and age group.

CONCLUSION AND RECOMMENDATIONS

Cardio-respiratory motion affects the acquisition of X-ray images during image-based cardiac interventions. Due to the potential risks of ionizing radiation, reducing the pulse rate or X-ray dose would be beneficial for patients and the staff during cardiovascular interventions. A high amount of images per second will produce high-quality sequences for cardiologists at the cost of higher radiation exposure to the patients and staff. The insufficient number of image per second provide poor navigation information.

Our first contribution was to reduce the required radiation dose for the patient and staff while preserving the image quality for the pediatric interventions by developing an end-to-end deep learning model. The learning-based model can predict realistic image frames, including motion and content in an X-ray sequence. Hence, the required number of frames per second (frame rate) and, consequently, the required number of pulses per second in X-ray image acquisitions can be reduced by predicting and generating multiple new frames.

Our second contribution was to estimate and predict the cardio-respiratory motion while considering the sudden and unexpected motions during the interventions. We developed a simple patient-specific LSTM model that tracks the motion signal in 2D X-ray images in a sequence accurately without sensitivity to irregularity in the motion signal. This model-free cardio-respiratory motion prediction approach can facilitate the navigation process but also potentially aids in reducing the required amount of contrast agent and radiation dose for cardiac interventions.

Our last contribution was complementary to the second objective and aimed to find a sex-specific cardiac motion pattern. A more general motion model based on patients' genders can aid the navigation guidance of cardiac interventions. We investigated male and female motion patterns using the simulated dataset with XCAT. We analyzed our previous experiments to find a general solution that provides an accurate sex-specific motion prediction and estimation.

Non-invasive imaging techniques in diagnosing, monitoring, and treating cardiac diseases are essential nowadays. Manufacturers are designing their X-ray systems to minimize the dose. The next generation of X-ray systems might be equipped to expose less radiation dose to the patients and staff and a more accurate navigation process with a predict-ahead X-ray acquisition system. Radiation dose management based on realistic X-ray frame prediction can be a breakthrough improvement for X-ray device technology. Using the predicted images can reduce the required pulses without affecting the image quality. The average required acquisition can drop dramatically by generating new images without radiation exposure. Moreover, an accurate cardio-respiratory motion prediction approach not sensitive to motion irregularities can assist the interventional navigation process and play a key role in radiation dose reduction. For future work, the inclusion and analysis of other radiation dose parameters, such as air kerma at the reference point (KAP) and DAP in dose reduction measurements, must be considered. Second is the exploration of a precise automatic segmentation of the visible objects (arteries and medical devices) for better extracting the motion features that affect the accuracy of the prediction. The third is investigating how the arteries' movements can be different with a medical device like a catheter or guide wire inside the arteries. Forth, video super-resolution methods can be included in the content predictor to improve the image quality of the predictions.

BIBLIOGRAPHY

Ambrosini, P., Ruijters, D., Niessen, W. J., Moelker, A. & Walsum, T. v. (2017). Fully automatic and real-time catheter segmentation in X-ray fluoroscopy. *International conference on medical image computing and computer-assisted intervention*, pp. 577–585.

Aufrichtig, R., Xue, P., Thomas, C. W., Gilmore, G. C. & Wilson, D. L. (1994). Perceptual comparison of pulsed and continuous fluoroscopy. *Medical physics*, 21(2), 245–256.

Azizmohammadi, F., Martin, R., Miro, J. & Duong, L. (2019). Model-free cardiorespiratory motion prediction from X-ray angiography sequence with LSTM network. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7014–7018.

Azizmohammadi, F., Castellanos, I. N., Miró, J., Segars, P., Samei, E. & Duong, L. (2022). Generative learning approach for radiation dose reduction in X-ray guided cardiac interventions. *Medical Physics*.

Azzouni, A. & Pujolle, G. (2017). A long short-term memory recurrent neural network framework for network traffic matrix prediction. *arXiv preprint arXiv:1705.05690*.

Baka, N., Lelieveldt, B., Schultz, C., Niessen, W. & Van Walsum, T. (2015a). Respiratory motion estimation in x-ray angiography for improved guidance during coronary interventions. *Physics in Medicine & Biology*, 60(9), 3617.

Baka, N., Lelieveldt, B., Schultz, C., Niessen, W. & van Walsum, T. (2015b). Respiratory motion estimation in x-ray angiography for improved guidance during coronary interventions. *Physics in Medicine & Biology*, 60(9), 3617.

Balter, S. (2014). Fluoroscopic frame rates: not only dose. *American Journal of Roentgenology*, 203(3), W234–W236.

Baumgartner, C., Kolbitsch, C., McClelland, J., Rueckert, D. & King, A. (2017). Autoadaptive motion modelling for MR-based respiratory motion estimation. *Medical image analysis*, 35, 83–100. doi: 10.1016/j.media.2016.06.005.

Chida, K., Ohno, T., Kakizaki, S., Takegawa, M., Yuuki, H., Nakada, M., Takahashi, S. & Zuguchi, M. (2010). Radiation dose to the pediatric cardiac catheterization and intervention patient. *American Journal of Roentgenology*, 195(5), 1175–1179.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1), 53-65. doi: 10.1109/MSP.2017.2765202.

Dauer, L. T. (2011). Radiation Dose Management for Fluoroscopically-Guided Interventional Procedures. LWW.

de Roos, A. & Higgins, C. B. (2014). Cardiac radiology: centenary review. *Radiology*, 273(2S), S142–S159.

Doss, M. (2013). The importance of adaptive response in cancer prevention and therapy. *Medical physics*, 40(3), 030401–030401.

Dürichen, R., Wissel, T., Ernst, F., Schlaefer, A. & Schweikard, A. (2014). Multivariate respiratory motion prediction. *Physics in Medicine & Biology*, 59(20), 6043.

El-Sayed, T., Patel, A. S., Cho, J. S., Kelly, J. A., Ludwinski, F. E., Saha, P., Lyons, O. T., Smith, A., Modarai, B., Tyrrell, M. et al. (2017). Radiation-induced DNA damage in operators performing endovascular aortic repair. *Circulation*, 136(25), 2406–2416.

Ernst, F., Dürichen, R., Schlaefer, A. & Schweikard, A. (2013). Evaluating and comparing algorithms for respiratory motion prediction. *Physics in Medicine & Biology*, 58(11), 3911.

Ernst, F. & Schweikard, A. (2009). Forecasting respiratory motion with accurate online support vector regression (SVRpred). *International journal of computer assisted radiology and surgery*, 4(5), 439–447.

Fang, H., Li, H., Song, S., Pang, K., Ai, D., Fan, J., Song, H., Yu, Y. & Yang, J. (2020). Motion-flow-guided recurrent network for respiratory signal estimation of x-ray angiographic image sequences. *Physics in Medicine & Biology*, 65(24), 245020.

Fischer, P., Faranesh, A., Pohl, T., Maier, A., Rogers, T., Ratnayaka, K., Lederman, R. & Hornegger, J. (2017). An MR-based model for cardio-respiratory motion compensation of overlays in X-Ray fluoroscopy. *IEEE transactions on medical imaging*, 37(1), 47–60.

Frangi, A. F., Niessen, W. J., Vincken, K. L. & Viergever, M. A. ((1998)). Multiscale vessel enhancement filtering. *International conference on medical image computing and computer-assisted intervention*, pp. 130–137.

Freer, P. E., Riegert, J., Eisenmenger, L., Ose, D., Winkler, N., Stein, M. A., Stoddard, G. J. & Hess, R. (2017). Clinical implementation of synthesized mammography with digital breast tomosynthesis in a routine clinical practice. *Breast cancer research and treatment*, 166(2), 501–509.

Gao, Y., Song, Y., Yin, X., Wu, W., Zhang, L., Chen, Y. & Shi, W. (2019). Deep learning-based digital subtraction angiography image generation. *International Journal of Computer Assisted Radiology and Surgery*, 14(10), 1775–1784.

Gautam, R., Kaur, P. & Sharma, M. (2019). A comprehensive review on nature inspired computing algorithms for the diagnosis of chronic disorders in human beings. *Progress in Artificial Intelligence*, 8(4), 401–424.

Gierga, D. P., Brewer, J., Sharp, G. C., Betke, M., Willett, C. G. & Chen, G. T. (2005). The correlation between internal and external markers for abdominal tumors: implications for respiratory gating. *International Journal of Radiation Oncology* Biology* Physics*, 61(5), 1551–1558.

Gislason-Lee, A. J., Kumcu, A., Kengyelics, S. M., Brettle, D. S., Treadgold, L. A., Sivananthan, M. & Davies, A. G. (2015). How much image noise can be added in cardiac X-ray imaging without loss in perceived image quality? *Journal of Electronic Imaging*, 24(5), 051006.

Gislason-Lee, A. J., Keeble, C., Malkin, C. J., Egleston, D., Bexon, J., Kengyelics, S. M., Blackman, D. & Davies, A. G. (2016). Impact of latest generation cardiac interventional X-ray equipment on patient image quality and radiation dose for trans-catheter aortic valve implantations. *The British Journal of Radiology*, 89(1067), 20160269.

Gislason-Lee, A. J., Keeble, C., Egleston, D., Bexon, J., Kengyelics, S. M. & Davies, A. G. (2017). Comprehensive assessment of patient image quality and radiation dose in latest generation cardiac x-ray equipment for percutaneous coronary interventions. *Journal of Medical Imaging*, 4(2), 025501.

Goo, H. W. (2015). Coronary artery imaging in children. *Korean journal of radiology*, 16(2), 239–250.

G.V. Puskorius, L. F. (1994). Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks. *IEEE Trans. Neural Netw. 5(2), 279–297 (1994)*, 5(2), 279–297.

Ha, I. Y., Wilms, M., Handels, H. & Heinrich, M. P. (2018). Model-based sparse-to-dense image registration for realtime respiratory motion estimation in image-guided interventions. *IEEE Transactions on Biomedical Engineering*, 66(2), 302–310.

Haas, N. A., Happel, C. M., Mauti, M., Sahyoun, C., Tebart, L. Z., Kececioglu, D. & Laser, K. T. (2015). Substantial radiation reduction in pediatric and adult congenital heart disease interventions with a novel X-ray imaging technology. *IJC Heart & Vasculature*, 6, 101–109.

Hadida, J., Desrosiers, C. & Duong, L. (2012). Stochastic 3D motion compensation of coronary arteries from monoplane angiograms. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, 651-658.

Hirshfeld, J. W., Balter, S., Brinker, J. A., Kern, M. J., Klein, L. W., Lindsay, B. D., Tommaso, C. L., Tracy, C. M., Wagner, L. K., Creager, M. A. et al. (2004). ACCF/AHA/HRS/SCAI clinical competence statement on physician knowledge to optimize patient safety and image quality in fluoroscopically guided invasive cardiovascular procedures: a report of the American College of Cardiology Foundation/American Heart Association/American College of Physicians Task Force on Clinical Competence and Training. *Journal of the American College of Cardiology*, 44(11), 2259–2282.

Hsieh, J.-T., Liu, B., Huang, D.-A., Fei-Fei, L. F. & Niebles, J. C. (2018). Learning to decompose and disentangle representations for video prediction. *Advances in Neural Information Processing Systems*, pp. 517–526.

Hu, Z. & Wang, J. T. (2019). Generative adversarial networks for video prediction with action control. *International Joint Conference on Artificial Intelligence*, pp. 87–105.

Hunold, P., Vogt, F. M., Schmermund, A., Debatin, J. F., Kerkhoff, G., Budde, T., Erbel, R., Ewen, K. & Barkhausen, J. (2003). Radiation exposure during cardiac CT: effective doses at multi–detector row CT and electron-beam CT. *Radiology*, 226(1), 145–152.

Jung, B.-H., Kim, B.-H. & Hong, S.-M. (2013). Respiratory motion prediction with extended kalman filters based on local circular motion model. *International Journal of Bio-Science and Bio-Technology*, 5(1), 51–58.

Kakar, M., Nyström, H., Aarup, L. R., Nøttrup, T. J. & Olsen, D. R. (2005). Respiratory motion prediction by using the adaptive neuro fuzzy inference system (ANFIS). *Physics in Medicine & Biology*, 50(19), 4721.

Kalet, A., Sandison, G., Wu, H. & Schmitz, R. (2010). A state-based probabilistic model for tumor respiratory motion prediction. *Physics in Medicine & Biology*, 55(24), 7615.

Kesner, S. B. & Howe, R. D. (2010). Design and control of motion compensation cardiac catheters. *2010 IEEE international conference on robotics and automation*, pp. 1059–1065.

King, A. P., Rhode, K. S., Razavi, R. S. & Schaeffter, T. R. (2009). An adaptive and predictive respiratory motion model for image-guided interventions: theory and first clinical application. *IEEE transactions on medical imaging*, 28(12), 2020–2032.

Kobayashi, T. & Hirshfeld Jr, J. W. (2017). Radiation exposure in cardiac catheterization: operator behavior matters. *Circulation: Cardiovascular Interventions*, 10(8), 005689.

L. Ma,C. Herrmann, K. S. (2007). Modeling and prediction of lung tumor motion for robotic assisted radiotherapy. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 189-194.

Liu, Y., Chen, S., Zühlke, L., Black, G. C., Choy, M.-k., Li, N. & Keavney, B. D. (2019a). Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. *International journal of epidemiology*, 48(2), 455–463.

Liu, Y., Chen, S., Zühlke, L., Black, G. C., Choy, M.-k., Li, N. & Keavney, B. D. (2019b). Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. *International journal of epidemiology*, 48(2), 455–463.

Lohse, I., Lang, S., Hrbacek, J., Scheidegger, S., Bodis, S., Macedo, N. S., Feng, J., Lütolf, U. M. & Zaugg, K. (2011). Effect of high dose per pulse flattening filter-free beams on cancer cell survival. *Radiotherapy and Oncology*, 101(1), 226–232.

Lossau, T., Nickisch, H., Wissel, T., Bippus, R., Schmitt, H., Morlock, M. & Grass, M. (2019). Motion estimation and correction in cardiac CT angiography images using convolutional neural networks. *Computerized Medical Imaging and Graphics*, 76, 101640.

Lyu, Q., Shan, H., Xie, Y., Kwan, A. C., Otaki, Y., Kuronuma, K., Li, D. & Wang, G. (2021). Cine cardiac MRI motion artifact reduction using a recurrent neural network. *IEEE Transactions on Medical Imaging*, 40(8), 2170–2181.

Ma, H., Smal, I., Daemen, J. & van Walsum, T. (2020). Dynamic coronary roadmapping via catheter tip tracking in X-ray fluoroscopy with deep learning based Bayesian filtering. *Medical image analysis*, 61, 101634.

Mahesh, M. (2001). Fluoroscopy: patient radiation exposure issues. *Radiographics*, 21(4), 1033–1045.

Manke, D., Rösch, P., Nehrke, K., Börnert, P. & Dössel, O. (2002). Model evaluation and calibration for prospective motion correction in coronary MR angiography based on 3D image registration. *IEEE transactions on medical imaging*, 21, 1132–41.

Mathieu, M., Couprie, C. & LeCun, Y. (2016). Deep multi-scale video prediction beyond mean square error. *ICLR (Poster)*.

McClelland, J. R., Hawkes, D. J., Schaeffter, T. & King, A. P. (2013). Respiratory motion models: a review. *Medical image analysis*, 17(1), 19–42.

McCullough, P. A. (2007). Coronary artery disease. *Clinical Journal of the American Society of Nephrology*, 2(3), 611–616.

McLeish, K., Hill, D. L., Atkinson, D., Blackall, J. M. & Razavi, R. (2002). A study of the motion and deformation of the heart due to respiration. *IEEE transactions on medical imaging*, 21(9), 1142–1150.

McNeice, A. H., Brooks, M., Hanratty, C. G., Stevenson, M., Spratt, J. C. & Walsh, S. J. (2018). A retrospective study of radiation dose measurements comparing different cath lab X-ray systems in a sample population of patients undergoing percutaneous coronary intervention for chronic total occlusions. *Catheterization and Cardiovascular Interventions*, 92(3), E254–E261.

Merz, A. A. & Cheng, S. (2016). Sex differences in cardiovascular ageing. *Heart*, 102(11), 825–831.

M'hiri, F., Duong, L., Desrosiers, C., Leye, M., Miró, J. & Cheriet, M. (2016). A graph-based approach for spatio-temporal segmentation of coronary arteries in X-ray angiographic sequences. *Computers in biology and medicine*, 79, 45–58.

M.J. Murphy, D. P. (2009). Optimization of an adaptive neural network to predict breathing. *Physics in medicine and biology*, 40-47.

M.J. Murphy, S. D. (2006). Comparative performance of linear and nonlinear neural networks to predict irregular breathing. *Physics in medicine and biology*, 5903–5914.

motion prediction by using the adaptive neuro fuzzy inference system (ANFIS), R. (2005). Respiratory motion prediction by using the adaptive neuro fuzzy inference system (ANFIS). *Physics in Medicine & Biology*, 50, 4721.

Myronenko, A. & Song, X. (2010). Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12), 2262–2275.

Nagata, Y., Hiraoka, M., Mizowaki, T., Narita, Y., Matsuo, Y., Norihisa, Y., Onishi, H. & Shirato, H. (2009). Survey of stereotactic body radiation therapy in Japan by the Japan 3-D Conformal External Beam Radiotherapy Group. *International Journal of Radiation Oncology* Biology* Physics*, 75(2), 343–347.

Nehrke, K., Bornert, P., Manke, D. & Bock, J. C. (2001). Free-breathing cardiac MR imaging: study of implications of respiratory motion—initial results. *Radiology*, 220(3), 810–815.

Olcay, A., Güler, E., Karaca, İ. O., Omaygenc, M. O., Kızılırmak, F., Olgun, E., Yenipınar, E., Çakmak, H. A. & Duman, D. (2015). Comparison of fluoro and cine coronary angiography: balancing acceptable outcomes with a reduction in radiation dose. *The Journal of invasive cardiology*, 27(4), 199.

Oshinski, J. N., Hofland, L., Mukundan Jr, S., Dixon, W. T., Parks, W. J. & Pettigrew, R. I. (1996). Two-dimensional coronary MR angiography without breath holding. *Radiology*, 201(3), 737–743.

Ostadal, B. & Ostadal, P. (2014). Sex-based differences in cardiac ischaemic injury and protection: therapeutic implications. *British journal of pharmacology*, 171(3), 541–554.

Pasternak, J. J. & Williamson, E. E. (2012). Clinical pharmacology, uses, and adverse reactions of iodinated contrast agents: a primer for the non-radiologist. *Mayo Clinic Proceedings*, 87(4), 390–402.

Pavlidis, A. N., Jones, D. A., Sirker, A., Mathur, A. & Smith, E. J. (2016). Reducing radiation in chronic total occlusion percutaneous coronary interventions. *Current cardiology reviews*, 12(1), 12–17.

Pierre, S. R. S., Peirlinck, M. & Kuhl, E. (2022). Sex matters: a comprehensive comparison of female and male hearts. *Frontiers in Physiology*, 13, 831179.

Pyne, C. T., Gadey, G., Jeon, C., Piemonte, T., Waxman, S. & Resnic, F. (2014). Effect of reduction of the pulse rates of fluoroscopy and CINE-acquisition on X-ray dose and angiographic image quality during invasive cardiovascular procedures. *Circulation: Cardiovascular Interventions*, 7(4), 441–446.

Rana, R., Singh, V., Jain, A., Bednarek, D. & Rudin, S. (2015). Anti-scatter grid artifact elimination for high resolution X-ray imaging CMOS detectors. *Medical Imaging 2015: Physics of Medical Imaging*, 9412, 941243.

Riaz, N., Shanker, P., Wiersma, R., Gudmundsson, O., Mao, W., Widrow, B. & Xing, L. (2009). Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression. *Physics in Medicine & Biology*, 54(19), 5735.

Rigsby, C. K., McKenney, S. E., Hill, K. D., Chelliah, A., Einstein, A. J., Han, B. K., Robinson, J. D., Sammet, C. L., Slesnick, T. C. & Frush, D. P. (2018). Radiation dose management for pediatric cardiac computed tomography: a report from the Image Gently 'Have-A-Heart'campaign. *Pediatric radiology*, 48(1), 5–20.

Rodgers, J. L., Jones, J., Bolleddu, S. I., Vanthenapalli, S., Rodgers, L. E., Shah, K., Karia, K. & Panguluri, S. K. (2019). Cardiovascular risks associated with gender and aging. *Journal of cardiovascular development and disease*, 6(2), 19.

Roland, T., Mavroidis, P., Shi, C. & Papanikolaou, N. (2010). Incorporating system latency associated with realtime target tracking radiotherapy in the dose prediction step. *Physics in Medicine and Biology*, 55(9), 2651–2668.

Ruan, D., Fessler, J. A. & Balter, J. M. (2007). Real-time prediction of respiratory motion based on local regression methods. *Physics in Medicine & Biology*, 52(23), 7137.

Ruan, D. & Keall, P. (2010). Online prediction of respiratory motion: multidimensional processing with low-dimensional feature learning. *Physics in Medicine & Biology*, 55(11), 3011.

Schneider, M., Sundar, H., Liao, R., Hornegger, J. & Xu, C. (2010). Model-based respiratory motion compensation for image-guided cardiac interventions. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2948–2954.

Schweikard, A., Glosser, G., Bodduluri, M., Murphy, M. J. & Adler, J. R. (2000). Robotic motion compensation for respiratory movement during radiosurgery. *Computer aided surgery*, 5(4), 263–277.

Segars, W., Sturgeon, G., Mendonca, S., Grimes, J. & Tsui, B. M. (2010). 4D XCAT phantom for multimodality imaging research. *Medical physics*, 37(9), 4902–4915.

Segars, W., Bond, J., Frush, J., Hon, S., Eckersley, C., Williams, C. H., Feng, J., Tward, D. J., Ratnanather, J., Miller, M. et al. (2013). Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. *Medical physics*, 40(4), 043701.

Segars, W., Norris, H., Sturgeon, G. M., Zhang, Y., Bond, J., Minhas, A., Tward, D. J., Ratnanather, J., Miller, M., Frush, D. et al. (2015). The development of a population of 4D pediatric XCAT phantoms for imaging research and optimization. *Medical physics*, 42(8), 4719–4726.

Sharp, G. C., Jiang, S. B., Shimizu, S. & Shirato, H. (2004). Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Physics in Medicine & Biology*, 49(3), 425.

Shechter, G., Ozturk, C., Resar, J. R. & McVeigh, E. R. (2004). Respiratory motion of the heart from free breathing coronary angiograms. *IEEE transactions on medical imaging*, 23(8), 1046–1056.

Suhermi, N., Prastyo, D. D., Ali, B. et al. (2018). Roll motion prediction using a hybrid deep learning and ARIMA model. *Procedia computer science*, 144, 251–258.

Sun, Z. (2010). Multislice CT angiography in coronary artery disease: Technical developments, radiation dose and diagnostic value. *World journal of cardiology*, 2(10), 333.

Tanner, C., Eppenhof, K., Gelderblom, J. & Székely, G. (2014). Decision fusion for temporal prediction of respiratory liver motion. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 698–701.

Taylor, A. M., Jhooti, P., Wiesmann, F., Keegan, J., Firmin, D. N. & Pennell, D. J. (1997). MR navigator-echo monitoring of temporal changes in diaphragm position: implications for MR coronary angiography. *Journal of Magnetic Resonance Imaging*, 7(4), 629–636.

Thürauf, S., Hornung, O., Körner, M., Vogt, F., Knoll, A. & Nasseri, M. A. (2018). Model-based calibration of a robotic C-arm system using x-ray imaging. *Journal of Medical Robotics Research*, 3(03n04), 1841002.

Timinger, H., Krueger, S., Dietmayer, K. & Borgert, J. (2005). Motion compensated coronary interventional navigation by means of diaphragm tracking and elastic motion models. *Physics in Medicine and Biology*, 50(3), 491. Retrieved from: http://stacks.iop.org/0031-9155/50/i=3/a=007.

Tulyakov, S., Liu, M. Y., Yang, X. & Kautz, J. (2018). Mocogan: Decomposing motion and content for video generation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1526–1535.

Vedam, S., Keall, P., Docef, A., Todor, D., Kini, V. & Mohan, R. (2004). Predicting respiratory motion for four-dimensional radiotherapy. *Medical physics*, 31(8), 2274–2283.

Vernikouskaya, I., Bertsche, D., Rottbauer, W. & Rasche, V. (2022). Deep learning-based framework for motion-compensated image fusion in catheterization procedures. *Computerized Medical Imaging and Graphics*, 98, 102069.

Villegas, R., Yang, J., Hong, S., Lin, X. & Lee, H. (2017). Decomposing motion and content for natural video sequence prediction. *International Conference on Learning Representations (ICLR)*.

Werner, R., Ehrhardt, J., Schmidt, R. & Handels, H. (2009). Patient-specific finite element modeling of respiratory lung motion using 4D CT image data. *Medical physics*, 36(5), 1500–1511.

Wilson, D. L., Xue, P. & Aufrichtig, R. (1994). Perception of fluoroscopy last-image hold. *Medical physics*, 21(12), 1875–1883.

Yamagata, K., Aldhoon, B. & Kautzner, J. (2016). Reduction of fluoroscopy time and radiation dosage during catheter ablation for atrial fibrillation. *Arrhythmia & electrophysiology review*, 5(2), 144.