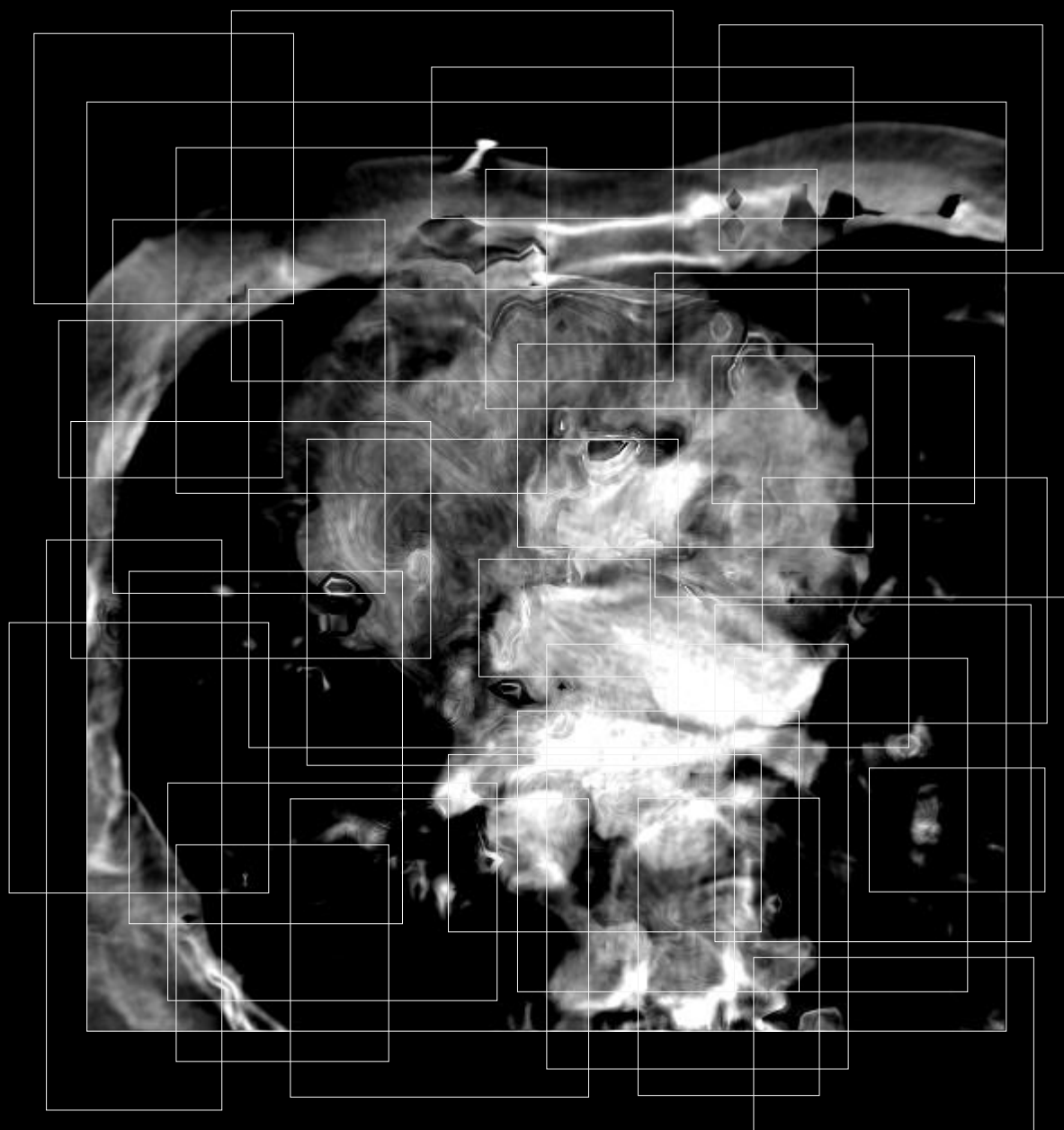

Machine Learning in Cardiac CT Image Reconstruction

Labeled Data Synthesis for the Removal of Motion and Metal Artifacts



MACHINE LEARNING IN CARDIAC CT IMAGE RECONSTRUCTION

**Vom Promotionsausschuss der
Technischen Universität Hamburg**

zur Erlangung des akademischen Grades
Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von
Tanja Loßau

aus
Lübeck

2020

Promotionskomitee:

Prof. Otto von Estorff	Technische Universität Hamburg, <i>Vorsitzender des Prüfungsausschusses</i>
Prof. Michael M. Morlock	Technische Universität Hamburg, <i>Erster Gutachter</i>
PD Dr. Michael Graß	Technische Universität Hamburg, Philips Research Hamburg, <i>Zweiter Gutachter</i>

Tag der mündlichen Prüfung: 03.03.2020

Independent contributors:

Dr. Hannes Nickisch	Philips Research Hamburg
Dr. Tobias Wissel	Philips Research Hamburg
Dr. Rolf Bippus	Philips Research Hamburg
Dr. Holger Schmitt	Philips Research Hamburg
Dr. Frank Bergner	Philips Research Hamburg

TUHH
Hamburg University of Technology

bmh



PHILIPS

Keywords: Cardiac Computed Tomography, Image Reconstruction, Artifact Quantification, Motion Compensation, Metal Artifact Removal, Machine Learning, Convolutional Neural Network

SUMMARY

In the last decade, deep learning techniques have revolutionized the research field of computer vision and reinforced data as the key element for predictive model generation. Also in the medical domain, image processing solutions are increasingly data-driven. However, the required quantity and quality of image and corresponding label data is often a challenge in practice.

This dissertation describes a methodology to leverage the power of state-of-the-art deep learning algorithms bypassing time-consuming, potentially noise-affected and in its complexity limited manual data annotation. The main application focus is the removal of cardiac computed tomography (CT) imaging artifacts. So-called forward models for virtual artifact introduction are developed by incorporating prior knowledge about the cardiac anatomy and CT imaging physics. They form the counterpart of the desired deep-learning-based backward models for image enhancement. Artifact-free clinical data is transformed by the forward models to produce pairs of artifact-perturbed image data and underlying artifact parameters which serve as basis for predictive model training. Estimation of artifact parameters is exclusively performed by convolutional neural networks (CNNs) as these models exploit the low-level statistics of the underlying medical images. The learned networks are used to detect, quantify and remove artifacts.

The proposed methodology is applied to two clinical relevant problems: coronary motion and pacemaker metal artifacts. Due to potential blurring and concealing of anatomies and anomalies in reconstructed CT image volumes, artifact reduction is defined as primary goal. In the first application, a forward model is developed to retrospectively simulate motion during the CT acquisition. Pairs of motion-perturbed images and motion parameters are generated. Based on this data, backward models for motion artifact measurement and motion compensation are learned.

In the second application, a forward model inserts synthetic pacemaker leads into clinical data without pacemakers. Based on the resulting pairs of metal-free and metal-affected sinograms, CNNs are trained for metal removal directly in the projection domain. Furthermore, the backward model is extended to localize metal positions inside the image volume. In both applications, generalization capabilities of the learned models are verified on data with real artifacts and with the aid of human observer ratings. In comparison to existing model-based approaches for artifact detection and removal, similar or even higher performances are achieved.

Both applications demonstrate that predictive models trained on synthetic data only can generalize to real-world problems without the need of additional fine-tuning. The dissertation provides a thorough analysis regarding strengths and challenges of labeled data synthesis based on findings made in the addressed applications. The ability of high-level

label generation, the data- and the time-efficiency are the main benefits compared to traditional manual annotation. The understanding of the data acquisition physics and the system processing enables efficient and high quality data generation. The proposed general concept of knowledge-driven forward modeling and deep-learning-based predictive backward modeling is extendable to different imaging modalities and clinical applications.

CONTENTS

1	Introduction	1
1.1	Purpose of the study	2
1.2	Structure of the work	5
2	Cardiac CT imaging	9
2.1	Acquisition.	10
2.1.1	3D cone-beam projection.	10
2.1.2	ECG-gated acquisition:.	11
2.2	Reconstruction	13
2.2.1	Filtered back-projection	13
2.2.2	Aperture weighted cardiac reconstruction	14
2.2.3	Motion compensated filtered back-projection	15
2.3	Imaging artifacts	16
2.3.1	Coronary motion artifacts	17
2.3.2	Pacemaker metal artifacts	20
3	Machine learning	25
3.1	Statistical learning theory	26
3.1.1	Supervised Learning	27
3.1.2	Convolutional Neural Networks	29
3.2	Applications in cardiac CT	31
3.3	Forward models for labeled data synthesis	33
4	Recognition and quantification of coronary motion artifacts	37
4.1	Material	39
4.1.1	Clinical reference data without artifacts	39
4.1.2	Clinical test data	39
4.2	Forward model.	39
4.2.1	Synthetic motion vector field.	40
4.2.2	Piecewise linear motion model	41
4.2.3	Constant linear motion model	42
4.3	Learned method	44
4.4	Experiments and results	46
4.4.1	Relative artifact measurement	47
4.4.2	Absolute artifact measurement	49
4.5	Discussion	52

5	Removal of coronary motion artifacts	55
5.1	Material	57
5.1.1	Clinical reference data without artifacts	57
5.1.2	Phantom reference data without artifacts.	58
5.1.3	Clinical test data with real artifacts.	58
5.2	Forward model.	58
5.3	Learned method	59
5.3.1	Supervised learning	59
5.3.2	Motion compensation pipeline	62
5.4	Experiments and results	64
5.4.1	Quantitative analysis on synthetic artifacts.	64
5.4.2	Qualitative analysis on real artifacts	65
5.5	Discussion	69
6	Removal of pacemaker metal artifacts	73
6.1	Material	75
6.1.1	Learning data	75
6.1.2	Clinical test data	76
6.2	Forward Model.	76
6.3	Learned Method	77
6.3.1	Shared learning framework	78
6.3.2	SegmentationNets	79
6.3.3	InpaintingNets	80
6.3.4	ModelingNets	81
6.4	Experiments and Results	84
6.4.1	Evaluation on synthetic learning data	84
6.4.2	Evaluation on clinical test data	85
6.5	Discussion	88
7	Discussion, synthesis and conclusions	91
7.1	Labeled data synthesis vs. manual annotation	92
7.2	Outlook and conclusion	95
	References	99
	List of Abbreviations	112
	List of Symbols	114
	List of Figures	118
	List of Tables	120
	List of Publications	121
	Curriculum Vitae	126

1

INTRODUCTION

ISCHEMIC heart disease has been the leading global cause of death since many years according to the World Health Organization¹ and took more than 9.4 million lives in 2016 alone. This represents about 16.3% of all deaths within this year. In order to examine what happens inside the patient's body or, more precisely, in the cardiovascular system, computed tomography (CT) is routinely applied. This imaging modality combines radiographs from different recording angles for the calculation of a three-dimensional image volume. Therein, body components like bone structures, soft tissues, blood vessels and air are visible in characteristic intensities and high spatial resolution. Various hardware and software developments continually increased the CT image quality within the last decades, enabling challenging clinical applications like, for instance, coronary CT angiography (CCTA). Here, clinical risk markers for coronary artery disease (CAD) like plaque densities, locations and volumes are extracted directly from the contrast-enhanced CT image without the need of invasive catheterization [28].

During CT reconstruction, consistency of the acquired projection data across all recording angles is implicitly assumed. This assumption is frequently violated, for instance, due to cardiac motion or high-density metal implants inside or close to the heart. Resulting image errors, so-called artifacts, limit the diagnostic value and potentially cause misinterpretations. In this work, two types of artifacts are addressed - namely coronary motion artifacts and pacemaker metal artifacts.

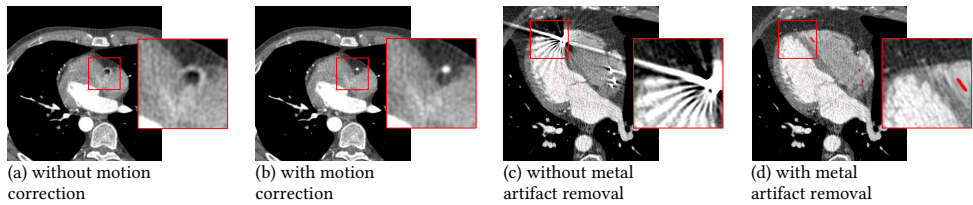


Figure 1.1: Example pairs of artifact-perturbed and artifact-reduced clinical data. Both types, motion (a) and metal artifacts (c) may preclude the analysis of CT images with respect to cardiac diseases. This dissertation describes the development of two software solutions for post-processing inconsistent raw projection data in such a way that related artifacts are suppressed (b,d).

1.1. PURPOSE OF THE STUDY

Coronary motion artifacts manifest in arc-shaped blurring and intensity undershoots (see Figure 1.1a,b). As plaque densities and volumes cannot be reliably determined, detection and treatment planning of CAD is frequently precluded. The purpose of this study is to improve imaging quality in order to provide the basis for a better diagnosis. More precisely, software solutions to identify vessel segments with motion-perturbation, quantify artifact strengths and compensate coronary motion are aimed for. The thesis, furthermore, deals with pacemaker metal artifacts which manifest in streak-shaped intensity under- and overshoots combined with motion artifact pattern (see Figure 1.1c,d). Due to associated concealing of neighboring anatomy, artifact removal is formulated as primary goal. In addition, localization of metal positions inside the image volume is intended.

¹<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>

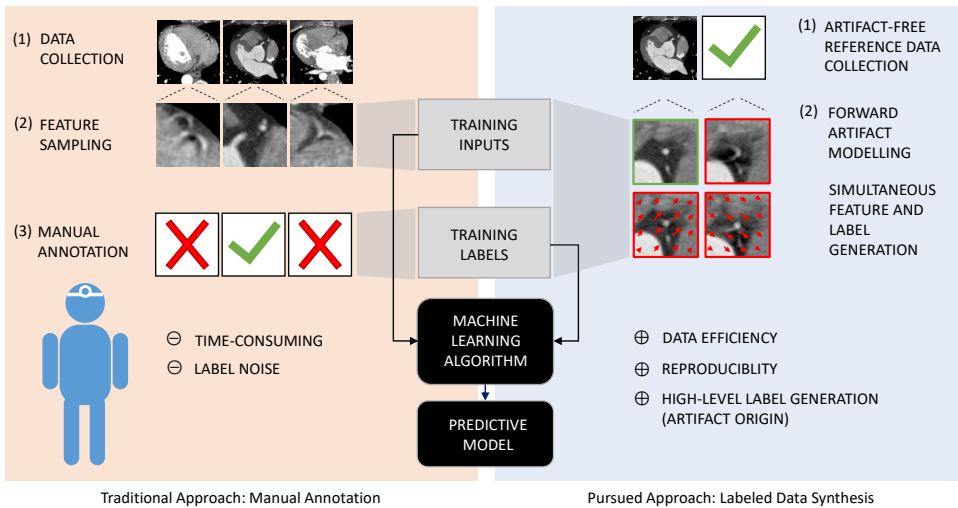


Figure 1.2: Machine learning algorithms build predictive models in a data-driven way. Two approaches to receive the required feature and label data for supervised learning are compared. As an alternative to traditional manual labeling, the software-based synthesis of labeled data is proposed. Artifact-free data is collected as input of a forward model which retrospectively simulates motion during CT acquisition. By this procedure, tuples of artifact-perturbed image data, underlying motion vectors and the corresponding artifact-free version are created.

This work focuses on deep-learning-based approaches to address the defined objectives. Deep learning (DL) started to dominate the field of computer vision around 2012 [59] and reinforced data as the key to artificial intelligence (AI). Based on pairs of input and output samples, a mapping function is trained in an iterative way to fit the given data. During test time, the resulting predictive model estimates for each novel input sample the corresponding unknown output. The accuracy of a predictive model depends on the representativity of the training data. Especially in the medical domain, correctly labeled and properly balanced training data in sufficient quality and quantity are hard to get.

Let us, for instance, consider the traditional approach to learn a predictive model for the detection of coronary motion artifacts (see Figure 1.2, left). First of all, the required input data for supervised learning needs to be provisioned. Due to privacy and data security regulations which restrict the medical data transfer, collection of sufficient clinical data may already represent a challenge. As processing of entire 3D data sets may be associated with memory and data imbalance issues, cross-sectional image slices can alternatively be used as input features of the predictive model. Corresponding labels have to be manually assigned by an expert human observer, e.g. a radiologist. Assessing which vessel segments of the coronary artery tree are motion-affected is time-consuming and often not clearly differentiable. Bifurcations and blurring artifacts have, for instance, similar appearances. High intra- and interobserver variabilities are the consequence. Besides these drawbacks, manual annotation is furthermore limited to low-level labels which do not enable learning of artifact reduction pipelines. For a human observer, annotating ground truth for underlying motion trajectories based on artifact-perturbed image data is not feasible.

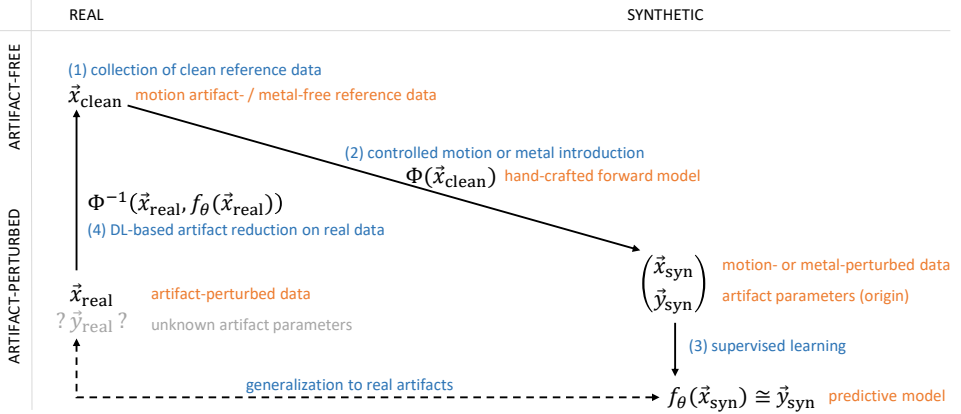


Figure 1.3: A prior knowledge-driven forward model ϕ introduces synthetic motion or metal to clean reference cases \vec{x}_{clean} . The resulting tuples of artifact-perturbed data \vec{x}_{syn} and associated artifact parameters \vec{y}_{syn} are used for supervised learning of f_{θ} . The predictive model f_{θ} is integrated into the learned backward model ϕ^{-1} for artifact removal on real data. Furthermore, predictive models are trained for the tasks of motion artifact recognition, artifact strength quantification and metal reinsertion (not shown in this Figure).

Labeled data synthesis In this work, the approach of labeled data synthesis is pursued, as an alternative to traditional manual annotation. Figure 1.2 (right) illustrates the software-based generation of paired data for the particular task of coronary motion artifact detection. The labeled data synthesis is built on a transformation function for synthetic artifact introduction, hereafter called the forward model. It is developed based on an exhaustive analysis of artifact-perturbed data with respect to artifact origins, different appearances and potential influencing factors. Physical and mathematical prior information like the CT acquisition geometry are included into the forward model.

As visualized in Figure 1.3, the forward model Φ takes artifact-free clinical data \vec{x}_{clean} as input and delivers a tuple of artifact-introduced data \vec{x}_{syn} and corresponding target labels \vec{y}_{syn} as output. The target labels can, for instance, be the ground truth motion vectors. Based on the synthetic data $(\vec{x}_{\text{syn}}, \vec{y}_{\text{syn}})$, a predictive model f_{θ} is trained to estimate such underlying artifact parameters. The predictive model is integrated into a processing pipeline which can also be applied on real data \vec{x}_{real} with unknown underlying artifact parameters \vec{y}_{real} . The processing can include, for instance, artifact detection, quantification and removal. The next chapters will detail the individual steps of this described concept.

As learning of the predictive model is restricted to clinical data with synthetic artifacts, applicability to real artifacts is crucial. The working hypothesis is formulated that the predictive models trained on the synthesized data will also generalize to real-world examples. Based on the applications of coronary motion and pacemaker metal artifact removal, this hypothesis is to be verified. The transferability to clinical practice is evaluated by manual image quality assessment of CT image volumes with and without application the resulting DL-based correction procedures. Furthermore, challenges and benefits of the labeled data synthesis shall be determined within this work.

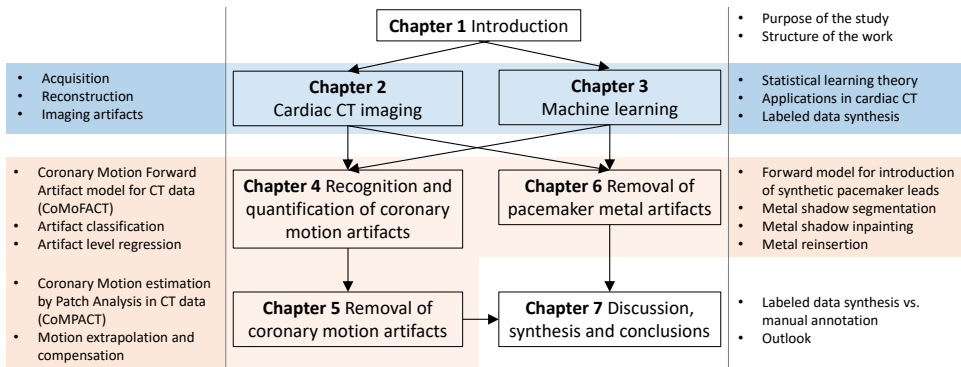


Figure 1.4: The chapter organization enables different potential orders of reading. An arrow from one chapter to another indicates that the latter chapter builds upon the former.

1.2. STRUCTURE OF THE WORK

This work is organized into seven chapters. Figure 1.4 provides a flowchart depicting the structure of the work and reading dependencies. Following this chapter, introductions of cardiac CT imaging (see Chapter 2) and machine learning (see Chapter 3) are given. Readers with extensive knowledge in these fields may skip the block of fundamentals (highlighted in blue) in part. However, reading of Section 2.3, Section 3.2 and Section 3.3 is recommended. Section 2.3 has to be highlighted, as therein, mathematical and physical backgrounds of coronary motion artifacts and moving metal artifacts are presented. Furthermore, the section points out the clinical relevance of the defined objectives and provides an exhaustive literature overview on existing approaches addressing these types of artifacts. In Section 3.2, trends in predictive model design and data handling are derived based on a review of current publications in the research field of machine learning in cardiac CT imaging. Section 3.3 describes the pursued approach of labeled data synthesis in detail.

The methodology block (highlighted in orange) is based on previous publications. A full list of publications incl. invention disclosures and patent applications can be found in the Appendix. In agreement with the publisher Elsevier, contents of the journal articles are partially adopted in this dissertation. Tables and Figures taken from published literature are marked in the List of Tables and the List of Figures, respectively. In order to increase clarity and comprehensibility of this dissertation, a List of Abbreviations and a List of Symbols are provided in the Appendix.

The methodological part on coronary motion artifacts is split into two chapters. The Chapter 4 is mainly inspired by the journal article [69]. It deals with artifact recognition and quantification and initially describes the *Coronary Motion Forward Artifact model for CT data* (CoMoFACT). Building on this, the pipeline for artifact removal based on *Coronary Motion estimation by Patch Analysis in CT data* (CoMPACT) is introduced in Chapter 5. The corresponding journal article is [68]. In the second part of the thesis, a fully automatic pipeline for *Dynamic Pacemaker Artifact Removal* and metal reinsertion is developed. The so-called DyPAR+ is introduced in Chapter 6 and in the journal article [71].

Each methodology chapter $X \in \{4,5,6\}$ is structured as follows:

X.1 Material

Acquisition procedures and pre-processing steps of the artifact-free reference data \vec{x}_{clean} and the clinical test data with real artifacts \vec{x}_{real} are described.

X.2 Forward model

The generation process of the synthetic learning data by means of the forward model Φ is detailed.

X.3 Learned method

This Section includes information about feature sampling, target label assignment, the supervised learning setup and the integration of the predicted model f_{θ} into the processing pipeline.

X.4 Experiments and results

An evaluation of the predictive model f_{θ} and the learned processing pipeline on synthetic \vec{x}_{syn} and real data \vec{x}_{real} is performed.

X.5 Discussion

Limitations and potential extensions of the proposed method are discussed.

Strengths and challenges of the proposed procedure on labeled data synthesis compared to common manual annotation are discussed in Chapter 7 based on the findings described in the previous chapters. Finally, an outlook on potential further applications in medical imaging is provided.

2

CARDIAC CT IMAGING

CARDIAC computed tomography (CCT) is a non-invasive diagnostic imaging modality for the assessment of cardiovascular diseases. Clinical applications include

- detection/diagnosis of coronary artery diseases (CAD), aortic dissections, valvular defects, tumors and congenial heart defects,
- evaluation of vascular integrity, vascular volumetry and cardiac functions,
- preoperative assessment and planning for coronary artery bypasses,
- preoperative planning for device insertion or removal (e.g. coronary stents, pacemakers or artificial valves).

Due to cardiac motion, a high spatial and temporal resolution is crucial for the diagnostic value of CCT images [2, 13]. This chapter gives an introduction to CT acquisition in general and dedicated scan modes for cardiac imaging (see Section 2.1). The image reconstruction methods utilized in this work are described in Section 2.2 and a literature overview on the research fields of coronary motion artifacts and pacemaker metal artifacts is provided in Section 2.3.

2.1. ACQUISITION

Photon-matter interaction leads to exponentially decrease of X-ray radiation intensity. The Lambert-Beer law provides a mathematical model of monochromatic X-ray attenuation. Assuming that photons are passing through an object of thickness η with a constant attenuation coefficient μ , the radiation intensity at the detector I_D is determined by

$$I_D = I_S e^{-\mu\eta}, \quad (2.1)$$

where I_S denotes the source intensity. It has to be mentioned that scatter effects and energy dependencies of attenuation values are not considered in this formulation. However, this simplified model is the basis for the computation of the unknown absorption coefficients by CT systems.

2.1.1. 3D CONE-BEAM PROJECTION

State of the art CT scanners are equipped with cone-beam detector systems as schematically depicted in Figure 2.1. The projection integral is defined by the negative logarithm of the ratio of the incoming number of photons $I_D(t, u, v)$ at the detector element and the emitted number of photons $I_S(t, u, v)$ at the X-ray source

$$p(t, u, v) = -\ln\left(\frac{I_D(t, u, v)}{I_S(t, u, v)}\right) = \int_{[0,1]} \mu\left(\vec{S}(t) + \eta(\vec{D}(t, u, v) - \vec{S}(t))\right) d\eta. \quad (2.2)$$

The line integrals represent the sum of attenuation coefficients along X-ray beams. $\vec{S}(t)$ determines the source position and $\vec{D}(t, u, v)$ the position of detector element (u, v) at time point $t \in [t_{\text{start}}, t_{\text{end}}]$. The variables t_{start} and t_{end} denote start time and end time of the CT scan. The continuous model of line integrals is replaced in computer discretization by projection sums. During image reconstruction, the calculation of attenuation coefficients on a grid of voxel positions $\vec{v} \in \Omega \subset \mathbb{R}^3$ is aimed for. Acquisition is performed for a limited number of projection views P_j . Each projection view P_j is associated with a specific acquisition time point t_j and gantry rotation angle $\gamma_j = \gamma(t_j)$. It is defined as an array of line integrals of

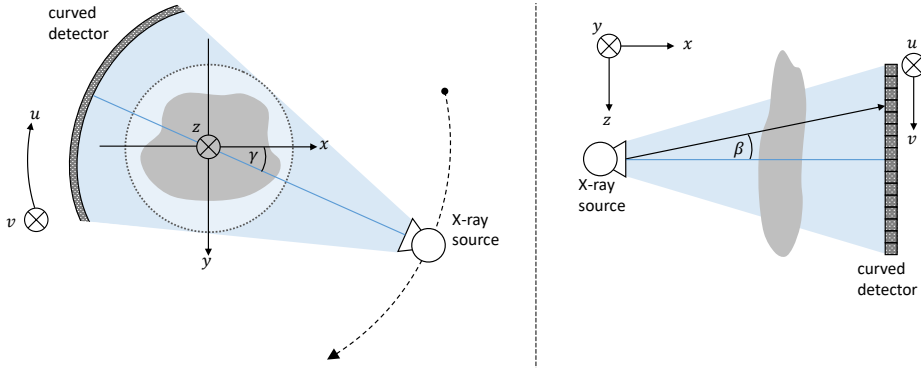


Figure 2.1: Schematic drawings of the 3D cone-beam projection geometry. Left: The gantry angle γ is given by the angular position of the rotating X-ray source in relation to the axial plane (x, y) of the fixed patient coordinate system. Right: The cone-angle β influences the length of a line integral.

detector shape. By vectorizing projection values and unknown attenuation coefficients, the forward projection, i.e. the Radon transformation \mathcal{R} , can be described as a matrix multiplication:

$$\begin{aligned} \vec{p} &= R \vec{\mu} \\ \downarrow & \quad \downarrow \downarrow \\ p(t, u, v) &= \mathcal{R}\{\mu(\vec{v})\}. \end{aligned}$$

The system matrix R specifies the relation between X-ray beams and voxels. For 3D cone-beam geometry, r_{ij} is assigned as the irradiated volume of voxel j by beam i . The size of the system matrix is, therefore, determined by *the number of voxels within the pre-defined grid* \times *(the detector height \cdot the detector width \cdot the number of acquired projection views)*.

2.1.2. ECG-GATED ACQUISITION:

In Equation (2.2), the attenuation coefficient is defined as location-dependent variable. The assumption that the object is static during acquisition does not hold in cardiac imaging. Due to motion, attenuation coefficients $\mu(\vec{v}, t)$ have to be modeled as time-dependent variables as well. In CT reconstruction, calculation of the image volume at a fixed time point is aimed for, comparable to a frozen snapshot of the scanned object. However, hardware constraints and the angular range required for reconstruction restrict the temporal resolution of the CT image volumes. In the following, requirements for performing cardiac CT scans are described and methods to limit motion artifacts are introduced.

The scan preparation typically includes administration of beta blockers which slow down the patients heart rate and the application of contrast agent which enhances the visibility of the blood volume in reconstructed CT images. Besides the cancer risk which increases with the radiation dose, allergic reactions to the usually iodine-based contrast dye may occur. Furthermore, limited renal function is handled as contraindication for contrast-enhanced CT. The acquisition is done during breath hold in order to inhibit pulmonary motion artifacts. Three different acquisition modes are known which deal with cardiac motion by exploiting its assumed periodicity using additional ECG information.

Retrospectively ECG-gated spiral acquisition A simultaneous acquisition of ECG signal and projection data enables the association of each projection view P_j with a specific phase point $t_{cc,j} \in [0\%, 100\%]$ within the cardiac cycle. The scan period covers multiple heart cycles and gantry rotations, as illustrated in Figure 2.2 and stated in [37]. Due to the data redundancy, reconstruction can be performed at an arbitrary reference heart phase r . The reference heart phase defines the center of the cardiac gating window which is described in detail in Section 2.2.2. By means of retrospective ECG-gating, either image volumes at heart phases of least motion, e.g. in diastolic rest phase, or full-cycle cine-sequences can be generated.

The spiral acquisition mode combines continuous rotation of the sampling unit in the x-y plane and a continuous table feed along the scanners z-axis, therefore resulting in a helical arrangement of the raw projection data. The slice thickness in z depends on the z -resolution of the detector and the helical pitch, i.e. the table displacement during 360° gantry rotation. A lower pitch corresponds to a higher data redundancy, a longer contrast agent bolus and higher clinical risk as the same anatomic regions are repeatedly exposed to radiation. Dose modulation, also known as ECG pulsing, is frequently performed to limit the radiation dose. Under condition of a stable heart rate, 100% dose is emitted during a selected cardiac window while the tube output is reduced during the remaining parts of the cardiac circle.

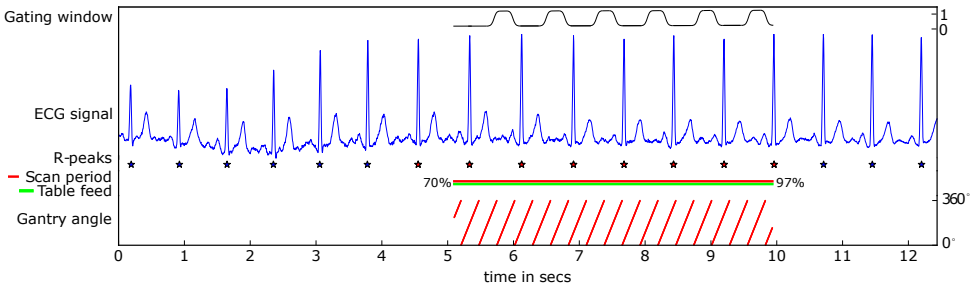


Figure 2.2: Scan information plot of a clinical case with retrospectively ECG-gated spiral acquisition. Projection data is retrospectively weighted depending on its associated cardiac phase point during image reconstruction.

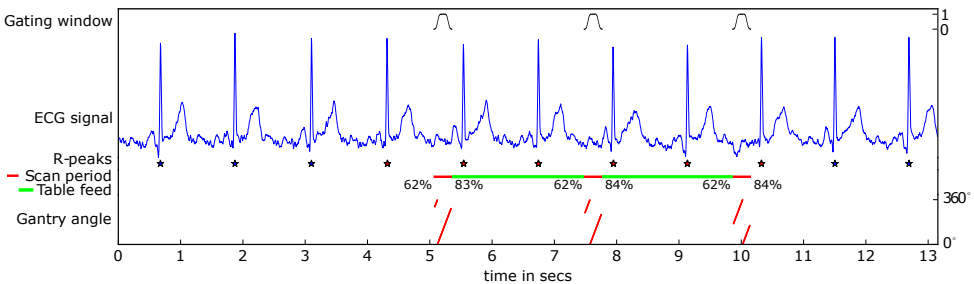


Figure 2.3: Scan information plot of a clinical case with prospectively ECG-triggered sequential acquisition. Scanning periods are determined by consideration of previous R-peaks times.

Prospectively ECG-triggered sequential acquisition In the sequential acquisition mode, table movement and CT acquisition are performed alternately, therefore resulting in a circular arrangement of the raw projection data [45, 53]. In a so-called step-and-shoot scan, the image slice thickness solely depends on the z -resolution of the detector, since the z -coordinate remains constant for each scanning period. Depending on the detector height and the patients anatomy, 2-3 scanning shots are commonly combined to achieve a gapless volume coverage of the whole heart. For each axial slice, the projection data of the scanning shot closest in z is used during reconstruction. Therefore, in contrast to spiral acquisition, stack transition artifacts occur.

Scanning periods are determined by a pre-selected target heart phase (see Figure 2.3). This limits the application scope to patients with low heart rate (≤ 65 bpm recommended) and without tachyarrhythmia. As minimal angular scan range, 180° plus the fan angle are required for image reconstruction in cone-beam CT. So-called padding, i.e. acquisition over a wider angular range increases the flexibility in retrospectively choosing the reference heart phase r for gated reconstruction. Prospective ECG-triggering with its significantly reduced radiation doses is currently the most widely used acquisition mode for cardiac CT [78].

Prospectively ECG-triggered high-pitch spiral acquisition Under condition of a stable sinus rhythm with low heart rate, CT acquisition with a relative pitch of more than 1.5 for single-source systems or 3.0 for dual-source systems enable gapless volume coverage of the whole heart within a fraction of a single cardiac circle [62, 97]. In contrast to the previous approaches, the image volumes are not reconstructed at a fixed motion state, but exhibit a phase shift along the z -axis. This acquisition mode is listed for the sake of completeness only and not applied in this work.

2.2. RECONSTRUCTION

The task of CT image reconstruction corresponds to the inversion of the Radon transformation $\mathcal{R}\{\mu(\vec{v})\}$. Reconstruction algorithms can be grouped into analytical and iterative ones. Analytical approaches provide a direct solution of the reconstruction problem by inverting the Radon transformation. Iterative approaches like algebraic reconstruction techniques [87] or maximum likelihood reconstruction techniques [60] handle the image reconstruction as an optimization problem by seeking the optimal fit given the measured projection data. In comparison, iterative reconstruction approaches are more robust with regard to noisy, missing or inconsistent projection data, but computationally expensive. In the following, the thesis focuses on analytical reconstruction methods based on the filtered back-projection. However, it should be mentioned here that AI-based artifact detection and correction methods can also be combined with iterative reconstruction methods.

2.2.1. FILTERED BACK-PROJECTION

The filtered back-projection (FBP) is the most popular CT reconstruction technique due to its high computational efficiency. It is based on the Fourier Slice Theorem (or central slice theorem for 3D reconstruction from 2D parallel projections) which states that the 2D Fourier transform of a projection view is equal to a plane through the origin in the 3D

frequency domain of the object function μ . The Fourier slice theorem is usually not directly applied as inverse Radon transform, because parallelization of data acquisition and reconstruction is not feasible. Furthermore, the regridding might produce bad interpolation results for the edge-encoding high frequencies. However, the Fourier Slice Theorem forms the basis for the derivation of the analytic reconstruction methods [14].

The FBP consists of four steps: (1) Fourier transform of the projection data (2) High-pass filtering in the frequency domain (3) Inverse Fourier transform (4) Back-projection. In practice, high-pass filtering is performed by a 1D fast Fourier transformation (FFT) along the detector rows, subsequent application of a band-limited ramp filter and final inverse FFT. The voxel-driven back-projection operator \mathcal{B} is defined as

$$\mathcal{B}\{p\}(\vec{v}) = \int_{t_{\text{start}}}^{t_{\text{end}}} p(t, \vec{v}) dt \quad (2.3)$$

where $p(t, \vec{v})$ is the projection integral, which passes through the voxel \vec{v} at time point t . A zero is assigned to $p(t, \vec{v})$, if the voxel \vec{v} is not hit at time point t . The FBP can also be expressed as matrix-vector multiplication $\vec{\mu} = BH\vec{p}$, where the back-projection matrix $B = R^T$ is equal to the transposed system matrix. The high-pass filtering matrix $H = F_{1D}^{-1}GF_{1D}$ is a composition of the discrete Fourier transform (DFT) matrix F_{1D} , its inverse F_{1D}^{-1} and the ramp filter matrix G . The FBP is associated with the assumptions of

- (1) **orthogonal projections** X-ray beams are of the same length and in parallel-beam projection geometry,
- (2) **complete projections** The projection data is uniformly sampled over the full required angular range of 180° (in parallel-beam geometry),
- (3) **consistent projections** The projection data is consistent in time, i.e. no motion, beam hardening, scatter, Poisson noise etc. occurs during acquisition.

2.2.2. APERTURE WEIGHTED CARDIAC RECONSTRUCTION

As the previously mentioned assumptions (1) and (3) contradict reality, additional weighting functions are introduced in the aperture weighted cardiac reconstruction (AWCR) to prevent or suppress corresponding imaging artifacts. The first step of the AWCR is the rebinning $p \rightarrow p_{\text{wedge}}$ of the input 3D cone-beam projection data into virtual wedge (parallel-beam) geometry. Subsequently, the attenuation coefficient $\mu(\vec{v})$ of each voxel $\vec{v} \in \Omega$ in the field of view (FOV) $\Omega \subset \mathbb{R}^3$ can be calculated by

$$\mu(\vec{v}) = \int_{t_{\text{start}}}^{t_{\text{end}}} w_{\text{AWCR}}(t, r, \vec{v}) p_{\text{filt}}(t, \vec{v}) dt \quad , \text{ where} \quad (2.4)$$

$$p_{\text{filt}} = \mathcal{F}_{1D}^{-1} \{ w_{\cos} p_{\text{wedge}} \} g \quad \text{and} \quad (2.5)$$

$$w_{\text{AWCR}}(t, r, \vec{v}) = \frac{w_c(t, r) w_{ap}(\beta(t, \vec{v}))}{\sum_{t_j \in \Pi(t)} w_c(t_j, r) w_{ap}(\beta(t_j, \vec{v}))}. \quad (2.6)$$

The high pass filtering in Equation (2.5) is performed in the frequency domain by multiplying the fourier-transformed, pre-weighted and rebinned projection data with a band limited ramp filter g . Purpose and design of the weighting functions are described in Table

Table 2.1: Weighting functions in aperture weighted cardiac reconstruction.

Variable	Name	Description	Function
$w_{\cos}(\beta(t, \vec{v}))$	cosine weighting	The cosine pre-weighting normalizes line integrals with respect to their length.	cosine
$w_{ap}(\beta(t, \vec{v}))$	aperture weighting	The aperture weighting suppresses large cone-angles to avoid cone-beam artifacts.	trapezoid
$w_c(t, r)$	cardiac weighting	The gating window function w_c weights line integral depending on its associated cardiac phase point.	squared cosine with compact support

2.1. For each cardiac cycle, the center of the cardiac gating window is defined by the projection view with heart phase closest to the reference phase r . The width of the cardiac gating window is typically chosen as narrow as possible (180° in parallel-beam geometry) to prevent motion artifacts. For a more detailed description, please refer to [58] for retrospectively ECG-gated spiral acquisition and [110] for prospectively ECG-triggered sequential acquisition.

In case of retrospective ECG-gating, the trajectory covers multiple gantry rotations. Projections that contribute to a voxel under the identical viewing angle are called pi-partners. For a given acquisition time point t , $\Pi(t) = \{t_j | \gamma(t_j) = \gamma(t) + i\pi, i \in \mathbb{Z}\}$ is defined as the set of pi-partners. In Equation (2.6), pi-partner normalization is included, i.e. the contributions from all redundant projections sum up to one for each angle. For ungated CT scans, aperture weighted reconstruction (AWR) is defined by calculating attenuation coefficients according to Equation (2.4) with a constant cardiac weighting function $w_c(t, r) := 1$.

2.2.3. MOTION COMPENSATED FILTERED BACK-PROJECTION

Despite ECG-controlled imaging, cardiac motion frequently leads to artifacts in reconstructed CT images. Several motion compensation (MC) approaches have been developed in the last years which deal with motion vector field (MVF) estimation and subsequent motion compensated filtered back-projection (MC-FBP). The MC-FBP algorithm is an extension of the AWR/AWCR method which takes estimated displacements during CT acquisition $\vec{d}(t, \vec{v}) \in \mathbb{R}^3$ into account:

$$\mu(\vec{v}) = \int_{t_{\text{start}}}^{t_{\text{end}}} w_{\text{AWCR}}(t, \vec{v} + \vec{d}(t, \vec{v})) p_{\text{fit}}(t, \vec{v} + \vec{d}(t, \vec{v})) dt. \quad (2.7)$$

The MVF \vec{d} is usually calculated by: (1) Reconstructing a time-series of CT image volumes, e.g. multi-phase data (2) Enhancement of relevant features, e.g. by vesselness or edge filtering (3) Image registration to estimate the displacement each voxel has undergone between a selected reference time point t_0 and the remaining reconstruction time points (4) Interpolation in time domain to cover the temporal projection range required for MC-FBP. Enhancement features for motion correction can e.g. target coronary arteries [6, 48, 109], the aortic valve [24] or lung structures [36]. Figure 2.4 compares the concepts of AWR/AWCR with the MC-FBP and provides an example of motion compensated aortic

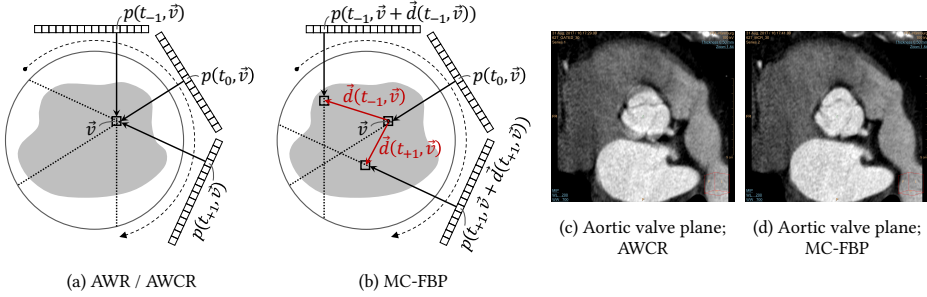


Figure 2.4: Schematic drawing of voxel-driven back-projection without (a) and with (b) motion compensation. The back-projection procedure in AWR/AWCR relies on the assumption that the object is static during acquisition. (c) Inconsistent projection data caused by motion leads to artifacts in the reconstructed CT image volume. (d) In case of MC-FBP, moving voxel positions are considered and line integrals are spatially corrected in the back-projection procedure in order to reduce associated blurring artifacts.

valve imaging. The MVF contains motion information of each voxel between the reference time t_0 and the time t_j each specific projection was acquired. During MC-FBP, each voxel is moved accordingly before back-projection is actually done, therefore leading to a compensation of correctly estimated motion as demonstrated by Roux [95] and Taguchi [105]. More detailed descriptions on MC-FBP are provided in [96] for ECG-triggered data with step-and-shoot acquisition, [109] for ECG-gated data with helical acquisition and [54] for ungated acquisition modes.

2.3. IMAGING ARTIFACTS

CT imaging artifacts, i.e. systematic discrepancies between measured and real attenuation coefficients, occur whenever the assumptions made in Section 2.2.1 are violated. Different types of artifacts are known which vary in appearance and origin [5] like

- ring artifacts due to miscalibrated detector elements,
- beam hardening artifacts due to more rapid absorption of low-energy photons,
- photon starvation due to insufficient photons reaching the detector through paths of high X-ray attenuation,
- Poisson noise due to the quantum nature of X-ray photons,
- partial volume artifacts due to a limited image resolution corresponding to averaged attenuations within voxels,
- aliasing artifacts due to the limited number of projection views used for image reconstruction,
- truncation artifacts due to objects extending the scan field of view,
- stack transition artifacts due to temporal inconsistencies between scanning shots.

This work focuses on coronary motion artifacts and pacemaker metal artifacts. Section 2.3.1 and 2.3.2 detail characteristics of these artifacts and existing software-based solutions for their suppression.

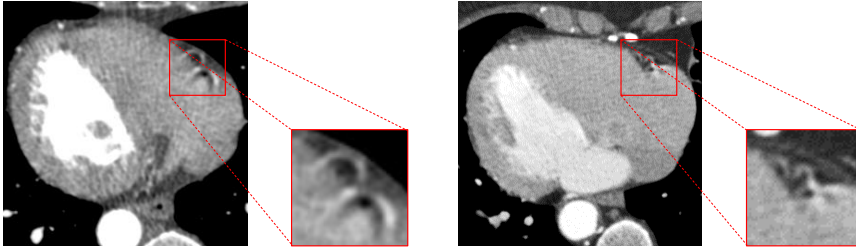


Figure 2.5: Axial planes of a step-and-shoot case (left) and a helical case (right) reconstructed at diastolic rest phase. Severe motion artifacts at the right coronary arteries (RCAs) are highlighted in red. Cardiac motion leads to differently shaped artifacts in helical and step-and-shoot CT scans. Motion artifacts in step-and-shoot CT scans are less complex as merely one coherent angular segment is used for the reconstruction of each voxel. Due to table movement and multi-cycle reconstruction, motion artifacts have a different appearance in CT scans with helical acquisition mode.

2.3.1. CORONARY MOTION ARTIFACTS

The majority of about two-thirds of all cardiac CT scans is performed for diagnosis and treatment planning of coronary artery diseases (CAD) according to the European MR/CT Registry 2019 of the European Society of Cardiovascular Radiology¹. High quality imaging of the vascular system is a prerequisite for non-invasive coronary computed tomography angiography (CCTA) [12, 15, 28, 66]. Here, the main challenge is the suppression of motion as hardware constraints like the gantry rotation speed and the required angular acquisition range restrict the temporal resolution of reconstructed CT image volumes. Despite application of dual source CT systems [92] and ECG-controlled acquisition (see Section 2.1), cardiac motion frequently leads to artifacts in the reconstructed CT image volumes [31]. Such motion artifacts manifest in characteristic patterns of arc-shaped blurring and intensity undershoots due to the CT reconstruction geometry (see Figure 2.5) and potentially limit diagnosis or even cause misinterpretations.

Multitudes of software-based solutions for motion artifact detection, quantification and reduction have been developed in the last decade. A selection of related publications is listed and compared in Table 2.2. Motion vector field (MVF) estimation and subsequent motion compensated reconstruction (MCR) are the main components of most motion compensation (MC) algorithms. A variety of MCR methods including motion compensated filtered back-projection (MC-FBP) [96, 109], motion compensated iterative reconstruction [48] and backproject-then-warp (BPW) strategies [6, 10] are known. The MC algorithms in Table 2.2 are assigned into registration-based, PAR-based (partial angle reconstruction), metric-based and image-based approaches.

Registration-based Motion estimation by 3-D/3-D registration of multiple heart phases has shown great results in the reduction of moderate and severe motion artifacts [6, 48, 106, 109] but requires an extended temporal scan range which corresponds to increased radiation doses. In [38] a series of CT image volumes with reduced angular range of 75° is reconstructed at the angular positions -120° , -60° , $+60^\circ$ and $+120^\circ$ around a

¹https://www.mrct-registry.org/images/ESCR_2019_CardiacMRCTbooklet.pdf

Table 2.2: Literature survey in the research field of coronary motion artifacts in CCTA images. Publications are clustered into artifact reduction approaches, artifact detection/quantification approaches and the proposed approaches which are based on synthetically motion-perturbed data generated with the *Coronary Motion Forward Artifact model for CT data* (CoMoFACT). The artifact removal pipeline relies on *Coronary Motion estimation by Patch Analysis in CT data* (CoMPACT). In the second row, symbols (first column) and abbreviations (last column) are explained.

Publications	Purpose				Approach				Data	Term	Keywords			
Symbols R: rule-based D: data-driven Γ : required angular scan range	Artifact Detection	Artifact Quantification	Motion Estimation	Artifact Reduction	Registration-Based	PAR-Based	Metric-Based	Purely Image-Based	Algorithm	Manually Labeled	Forward Model	Approximate Centerline	$\Gamma > 180^\circ + \text{Fan Angle}$	Abbreviations PAR partial angle reconstruction MC-FBP motion compensated filtered back-projection MC-IR motion compensated iterative reconstruction BPW backproject-then-warp ME motion estimation TRIM temporal resolution improvement method MAM motion artifact measure
[109] Stevendaal			✓	✓	✓				R				✓	model-based surface reg.; MC-FBP
[48] Isola			✓	✓	✓				R				✓	multi-phase elastic image reg.; MC-IR
[6] Bhagalia			✓	✓	✓				R			✓	✓	centerline reg.; subphasic warp and add; BPW
[106] Tang			✓	✓	✓				R				✓	multi-phase image reg.; alternating ME and MC-FBP
[38] Grass			✓	✓	✓	✓			R				✓	vesselness filtering; opposite PAR reg.
[56] Kim			✓	✓	✓	✓	✓		R				✓	linear PAR reg.; metric-based MVF refinement
[98] Schöndube				✓		✓			R					iterative PAR; histogram constraint; TRIM
[93] Rohkohl			✓	✓			✓		R				✓	iterative MAM optimization
[40] Hahn			✓	✓		✓	✓		R				✓	PAR; MAM optimization; BPW
[50] Jung				✓				✓	D		✓	✓		cross-phase style transfer; image-to-image translation
[112] Šprem	✓							✓	D	✓		✓		coronary artery calcification; deep learning
[72] Ma	✓	✓						✓	R			✓		fold overlap ratio; low-intensity region score
[26] Elss	✓							✓	D		✓	✓		CoMoFACT; deep learning
[25] Elss			✓					✓	D		✓	✓		CoMoFACT; deep learning
[69] Lossau	✓	✓						✓	D		✓	✓		CoMoFACT; deep learning
[68] Lossau			✓	✓				✓	D		✓	✓		CoMoFACT; deep learning; CoMPACT

selected center phase. The resulting partial angle volumes are post-processed by first combining high frequencies from the partial scans with low frequencies from a central full scan and subsequent vessel feature enhancement according to [116]. Dense MVFs are obtained by elastic image registration of diametrically opposed partial scans as described by Kim et al. [55]. As described in Section 2.2.3, the original raw projection data and the estimated MVFs are integrated into MC-FBP to yield the motion-corrected image volume. By this procedure, Grass et al. [38] reduced the required angular scan range to 315° plus fan angle of the reconstruction field of view. The Table 2.2 provides an overview which methods are associated with the constraint of $\Gamma > 180^\circ + \text{Fan Angle}$. The remaining approaches only require the usual angular scan range for reconstruction via FBP, i.e. $\Gamma = 180^\circ + \text{Fan Angle}$.

PAR-based Partial angle reconstructions exhibit an increased temporal resolution compared to short scan reconstruction, but suffer from streak-shaped artifacts. Several MC methods like [38, 40, 56] exploit the increased temporal resolution during motion estimation. Schöndube et al. [98] introduced the temporal resolution improvement method which is based on an iterative PAR with an additional histogram constraint.

Metric-based Rohkohl et al. [93] presented an initial metric-based approach, whereby iterative minimization of handcrafted motion artifact measures (MAMs) enables MVF estimation. This method was extended in [40] by introducing a novel motion model parametrization and application of estimated MVFs to PARs. A hybrid approach has been proposed by Kim et al. [56]. They combine the mentioned three strategies by first estimating linear motion using registration of PARs and subsequently refine the resulting MVF by information potential minimization.

Purely image-based In image-to-image translation, a direct mapping of artifact-perturbed image data to artifact-free image data is performed. An artifact suppression method using deep residual convolutional neural networks (CNNs) has, for instance, been developed by Jung et al. [50]. The main advantage of image-based approaches is the independence of the raw projection data. On the downside, the information content of the motion perturbed input patches represents a limiting factor.

A few research articles also deal with image-based detection and quantification of coronary motion artifacts. Assessment of the image quality might be useful for the application of automatic analysis tools, e.g. by reporting on the reliability of coronary artery calcium scores. Šprem et al. [112] introduced a DL-based method to identify coronary artery calcifications which are strongly affected by cardiac motion artifacts. Furthermore, motion correction methods can benefit from reliable motion artifact measures due to possible failure modes and their substantial computational footprint. In a first step, artifact recognition enables the decision whether and where motion correction is required for accurate diagnosis of CAD. Second, artifact quantification measures could be used to verify the success of a motion compensation method. Finally, an integration of motion artifact measures in the motion compensation process as shown by Rohkohl et al. [93] is conceivable.

Most handcrafted measures for motion artifacts [76, 93] such as entropy and positivity are best suited for relative assessment, i.e. for the comparison of the same image region at different motion states. An absolute measure for consistent artifact quantification across patients and vessel segments has to be robust with regard to possible variations in noise level, background intensity, vessel structure and contrast agent enhancement. The ability of five handcrafted metrics to quantify absolute motion artifact levels at the coronary arteries has been investigated by Ma et al. [73]. Besides the MAMs entropy and positivity, the three metrics normalized circularity, Fold Overlap Ratio and Low-Intensity Region Score are considered. These rely on a prior segmentation of the blurring artifacts or the intensity undershoot areas. During this work, deep-learning-based solutions for

- the recognition of coronary motion artifacts,
- the quantification of the absolute motion artifact levels and
- the estimation of coronary motion directions and magnitudes,

directly from coronary cross-sectional image patches, are aimed for. The content of the related conference and journal articles [25, 26, 68, 69] is reprocessed in Chapter 4 and 5.

Table 2.3: Comparative review of CT metal artifact reduction methods in terms of approach, properties and data in chronological order. In the second row, symbols (first column) and abbreviations (last column) are explained. The proposed approach for *Dynamic Pacemaker Artifact Removal* (DyPAR) is robust to cardiac motion.

Publications	Approach			Properties				Data Source			Keywords		
Symbols ✓ : applicable ~ : not examined T : test data L : learning data	Sinogram Completion	Iterative Reconstruction	Image-To-Image Transfer	Motion-Robust	Multi-Pass	Rule-Based	Data-Driven	Hardware Phantom	Digital Phantom	Computer-Simulated Metal	Clinical Data	Abbreviations LI linear interpolation; EM expectation maximization ART algebraic reconstruction technique ML maximum likelihood NMAR normalized metal artifact reduction FBP filtered back-projection BHC beam hardening correction; DLB deep-learning-based CNN convolutional neural network MC motion compensation	
[51] Kalender	✓				✓	✓		T			T	semi-automatic metal boundary determination; LI	
[113] Wang		✓			✓	✓			T			iterative deblurring; incomplete projections; EM; ART	
[89] Oehler	✓	✓			✓	✓					T	directional interpolation; weighted ML-EM	
[79] Meyer	✓				✓	✓			T		T	NMAR; prior generation; projection normalization	
[124] Zhang		✓			✓	✓		T	T			incomplete projections; constrained optimization	
[108] Slambrouck		✓		~	✓	✓		T	T			block-iterative scheme; local models of varying complexity	
[80] Meyer	✓				✓	✓			T		T	initial NMAR; frequency split; spatial weighting	
[107] Toftgaard	✓			✓	✓						T	marker segm. in projections; trajectory estimation	
[33] Gjestebj	✓		✓		✓	✓	✓	L				initial NMAR; image-based MAR	
[34] Gjestebj	✓				✓	✓	✓	L				initial NMAR; DLB sinogram correction	
[125] Zhang	✓		✓		✓	✓	✓			L	T	initial FBP; initial BHC; initial LI; DLB prior generation	
[120] Xu			✓	~				✓	T	L		deep residual CNN; metal artifact prediction	
[46] Huang			✓	~				✓			L	T	deep residual CNN; metal artifact prediction
[90] Park	✓				✓	✓	✓				L		image-based metal segm.; DLB sinogram correction
[39] Hahn	✓				✓	✓	✓				T	T	metal shadow refinement; LI; NMAR after MC
[70] Lossau	✓				✓			✓		L	T		DyPAR; DLB; metal shadow segmentation
[71] Lossau	✓				✓			✓		L	T		DyPAR+; DLB; metal shadow segm.; inpainting; modeling

2.3.2. PACEMAKER METAL ARTIFACTS

Beam hardening, photon starvation, scattered radiation and the partial volume effect are identified as potential causes of metal artifacts [20]. Beam hardening denotes the effect of a more rapid absorption of low-energy photons by metallic materials. Hence, the assumption of monochromatic X-ray attenuation made within the Lambert-Beer law in Equation (2.1) is violated leading to inconsistent projection data. Photon starvation describes the condition of insufficient photons reaching the detector surface through paths of high X-ray attenuation. For the calculation of the projection integrals in Equation (2.2), the negative logarithm of the ratio of incoming and outgoing numbers of photons is taken. In case of extremely high attenuation, it holds $I_D \rightarrow 0$ and therefore $p \rightarrow \infty$ which causes extremely noise-affected reconstruction. Metal artifacts manifest in streak-shaped intensity overshoots and shadings which frequently preclude the evaluation of neighboring anatomy e.g. with regard to inflammations or calcifications [74].

Besides the adaptation of acquisition to dual-energy protocols for material decomposition [3], several software-based solutions for CT metal artifact reduction (MAR) have been developed in the last decades [32, 82]. An exhaustive collection of related research is listed and compared in Table 2.3, where articles are grouped into three major approaches: sinogram completion, iterative reconstruction and image-to-image transfer.

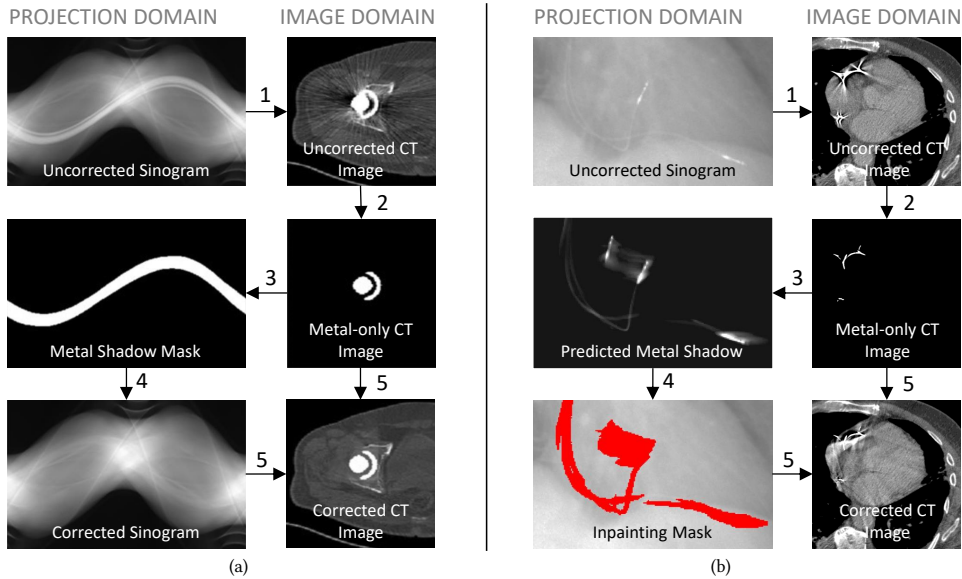


Figure 2.6: Flowcharts of simple sinogram completion (adapted from [77]; Figure 2). (a) MAR is performed by treating metal-affected line integrals as missing data. Incomplete projection data is filled based on surrounding line integrals, e.g. by means of linear interpolation. Sinogram completion implicitly assumes that the object is static during CT acquisition. (b) In case of moving metal objects, the initially reconstructed image volume is perturbed by a combination of metal and motion artifacts. Image-based metal segmentation and subsequent forward projection yields metal shadows of mixed motion states. Therefore, clear disagreements of predicted pacemaker shadow and real metal positions in the uncorrected sinogram are visible. After inpainting and reconstruction, image artifacts have even been aggravated.

Sinogram completion As visualized in Figure 2.6a, sinogram completion commonly comprises the following steps for the generation of consistent projection data by replacement of metal-affected line integrals:

1. Reconstruction of an initial image volume with metal artifacts using FBP
2. Metal segmentation in the initial image volume, e.g. by thresholding, delivers the metal mask in the reconstructed image geometry
3. Forward projection of the metal mask yields the metal shadow in the originally acquired projection geometry
4. Replacement of metal-affected line integrals, e.g. by interpolation or by incorporating prior knowledge
5. Reconstruction of the metal-free image volume using FBP and metal reinsertion using the metal-only image obtained in step 2.

A multitude of variants of this method is known which include different types of pre-processing, inpainting and post-processing procedures [39, 51, 79, 80].

Iterative Reconstruction In comparison to FBP, iterative reconstruction approaches are more robust with regard to noisy, missing or inconsistent projection data. Instead of refilling the holes of metal-affected line integrals, Zhang et al. [124] proposed to generate

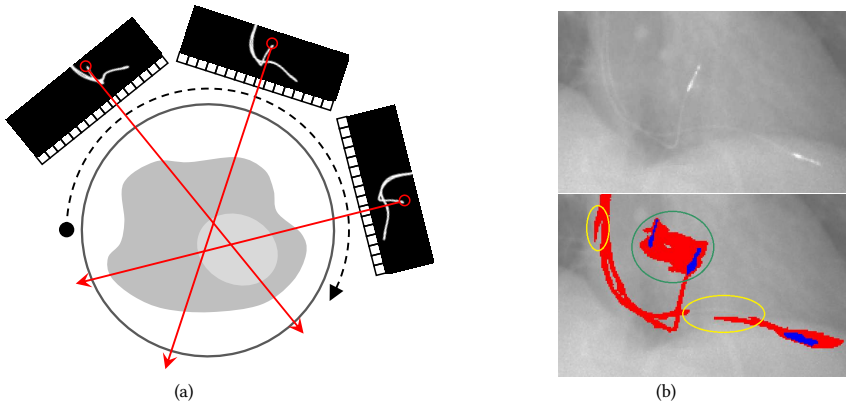


Figure 2.7: Second pass MAR approaches are not applicable for moving metal objects like pacemakers. (a) Each projection view contains the metal shadow at a specific motion state. These are mixed during reconstruction of the initial image volume. (b) Image-based metal segmentation and subsequent forward projection deliver inaccurate predictions of metal-affected line integrals. Blurred electrode shadows (green circle) as well as shifted and interrupted lead shadows (yellow circle) occur. Retrospective separation of the different motion states is not feasible.

the CT image volume directly from incomplete projection data by means of constrained optimization. However, iterative reconstruction is time-consuming. Therefore, van Slambrouck and Nuyts [108] introduced an image block-iterative scheme where metal-regions are reconstructed with a fully polychromatic model whereas non-metal regions are reconstructed with a model of reduced complexity. Also hybrid approaches using both, sinogram completion and iterative reconstruction, have been investigated [89].

Image-to-Image Transfer Direct mapping from MAR-perturbed image data to MAR-free image data, e.g. by means of deep residual convolutional neural networks, allows for artifact suppression without consideration of the raw projection data [46, 120]. However, these approaches are essentially restricted by the information content of the metal-affected input patches. In the articles [33, 125], image-to-image translation is applied as post-processing step after conventional projection-based MAR.

Most of the existing MAR methods are second pass approaches which fail in the presence of motion as they are based on metal segmentation in an initially reconstructed image volume (see Figure 2.6b). In the projection data, each recorded projection view corresponds to a specific motion state. Due to the cardiac motion, metal positions are well-defined in each projection view but not consistent across multiple views (see Figure 2.7a). During reconstruction the different motion states are mixed, thus precluding the extraction of a time-dependent 3D metal models. As already stated by Toftegaard et al. [107] and illustrated in Figure 2.7b, MAR approaches based on metal segmentation in the image domain mainly suffer from the following problems: (1) Due to the motion blur, hounsfield units (HU) of metal objects might get below the segmentation threshold resulting in incomplete metal shadow prediction. (2) Metal object sizes might be increased by the range of motion

resulting in overestimated metal shadow areas. (3) Also in case of reduced motion artifacts by means of ECG-gated reconstruction, predicted metal shadows are often slightly shifted compared to the real ones. This is caused by the temporal resolution required for reconstruction and concurrent metal segmentation in a specific motion state where most of the back-projected metal-affected line integrals are focused, e.g. capturing solely metal positions belonging to cardiac rest phase.

The first pass moving metal artifact reduction (MMAR) method of Toftegaard et al. [107] avoids such problems by automatic segmentation of cylindrical gold markers directly in the projection domain. Two methods for MMAR have been introduced by Hahn et al. [39]. The first approach utilizes image-based metal segmentation for coarse metal shadow determination. Inside the coarse mask, metal shadows are then refined by exploiting edge information in the projection data. The second approach assumes respiratory-gated CT data for the application of normalized MAR (NMAR) on a time series of previously motion compensated CT image volumes. So far, existing MMAR methods are rule-based and rely on the assumption of high-density metal objects as in the case of gold markers.

From the previous literature review and for the dedicated task of pacemaker artifact removal, the following requirements are identified. It is aimed for a MAR approach which

- works completely rawdata-based,
- is applicable to ECG-gated as well as ungated CT scans,
- is robust regarding extremely low differences between metal shadow and background line integrals as occurring in case of pacemaker leads,
- enables metal reinsertion at a preselected motion state.

A DL-based approach is chosen, in order to circumvent the challenging task of designing an appropriate rule-based system with, for instance, the necessary robustness regarding low-contrast projection data. The content of the related conference and journal articles [70, 71] is reprocessed in Chapter 6.

3

MACHINE LEARNING

IN order to address the defined objectives, machine learning approaches or, more precisely, deep learning approaches are applied. Deep learning is a subset of artificial intelligence which distinguishes from rule-based systems and classical machine learning through a purely data-driven methodology (see Figure 3.1). It thereby holds the promise to circumvent the challenging task of hand-designing appropriate rule-based feature extractors or mapping functions. The research field of deep learning is not novel and the concept of convolutional neural networks, for instance, has already been introduced 1980 by Fukushima [30]. However, it was the availability of more powerful and affordable graphics processing units (GPUs), the intense collection of large amounts of data and the development of an accompanying bag of tricks that led to the end of the so-called *AI winter* and enabled significant improvements in various data processing areas like visual perception, decision-making, speech recognition or language translation.

This chapter gives an introduction into the main concepts of supervised machine learning with particular attention to convolutional neural networks as the most relevant DL technique for the problem at hand (see Section 3.1). An overview of current applications of machine learning in cardiac CT imaging is provided in Section 3.2. Finally, the central idea of exploiting system knowledge for the synthesis of labeled data is introduced (see Section 3.3).

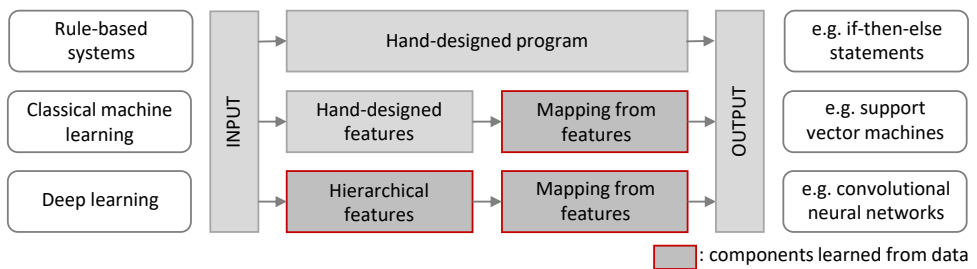


Figure 3.1: Overview on the different categories of artificial intelligence (adapted from [35]; Figure 1.5). Rule-based systems are composed of hand-designed rules which represent a fixed knowledge base, i.e. given data samples of inputs and target outputs are only used for testing purposes. In classical machine learning, relevant feature are first carefully extracted by means of manual feature engineering and subsequently fed to a statistical model such as a support vector machine. Deep learning models are based on the concept of hierarchical feature extraction and trained in an end-to-end fashion. They are characterized by an increased number of tunable parameters and not limited in their complexity.

3.1. STATISTICAL LEARNING THEORY

Several surveys, research monographs and textbooks including [9, 35, 83] provide a concise treatment of statistical learning theory. This section aims to give a rough overview and explain basic concepts applied in this work. In general, one can distinguish between supervised, unsupervised and reinforcement learning. Supervised learning corresponds to the task of finding a mapping that transforms a set of input features into a predefined set of output features. In unsupervised learning (e.g. clustering) one aims to recognize patterns in the input data and extract the corresponding, previously unknown features without the need of pre-existing target labels. Several mixed forms like self-supervision

by means of automatic label generation (e.g. in autoencoders), semi-supervision on partially labeled data sets (e.g. in pseudo labeling) and weak-supervision for higher-level output generation using low-level labels (e.g. segmentation from classification [22]) are known. Finally, in a reinforcement learning framework, a reward-penalty function needs to be provided to the algorithm, which enables a learning agent to assess whether previous actions were expedient and to interact accordingly with its environment. This work focuses on supervised learning approaches only.

3.1.1. SUPERVISED LEARNING

- **Inputs** $\vec{x} \in X$ are independent random variables drawn from the distribution $P(\vec{x})$.
- **Outputs** $\vec{y} \in Y$ are independent random variables assigned to \vec{x} with a conditional probability $P(\vec{y}|\vec{x})$.
- **Hypothesis space** \mathcal{H} defines a set of models $f \in \mathcal{H}$ with $f : X \rightarrow Y$.
- **Loss function** $L(\vec{y}, f(\vec{x}))$, $L : Y \times Y \rightarrow \mathbb{R}$ measures the difference between actual output $f(\vec{x})$ and desired output \vec{y} .

Supervised learning corresponds to the task of finding the function $f_* \in \mathcal{H}$ that minimizes the generalization error

$$E(f) = \int \int L(\vec{y}, f(\vec{x})) P(\vec{x}, \vec{y}) d\vec{x} d\vec{y} \quad (3.1)$$

based on a limited number of data pairs $\vec{z}_i = (\vec{x}_i, \vec{y}_i)$, $i = 1, \dots, m$. Input and output samples can be, for example, one-dimensional vectors or multi-dimensional image volumes. The definition of a proper loss function $L(\vec{y}, f(\vec{x}))$ and hypothesis space \mathcal{H} is part of the problem definition. In case of a parametric statistical model, one can write $\mathcal{H} = \{f_\theta : \theta \in \Theta\}$ where Θ is the parameter space. The model parameters θ are of finite number and can be discrete or continuous. One example of parametric modeling is polynomial regression. The input-output-mapping is defined by the coefficients θ . The degree of the polynomial function determines the model complexity and is called a hyper-parameter. As the probability $P(\vec{x}, \vec{y})$ is unknown, empirical risk minimization (ERM) is performed by searching the parameter setting

$$\theta_{\vec{z}} = \underset{\theta \in \Theta}{\operatorname{argmin}} E_{\vec{z}}(f_\theta) \quad , \text{whereby} \quad E_{\vec{z}}(f_\theta) = \frac{1}{m} \sum_{i=1}^m L(\vec{y}_i, f_\theta(\vec{x}_i)), \quad (3.2)$$

that minimizes the empirical error $E_{\vec{z}}(f_\theta)$ as an approximation of the generalization error. The inference from known samples to the general case is called inductive reasoning. The corresponding generalization error $E(f_{\theta_{\vec{z}}})$ can be separated into the sample and the structural error. The sample error is caused by the limited number of available data samples which might not be a good representation of the real underlying data distribution. The structural risk is determined by the hypothesis space \mathcal{H} which refers to the models capacity. Both components can be associated with the 'garbage in, garbage out' paradigm:

garbage data + appropriate hypothesis space = garbage model
 appropriate data + garbage hypothesis space = garbage model

Figure 3.2 shows the problem of choosing an appropriate hypothesis space and the relation to under- and overfitting. The generalization power of a model can be verified by

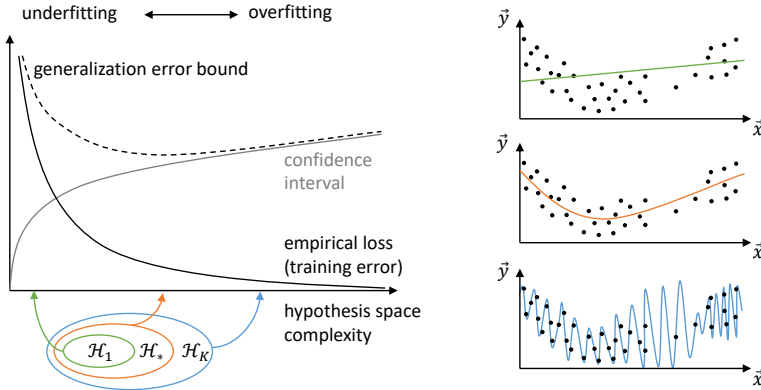


Figure 3.2: Underfitting is related to a hypothesis space of insufficient complexity, i.e. underlying models are not able to capture the given data. The underfitting example (green) shows the prediction of a linear relationship on non-linear data. Empirical risk minimization based on a hypothesis space of increased complexity (blue) is prone to overfitting, i.e. the risk of finding a function $f_{\theta_{\vec{z}}}$ that matches the known data nearly exactly but does not predict future output well. No generalization happens due to simple memorization of the given data pairs including outliers and noise. An appropriate predictive model (orange) represents a trade-off between complexity and empirical accuracy.

previously separating the known data into subsets for training and validation. Model optimization is performed solely on the training data, while the validation data is used to monitor the learning process. Low or even no decrease of the training and validation losses can be a sign of underfitting caused by an insufficient hypothesis space complexity or poor model optimization. Simultaneous decrease of the training loss and increase of the validation loss indicates overfitting, i.e. the model captures almost every information of the given training data including the noise and does not generalize to unseen test samples. Possible strategies to overcome the problem of overfitting are

- **regularization**, i.e. adaptation of the optimization function to $E_{\vec{z}}(f_{\theta}) + \lambda R(\theta)$ by adding a regularization term $R(\theta)$ which, for instance, imposes a smoothness constraint by penalizing high model coefficients,
- **structural risk minimization**, i.e. an implicit regularization where balancing between the empirical loss and the capacity of the approximating model is performed,
- **early stopping**, i.e. an implicit regularization in time where model optimization is stopped when the validation loss starts stagnating or decreasing,
- **pruning**, i.e. retrospective reduction of the model capacity by removing model parameters based on their contribution to the model output in an iterative fashion,
- **data augmentation**, i.e. enlarging the number of data pairs by applying either label-preserving or label-altering transformations that reflect real-world variations. By this procedure, prior knowledge is incorporated and invariants or equivariants (e.g. towards translations, rotations or mirroring) can be build into the models.

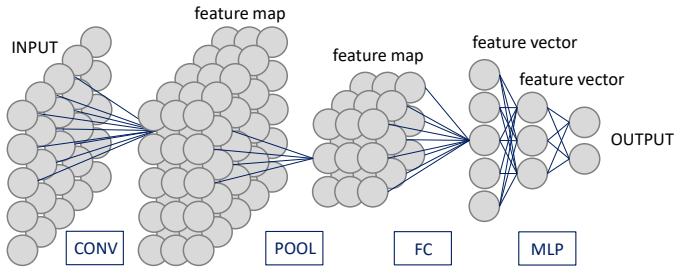


Figure 3.3: Exemplary visualization of a convolutional neural network (CNN). Basic components of CNNs are convolution layers (CONV) and pooling layers (POOL). Fully connected layers (FC and MLP), in contrast, do not exploit the topology of image data.

3

The choice of appropriate overfitting prevention strategies is problem-specific. Classification, regression, object detection, segmentation, registration, inpainting and style-transfer are common image-related tasks. For solving these problems, a wide range of supervised learning algorithms such as support vector machines (SVMs), Gaussian processes and artificial neural networks are available, each with its strengths and weaknesses. The so-called *no free lunch* theorem says that there is no single learning algorithm that works best on all supervised learning problems [117]. In their article, Wolpert and Macready state:

"[...] if an algorithm performs well on a certain class of problems then it necessarily pays for that with degraded performance on the set of all remaining problems."

3.1.2. CONVOLUTIONAL NEURAL NETWORKS

Deep learning is a kind of machine learning based on the concept of hierarchical feature extraction and refers to deep artificial neural networks or less frequently to deep reinforcement learning. During this work, supervised learning tasks are solved by means of convolutional neural networks (CNNs), a specific type of mapping function which defines the state-of-the-art in many computer vision applications.

Network architecture An artificial neural network (ANN) in its most simple form is structured as a cascade of layers

$$\tilde{x}^{(i)} = \phi\left(W^{(i)}\tilde{x}^{(i-1)}\right) \quad , \text{ for } i \in \{1, \dots, N\} \quad (3.3)$$

with input $\tilde{x}^{(0)}$ and predicted output $\hat{y} = \tilde{x}^{(N)}$. The hyper-parameter N defines the network depth and ϕ denotes a non-linear activation function like, for instance, the sigmoid function or the rectified linear unit (ReLU) [84]. The set of weight matrices $\theta = \{W^{(i)}\}_{i=1}^N$ represents the model parameters which are learned during the training process, i.e. the hypothesis space \mathcal{H} or rather the parameter space Θ is given by the network architecture while the learned weights define the mapping function f_{θ} .

The structure of CNNs is loosely inspired by the information processing in the primary visual cortex and therefore characterized by sparse connections and shared weights [30]. More precisely, the multiplication with a fully-connected weight matrix $W^{(i)}$ is restricted to a convolution of limited receptive field in at least one layer of a CNN leading

to translation invariant feature extraction. As visualized in Figure 3.3, pooling operators build another basic component of CNNs. Common methods for downsampling of feature maps are strided convolutions, max- and average-pooling [8]. In Figure 3.3 only one convolution layer is depicted. In practice, pooling layers alternate with stacks of convolution layers. By this procedure of alternating multiplication of the number of feature maps and spatial downsampling, representations are learned as a nested hierarchy of concepts. The abstraction level of extracted features increases with the network depth.

Depending on the learning task, further layer types are integrated into the network. In case of voxel-level predictions (e.g. segmentation), upsampling layers are frequently utilized in order to invert the spatial downsampling of the pooling operations [94]. In case of image-wise predictions (e.g. classification or regression), fully connected (FC) layers or rather multilayer perceptrons (MLPs) often define the final layers of the network [59]. By this procedure and in contrast to classic machine learning, feature extractor and mapping function are learned simultaneously, i.e. CNNs are trained in an end-to-end fashion.

Learning setup Stochastic gradient descent (SGD) solvers are popular neural network optimizers. In each training iteration, gradients are approximated based on a randomly selected subset of $m_{\text{batch}} \ll m$ training samples. Suitable smoothness properties (e.g. sub-differentiability) of the loss function and network layers are prerequisites for gradient descent. For this reason, supplemental loss functions, like the cross-entropy in classification tasks, commonly drive the learning process, while the real objective is defined as validation metric. In SGD, the set of weight matrices θ_i at iteration i is updated by

$$\theta_{i+1} = \theta_i + V_{i+1} \quad , \text{ where } \quad V_{i+1} = \zeta V_i - \epsilon(i) \nabla_{\theta_i} L \quad \text{ and } \quad V_0 = 0,$$

i.e. a linear combination of the previous weight change V_τ and the negative gradient of the loss function $-\nabla_{\theta_i} L$ which is determined by backpropagation [61]. These terms are weighted by the momentum ζ and the learning rate $\epsilon(i)$. Several adaptive learning rate methods like ADAGRAD [21], ADADELTA [122] and Adam [57] have been developed within the last decade. Besides the learning rate, the choice of appropriate starting values θ_0 is crucial for the convergence behavior. In [59] network weights are initialized by random values following a zero-mean Gaussian distribution. Weight normalization according to the activation function ReLU has been proposed by He et al. [41] and is applied in the following. Fine-tuning of a pre-trained model, either partially or end-to-end, constitutes an alternative to learning from scratch. The network optimization is performed for a fixed number of training epochs, where one epoch is defined as a run through all m training samples. Typically, the predictive model with the highest validation accuracy over all epochs is selected for further application.

It has to be noted that only a rough introduction to the fascinating world of machine learning is provided here. In practice, a multitude of additional tricks is used to facilitate learning processes like the deployment of batch normalization to stabilize activation distributions [47], skip connections for improved gradient flows [42] or dropout as strategy to prevent overfitting [104]. Due to the accordingly high number of adjusting screws which include all hyper-parameters within network design and learning setup, CNN optimization can be quite complex and often requires multiple tuning runs.

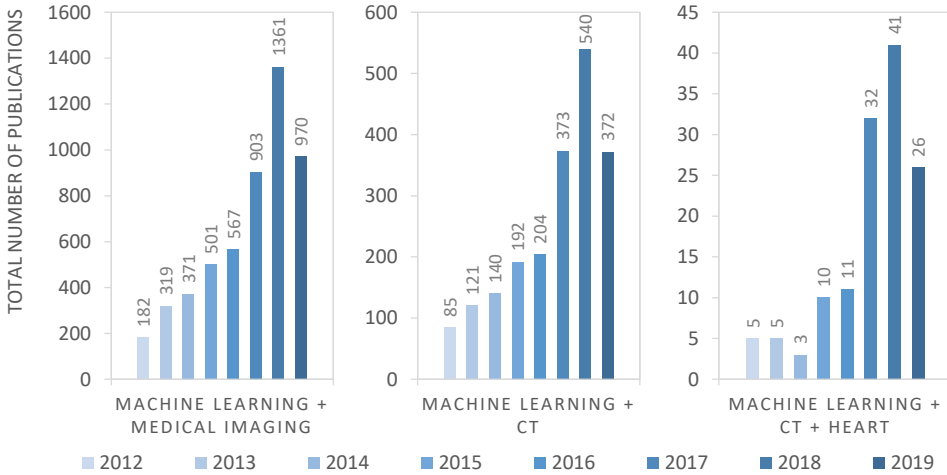


Figure 3.4: Total number of publications listed in Web of Science Core Collection database for the years 2012 to 2019. Statistics are determined using the basic search on all fields. (Date: September 2019)

3.2. APPLICATIONS IN CARDIAC CT

Figure 3.4 shows the general increase of publications in the research field of machine learning in medical imaging and the share of articles particularly dealing with machine learning in cardiac CT. The number of publications in 2019 cannot yet be accurately estimated, but it is indicated that a peak is reached. An overview of further existing studies on the application of machine learning methods in cardiac CT imaging besides the detection and removal of coronary motion artifacts and pacemaker metal artifacts is provided in this section. Please refer to the review article [103]. In addition, a selection of publications is listed in Table 3.1. On the basis of this literature collection, the following trends can be derived.

Purposes: Table 3.1 reveals the diversity of processing tasks like risk assessment, landmark detection, anatomical segmentation, noise reduction, domain transfer and registration. Due to the high clinical relevance, a multitude of articles focuses on the analysis of the coronary arteries [29, 86, 118, 119, 127].

Methods: CNNs are predominant and frequently applied in an end-to-end fashion [1, 11]. Further examples which are not listed in Table 3.1, are the direct Agatston score regression proposed by Cano-Espinosa et al. [16] and the global coronary lumen segmentation via paired 3D multi-scale CNNs as introduced by Chen et al. [18]. Combinations with or integrations into model-based pipelines can additionally provide a benefit. In [100], coarse U-Net vessel segmentation outputs are refined by means of a level set approach. Wolterink et al. [119] address the task of lumen segmentation by an iterative vessel tracking algorithm. It is based on a CNN which predicts the coronary centerline direction and lumen radius directly from local image patches. By this procedure, the tubular structure

Table 3.1: Representative selection of publications in the research field of machine learning in cardiac CT.

Article	Purpose	Method	Data
Nickisch 2015 [86]	Coronary blood flow simulation	Lumped model for FFR simulation; learned translation of local vessel geometries into parameters of nonlinear vessel segment resistor; hydraulic effect coefficient prediction via NNLS	32 CCTA images; 3D finite element simulation; 59 FFR invasive measurements
Wolterink 2016 [118]	Automatic CAC scoring	Bounding box detection by three independent CNNs on orthogonal planes; paired CNNs for voxel-wise proposal extraction and classification	250 CSCT and CCTA images; manually annotated bounding box center and calcification positions; automatic CAC score calculation
Motwani 2016 [81]	5-year ACM prediction	Automated feature selection by information gain attribute ranking; LogitBoost for ACM classification	10030 patients; 25 clinical and 44 CCTA parameters
Chartsias 2017 [17]	MR synthesis from cardiac CT data	Circle GAN for unpaired image-to-image translation; improved myocardium segmentation using synthetic MR data for U-Net training	20 CT images; 20 MR images (unpaired); manual heart segmentation
Noothout 2018 [88]	Landmark detection	Multi-task FCN with shared weights; 3D displacement vector regression from patch center to eight landmarks; classification whether patch contains landmark	198 CCTA images; manually annotated landmark positions
Joyce 2018 [49]	Multi-class anatomical segmentation	Train GAN to synthesize realistic segmentation masks; reconstruction error, region variance (minimization) and region size (maximization) as additional unsupervised correspondence costs	20 CT images; 20 MR segmentation masks (unpaired)
Zreik 2018 [127]	Stenosis classification	LV myocardium segmentation by multiscale CNN; unsupervised CAE and spatial clustering for feature extraction; SVM classifier for significant stenosis	166 CCTA images; manual LV segmentation; invasive FFR measurements
Wolterink 2019 [119]	Coronary artery centerline extraction	Vessel direction and radius prediction from local image patch by 3D dilated CNN; iterative tracking stopped based on uncertainty of direction classifier	8+24 CCTA images; manually annotated centerline location and radius
Ahn 2019 [1]	Low-dose image denoising	Residual U-Net for low-dose noise prediction	25 CTA images; pairs of low-dose CTA and noise images generated by low-dose simulation tool
Freiman 2019 [29]	Abnormality detection in coronary arteries	Unsupervised anomaly detection in coronary cross-sectional image patches using deep sparse autoencoders	90 CCTA images; manually annotated centerlines; automatic lumen and wall segmentation; normal cross-section defined by a lumen-wall-ratio > 0.8
Bruns 2019 [11]	Whole-heart segmentation in NCCT	FCN training on multi-slice VNC images with propagated labels	18 contrast-enhanced CT scans; automatic segmentation in CCTA images; label propagation to VNC images; 218 NCCT images
Hering 2019 [43]	Inter-patient CT-MR whole-heart registration	MVF estimation using three independent 2D CNNs on orthogonal planes; NGF for unsupervised distance measuring; weak supervision using segmentation labels	20 CT images; 20 MR images (unpaired); manual heart segmentation

Abbreviations: ACM all-cause mortality; CAC coronary artery calcium; CAE convolutional autoencoder; CSCT calcium scoring CT; FCN fully convolutional network; FFR fractional flow reserve; GAN generative adversarial network; LV left ventricle; NCCT non-contrast CT; NGF normalized gradient field; NNLS non-negative least-squares; VNC virtual-non-contrast

of the vessel segments is ensured. One can observe that sampling of *rich* input and output features which accurately represent the underlying data structure is still essential, despite the feasibility of model training in an end-to-end fashion. In several articles, the main methodological focus lies in the feature engineering or the data generation itself.

Data: Several smart ways to circumvent the need of manual annotation are introduced by either application of unsupervised learning procedures [49] or development of software solutions for automatic paired data generation [1, 11, 29]. Furthermore, different strategies to get the highest gain out of given information are described. Hering et al. [43], for instance, incorporate modality-independent prior information in form of segmentation distance measuring and a second order curvature regularizer to guarantee smooth deformation fields. In [88], the CNN is forced to learn additional anatomical context by defining the supervised learning problem as combined landmark classification and displacement vector regression task.

3.3. FORWARD MODELS FOR LABELED DATA SYNTHESIS

As already mentioned in the previous section, the application of machine learning and especially deep learning leads to a focus shift from algorithm design to data handling. Several public tools and pre-trained networks are available which enable learning of predictive models without proper knowledge in statistical learning theory. An example is the interactive deep learning GPU Training System (NVIDIA DIGITS 6, NVIDIA Corporation, Santa Clara, CA, USA). Data collection, curation, preparation and handling is often more time-consuming and challenging than the network design and the hyper-parameter tuning. Hornik et al. [44] proved that an ANN with just one hidden layer can approximate any continuous function to any degree of accuracy as long as it comprises a sufficient number of neurons. That means, deep learning holds the promise to solve tasks of any complexity. However, the universal approximator characteristic can only be exploited with large amounts of appropriately labeled, representative data.

This can be crucial, especially in the medical domain, as data collection is restricted by multiple types of regulations. Label noise, high time expenditure and the limitation to low-level labels are the main disadvantages of manual annotation as explained in Chapter 1. Besides the use of hardware phantoms (e.g. the Dynamic Cardiac Phantom Model 008C from CIRS¹) and additional data acquisition (e.g. by ECG units or respiratory belts), software solutions form an alternative way of data collection. Synthetic sample generation has been investigated from different perspectives:

- **Over-sampled data** is generated in [91] by convex combination of seed patterns belonging to the same class. A pre-existing labeled database is required here.
- **Paired data** is generated in [126] based on two classes of unpaired image data using cycle GANs. Unsupervised image-to-image translation enables label propagation for subsequent supervised learning as performed by Chartsias et al. [17].
- **Fully-synthesized data** is generated in [75] among others by photo-realistic 3D rendering for subsequent network training on optical flow estimation.

¹<https://www.cirsinc.com/products/all/128/dynamic-cardiac-phantom/>

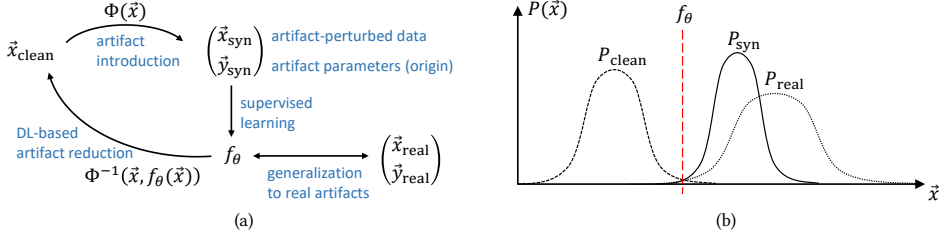


Figure 3.5: (a) Development of a prior knowledge-driven forward model Φ is proposed in order to train an AI-enabled backward model Φ^{-1} . The forward model requires artifact-free input data and delivers the feature and label data $(\vec{x}_{\text{syn}}, \vec{y}_{\text{syn}})$ required for supervised learning as output. It aims to mimic the transition from artifact-free data \vec{x}_{clean} to artifact-perturbed data \vec{x}_{real} . (b) Example of a classification network trained on artifact-free data $\vec{x}_{\text{clean}} \sim P_{\text{clean}}$ and data with synthetic artifacts $\vec{x}_{\text{syn}} \sim P_{\text{syn}}$. The decision boundary is illustrated as dashed red line and also generalizes to real artifact-perturbed data $\vec{x}_{\text{real}} \sim P_{\text{real}}$, despite mismatching data distributions P_{syn} and P_{real} .

3

- **Transformed data** is generated in [27] by clipping of projection data for subsequent network training on FOV extension.
- **Augmented data** is generated in several machine learning articles by performing, for instance, random cropping, rotation or mirroring. In contrast to the previous category of transformed data, a pre-existing labeled database is required here.

In this work, the fourth category of data synthesis approaches - the transformation of existing clinical data by a prior-knowledge-based forward model is applied. As visualized in Figure 3.5a, the forward model Φ takes artifact-free clinical data \vec{x}_{clean} as input and delivers a tuple of artifact-introduced data \vec{x}_{syn} and corresponding target labels \vec{y}_{syn} as output. The labels are task-specific and can, for instance, be ground truth motion vectors or metal shadow masks. Within the forward model, existing model-based approaches and the given knowledge about the cardiac anatomy, the CT projection geometry and task-specific measurements like vessel velocities [111] or metal densities are incorporated.

A predictive model f_θ is trained to estimate underlying artifact parameters based on the synthetic artifact-perturbed data, i.e. $f_\theta(\vec{x}_{\text{syn}}) \cong \vec{y}_{\text{syn}}$. At this point, it is assumed that the information \vec{y}_{syn} is incorporated in the provided data \vec{x}_{syn} . The main goal is that the predictive model f_θ also generalizes to real-world data \vec{x}_{real} for which corresponding artifact parameters \vec{y}_{real} are unknown. The predictive model f_θ can, for instance, be trained and applied for artifact recognition and quantification, or be integrated into an AI-enabled backward model. Given a predictive model which produces sensible results for real data, the inverse problem of artifact removal can be addressed as follows

$$\phi^{-1}(\vec{x}_{\text{real}}, f_\theta(\vec{x}_{\text{real}})) \approx \vec{x}_{\text{clean}}. \quad (3.4)$$

It has to be noted that the methodological approach is only sketched here. The reference data \vec{x}_{clean} consists of projection data and additional information like ECG measurements. But, in case of artifact removal, the output of the backward model is the reconstructed image volume without artifacts. Furthermore, some processing steps like the feature sampling will be considered later.

In general, it is not mandatory that the underlying data distribution P_{real} is accurately approximated by the forward model. Figure 3.5b shows one example of sensible artifact classification despite mismatching data distributions P_{real} and P_{syn} . However, representative synthetic data, i.e. realistic features with nicely calibrated target labels, is always beneficial. The quality of the learning data is governed by the complexity of the forward model. Detailed descriptions of the proposed coronary motion and the pacemaker metal forward models are provided in the following Chapters 4 and 6.

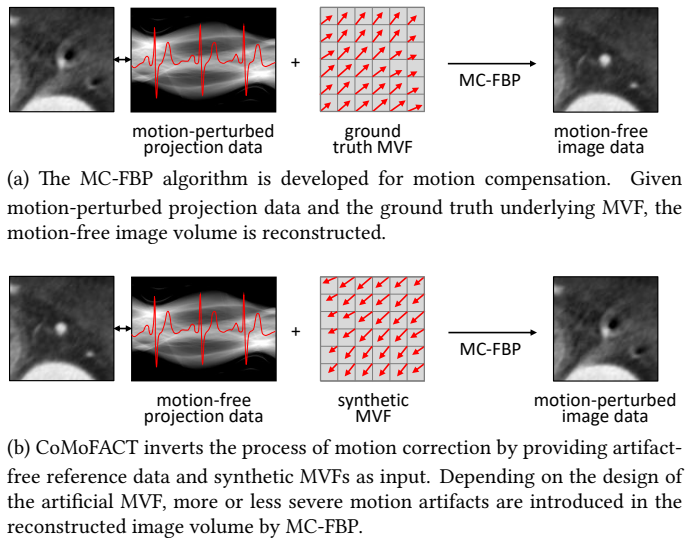
4

RECOGNITION AND QUANTIFICATION OF CORONARY MOTION ARTIFACTS

FIGURE 4.1 illustrates the basic concept of the *Coronary Motion Forward Artifact model for CT data* (CoMoFACT). Synthetic motion is introduced to reference cases with excellent image quality by means of the MC-FBP algorithm described in Section 2.2.3. Development of the forward model went through different levels of complexity. First, the CoMoFACT was designed in such a way that piecewise linear motion is introduced in a limited area around the vessel tree. In the corresponding conference article [26], the feasibility of motion artifact recognition using CNNs has initially been demonstrated. However, testing is performed on data with synthetic artifacts only. Extending this work, the journal article [69] furthermore presents an approach for DL-based artifact quantification. The CoMoFACT is modified for local motion introduction in a limited area around a single coronary centerline point. By this procedure, the relation between motion direction, angular reconstruction range and coronary orientation can be steered. In addition to the piecewise linear motion model, a constant linear motion model is applied for data generation. The article also provides an extensive evaluation on clinical data with real artifacts. This chapter is mainly inspired by [69] and structured as follows:

- 4.1 CCTA cases are collected and grouped into clinical reference or test data.
- 4.2 The CoMoFACT generates the required learning data by introducing simulated and hence controlled motion to the artifact-free reference data.
- 4.3 An ensemble of CNNs is trained for the classification task of separating motion-free and motion-perturbed coronary cross-sectional image patches. In addition, an ensemble of CNNs is trained for the task of artifact level regression.
- 4.4 Accuracy and generalization capabilities of the resulting predictive models in relative and absolute motion artifact measurement are investigated.
- 4.5 Limitations related to label calibration, centerline extraction and amount of learning data are discussed.

Figure 4.1: The forward model for coronary motion introduction called CoMoFACT is based on the MC-FBP algorithm. The MC-FBP takes raw projection data and a motion vector field (MVF) defined on the image grid as input. It spatially displaces line integrals during back-projection and delivers the motion-corrected image volume as output. MC-FBP leads to a compensation of correctly estimated motion (a), whereas the application of the MC-FBP with an artificial MVF on high quality cases induces motion artifacts (b).



4.1. MATERIAL

The CoMoFACT uses artifact-free CCTA data with step-and-shoot acquisition protocol as a reference point to introduce motion. The restriction to step-and-shoot data offers the advantage of artifact generation in a well-controlled situation, i.e. without table movement and multi-cycle reconstruction. Besides the artifact-free reference data, additional clinical data with real motion-perturbation is collected for testing purposes.

4.1.1. CLINICAL REFERENCE DATA WITHOUT ARTIFACTS

Slice-by-slice visual inspection is performed to gather contrast-enhanced cardiac CT data sets which exhibit no coronary motion artifacts in the reconstructed CT image volume. In total, 17 prospectively ECG-triggered clinical data sets from different patients are collected. The reference cases are acquired by a 256-slice CT scanner (Brilliance iCT, Philips Healthcare, Cleveland, OH, USA) with a gantry rotation speed of 0.272 sec per turn. The mean heart rates of the patients HR_{mean} ranged from 45.2 bpm to 66.0 bpm during acquisition. Cardiac CT image volumes are reconstructed at the mid-diastolic quiescent phase by AWCR (see Section 2.2.2). The center of the cardiac gating window, hereafter called the reference cardiac phase r , is chosen between 70% and 80% R-R interval, respectively. In addition to the reconstructed CT image volumes which determine the *no motion* state, the corresponding ECG and raw projection data is required. Furthermore, the coronary artery tree of each case is segmented by means of the Comprehensive Cardiac Analysis Software (IntelliSpace Portal 9.0, Philips Healthcare, Cleveland, OH, USA). As the image volumes are of excellent quality, the software delivers a set of accurately determined centerline points $\vec{c} \in C$ as output. Associated information on the corresponding cross-sections including lumen contour and the normal vector $\vec{n}_{\vec{c}}$ in centerline direction are provided as well.

4.1.2. CLINICAL TEST DATA

In order to test the transferability to non-synthetic data, eight clinical cases which exhibit real motion artifacts are collected. Step-and-shoot data (five vessels) as well as helical data (three vessels) from different patients is included. The clinical test cases are acquired by a Brilliance iCT and coronary centerlines are extracted by the Comprehensive Cardiac Analysis Software. The step-and-shoot scans are performed using the same acquisition protocol as in the reference data. The extended temporal scan range in ECG-gated spiral acquisition enables retrospective multi-phase reconstruction. Centerlines are extracted for these cases from the CT image volume reconstructed at 75% R-R.

4.2. FORWARD MODEL

As visualized in Figure 4.2, the CoMoFACT takes a reconstructed CT image volume with corresponding ECG-triggered raw projection data and segmented coronary artery tree as input. It delivers locally motion-perturbed CT image volumes as output, whereby the introduced motion magnitude is determined by the control parameter $s \in \mathbb{R}^+$, hereafter called the target motion strength. As already mentioned, the CoMoFACT introduces simulated motion by application of the MC-FBP algorithm using a synthetic MVF. For each centerline point $\vec{c} \in C$ in the coronary artery tree, a continuous MVF $\vec{d}_{\vec{c}}$ with pre-selected

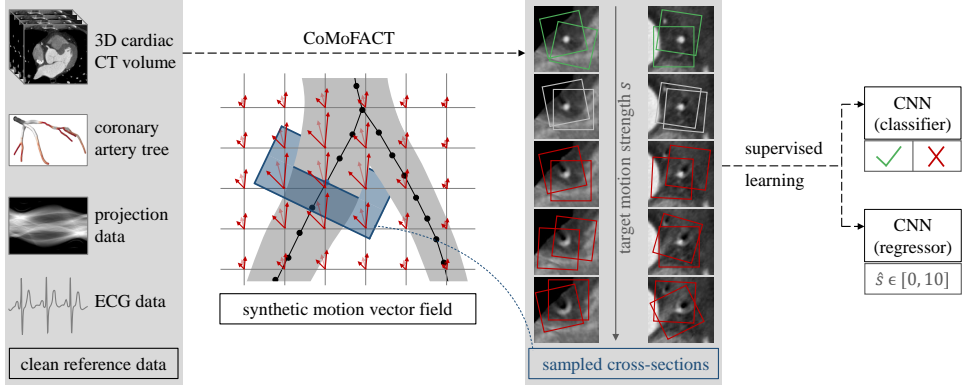


Figure 4.2: A *Coronary Motion Forward Artifact model for CT data* (CoMoFACT) is developed which enables to transform cardiac CT data sets with excellent image quality to locally motion-perturbed CT image volumes. Motion is introduced around each centerline point of the coronary artery tree by respectively creating a synthetic continuous MVF and applying the MC-FBP algorithm. The target motion strength s is a control parameter which scales the length of the displacement vectors in the CoMoFACT. A coronary cross-sectional patch (highlighted in blue) is sampled perpendicular to the centerline from the locally motion-perturbed image sub-volume and used as input data for supervised learning. Corresponding ground truth labels are defined by means of the target motion strength s . Finally, CNNs are trained for motion artifact classification and artifact level regression on randomly rotated, mirrored and cropped cross-sectional patches (highlighted in green, gray and red).

target motion strength s is created. The Subsections 4.2.1, 4.2.2, 4.2.3 detail the design of the synthetic MVF and underlying motion models. Subsequent MC-FBP yields the CT image volume locally motion-perturbed around the corresponding centerline point. The reversing motion trajectory $\vec{d}_{\vec{c}}^{-1}$ corresponds to the simulated heart motion during acquisition. After motion introduction via CoMoFACT, one cross-sectional image patch is sampled perpendicular to the centerline as input data for supervised learning. Corresponding target labels are defined by the underlying target motion strength s (see Section 5.3.1).

4.2.1. SYNTHETIC MOTION VECTOR FIELD

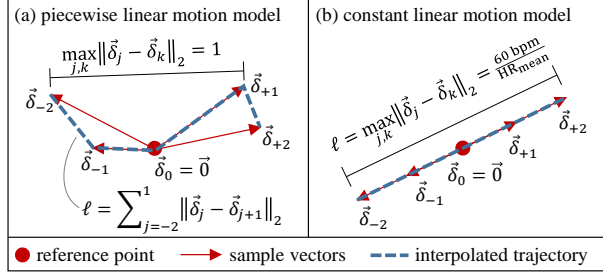
The continuous MVF $\vec{d}_{\vec{c}}: [0\%, 100\%] \times \Omega \rightarrow \mathbb{R}^3$ describes the displacement of each voxel coordinate $\vec{v} \in \Omega \subset \mathbb{R}^3$ in the CT volume at each time point $t_{cc} \in [0\%, 100\%]$ in millimeters. Due to the simulation of periodic motion, time is measured in percent cardiac cycle here. The artificial MVF is defined by three separable components:

$$\vec{d}_{\vec{c}}(t_{cc}, \vec{v}) = s \cdot m_{\vec{c}}(\vec{v}) \cdot \vec{\delta}_{\vec{c}}(t_{cc}) \quad (4.1)$$

The first component is the pre-selected target motion strength s which scales the length of each displacement vector in the continuous MVF. The role of s as motion level regulator is illustrated in Figure 4.2. Whenever $s = 0$, no motion is introduced and the CoMoFACT delivers the input CT image volume without motion artifacts as output.

The second component is the location-dependent weighting mask $m_{\vec{c}}: \Omega \rightarrow [0, 1]$ which restricts the motion to a limited area around the currently processed centerline point \vec{c} and additionally forces the MVF to be spatially smooth. It is defined as a 3D trapezoidal function generated by binary dilation of the centerline point and subsequent

Figure 4.3: Schematic drawing of the time-dependent motion trajectories (dashed blue lines) determined by the sample vectors $\vec{\delta}_i$ (red arrows) for both sub-models. (a) The piecewise linear motion model comprises random displacement directions and varying velocities. (b) The constant linear motion model is restricted to a predefined motion direction and equidistant sample vectors.



uniform filtering. A kernel radius of 15 mm for dilation and a uniform filter size of 12.4 mm \times 12.4 mm are chosen in the following experiments. The limitation of the motion area is required to avoid undesired motion artifacts from peripheral structures like bones. The smoothing is necessary to avoid reconstruction artifacts as elastic tissue structure forbids abrupt changes of motion in a local neighborhood.

The third component $\vec{\delta}_{\vec{c}}: [0\%, 100\%) \rightarrow \mathbb{R}^3$ defines the motion direction for each point in time. It is obtained by piecewise linear interpolation between five sample vectors $\vec{\delta}_i \in \mathbb{R}^3$, $i \in \{-2, -1, 0, +1, +2\}$. The corresponding phase points $t_i \in \{r - 10\%, r - 5\%, r, r + 5\%, r + 10\%\}$ are assigned around the reference heart phase r of the input CT volume with a temporal distance of 5% cardiac cycle. The temporal projection range required for reconstruction depends on the heart rate and the gantry rotation speed. For the given data sets, the angular weighting window is narrower than 20% cardiac cycle respectively, so no extrapolation has to be performed.

A schematic drawing of the artificial MVF is given in Figure 4.2. The displacement vectors (light red arrows) are linearly interpolated in time domain from the sample vectors (dark red arrows) to obtain the motion state at some $t_{cc} \in [r - 10\%, r + 10\%]$. For a phase point t_{cc} , the motion directions are spatially constant, while the displacement length decreases with increasing distance to the currently processed centerline point \vec{c} (highlighted in blue). Two model variants are presented in the following Subsections 4.2.2 and 4.2.3 which differ in terms of the sample vector definition. Both concepts are compared in Figure 4.3.

4.2.2. PIECEWISE LINEAR MOTION MODEL

The first model variant of piecewise linear motion has been introduced in Elss et al. [26] and was developed for the classification task of separating *no-artifact* and *artifact* coronary cross-sectional patches. The sample vectors are calculated by:

$$\vec{\delta}_i = \frac{\vec{\rho}_i}{\max_{j,k} \|\vec{\rho}_j - \vec{\rho}_k\|_2} \quad (4.2)$$

The motion directions are given by random uniform vectors $\vec{\rho}_i \sim \mathcal{U}[-1, 1]^3$ for $i \in \{-2, -1, +1, +2\}$ and the center heart phase defines the reference state of no motion, i.e. $\vec{\rho}_0 = \vec{0}$. Normalization of the random uniform vectors $\vec{\rho}_i$ is performed so that the target motion strength s finally corresponds to the maximal displacement during 20% R-R interval in millimeters (see Figure 4.3a).

The choice of random uniform vectors for $\vec{\rho}_i$ enables complex motion trajectories and the application of the CoMoFACT leads to realistic-looking motion artifact pattern with differently shaped blurring and intensity undershoots (see Figure 4.4a). The target motion strength s is a weak surrogate of the visual artifact level. For DL-based motion artifact quantification, the motion model is adapted with the objective of increasing the correlation between selected target motion strength s and the resulting visual artifact level. Compared to [26], the CoMoFACT is extended for cross-section-wise motion corruption under consideration of the mean heart rate and the angular reconstruction range.

4.2.3. CONSTANT LINEAR MOTION MODEL

The second model variant is developed for the regression task of predicting the artifact level in coronary cross-sectional patches. Multiple factors besides the motion level during acquisition have an impact on the artifact level. The phantom study in Figure 4.5 shows that the relation between motion direction and the angular reconstruction range is essential. Most severe artifacts occur in case of motion which is orthogonal to the mean reconstruction direction (highlighted in gray). In addition, the visual artifact level depends on surrounding background intensities, the temporal resolution required for reconstruction and the relation between motion direction and vessel orientation.

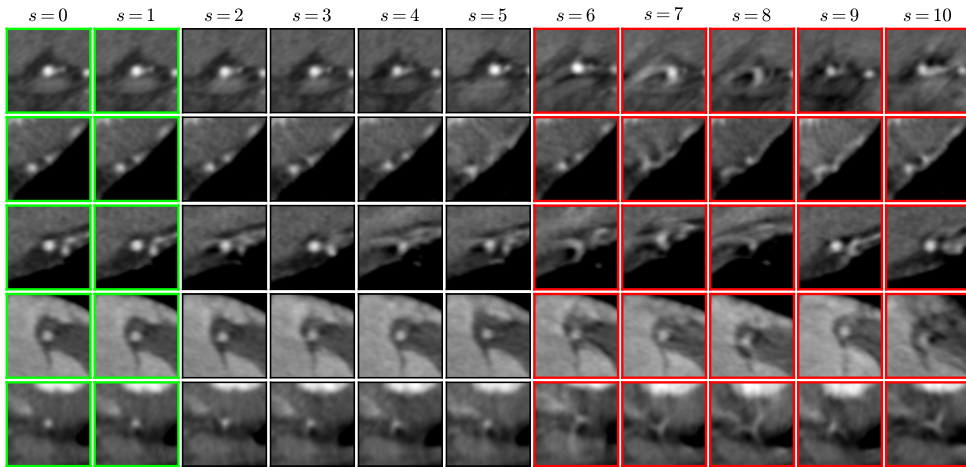
In contrast to the classification model variant presented in Section 4.2.2, severe restrictions are made to consider each of the aforementioned influencing factors except for surrounding background intensities. The sample vectors are calculated now by the following formula:

$$\vec{\delta}_i = \frac{60 \text{ bpm}}{\text{HR}_{\text{mean}}} \frac{i}{4} \cdot \frac{\vec{\rho}_{\text{orth}}}{\|\vec{\rho}_{\text{orth}}\|_2} \quad (4.3)$$

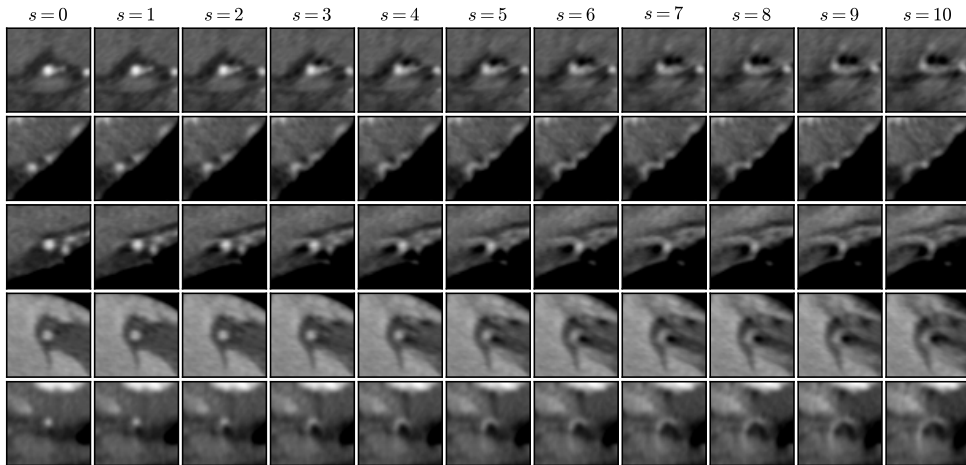
As illustrated in Figure 4.3b, the regression model is limited to constant linear motion. It takes the mean heart rate HR_{mean} of each data set during acquisition into account to force homogeneous velocities among the clinical input cases. In contrast to the previous classification model, the motion direction now depends on the currently processed centerline point \vec{c} .

The motion direction determined by $\vec{\rho}_{\text{orth}}$ is defined as the cross product of the normal vector $\vec{n}_{\vec{c}}$ of the corresponding centerline segment and the mean reconstruction direction in axial plane (see Figure 4.5, right). The mean reconstruction direction is computed by means of the gantry rotation angle at the center of the cardiac gating window and is constant for each voxel reconstructed by the same circular scanning shoot. In case of helical acquisition trajectories, voxels are not necessarily reconstructed by one coherent angular segment. So, in contrast to the classification model, the regression model with its orthogonal displacement directions is not directly transferable to helical cases.

Figure 4.4b shows coronary cross-section images with varying target motion strength s generated by the constant linear motion model. In comparison to the corresponding outputs of the piecewise linear model in Figure 4.4a, the data looks more consistent and the target motion strength s can be interpreted as an approximate measure for the artifact level. However, the severely restricted constant linear motion model only allows for a specific artifact appearance (banana-shaped blurring) whereas the more complex motion trajectories of the piecewise linear motion also generates bird-shaped artifacts. Therefore, the risk of overfitting should be considered in the evaluation of the regression networks.



(a) Example output patches of the piecewise linear motion model from Section 4.2.2.



(b) Example output patches of the constant linear motion model from Section 4.2.3.

Figure 4.4: Coronary cross-sectional image patches are sampled from motion-perturbed CT image volumes, which are generated by means of the CoMoFACT. Each row shows the same cross-section of size 60×60 pixels in different motion states. Both sub-models of the CoMoFACT are compared and reveal the trade-off between motion model complexity and suitability of s as artifact level surrogate.

(a) The piecewise linear motion model from Section 4.2.2 leads to a wide range of coronary artifact appearances. But, the visual coronary artifact level is not monotonically increasing with the underlying target motion strength s . Patches highlighted in green and red are assigned to the classes *no artifact* or *artifact*, respectively. Non-highlighted patches with a target motion strength s between two and five cannot be uniquely assigned to a certain class and are, therefore, excluded from the learning process.

(b) The constant linear motion model from Section 4.2.3 leads to a limited range of arc-shaped coronary artifact appearances. But, the underlying target motion strength s highly correlates with the visual artifact level at the coronary arteries. All patches are included as input data in the regression learning process with s as their corresponding label.

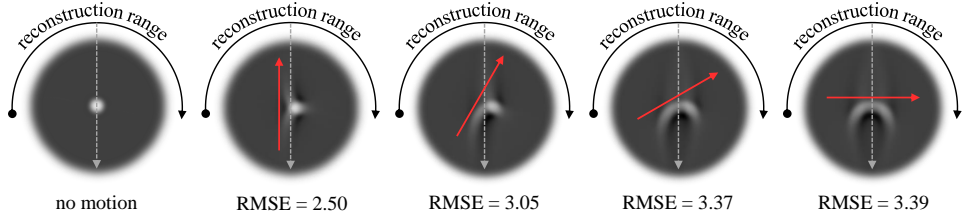


Figure 4.5: Constant linear motion is introduced in the axial plane of a vessel phantom using the CoMoFACT. Depending on the relation of motion direction (red arrow) and mean reconstruction direction (gray dashed arrow), motion artifacts with varying level occur. The artifact level is measured by the root mean squared error (RMSE) of the image intensities with respect to the motion-free phantom plane (left). Orthogonal motion (right) leads to most severe motion artifacts in the reconstructed image.

4

4.3. LEARNED METHOD

The proposed CoMoFACT enables the generation of multiple motion-perturbed CT image volumes $I_{\Omega}^{\vec{c},s}$ with controlled motion level s at specific coronary centerline points $\vec{c} \in C$. On the basis of velocity measurements at the coronary arteries by Vembar et al. [111], the data generation process for the supervised learning task is limited in the following experiments to maximal displacements of 10 millimeters during 20% cardiac circle.

Patch sampling The sampling process of the input data is illustrated in Figure 4.2. One cross-sectional image patch $I_{100}^{\vec{c},s}$ of size $100 \times 100 \times h$ voxels (blue box) is sampled by trilinear interpolation from each output CT image volume $I_{\Omega}^{\vec{c},s}$ of the CoMoFACT with a resolution of $0.4 \times 0.4 \times 0.4$ millimeters per voxel. The first two patch dimensions are aligned perpendicular to the centerline segment while the third dimension is oriented along the normal vector $\vec{n}_{\vec{c}}$. The center slice of each image patch covers the processed centerline point \vec{c} . Single-slice ($h = 1$) and multi-slice ($h > 1$) patches are tested as input data in the following experiments. The grey values of each patch are clipped to the relevant intensity range (window/level setting of 900/200 HU) and additionally normalized (from $[-250, 650]$ to $[-1, 1]$).

Sub-volume reconstruction For the final extraction of one cross-sectional image patch $I_{100}^{\vec{c},s}$ per motion-perturbed image volume $I_{\Omega}^{\vec{c},s}$, merely a limited area of the image grid is of interest. The CoMoFACT can thus be accelerated and be made more memory efficient, by restricting the FOV $\Omega = \Omega(\vec{c})$ for the reconstruction of a sub-volume only. The selected centerline point \vec{c} defines the center of the restricted FOV and the FOV size is determined by the patch sampling size.

Data augmentation Due to the patch similarity of adjacent centerline points, the data for training, validation and testing are case-wise separated with a ratio of 9 : 4 : 4. The database during the training process is extended by online data augmentation. Motion artifacts are variable in shape, orientation and position. In order to build this invariance into the neural network, the following transformations are performed in cross-sectional plane, i.e. limited to the first two dimensions:

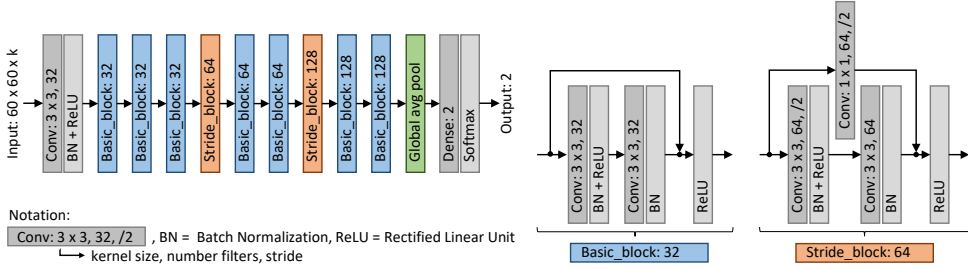


Figure 4.6: The 20-layer ResNet takes coronary cross-sectional image patches of the size $60 \times 60 \times h$ as input. In the last layer the soft-max function is used as activation function of the two (positive and negative) output nodes. The projection shortcuts are realized as 1×1 convolutions. In case $h > 1$ the convolutional kernel sizes are enlarged from $(3 \times 3 \times 1)$ to $(3 \times 3 \times 3)$.

4

- (1) **Cropping** The CoMoFACT may cause small vessel shifts compared to the original coronary centerline position. Image translation by cropping is necessary to avoid a bias originating from the in-plane coronary position. Therefore, sub-patches of the size $85 \times 85 \times h$ are randomly cropped from $I_{100}^{\vec{c},s}$.
- (2) **Rotation** The sub-patches of the size $85 \times 85 \times h$ are randomly rotated by 0 to 360 degrees. The center patch of the size $60 \times 60 \times h$ is finally cropped to ensure full image contents.
- (3) **Mirroring** Horizontal mirroring is performed with a probability of 0.5.

The final image patches $I_{60}^{\vec{c},s}$ of size $60 \times 60 \times h$ voxels are used as input data for supervised learning. During validation and testing only center cropping is performed.

Artifact classification Following Elss et al. [26], the database for the classification task of separating *artifact* and *no artifact* cross-sectional patches is generated by applying the proposed piecewise linear motion model of Section 4.2.2 seven times (with different target motion strength s) per coronary centerline point. The target label y_{class} (0: *no artifact*, 1: *artifact*) of each image patch $I_{60}^{\vec{c},s}$ is defined by the corresponding utilized target motion strength:

$$y_{\text{class}} = \begin{cases} 0, & \text{if } s \in \{0, 1\} \\ 1, & \text{if } s \in \{6, 7, 8, 9, 10\} \end{cases} \quad (4.4)$$

The gap in s is chosen to assure better class separation. Merely a subset of two fifths of the samples from class *artifact* is randomly selected and included in the learning process. This ensures balanced classes during training and testing. A total amount of 14 724 samples is collected as classification database. Multiple hyper-parameter settings and network architectures including the ResNet [42] and VGG inspired networks [102] were tested by extensive cross-validation. Highest validation performances are achieved by a 20-layer ResNet which is employed in all subsequent experiments. Figure 6.4 illustrates the network architecture. The learning process is driven by the cross-entropy loss. The Adam optimizer [57] with an initial learning rate of 0.05, a minibatch size of 32 and a momentum of 0.8 is defined as the learning setup. Finally, the trained neural network NN_{class} takes a cross-sectional patch $I_{60}^{\vec{c},s}$ as input and delivers a predicted artifact probability as output.

Artifact level regression The database for the regression task of predicting the artifact level in cross-sectional patches is generated by applying the proposed constant linear motion model of Section 4.2.3 eleven times per coronary centerline point. The target label $y_{\text{regr}} = s \in \{0, 1, \dots, 10\}$ of each image patch $I_{60}^{\tilde{c},s}$ is set equal to the corresponding utilized target motion strength. By this discrete equidistant labeling procedure, a total amount of 40 491 samples is collected as regression database. Except for the reduction of output neurons in the last layer from two to one and the replacement of the soft-max function by simple linear activation, no adaption of the network architecture is done compared to Figure 6.4. The initial learning rate is changed to $5 \cdot 10^{-4}$ while the remaining hyper-parameters remain unchanged. The neural network NN_{regr} takes a cross-sectional patch $I_{60}^{\tilde{c},s}$ as input and delivers a prediction $\hat{y} \in \mathbb{R}$ as output. The learning process is driven by a piecewise L1 loss function:

4

$$L_{\text{leaky}}(y_{\text{regr}}, \hat{y}) = \begin{cases} \max(0, \hat{y}), & \text{if } y_{\text{regr}} = 0 \\ |y_{\text{regr}} - \hat{y}|, & \text{if } 0 < y_{\text{regr}} < 10 \\ \max(0, 10 - \hat{y}), & \text{if } y_{\text{regr}} = 10 \end{cases} \quad (4.5)$$

This loss function considers adaptive penalization at the boundaries for predictions outside the target interval $[0, 10]$ to avoid too conservative predictions. In comparison to network training with the simple L1 loss, the regressors more often dare output values near zero or ten. Clipping of the network output finally delivers the predicted artifact level $\hat{s} = \min(\max(0, \hat{y}), 10) \in [0, 10]$ which is used for the following evaluation.

Bagging approach An ensemble of five CNNs is learned using the aforementioned network design and hyper-parameter setting by the following bagging approach:

1. Four validation cases and four test cases are randomly sampled.
2. Network training is performed based on the remaining nine clinical cases.
3. After every epoch of the learning process the generalization capability is examined by means of the validation set.
4. The model with the highest performance on the validation set during 60 epochs of training is selected for calculation of the test accuracy (or test error).
5. Steps 1.-4. are performed five times in total.
6. The mean test accuracy (or the mean test error) over the five splits is calculated.
7. Steps 5.-6. are performed for $h \in \{1, 3, 5, 7\}$ (with the same separations in training, validation and testing for comparability).

4.4. EXPERIMENTS AND RESULTS

For all experiments, the Microsoft Cognitive Toolkit (CNTK v2.5, Microsoft Research, Redmond, WA, USA) is used as deep learning framework.

Quantitative error analysis The test results of the classification and the regression networks are summarized in Table 4.1. The network performances increase with the number of input slices. The additional information in multi-slice patches seems to provide a benefit, e.g. in the differentiation between bifurcations and blurring artifacts. But,

Table 4.1: Mean and standard deviation of the classification accuracy and the absolute regression error on the test subsets with synthetic motion artifacts for single-slice ($h = 1$) and multi-slice ($h > 1$) input data.

number slices	classification accuracy	regression error
$h = 1$	$91.64\% \pm 1.63\%$	1.38 ± 0.17
$h = 3$	$92.08\% \pm 2.12\%$	1.16 ± 0.06
$h = 5$	$92.70\% \pm 2.18\%$	1.14 ± 0.07
$h = 7$	$93.26\% \pm 1.82\%$	1.12 ± 0.07

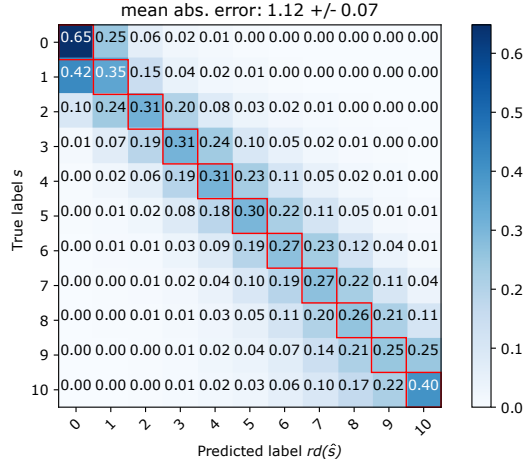


Figure 4.7: Confusion matrix of the regression network for multi-slice inputs ($h = 7$). rd denotes the rounding operator.

higher memory requirements and execution time have to set against it. In case of $h = 7$, the classification result splits into a ratio of 46.90% : 46.36% : 3.64% : 3.10% for the rates TN : TP : FN : FP, where *positive* refers to the class *artifact*. Figure 4.7 shows the confusion matrix of the corresponding regression network. A clear diagonal structure with few scattering of the predicted labels is observable.

The presented quantitative results prove that CNNs are able to identify artifact pattern from synthetically introduced motion. To further evaluate generalization capabilities and the performance of the learned motion artifact measures in clinical practice additional qualitative experiments are performed. In comparison with the handcrafted MAMs entropy and positivity from Rohkohl et al. [93], the abilities for measuring relative and absolute levels of motion artifacts are verified.

4.4.1. RELATIVE ARTIFACT MEASUREMENT

In the following, it is investigated whether the DL-based artifact measures are able to identify the cardiac phase of a clinical data set with least motion artifacts. A quantitative study by Vembar et al. [111] has shown that minimum velocities at the right coronary artery (RCA) can be observed in the mid-diastolic cardiac phase (between 70% and 80% R-R). Therefore, increasing temporal distance to this cardiac phase should go along with increasing artifact levels.

The three helical test cases (see Section 4.1.2) are reconstructed at multiple cardiac phases using the AWCR algorithm. The selected cardiac phases are arranged around 75% R-R interval (mid-diastole) with a temporal distance of 2% R-R interval. Cross-sectional patches of size $60 \times 60 \times 7$ are sampled along the RCA as input for the DL-based and the handcrafted motion artifact measures. Finally, the mean motion artifact measures across the entire vessel are computed for each reconstructed phase image. In this experiment, the bagging ensembles of the five classification and the five regression networks with $h = 7$ are selected. The calculated motion artifact measures are scaled to the value range

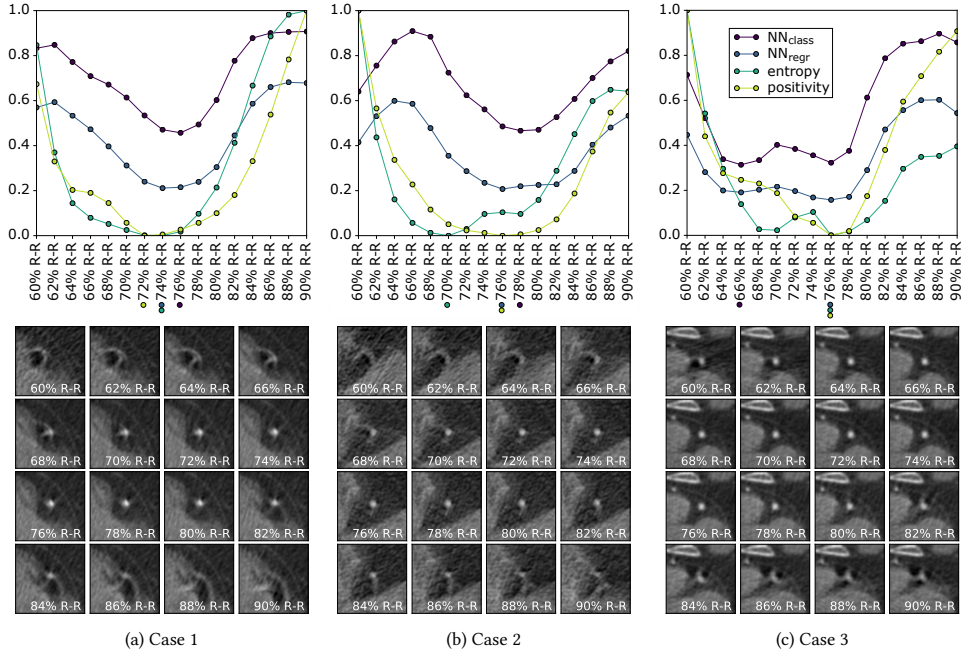


Figure 4.8: The mean motion artifact level of the RCA is analyzed in three cardiac CT cases with helical acquisition mode by the DL-based and the handcrafted motion artifact measures. For each reconstructed CT image volume within the phase interval [60%, 90%], cross-sectional image patches are sampled along the centerline segmented at the reference phase 75% R-R. The outputs of the motion artifact measures are averaged across the vessel, scaled to the interval [0, 1] and visualized in the upper row. The predictions of the best cardiac phases are indicated by colored dots, respectively. One axial slice per selected cardiac phase is given below for visual inspection.

[0, 1], to provide comparability. Therefore, the predicted artifact levels $\hat{s} \in [0, 10]$ of the regression network NN_{regr} are down-scaled by a factor of ten. The handcrafted MAMs are normalized to the interval [0, 1] by the minimal and maximal output over all motion states. As the classification network NN_{class} already delivers predicted artifact probabilities, these remain unchanged.

Figure 4.8 shows the results of the proposed multi-phase experiment. In the first case (Figure 4.8a), all motion artifact measures (handcrafted and DL-based) provide similar results. The predicted best cardiac phases around 72–76% R-R comply with the visual inspection. In the second case (Figure 4.8b), only the positivity and the DL-based measures deliver predictions of the best cardiac phase which concur with the visual impression. However, a weakness of the neural networks can be discerned. The modulation of the radiation dose leads to lower signal-to-noise ratios (SNRs) at the marginal cardiac phases around 60% R-R. The trained neural networks seem to be more vulnerable to such SNR fluctuations than the handcrafted measures. In the third case (Figure 4.8c), the regression network is in agreement with both handcrafted MAMs with respect to the predicted best cardiac phase at 76% R-R. The classification network selects an earlier stage around 66% R-R in which also hardly artifacts occur. A temporally extended rest phase is observable which is discovered by the trained neural networks.

The results of this multi-phase experiment are promising given the fact that the DL-based measures are solely trained on step-and-shoot data which is perturbed by constant or piecewise linear motion. Generalization capabilities of the CNNs and transferability to helical data sets with real motion artifacts are demonstrated by this experiment.

4.4.2. ABSOLUTE ARTIFACT MEASUREMENT

In the following, it is investigated whether the DL-based artifact measures are able to detect a region of motion, given the approximate location of the coronary artery.

EVALUATION ON SYNTHETIC MOTION ARTIFACTS

Local motion is introduced to test reference cases at arbitrary points in the coronary tree by means of the piecewise linear motion model proposed in Section 4.2.2. A target motion strength of $s = 8$ is selected in the following experiments. For each centerline point of the corresponding vessel, a cross-sectional patch of size $60 \times 60 \times 7$ is sampled from the locally motion-perturbed CT image volume $I_{\Omega}^{E,s}$ and the corresponding motion artifact measures are calculated. The classification and regression networks are selected so that the currently processed test case has neither been used for training nor for validation.

Figure 4.9 shows the results of this local motion experiment for two test cases. The left anterior descending artery (LAD) and the right coronary artery (RCA) are processed, respectively. In each subplot, the original no motion state is visualized as reference above the local motion state. The x-axis corresponds to the spatial coordinate along the centerline. The value range $[0, 1]$ of the y-axis is determined by the weighting mask values $m_{\tilde{c}} \in [0, 1]$ which correspond to the level of introduced motion. In each subplot, calculated motion artifact measures are scaled accordingly, to provide comparability again.

As expected, the handcrafted MAMs are not suitable for section-wise motion artifact quantification, due to limited robustness regarding variations in background intensities. The DL-based measures, by contrast, accurately detect regions of motion with few exceptions. Both networks are robust towards vessel shifts at stack transitions (see Figure 4.9a). Bifurcations (see Figures 4.9b), calcifications (see Figures 4.9c) and varying contrast levels between scanning sequences (see Figure 4.9d) do not affect the DL-based measures either. The aforementioned image areas are highlighted in the multi-planar reformations (MPRs) by white arrows. The lowest artifact level in the motion area is predicted by the regression network in Figure 4.9d, which also confirms with visual inspection of the four motion-perturbed MPRs. This experiment already demonstrates generalization capability of the regression network which is trained on perturbed data with constant linear motion and tested on data with more complex piecewise linear motion trajectories.

EVALUATION ON REAL MOTION ARTIFACTS

The ability for absolute motion artifact measurement is additionally tested on eight clinical cases with real motion artifacts (see Section 4.1.2). Ensemble averaging ($h = 7$) is performed for the evaluation. Figure 4.10 shows the resulting artifact measurements and corresponding cross-sectional patches. The vessels are sorted by the maximum artifact level predicted by the regression network. Artifact areas identified by the classification networks (with running average for outlier removal) are highlighted in red.

Four separate observer studies were performed to rate the 120 cross-sectional image patches visualized in Figure 4.10 with respect to diagnostic reliability in a five point Likert

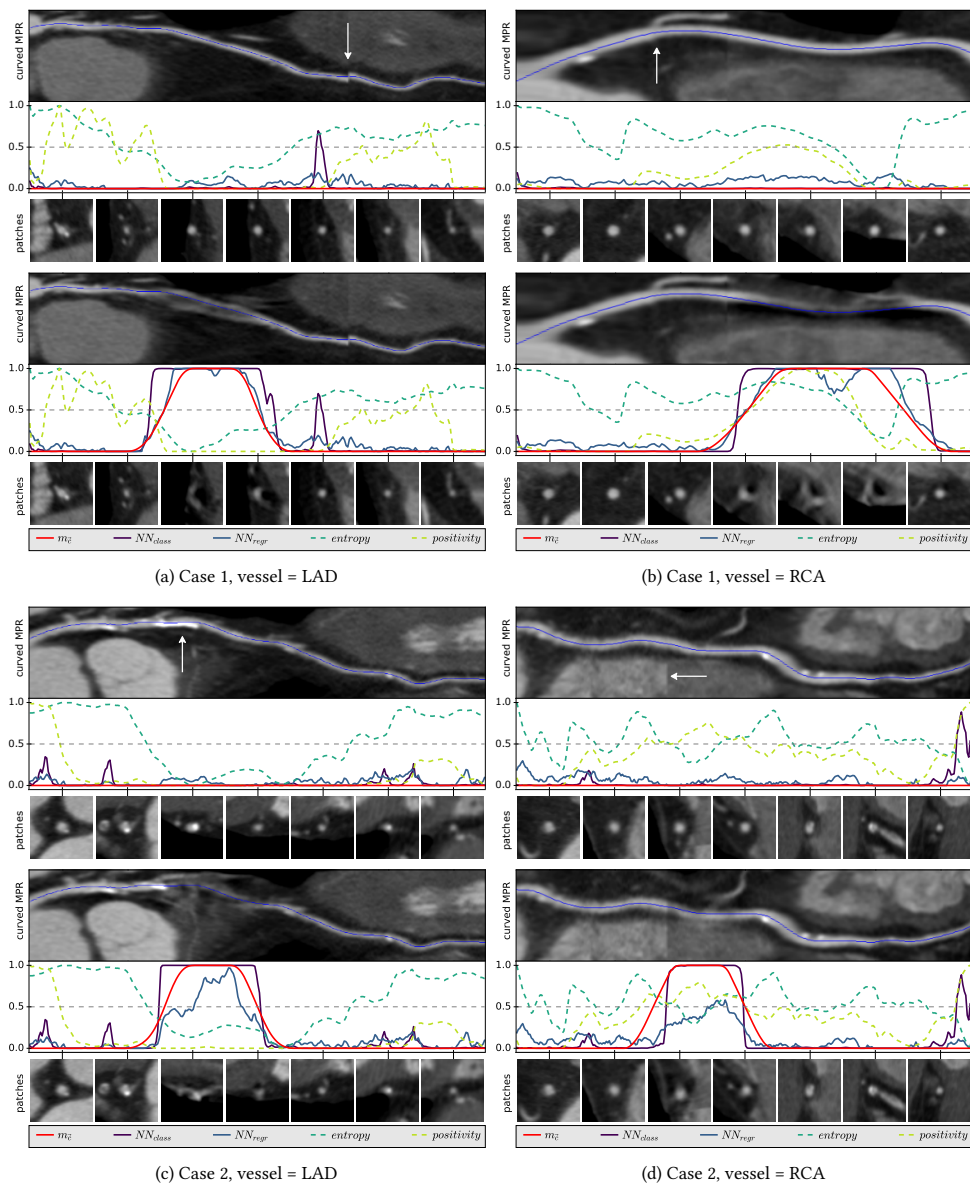


Figure 4.9: Section-wise outputs of the DL-based and the handcrafted motion artifact measures are calculated for four vessels without and with synthetic motion artifacts. In each subplot, the no motion state is given as reference above the local motion state. Stack transition artifacts (a)&(d), bifurcations (b) and severe calcifications (c) are indicated by white arrows in the MPRs. The weighting mask value m_z marked in red, corresponds to the true relative displacement width of each centerline point and determines the area of introduced motion. Corresponding cross-sectional image patches are given below for visual inspection. In contrast to the handcrafted MAMs, high predictions made by the neural networks are mostly correctly located at the regions of motion influence.

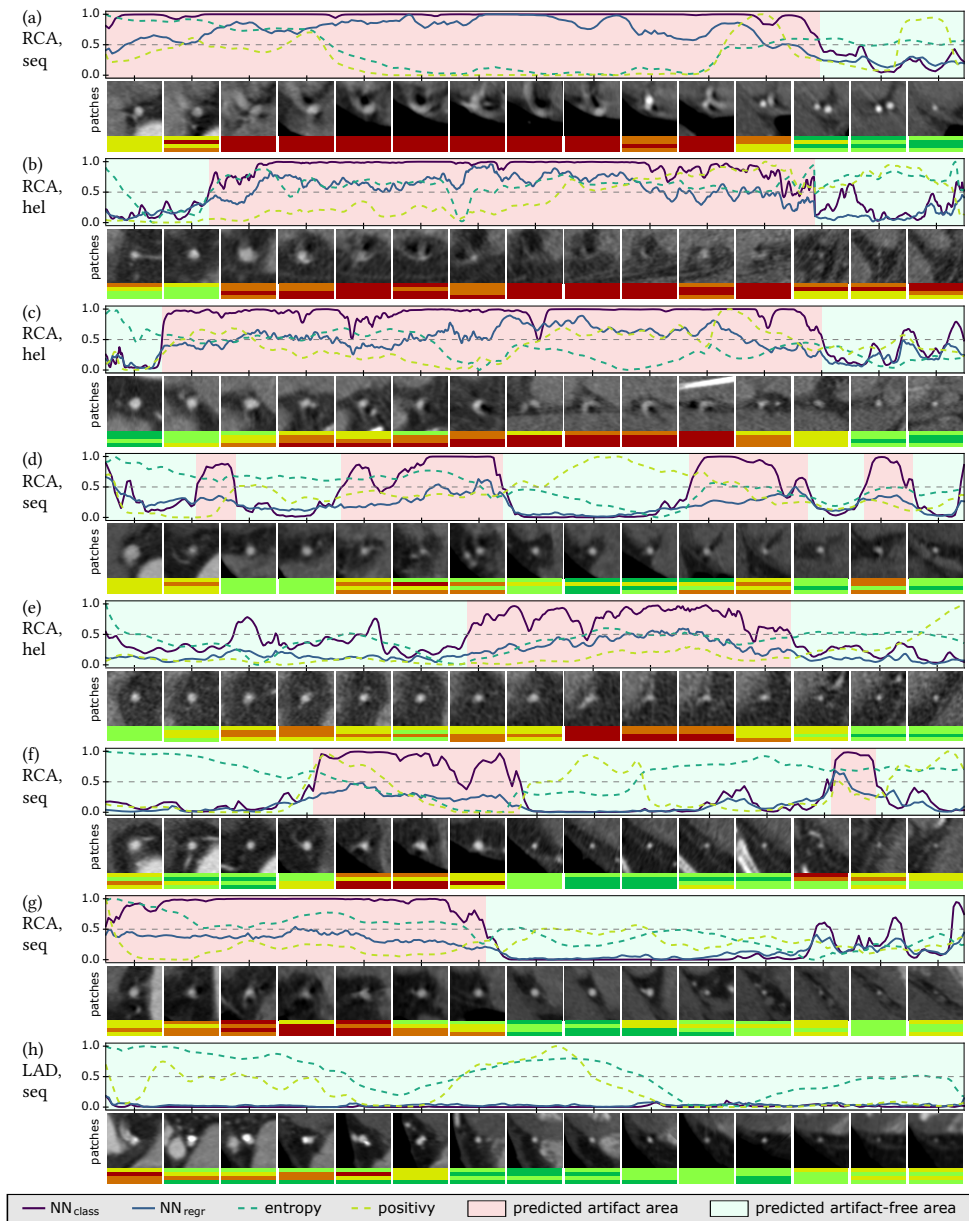


Figure 4.10: Section-wise outputs of the DL-based and the handcrafted motion artifact measures are calculated for eight vessels with real motion artifacts. The predicted artifact level \hat{s} of NN_{regr} , the entropy and the positivity are down-scaled to $[0, 1]$. Corresponding cross-sectional image patches with four human observer ratings each (from green : excellent to red : non-diagnostic) are given below for visual inspection. The vessels are sorted by the maximum artifact level predicted by the regression network. Vessel type (RCA or LAD) and acquisition mode (seq: step-and-shoot or hel: helical) are specified.

scale (— : excellent, — : good, — : mixed, — : strong artifact, — : non-diagnostic). The eight vessels were presented in random order without indication of the motion artifact measures to the readers. It has to be noted that the readers were no radiologists, but research scientists with high level of expertise in reading cardiac CT images. The resulting annotations are visualized as color bars in the Figure 4.10.

In contrast to the handcrafted MAMs, the DL-based measures deliver sensible results. The classification networks enable the rough detection of artifact areas and the regression network additionally allows one to assess the artifact severity. Noise (see Figure 4.10e) and vessel segments with narrow lumen (see Figure 4.10g) can be identified as potential sources of uncertainty. The human observer ratings show disagreements in these areas as well. Furthermore, confusions between bifurcations and blurring artifacts are observable in the reader scores (see first patch of Figure 4.10b and Figure 4.10h). Image patches with consistent artifact labels (red or orange) are correctly identified as motion-perturbed by the neural networks. This also holds for image patches with consistent artifact-free labels (light green and dark green). Hence, transferability from synthetic to real motion artifacts is demonstrated. The proposed experiments reveal both, potential and current limitations of DL-based artifact measurement in clinical practice.

4.5. DISCUSSION

The feasibility of accurate motion artifact quantification in coronary arteries using deep learning is demonstrated. The results of the quantitative error analysis only provide indications, as the target labels do not always constitute exact ground truth. Figure 4.4a shows multiple image patches with false positive target label, for instance originating from motion along the vessel orientation. The thresholds defining the margin in s are crucial parameters which might affect the classification performance. The reduced complexity of the constant linear motion model presented in Section 4.2.3 enables more consistent target labels, but these might also exhibit slight inaccuracies. Some label noise originates from approximations and simplifications in the CoMoFACT. The centerline and its normal vectors are only estimates and the mean reconstruction direction is limited to the axial plane, i.e. the z -axis is not considered. Furthermore, motion in the axial plane has different effects on the reconstructed image data than motion in z -direction. Also other factors like background intensities and image noise influence the visual artifact level. The majority of the generated image and corresponding label data constitute good approximations of the reality. For network training purposes, the label quality is sufficient, but exact test performance can not be determined. As a next step, quantitative comparison studies to hand-labeled data from radiologists should be performed.

For the quality assessment of CCTA images using the proposed motion artifact measures, the approximate locations of the coronary arteries have to be known. Since motion artifacts frequently inhibit fully automatic centerline segmentation, alternative approaches are required. Many tools for coronary centerline extraction are semi-automatic, i.e. they allow the user to guide the segmentation process. Furthermore, the creation of a coronary artery atlas which involves the probability density for the position of each vessel segment with respect to the heart segmentation and the deployment of deep-learning based centerline extraction are options which should be investigated. The proposed CoMoFACT enables an evaluation of the robustness of centerline extraction methods with

regard to motion artifacts. In general, motion introduction by the CoMoFACT might be useful as a data augmentation strategy in several other CT data-driven learning tasks.

So far, the proposed measures are based on 17 clinical data sets only. CT images for non-invasive coronary angiography are acquired with a wide variety of scanner types, imaging protocols and reconstruction algorithms. In order to increase robustness of the CNNs, collection of more data and network fine-tuning should be performed. Especially in order to increase the networks insensitivity to noise, an extension of the training database by including clinical cases with lower SNR or synthetic noise introduction would be required. Robustness might additionally be improved by measurement smoothing of adjacent centerline points in order to reduce appearing scatter and to avoid outliers.

5

REMOVAL OF CORONARY MOTION ARTIFACTS

THE CoMoFACT introduced in [69] and Section 4.2 enable the generation of paired data which consists of motion-perturbed image patches, corresponding artifact-free versions and underlying motion trajectories. Two possible options to exploit the given data for supervised learning of a motion compensation method are illustrated in Figure 5.1 and discussed in the following.

Option 1: Image-to-image translation

Initial experiments for direct mapping of motion-perturbed image data to artifact-free image data using deep residual networks were performed. However, the information content of the corrupted input patches mostly seems to be insufficient to restore the original reference data. In the following, the second option is investigated.

Option 2: DL-based motion estimation + motion correction

The steps of motion estimation and correction are separated. First, previously learned regression models take the motion-perturbed image data as input to estimate the underlying motion parameters. Second, based on the predicted motion parameters, a dense MVF is calculated and integrated into the MC-FBP. The MC-FBP is applied in its usual way and motion compensation is performed. By this procedure, the concomitant projection data is also taken into account.

Phantom studies showed that motion artifacts manifest in characteristic pattern depending on the relation between motion direction and angular reconstruction range (see Figure 4.5). However, there is no guarantee that the information of underlying motion parameters such as the motion direction, is incorporated in the motion-perturbed image data. Possibly artifact types are ambiguous. This chapter addresses the question of how well motion estimation can be performed from a single reconstructed CT image patch based on the coronary artifact appearance.

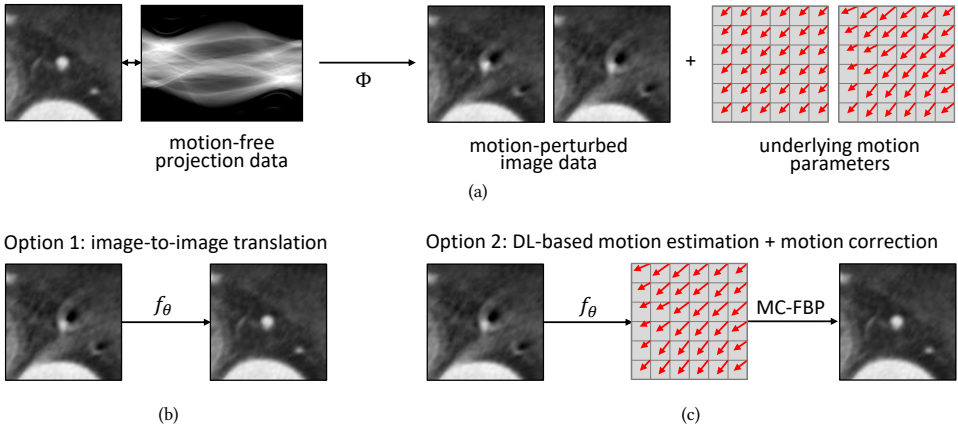


Figure 5.1: (a) The forward model Φ , also called CoMoFACT, takes high quality clinical cases \vec{x}_{clean} as input and delivers more or less motion-perturbed image data \vec{x}_{syn} and corresponding underlying motion parameters \vec{y}_{syn} as output. The synthesized paired data enables learning of direct image-to-image translation (b) or DL-based motion estimation combined with common MCR (c).

The previous chapter reveals that CNNs are capable of identifying coronary motion artifact patterns. In the conference article [25], the feasibility of axial motion estimation using CNNs trained on synthetic data has initially been demonstrated. However, only centerline segments with a maximal inclination of 45 degree to the z-axis of the fixed patient system are considered here. A local motion compensation experiment is performed on clinical data with real artifacts in order to investigate generalization capabilities.

Extending this work, the journal article [68] introduces the *Coronary Motion estimation by Patch Analysis in CT data* (CoMPACT) method which is applicable on coronary artery segments with arbitrary orientations. The CoMPACT consists of an ensemble of five CNNs each taking cross-sectional image patches as input and delivering 2D axial motion vectors as output. The patch sampling procedure and the network architecture are adapted, accordingly. Furthermore, a novel iterative motion compensation pipeline, called the CoMPACT MC method, is developed which combines the DL-based motion estimation, MVF extrapolation and MC-FBP. This chapter is mainly inspired by [68] and structured as follows:

- 5.1 CCTA cases are collected and grouped into clinical reference and test data. Based on the reference data, phantom vessel trees are generated to investigate the feasibility of DL-based motion estimation under well-controlled conditions, i.e. without variation in background intensities, contrast agent density or noise level.
- 5.2 The CoMoFACT generates the required learning data by introducing simulated and hence controlled motion to the artifact-free clinical and phantom reference data.
- 5.3 Ensembles of CNNs are trained for the task of motion vector regression based on the artifact appearance. The CoMPACT networks trained on clinical data only are integrated into an iterative motion compensation pipeline which uses alternating MVF estimation and MC-FBP.
- 5.4 The proposed MC pipeline is tested on clinical test cases with real artifacts and compared to the registration-based MC approach from Grass et al. [38] in order to analyze the transferability of CoMPACT MC to clinical practice.
- 5.5 Advantages and disadvantages of the proposed CoMPACT MC compared to registration-based MC are discussed.

5.1. MATERIAL

The CoMoFACT uses artifact-free CCTA cases with step-and-shoot acquisition protocol as reference point for the motion introduction process. In the following, phantom as well as patient data studies are performed. Section 5.1.1 details the expansion of the reference database compared to the previous chapter. The design of the computer-simulated vessels is described in Section 5.1.2. Additional clinical cases with real motion-perturbation are collected in order to test the transferability to non-synthetic artifacts (see Section 5.1.3).

5.1.1. CLINICAL REFERENCE DATA WITHOUT ARTIFACTS

The database described in Section 4.1.1 is expanded by two additional clinical cases, resulting in a total amount of 19 reference data sets which exhibit no coronary motion artifacts. The novel contrast-enhanced CT data is acquired using the same scanner type (Brilliance iCT, Philips Healthcare, Cleveland, OH, USA) and scan protocol (prospectively

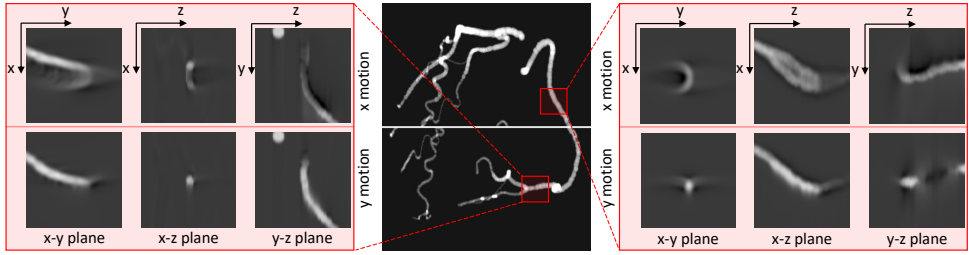


Figure 5.2: Constant linear motion is introduced to phantom vessel trees using the CoMoFACT. Depending on the relation of motion direction, reconstruction direction and coronary orientation, motion artifacts of different appearance occur. In case of motion in direction of the coronary artery, artifacts are hardly visible due to blurring within the vessel (left box, bottom row). The forward projected coronary artery mask depicted in the center is split into two scanning shots. Patch-based motion prediction models have to be robust regarding related stack transition artifacts.

5

ECG-triggered). AWCR and the segmentation of the coronary artery trees are executed as before. Altogether the patients mean heart rates HR_{mean} range now from 45.2 bpm to 68.9 bpm.

5.1.2. PHANTOM REFERENCE DATA WITHOUT ARTIFACTS

From each clinical reference case, one binary phantom mask is extracted which contains the segmented lumen contour of the entire vessel tree (see Figure 5.2). Ray-driven forward projection [7] and subsequent high-pass filtering delivers the projection data required for the application of the CoMoFACT. The projection geometry, the ECG-data and the coronary centerline points are adopted from the corresponding clinical data set. The phantom study allows identifying the limits of DL-based motion estimation in a well-controlled scenario without variation in background intensities, contrast agent density or noise level. It focuses on variations in the vessel structure comprising different orientations, curvatures, radii and bifurcations.

5.1.3. CLINICAL TEST DATA WITH REAL ARTIFACTS

Twelve additional clinical cases which belong to different patients and exhibit real motion artifacts at the coronary arteries are collected for testing purposes. Acquisition is performed by a Brilliance iCT scanner using the same scan protocol as in the reference data. The mean heart rates of the patients HR_{mean} ranged from 57.9 bpm to 83.0 bpm.

5.2. FORWARD MODEL

In this part of the work, CNNs are trained for motion vector estimation in coronary image patches. The required data for supervised learning is generated using an adapted version of the previously introduced CoMoFACT. The CoMoFACT simulates inconsistent projection data caused by artificial motion near the coronary arteries.

Synthetic MVF: In principle, arbitrary motion trajectories can be simulated using this approach by adjusting the synthetic MVF. For simplicity, the model is restricted to constant

linear motion again, but with the additional restriction to arbitrary motion directions in the axial plane. Therefore, minor adaptations of the synthetic MVF from Section 4.2.1 are performed. In this CoMoFACT variant, the displacement $\vec{d}_{\vec{c}}: [0\%, 100\%] \times \Omega \rightarrow \mathbb{R}^3$ of each voxel \vec{v} at time point $t_{cc} \in [0\%, 100\%]$ in millimeters is calculated by:

$$\vec{d}_{\vec{c}}(t_{cc}, \vec{v}) = s \cdot m_{\vec{c}}(\vec{v}) \cdot \vec{\delta}_{\vec{c}}(t_{cc}, \alpha) \quad (5.1)$$

As described before, t_{cc} is measured in percent cardiac cycle, $s \in \mathbb{R}_+$ denotes the target motion strength and $m_{\vec{c}}: \Omega \rightarrow [0, 1]$ indicates a weighting mask which limits the motion to the area of the currently processed centerline point $\vec{c} \in \Omega$. The motion direction is determined by $\vec{\delta}_{\vec{c}}: [0\%, 100\%] \times (-180^\circ, 180^\circ]$ which is adapted to:

$$\vec{\delta}_{\vec{c}}(t_{cc}, \alpha) = \frac{60\text{bpm}}{\text{HR}_{\text{mean}}} \cdot \frac{\vec{\rho}_{\vec{c}}(\alpha)}{\|\vec{\rho}_{\vec{c}}(\alpha)\|_2} \cdot \begin{cases} -0.5 & \text{if } t_{cc} < r - 10\% \\ \frac{(t_{cc}-r)}{20\%} & \text{if } r - 10\% \leq t_{cc} \leq r + 10\% \\ +0.5 & \text{if } t_{cc} > r + 10\% \end{cases} \quad (5.2)$$

The motion direction determined by $\vec{\rho}_{\vec{c}}(\alpha)$ is limited to the axial plane, i.e. the z-component is set to zero. The x- and y-components of $\vec{\rho}_{\vec{c}}(\alpha)$ are defined in such a way that α corresponds to the angle between mean reconstruction direction of the currently processed centerline point \vec{c} and the motion direction (see Figure 5.3). It has to be noted that the system rotation direction is consistent for all cases. This is important, since a reversed rotational direction would lead to a flipping of the artifact shapes.

In the previous chapter, the feasibility of motion artifact recognition and quantification by utilizing the parameter s for target value assignment is investigated. This forward model has one additional (angular) degree of freedom α , i.e. each MVF is now defined by the parameter tuple (s, α) . The target motion strength s scales the length of each displacement vector in the MVF and therefore determines the motion magnitude. On the basis of the velocity measurements at the coronary arteries by Vembar et al. [111], the target motion strength s is limited to the interval $[0, 10]$ in all experiments. The newly introduced angle parameter $\alpha \in (-180^\circ, 180^\circ]$ determines the in-plane motion direction.

5.3. LEARNED METHOD

The adapted CoMoFACT enables the generation of motion-perturbed image data with controlled motion magnitude s and motion direction α . Ensembles of CNNs are trained on the synthetic data to estimate underlying motion vectors based on the artifact appearance (see Section 5.3.1). Section 5.3.2 details the integration of the trained CNNs into iterative motion compensation pipeline for alternating MVF estimation and MC-FBP.

5.3.1. SUPERVISED LEARNING

The extended CoMoFACT enables the generation of multiple motion-perturbed CT image volumes $I_{\Omega}^{\vec{c}, s, \alpha}$ with controlled motion level s and motion direction α at a specific coronary centerline point \vec{c} . Both parameters s and α are randomly sampled from uniform distributions in the following experiments. Corresponding Cartesian coordinates $x_d = s \cos(\alpha)$ and $y_d = s \sin(\alpha)$ define the ground-truth labels for the supervised learning task.

Patch sampling: For each centerline point \vec{c} and parameter setting (s, α) , one 2.5D image patch $I_{2.5D}^{\vec{c}, s, \alpha}$ is sampled as input data for supervised learning. A 2.5D patch consists of three orthogonal image slices of size 60×60 pixels with an image resolution of 0.4×0.4 mm² per pixel, the so-called x-y plane, x-z plane and y-z plane. The centerline point \vec{c} defines the patch center and the patch orientation is contingent on the angular reconstruction range. The parameter γ_{mean} is defined as the gantry rotation angle at the reference heart phase r . As illustrated in Figure 5.3, γ_{mean} determines the relation between the static patient coordinate system $(x_{\text{patient}}, y_{\text{patient}}, z_{\text{patient}})$ and the rotated patch coordinate system (x, y, z) . More precisely, (x, y, z) represents a rotation of the coordinate system $(x_{\text{patient}}, y_{\text{patient}}, z_{\text{patient}})$ by γ_{mean} about the z_{patient} -axis. By this procedure, the information about the angular reconstruction range is embedded in the patch orientation.

In both studies, phantom and patient data, the CoMoFACT with subsequent patch sampling is applied 1000 times per reference case, thus, delivering a total amount of 19 000 samples as database for supervised learning. The artifact appearance with respect to the relation between motion direction and vessel orientation is illustrated in Figure 5.2. Figure 5.3 shows distinct blurring artifacts depending on angular reconstruction range and motion direction.

5

Data augmentation: Due to the patch similarity of adjacent centerline points, the data is case-wise separated for training, validation and testing with a ratio of 11 : 4 : 4 to avoid a bias. By this procedure, robustness of the trained networks is evaluated with regard to unknown variations in the vessel geometry and background intensity. The data basis during network training is extended by online data augmentation. Inherent symmetry properties of motion artifacts are exploited for random mirroring of each axis in the input patches. The target labels x_d, y_d are adapted accordingly. This procedure increases the amount of vessel geometry variations in the phantom study and background variations in the patient data study. Additionally, translation is performed as label-preserving augmentation strategy. The patch center is randomly shifted in a range of $[-10, 10]$ voxels in x, y and z direction in order to build translation invariance into the networks. This allows deviation between extracted and actual centerline positions. During testing and validation, no mirroring and no translation is performed.

Learning setup: Based on the synthetically motion perturbed data, CNNs are trained for patch-based motion estimation. The networks take one image patch $I_{2.5D}^{\vec{c}, s, \alpha}$ as input and deliver the predicted predominant motion vector (\hat{x}_d, \hat{y}_d) as output. Multiple patch sampling strategies (2D, 2.5D and 3D), network architectures and hyper-parameter settings were tested by extensive cross-validation. Best generalization capabilities are achieved by the CNN visualized in Figure 5.4 which is used in all subsequent experiments. The x-y plane, the x-z plane and the y-z plane are processed separately in one eight-layer ResNet [42] each. No weight sharing between the paths is performed since each plane exhibits individual, characteristic motion artifact pattern. The outputs of the global average pooling are concatenated and information are merged in a final dense layer with linear activation function and two output neurons to predict \hat{x}_d and \hat{y}_d . The learning process is driven by the squared error $l = (x_d - \hat{x}_d)^2 + (y_d - \hat{y}_d)^2$.

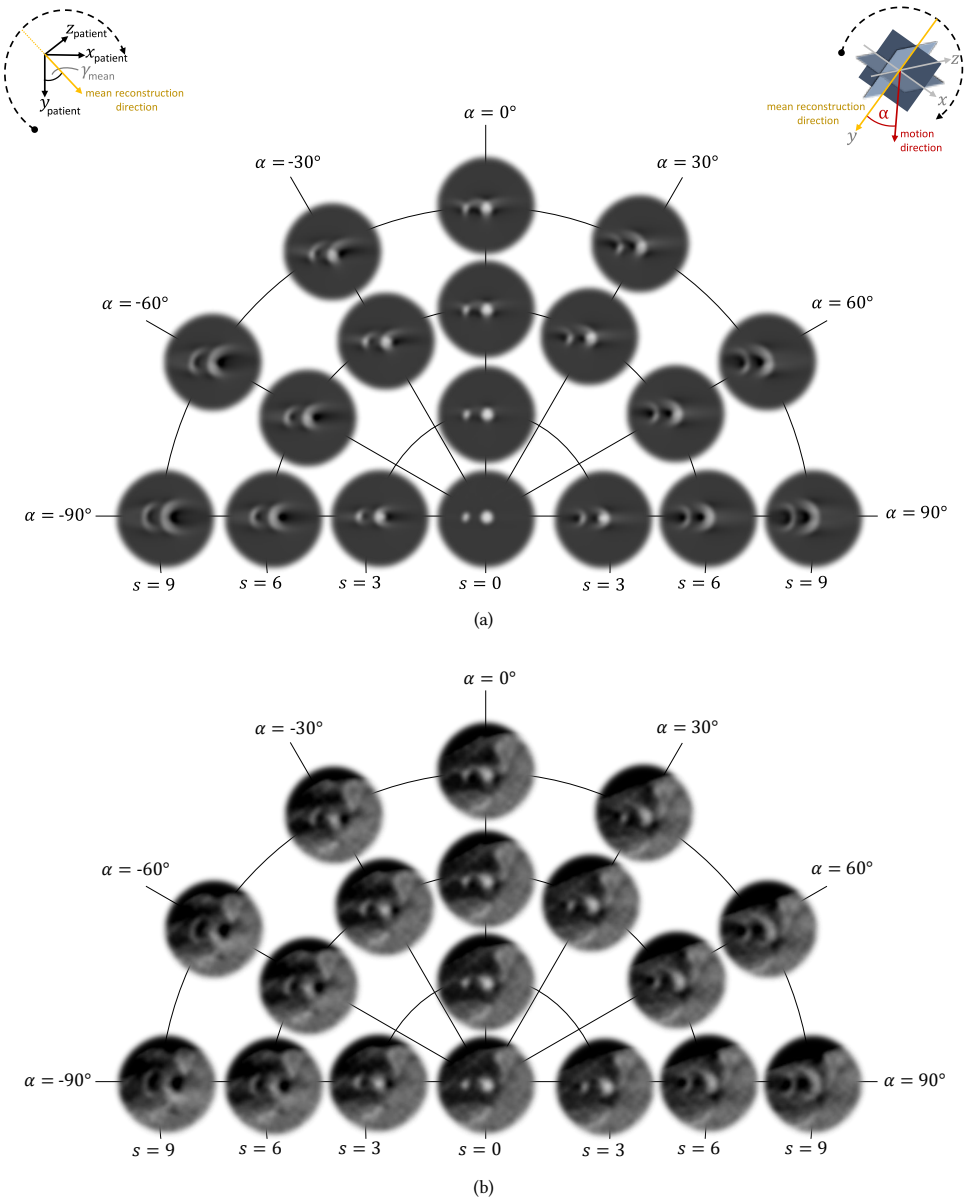


Figure 5.3: The x-y plane of an example centerline point is illustrated in a phantom (a) and a clinical scenario (b) for varying parameter settings (s, α). For better visualization, the patches of size 60×60 pixels are illustrated as circles. Depending on the motion angle α , distinct blurring artifacts occur. Orthogonal motion ($\alpha = \pm 90^\circ$) leads to the most severe banana-shaped artifacts while parallel motion ($\alpha = 0^\circ$ or $\alpha = 180^\circ$) causes bird-shaped blurring. In the clinical data, visibility of blurring artifacts and intensity undershoots are strongly influenced by surrounding background intensities, i.e. artifact types are visually more difficult to distinguish. As illustrated in the top left and top right corner of Subfigure (a), γ_{mean} defines the relation between the static patient coordinate system ($x_{\text{patient}}, y_{\text{patient}}, z_{\text{patient}}$) and the rotated patch coordinate system (x, y, z).

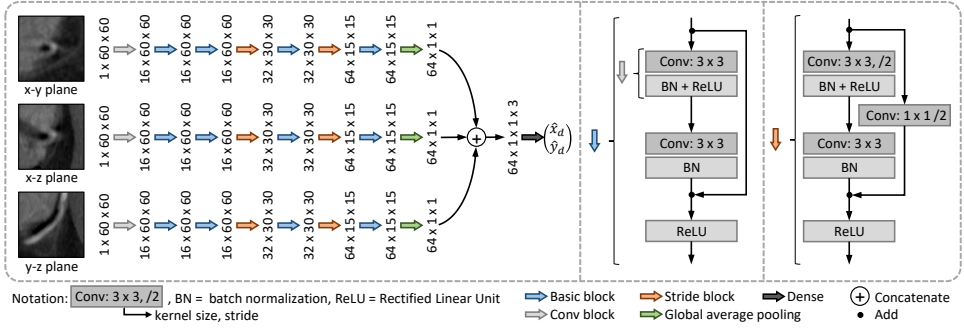


Figure 5.4: The orthogonal planes of each 2.5D input patch are processed separately in an individual path. Information about in-plane artifact pattern are finally fused in a dense layer with two output neurons to predict \hat{x}_d and \hat{y}_d . The network comprises 523 346 learnable weights.

5

The stochastic gradient descent solver Adam [57] with an initial learning rate of 0.01, a mini-batch size of 32 and a momentum of 0.8 is used for network optimization. Training is performed over 45 epochs while the learning rate halves after every 15th epoch. L2 regularization with a weight of 0.001 is used. Network training from scratch is performed on the phantom and the clinical data, separately. Furthermore, weight initialization based on previous phantom studies and subsequent fine-tuning on clinical data is investigated. This is motivated by the recent success of transfer learning approaches [67, 123] and is comparable with learning to recognize digits before learning to read house numbers.

Bagging approach: Three ensembles of five CNNs each are learned using the aforementioned network design and hyper-parameter setting by the following bagging approach:

1. Four validation cases and four test cases are randomly sampled.
2. Network training is performed based on the remaining eleven clinical cases.
3. After every epoch of the learning process the generalization capability is examined by means of the validation set.
4. The model with the lowest mean euclidean error on the validation set during 45 epochs of training is selected for calculation of the test metrics.
5. Steps 1.-4. are performed five times in total.
6. Mean and standard deviation of each test metric over the five splits are calculated.
7. Steps 5.-6. are performed three times; once for phantom data, once for the clinical data using training from-scratch and once for the clinical data with network initialization based on the learned weights from the phantom study. For comparability, the same case-wise data separation is used respectively.

5.3.2. MOTION COMPENSATION PIPELINE

CNNs trained on the clinical database become the main component of the following motion compensation pipeline. A cardiac CT image volume with corresponding set of approximate centerline positions $C \subset \mathbb{R}^3$ and the raw projection data are required as input. By patch sampling and subsequent application of the trained CNNs, a set of motion vectors $\hat{\vec{d}}_{\vec{c}} \in \mathbb{R}^3$ for $\vec{c} \in C$ is predicted along the centerline.

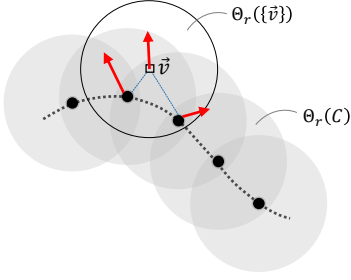


Figure 5.5: For each voxel $\tilde{v} \in \Theta_\tau(C)$ in the centerline bounding volume highlighted in gray, a motion vector is calculated by scattered data extrapolation. Distance weighting of adjacent centerline points $\tilde{c} \in \Theta_\tau(\{\tilde{v}\})$ is performed according to the equations (5.3) and (5.4).

Algorithm 5.1: Iterative MC in CCTA images using patch-based motion estimation.

Input: image volume I_0^Ω , raw projection data p , approximate centerline C , number iterations k_{\max}

Output: improved image volume $I_{k_{\max}}^\Omega$

set $\hat{d}_{\tilde{c}} = \vec{0}$ for $\tilde{c} \in C$

for $k = 1, 2, \dots, k_{\max}$ **do**

for $\tilde{c} \in C$ **do**

$I_{2.5D}^{\tilde{c},s,\alpha} = \text{sample}(I_{k-1}^\Omega, \tilde{c})$

$\hat{d}_{\tilde{c}} += \text{ensemble}(I_{2.5D}^{\tilde{c},s,\alpha})$

end for

 calculate $\hat{d}(t_{cc}, \tilde{v})$ according to (5.3)

$I_k^\Omega = \text{MC-FBP}(p, \hat{d}(t_{cc}, \tilde{v}))$

end for

The motion vector $(\hat{x}_d, \hat{y}_d, 0)$ estimated by the CNNs is back rotated by $-\gamma_{\text{mean}}$ about the z-axis to yield $\hat{d}_{\tilde{c}}$. Thus, $\hat{d}_{\tilde{c}}$ specifies the predicted displacement in the static patient coordinate system. For MC-FBP, these sparse motion information are transformed into a dense MVF $\hat{d}(t_{cc}, \tilde{v}): [0\%, 100\%) \times \Omega \rightarrow \mathbb{R}^3$ by distance-weighted extrapolation (see Figure 5.5).

MVF Extrapolation The estimated voxel displacements are calculated by:

$$\hat{d}(t_{cc}, \tilde{v}) = m_C(\tilde{v}) \frac{60\text{bpm}}{\text{HR}_{\text{mean}}} \cdot \begin{cases} 0 & \text{if } \tilde{v} \notin \Theta_\tau(C) \\ \sum_{\tilde{c} \in C: \tilde{c} \in \Theta_\tau(\{\tilde{v}\})} w(\tilde{v}, \tilde{c}) \hat{d}_{\tilde{c}} & \text{if } \tilde{v} \in \Theta_\tau(C) \end{cases} \quad (5.3)$$

$$w(\tilde{v}, \tilde{c}) = \frac{q(\tilde{c} - \tilde{v})}{\sum_{\tilde{c}_i \in C: \tilde{c}_i \in \Theta_\tau(\{\tilde{v}\})} q(\tilde{c}_i - \tilde{v})} \quad (5.4)$$

The bounding volume of a set $A \subset \mathbb{R}^3$ is defined as $\Theta_\tau(A) = \{\tilde{v} \mid \exists \vec{v} \in A: \|\tilde{v} - \vec{v}\| < \tau\}$. The distance weighting is performed using a 3D Gaussian kernel $q(\tilde{c} - \tilde{v}) = \exp(-\|\tilde{c} - \tilde{v}\|^2 / (2\sigma^2)) / \sqrt{2\pi\sigma^2}$ with $\sigma = 8$ (in millimeters). Besides MV extrapolation, the distance weighting also leads to MVF smoothing. The weighting mask $m_C: \mathbb{R}^3 \mapsto [0, 1]$ is generated by dilation of each centerline point in C with a kernel radius of 15 mm and subsequent uniform filtering with a kernel radius of 6.2 mm according to [69]. In order to yield a smooth transition to zero, the bounding volume radius τ for each Θ_τ is set to 21.2 mm in the following experiments. The MVF extrapolation in equation (5.3) is performed shot-wise, i.e. for each circular acquisition, as motion across different scanning shots is usually not smooth.

Table 5.1: Error metrics on the test cases with synthetic artifacts during the phantom study and the patient data study with and without fine-tuning (FT). The baseline corresponds to $\hat{d}_{\vec{c}} = \vec{0}$ which represents the mean ground truth MV determined in the CoMoFACT.

	$\epsilon_{x,y}$	ϵ_s	ϵ_x	ϵ_y	ϵ_α
Baseline	5.00	5.00	3.24	3.24	NaN*
Phantom	1.10 ± 0.12	0.77 ± 0.09	0.72 ± 0.05	0.66 ± 0.11	$13.37^\circ \pm 1.21^\circ$
Clinical (no FT)	2.92 ± 0.13	1.89 ± 0.07	1.75 ± 0.10	1.96 ± 0.07	$35.66^\circ \pm 1.57^\circ$
Clinical (with FT)	2.87 ± 0.16	1.86 ± 0.11	1.71 ± 0.08	1.93 ± 0.13	$34.85^\circ \pm 2.09^\circ$

* The angle error ϵ_α is 90.00° for arbitrary constant $\hat{d}_{\vec{c}} \neq \vec{0}$

Iterative MC In order to increase the robustness of the MV estimation, an ensemble of five CNNs (see bagging approach in Section 5.3.1) is utilized for gradual approximation in an iterative fashion. An additional input parameter k_{\max} is introduced which indicates the number of alternating MV estimation and MC-FBP steps. Algorithm 5.1 provides the pseudo code of the iterative CoPACT MC pipeline.

As an extension, alternative stopping criteria instead of a fixed number of iterations are conceivable. Termination with respect to the convergence behavior or exploitation of the artifact level quantification networks from [69] are possible solutions. The proposed CoPACT MC pipeline can be either applied on the full coronary artery tree, on individual segments or on a single centerline point. In case of $|C| = 1$ and accurate motion estimation, the dense MVF $\hat{d}(t_{cc}, \vec{v})$ equals to $\vec{d}_{\vec{c}}^{-1}(t_{cc}, \vec{v})$. The duration and the robustness regarding outliers can be regulated by the density of centerline points.

5.4. EXPERIMENTS AND RESULTS

The Microsoft Cognitive Toolkit (CNTK v2.5+, Microsoft Research, Redmond, WA, USA) is used as deep learning framework. In Section 5.4.1, networks accuracies are analyzed based on synthetic motion artifacts. Qualitative performance evaluation of the proposed CoPACT MC pipeline based on real motion artifacts, observer studies and a run-time analysis are provided in Section 5.4.2.

5.4.1. QUANTITATIVE ANALYSIS ON SYNTHETIC ARTIFACTS

The following error metrics are introduced for network evaluation:

$$\begin{aligned} \epsilon_{x,y} &= \sqrt{(x_d - \hat{x}_d)^2 + (y_d - \hat{y}_d)^2}, \\ \epsilon_o &= |o - \hat{o}|, \text{ for } o \in \{s, x, y\}, \\ \epsilon_\alpha &= \min(|\alpha - \hat{\alpha}|, 360^\circ - |\alpha - \hat{\alpha}|) \end{aligned}$$

Table 5.1 summarizes the results on the testing subsets. As expected, more accurate MV prediction is achieved during the phantom study. Fine-tuning shows a slight advantage over network optimization from scratch. However, during qualitative performance evaluation of the proposed CoPACT MC pipeline based on real motion artifacts, networks trained from scratch yield better results than fine-tuned ones. The fine-tuned networks deliver too conservative in clinical practice, i.e. they seldom venture predictions

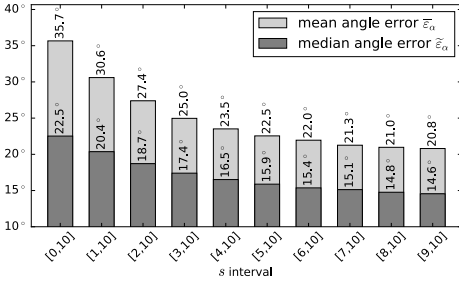


Figure 5.6: Bar plot of mean and median angle error evaluated for subsets determined by the selected s ranges.

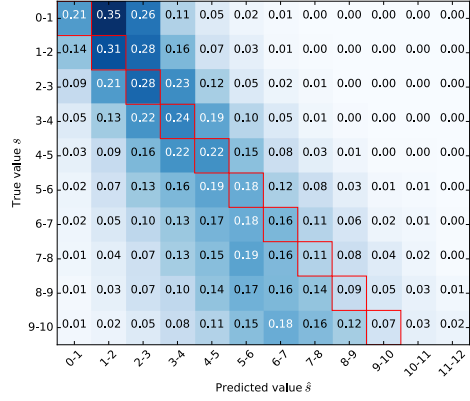


Figure 5.7: Confusion matrix of the target motion strength s .

far from the mean $\vec{0}$ also in the presence of severe artifacts. This could be explained by possible overfitting on the synthetic artifacts. For this reason, the following quantitative error analysis and qualitative experiments are performed based on the bagging ensemble of the five networks trained on the clinical data without fine-tuning.

Figure 5.6 illustrates the relationship between the accuracy of the predicted motion direction and the introduced motion strength s . High angle errors ϵ_α correlate with low s values, i.e. most accurate prediction of the motion direction is feasible for image patches with severe motion artifacts. Figure 5.7 shows the mean confusion matrix of the target motion strength s . The CNNs in the network ensemble frequently deliver too conservative predictions. Especially high motion levels tend to be underestimated. This network behavior supports the iterative MC scheme. Furthermore, a weakness in the differentiation of low motion levels $s \in [0, 3]$ is observable, which is also difficult for a human observer (see Figure 5.3).

5.4.2. QUALITATIVE ANALYSIS ON REAL ARTIFACTS

Local MC The first experiment investigates how well motion estimation and subsequent compensation can be done from a single 2.5D image patch, i.e. in case of $|C| = 1$. Centerline points for 24 test patches are manually selected from the twelve clinical test cases described in Section 5.1.3 at vessel segments of varying position, orientation and artifact level. In Figure 5.8, the corresponding x-y planes of size 60×60 pixels are visualized before and after $k \in \{1, 3, 10\}$ iterations of CoMPACT MC. For comparison, the registration-based MC approach from Grass et al. [38] is considered which exploits the entire 3D image field of view (FOV).

The CoMPACT MC shows gradual improvement of the image quality in a multitude of test cases (e.g. in Figure 5.8c,h,i,k) and sensible convergence properties (except for Figure 5.8m). It has to be noted that the registration-based approach also fails in the patch of Figure 5.8m. Indeed, the proposed pipeline has an advantage over the registration-based approach in several patches, like for instance in Figure 5.8d,p,t. The networks are robust regarding slight shifts between patch center and vessel position (see Figure 5.8b,e).

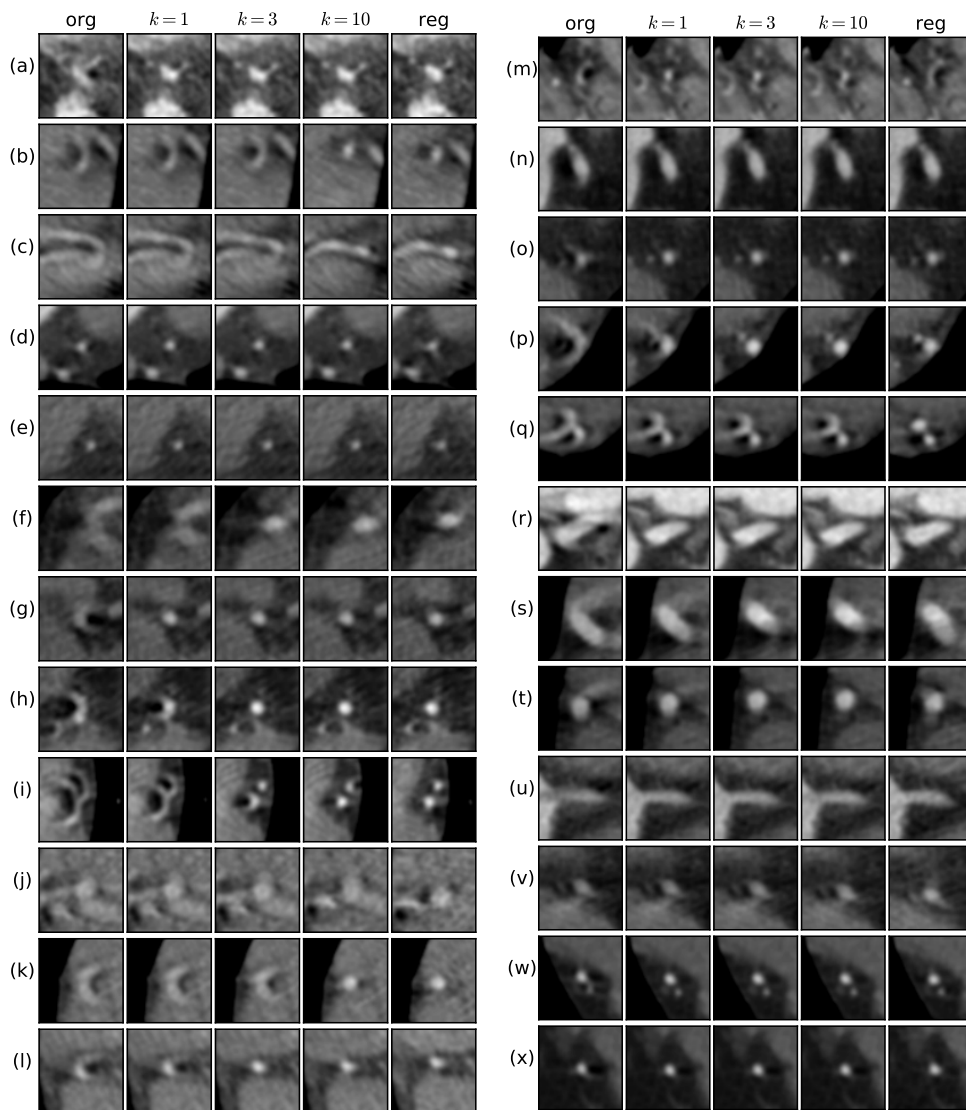


Figure 5.8: The x-y planes of 24 image patches belonging to twelve different patients are visualized before (org) and after $k \in \{1, 3, 10\}$ steps of CoMPACT MC. For comparison, the registration-based MC (reg) is considered. Significantly reduced artifact levels are observable in the majority of the test patches.

The main weakness of the method is the limited motion model complexity. Linear axial motion spatially constant in a local neighborhood is assumed. In some cases, these assumptions do not seem to be fulfilled, e.g. in the presence of more complex motion trajectories like turning motion (see Figure 5.8v) and spatially varying predominant motion directions (see Figure 5.8i,m,q). Nevertheless, the CoMPACT MC is remarkably successful in view of the little information content obtained from a single 2.5D image patch.

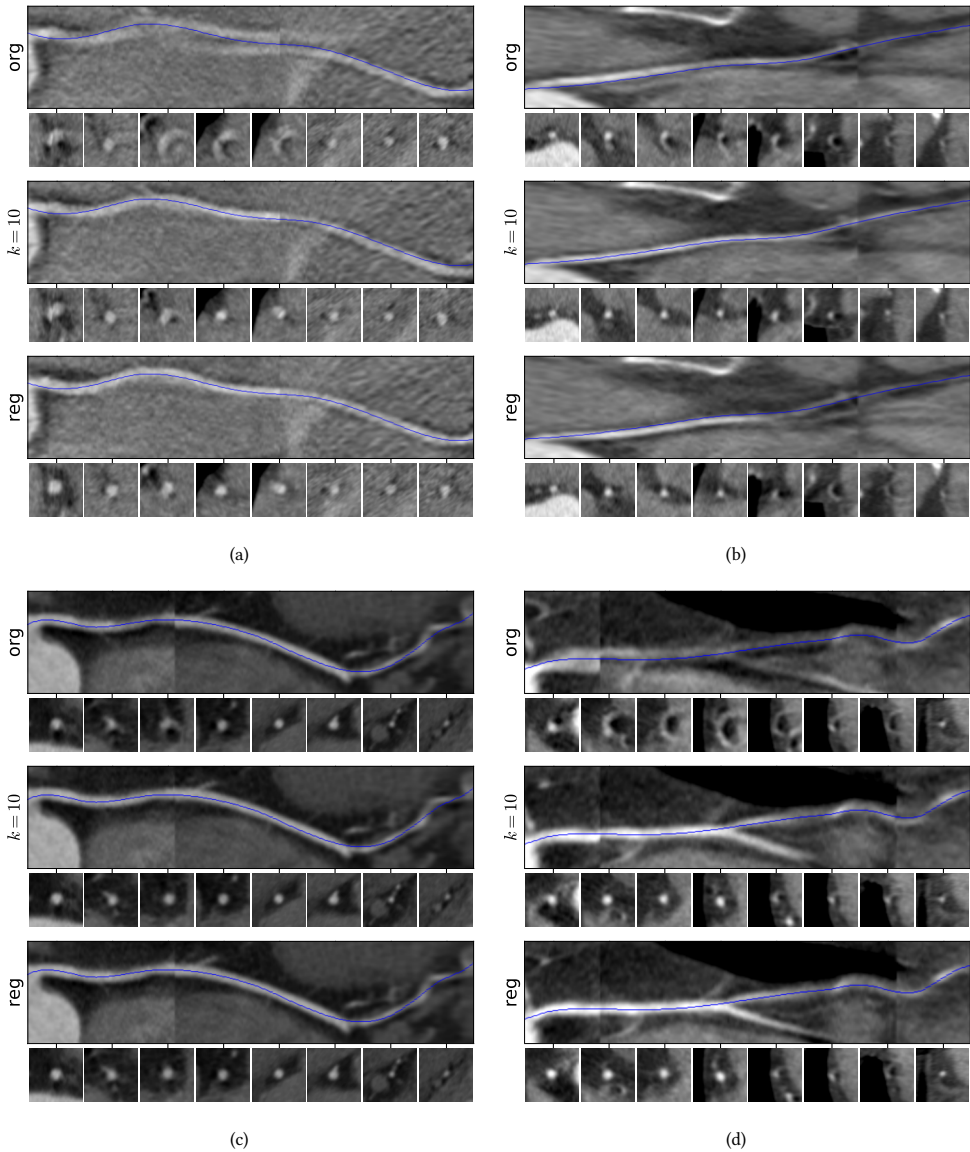


Figure 5.9: The multiplanar reformats of four vessels belonging to different patients and branches of the right coronary artery (RCA) are visualized before (org) and after $k = 10$ iterations of CoMPACT MC. Corresponding cross-sectional image patches which are perpendicular to the extracted centerline are given below for visual inspection. Furthermore, registration-based MC (reg) is considered for comparison. Both MC approaches, the registration-based and the proposed CoMPACT MC, lead to significant reduction of moderate and severe artifacts along the vessel.

Global MC The second experiment investigates how well motion estimation and subsequent compensation can be done in clinical practice by application on the whole coro-

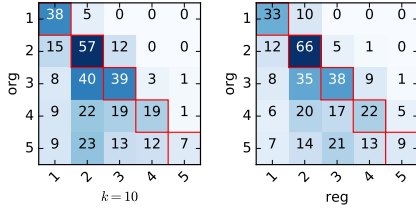


Figure 5.10: Cross-sectional image patches before and after MC are rated by four human observers in a five point Likert scale (from 1: excellent to 5: non-diagnostic).

Table 5.2: Mean duration of the CoPACT MC pipeline in case of $|C| = 1000$ and a FOV of size $512 \times 512 \times 300$ voxels.

Processing Step		Duration [secs]
Patch Sampling	$k = 1$	62.10 ± 3.61
	$k > 1$	11.54 ± 0.27
Ensemble Application		4.86 ± 0.03
MVF Extrapolation		172.17 ± 0.87
MC-FBP		116.16 ± 0.16
Total ($k_{\max} = 10$)		3196.44 ± 139.39

nary artery tree. The simultaneous consideration of various centerline points also allows for spatially irregular predominant motion directions. Figure 5.9 shows the multiplanar reformats of four vessels without MC, after 10 iterations of CoPACT MC and after registration-based MC. In Figure 5.9a,c,d, the corresponding centerlines were extracted from the output image volume of the registration-based MC using the Comprehensive Cardiac Analysis Software (IntelliSpace Portal 9.0, Philips Healthcare, Cleveland, OH, USA). The centerline in Figure 5.9b was determined based on the original image volume as the registration-based approach leads to increased artifact levels in the distal vessel segment. Significantly reduced artifact levels after CoPACT MC are observable in all cases, also in the presence of noise (see Figure 5.9a) or bifurcations (see Figure 5.9d). In Figure 5.9c, moderate artifacts in the proximal RCA are removed while the artifact-free mid and distal vessel segments captured by the second scanning shot remain unchanged.

Observer studies Four separate observer studies were performed to rate cross-sectional image patches before MC, after $k = 10$ iterations of CoPACT MC and after registration-based MC. Eight cross-sectional image patches are equidistantly sampled along the RCA (as illustrated in Figure 5.9) from eleven test cases resulting in a total number of $4 \cdot 3 \cdot 8 \cdot 11 = 1056$ labeled patches. The twelfth clinical test case was omitted as stack transition artifacts preclude the automatic coronary artery tree segmentation by means of the Comprehensive Cardiac Analysis Software. Rating is performed in a five point Likert scale (1: excellent, 2: good, 3: mixed, 4: strong artifact, 5: non-diagnostic). Vessel segments are presented in random order without indication of the underlying algorithm to the readers. It has to be noted that the readers were no radiologists, but research scientists with high level of expertise in reading cardiac CT images. The resulting annotations are summarized in Figure 5.10. Mean observer scores of 3.08 ± 0.24 , 2.28 ± 0.29 and 2.42 ± 0.23 are achieved on image volumes without MC, after $k = 10$ iterations of CoPACT MC and after registration-based MC. The performed experiments demonstrate the generalization capabilities of the trained neural networks on non-synthetic motion artifacts and a reasonable convergence behavior of the iterative MC scheme.

Run-time analysis Table 5.2 provides the results of a five-fold run-time analysis performed on a NVIDIA GeForce GTX 1080 Ti. In case of $k = 1$, patch sampling is performed with respect to the mean reconstruction direction and voxel positions are cached for faster

sampling in subsequent iterations ($k > 1$). Motion vector prediction by means of the networks, i.e. ensemble application, is the most time-efficient processing step, whereas MVF extrapolation and MC-FBP are quite time-consuming in the current implementation. Acceleration is possible by parallel processing of the individual scanning shots and restriction of the reconstruction region during MC-FBP to the cached voxel positions required for patch sampling. Furthermore, the run-time is controllable by adjusting the number of iteration steps k_{\max} and the centerline point density along the vessel.

5.5. DISCUSSION

The first single-phase motion estimation approach which works solely on reconstructed image data is proposed. The designed motion model which comprises linear trajectories in the axial plane, reveals potential and limitations of image-based motion estimation. Despite severe simplification of the actual, more complex heart motion, significant artifact reduction is achieved on clinical test data. More complex trajectories (e.g. turning motion) could be determined by performing constant linear motion estimation at multiple time points. Areas around the ostia exhibit 3D velocities with a noticeable contribution from the z-component. In contrast, mid- and distal RCA and mid-LCX segments have a dominant axial component and velocities in these segments are typically higher [111, 115]. The introduced procedure of data generation by the CoMoFACT and subsequent supervised learning is, in principle, extendable to arbitrary non-linear 3D motion trajectories. However, the information content of the reconstructed image volumes is a limiting factor in model extension to more complex motion. Initial tests on additional prediction of motion along the z-axis were not successful. These experiments are not described in the dissertation. The performed experiments on network fine-tuning from phantom to clinical data already reveal the problem of potential overfitting to synthetic artifacts.

For application of the proposed CoMPACT MC pipeline, the approximate locations of the coronary arteries have to be known. In case of incomplete or incorrect fully automatic centerline segmentation due to severe motion artifacts, semi-automatic approaches which require user-interaction have to be considered. This requirement constitutes the main disadvantage of patch-based MC in comparison to registration-based MC. Both approaches lead to significantly reduced artifact levels in the CCTA images. CoMPACT requires the minimal angular range of 180 degrees in parallel rebinned geometry while the registration-based MC demands 315 degrees for the partial image reconstructions. Furthermore, CoMPACT shows very promising results despite minimal spatial information which enables fast local processing of a few centerline points and their neighborhood.

Patch-based motion estimation offers a lot of potential for further research. Additional prediction of network uncertainty and integration into the distance-weighted MVF extrapolation might be useful in patches with little information content, e.g. in case of low contrast enhancement. A cascade process as visualized in Figure 5.11 is conceivable as well. In a first step, the artifact measures from Chapter 4 could decide whether and where motion correction is required. For remaining patches with mid-size or severe artifacts, more accurate prediction of the motion direction can be performed. As an alternative to the fixed number of iterations, artifact quantification can be applied to define a convergence stopping criteria. In general, motion orthogonal to the mean reconstruction direction leads to most severe artifacts in the reconstructed CT-image (see Figure 4.5). This

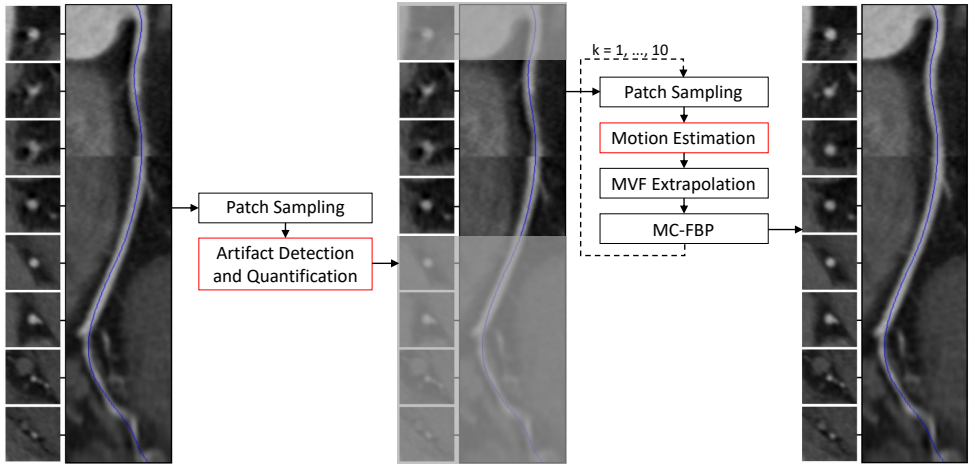


Figure 5.11: Schematic drawing of the cascaded MC pipeline, whereby DL-based components are highlighted in red. In a first step, artifact detection enables the decision whether and where motion correction is required for accurate diagnosis of CAD. Second, motion correction is locally performed by alternating CoMPACT, MVF extrapolation and MC-FBP. Finally, artifact quantification measures could be used to verify the success of the performed motion compensation and replace the fixed number of iterations as stopping criteria.

5

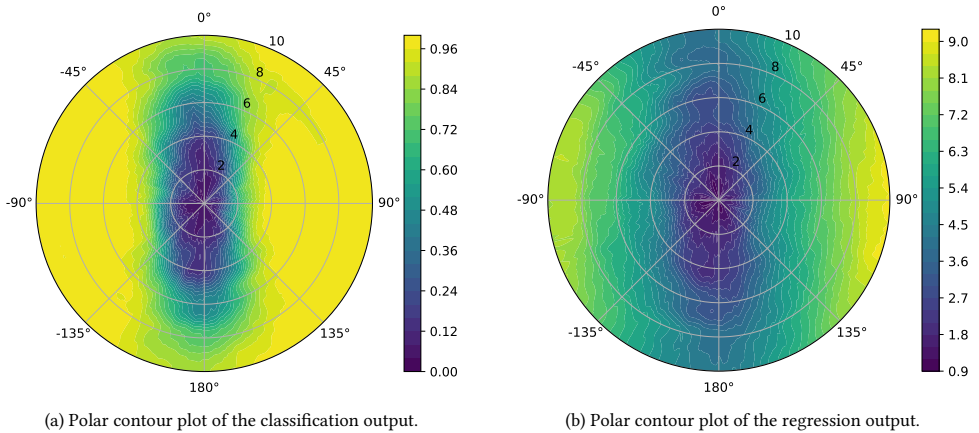


Figure 5.12: The predicted artifact probability of the classification network and the predicted artifact level of the regression network from [69] are calculated for the clinical validation data and visualized with regard to the underlying (s, α) coordinate. The networks diagnose least severe artifacts in case of parallel motion ($\alpha = 0^\circ$ or $\alpha = 180^\circ$). Especially, the output of the regression network is spatially smooth.

observation is also confirmed by the motion artifact measurements in Figure 5.12 which is taken from [25]. Hierarchical motion estimation by an initial line search, i.e. estimation of the motion magnitude with direction $\alpha = \pm 90^\circ$, and subsequent motion model extension is feasible as well.

So far, the proposed CoMPACT method is only based on 19 clinical data sets. In general, CCTA images are acquired with a wide variety of scanner types, imaging protocols

and reconstruction algorithms. In order to increase the network's robustness, collection of more data and network fine-tuning is required. The transferability of the proposed CoM-PACT MC pipeline to other scanner types and imaging protocols should be investigated.

The methodology of first introducing simulated motion to clinical cases with excellent quality and subsequent supervised learning of motion estimation models based on the artifact appearance is, in principle, not restricted to contrast-enhanced coronary arteries. By providing a set of reference cases without motion artifacts, patch-based motion estimation and compensation is on-site trainable for data of arbitrary contrast protocol and other parts of the human anatomy. Possible examples are motion artifact removal at the aortic valve or correction of calcium scores in non-contrast CT. The information content of the reconstructed image patches and overfitting to synthetic artifacts are again potentially limiting factors.

6

REMOVAL OF PACEMAKER METAL ARTIFACTS

EXISTING second-pass MAR methods are not applicable to cardiac CT data (see Figure 6.1a). In order to ensure robustness regarding motion, the approach of rawdata-based metal shadow segmentation as visualized in Figure 6.1b is pursued. The required feature and label data for supervised learning is generated by a forward model which inserts synthetic pacemaker leads into clinical cases without metal implants. The following differences to the previous forward model for coronary motion simulation are noticeable. The CoMoFACT can generate a smooth transition between artifact-free and artifact-perturbed data, whereas absence or presence of a pacemaker is a binary state. Furthermore, \vec{x}_{syn} contains projection data as well as image data in this application.

In the conference article [70], an initial *Dynamic Pacemaker Artifact Removal* (DyPAR) pipeline is developed. It is based on CNNs trained on clinical data with synthetic pacemaker leads for the task of metal shadow segmentation. This method is restricted to the insertion of static pacemaker leads, naive sinogram inpainting by inverse distance weighting (IDW) and no metal reinsertion is included. Extending this work, the journal article [71] introduces DyPAR+, whereby DL-based solutions for sinogram inpainting and 3D lead modeling are integrated as novel components into the pipeline. Furthermore, the forward model is adapted for dynamic lead insertion by taking concomitant ECG data into account.

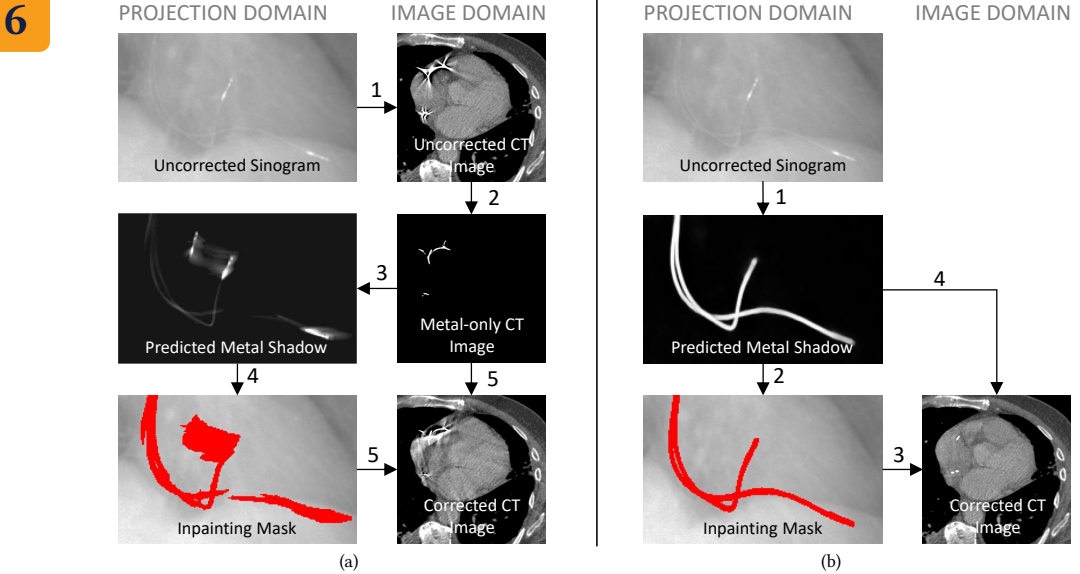


Figure 6.1: (a) Sinogram completion consists of (1) reconstruction, (2) metal segmentation, (3) forward projection, (4) inpainting, (5) reconstruction and metal reinsertion. Due to the metal segmentation in the image domain, these MAR approaches are not applicable in the presence of motion. (b) Circumvention of the detour via the image domain is proposed by (1) DL-based metal shadow segmentation directly in the projection domain. The purely rawdata-based methodology is independent of 3D motion blur, i.e. predicted pacemaker shadows are correctly located. Subsequent (2) inpainting and (3) reconstruction delivers the CT image volume without pacemaker associated artifacts. Based on the segmented metal shadow, (4) metal masks in the image domain can be estimated.

Table 6.1: Comparison of clinical test database and synthetic learning database with regard to pacemaker type, acquisition settings and scanner type (iCT: Brilliance iCT / B64: Brilliance 64, Philips Healthcare). The scanner type and the helical pitch determine the scan trajectory and thus also the reconstruction geometry. The rotation time [sec] and the number of recorded projection views per gantry turn define the temporal distance within the projection data.

	Clinical Test Data									Synthetic Learning Data			
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 1	Case 2-4	Case 5-7	Case 8-14
Defibrillator	yes	no	no	no	no	yes	no	no	no	no	no	no	no
Gated	yes	yes	yes	yes	yes	yes	no	no	no	yes	yes	yes	yes
Dose Modulation	no	no	yes	yes	yes	no	no	no	no	no	no	no	no
Scanner Type	iCT	iCT	iCT	B64	iCT	iCT	iCT	iCT	iCT	iCT	iCT	iCT	iCT
Helical Pitch	0.160	0.160	0.165	0.200	0.180	0.180	0.664	0.664	0.993	0.160	0.160	0.180	0.180
Rotation Time / Turn	0.272	0.272	0.272	0.420	0.330	0.330	0.330	0.330	0.750	0.272	0.272	0.272	0.272
Number Views / Turn	1800	1800	2400	2320	2400	2400	2400	2400	2400	2400	1800	1800	2400

This chapter is mainly inspired by [68] and structured as follows:

- 6.1 Clinical cases with and without implanted pacemakers are collected. The cases with pacemakers are grouped into template and clinical test data depending on availability of corresponding raw projection data.
- 6.2 Pacemaker lead models are build based on the template data. The synthetic leads are inserted into the pacemaker-free reference cases by means of thin plate spline smoothing, dilation and time-dependent forward projection. By this procedure, pairs of metal-free and metal-affected projection data in addition to corresponding metal masks in the image domain are generated.
- 6.3 Ensembles of CNNs are trained for the tasks of metal shadow segmentation, sinogram inpainting and 3D lead modeling and included into the fully-automatic DyPAR+ pipeline.
- 6.4 The proposed DyPAR+ pipeline is applied on the clinical test data and compared to common MAR and DyPAR.
- 6.5 Potential extensions of DyPAR+, e.g. for the establishment of consistent projection data or the application on other types of metal implants, are discussed.

6.1. MATERIAL

The learning data is generated by insertion of simulated pacemaker leads into the *reference data without pacemakers*. Reasonable lead positions and pathways are extracted from the *template data with pacemakers* as detailed in Section 6.2. Additional clinical cases with real pacemakers are collected in order to test the transferability of the proposed DyPAR+ pipeline in clinical practice.

6.1.1. LEARNING DATA

Reference cases without pacemakers The raw projection data of 14 contrast-enhanced cardiac CT data sets without pacemakers is collected for synthetic lead insertion. In all reference cases, acquisition was performed with a 256-slice CT scanner (Brilliance iCT, Philips Healthcare, Cleveland, Ohio, USA) using a retrospective gating protocol with helical trajectory. Details on the acquisition settings are summarized in Table 6.1.

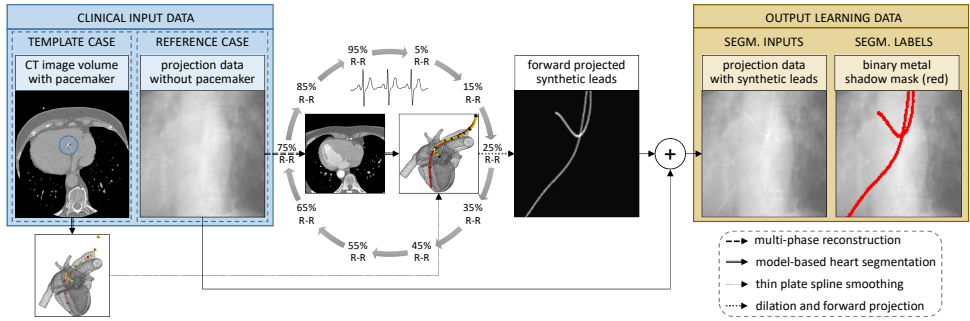


Figure 6.2: The data required for supervised learning is generated by a forward model which introduces synthetic pacemaker leads into the image and projection data of clinical cases without pacemakers. Surface meshes delineating the segmented heart during multi-phase reconstruction allow for sensible insertion positions and motion trajectories.

Template cases with pacemakers Seven reconstructed CT image volumes with pacemakers are collected for the extraction of pacemaker lead positions and pathways with respect to the cardiac anatomy. Corresponding raw projection data is not required. Dual as well as triple chamber pacemakers are included, i.e. synthesis of right atrial, right ventricular and coronary sinus leads is aimed for.

6

6.1.2. CLINICAL TEST DATA

In order to furthermore investigate generalization capabilities of the proposed DyPAR+ approach in clinical practice, the raw projection data of 9 additional cardiac CT data sets with real pacemakers is collected. The clinical test data allows evaluating the networks behavior in the presence of unseen features like electrodes or defibrillators and the robustness with regard to variations in contrast-enhancement, motion levels and acquisition settings. ECG-gated as well as ungated contrast-enhanced CT scans with helical acquisition trajectories are included (see Table 6.1). The ECG-gated test cases are reconstructed with a cardiac gating window around 75% R-R using AWCR, i.e. at mid-diastolic quiescent phase. For the ungated test cases which exhibit lower contrast agent densities, simple FBP is applied. Evaluation results on the clinical test data are presented in Section 6.4.2.

6.2. FORWARD MODEL

The data generation process is visualized in Figure 6.2. The forward model takes one template case (i.e. a reconstructed image volume with pacemaker) and one reference case (i.e. one ECG-gated sinogram without pacemaker P_{org}) as input and delivers synthetic data required for the supervised learning processes as output. First, a set of ten reference image volumes is obtained by multi-phase reconstruction with a temporal distance of 10% cardiac cycle using AWCR. For the resulting reference phase-volumes and the template image volume, corresponding heart meshes are determined by model-based heart segmentation according to [23]. Along each pacemaker lead in the template case, at least ten B-spline knots are manually selected. It has to be noted that the definition of these landmarks represents the only manual processing step within the dynamic forward model. Thin plate

spline smoothing based on point-to-point correspondences in the segmented heart meshes allows the transformation of the B-spline knots from the template case into each phase volume of the reference case.

Synthetic metal shadows in the originally acquired projection geometry of the reference case P_{metal} are gradually filled by the following procedure. For each projection view in the reference sinogram, the corresponding cardiac phase point t_{cc} is determined. Landmark positions in this specific motion state are calculated by linear interpolation of the B-spline knots associated with the two neighboring phase points within $\{5\%, 15\%, \dots, 95\%\}$ cardiac cycle. This approach ensures continuous movements across various projection views. The corresponding B-spline curve is determined by cubic B-spline interpolation for each pacemaker lead, separately. Dilation of the resulting lines with a chosen lead diameter of 2 millimeters and an attenuation value of 4500 HU yields the binary image volume $I_{\text{metal}}(t_{cc})$ in the reference image geometry. Subsequent forward projection delivers the metal shadow for the currently processed projection view. Clinical projection data with synthetic leads $P_{\text{input}} = P_{\text{org}} + P_{\text{metal}}$ is obtained by summation of the original projection data and the forward projected lead mask. Thresholding with zero defines the corresponding reference segmentation mask $P_{\text{mask}} = \mathbf{1}_{P_{\text{metal}} > 0}$.

Random combination of reference and template data enables generation of multiple metal-perturbed CT data sets. In the following experiments, the dynamic forward model is applied two times per template case as twice as many reference cases are available. With an iCT detector shape of 128×672 and an average number of 9 500 projection views per reference case, a total amount of $14 \cdot 128 \cdot 672 \cdot 9500 \approx 1.15 \cdot 10^{10}$ labeled line integrals is collected. The database required for the segmentation, inpainting and modeling learning tasks comprises for each provided reference case:

P_{input}	projection data with synthetic leads,
P_{mask}	binary mask of metal-affected line integrals,
P_{org}	original metal-free projection data,
$I_{\text{metal}}(t_{cc})$	time-dependent metal mask (image domain).

The learning data is case-wise separated into training, validation and testing subsets with a ratio of 8:4:2, or rather 4:2:1 with respect to the corresponding template cases in order to ensure disjoint pacemaker geometries and background line integrals among the subsets.

6.3. LEARNED METHOD

The proposed DyPAR+ pipeline is built of three CNN ensembles and takes the raw projection data of a metal-affected CT scan as input. The SegmentationNets identify metal-affected line integrals directly in the projection domain, i.e. independent of motion. The InpaintingNets treat metal-affected values as missing data and refill the projection data based on surrounding line integrals. Subsequent reconstruction of the inpainted sinogram delivers the CT image volume without metal. The ModelingNets finally determine metal positions in the image domain based on the segmented metal shadow mask. The resulting metal mask can optionally be visualized as overlay. The DyPAR+ processing pipeline is illustrated in Figure 6.3.

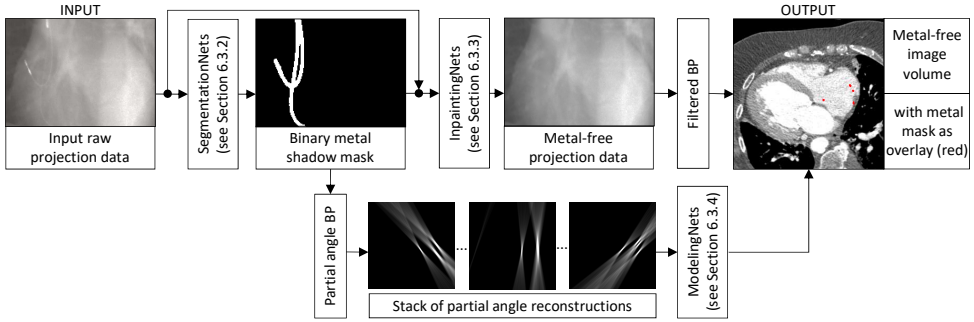


Figure 6.3: Neural networks trained for the tasks of metal shadow segmentation, sinogram inpainting and metal reinsertion represent the basic elements of the proposed dynamic pacemaker artifact removal (DyPAR+) pipeline. The back-projection operator is abbreviated to BP.

6.3.1. SHARED LEARNING FRAMEWORK

This section details components of the learning framework including network architecture and hyper-parameter settings which are shared across the different tasks of metal shadow segmentation, sinogram inpainting and metal modeling. Information on task-specific learning setups are provided in the following Sections 6.3.2, 6.3.3 and 6.3.4.

6

Network architecture During training, the neural networks take patches of size $b \times n \times n \times h$ as input and deliver patches of size $n \times n$ as output. The number of channels b , the number of slices h and the plane size $n \times n$ are task-specific parameters. Figure 6.4 illustrates the utilized U-Net architecture which is adapted from [94]. In case of multi-slice inputs ($h > 1$), feature extraction in the contracting path is performed for each slice separately using shared weights. Slice features are joint in the bottleneck and merely feature maps of the center slice are copied from the contracting to the expanding path in the skip connections. The network exhibits a receptive field size of $81 \times 81 \times h$. In contrast to [70], partial convolution based padding according to [65] is performed to keep in-plane input and output sizes equal. In general, arbitrary output shapes are enabled by the fully convolutional network, therefore, during validation and testing, metal shadow segmentation and inpainting is performed over the full detector size of 128×672 in a single step.

Bagging approach The stochastic gradient descent solver Adam [57] with an initial learning rate of 0.01 and a momentum of 0.8 is used for network optimization. The learning rate decreases with a factor of two after 33% and 66% of the overall training time. L2 regularization with a weight of 0.0002 is used. One training epoch is defined by 10^5 processed samples. Please remind, that the learning data is case-wise separated with respect to the corresponding template cases, in order to ensure disjoint pacemaker geometries. For each task, an ensemble of seven CNNs is trained by the following bagging approach:

1. Test data associated with one template case is selected.
2. Validation data (associated with two template cases) is randomly sampled from the remaining data.

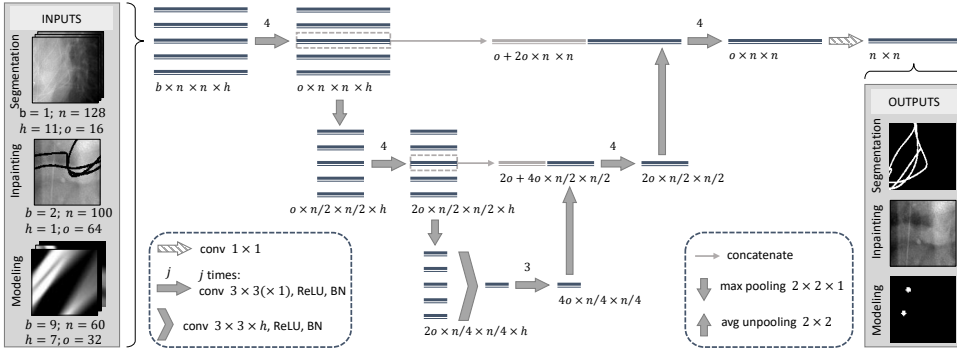


Figure 6.4: The fundamental network architecture is shared across the tasks of metal shadow segmentation, sinogram inpainting and metal modeling. The adapted U-Net design enables multi-channel and multi-slice input patches. Task-specific parameters include the number of input channels b , the number of input slices h , the plane size $n \times n$ and the number of feature maps at each layer determined by o . Each double line corresponds to a multi-channel feature map. A slice number of $h = 5$ is illustrated for clarity. The actual shape of each feature map is denoted below each line block. The arrows represent the different operations.

3. Network training is performed based on the remaining data (associated with four template cases).
4. After every epoch, the network's generalization capabilities on the validation data are assessed by a task-specific validation metric.
5. The model with the highest validation metric within all training epochs is selected for performance evaluation and application in DyPAR+.
6. Steps 1-5. are repeated seven times; every synthetic learning case is selected one time for testing.

Case-wise subsets for training, validation and testing remain unchanged across the tasks.

6.3.2. SEGMENTATIONNETS

SegmentationNets are trained to map projection data with metal leads P_{input} to corresponding binary masks P_{mask} of metal-affected line integrals. On average, 1.37% of the labeled projection data contain metal shadow voxels. To compensate for this foreground-background class imbalance, a patch-based learning approach is applied.

Patch sampling During training, the SegmentationNets take patches of size $1 \times 128 \times 128 \times 11$ as input and deliver patches of size 128×128 as output. The first dimension of the networks input corresponds to the number of channels (here $b = 1$). The second and the third dimension contain the information of the detector row and column. The fourth dimension indicates projection views which are equidistantly sampled with respect to the number of views per gantry turn so that 12 degrees gantry rotation are captured. The SegmentationNets target the segmentation mask of the central (sixth) view. By including neighboring projection views, the networks get additional information on the rotation velocity, i.e. the distance of supposed pacemaker leads to the rotation center. The sampling process is controlled such that 75% of the target output patches contain at least one object voxel while the remaining 25% are randomly sampled.

Learning framework In the contracting path of the SegmentationNets, 2D lead features are extracted for each view, separately and joint in the bottleneck to exploit the temporal information. In the expanding path, location information of the center slice to be segmented are copied from the contracting path via skip connections. In the last network layer, the soft-max function is used as activation. The SegmentationNets architecture with its shared weights in the contracting path has 423 730 learned parameters. Training is performed over 30 epochs using a mini-batch size of 32. The learning process is driven by the focal loss [63] which is well suited for imbalanced segmentation tasks using a focusing parameter of $\gamma = 2$. The Dice coefficient with a probability threshold of $T = 0.5$ is used as validation metric in the bagging approach.

Application in DyPAR+ In order to increase the robustness of the metal shadow segmentation, the entire ensemble of seven SegmentationNets yielded by the bagging approach is applied on the input raw projection data. The output probability maps are averaged across the ensemble and contain values in $[0, 1]$. The binary metal shadow mask (1: metal-affected, 0: background) is obtained by thresholding with $T = 0.15$. The choice of the relatively low threshold is motivated by the fact that incompletely segmented metal shadows may lead to newly introduced artifacts after inpainting and reconstruction. Therefore, sensitivity is judged as more important performance measure than precision. The threshold of $T = 0.15$ corresponds to a maximal false negative rate of 1% during testing on synthetic learning data.

During DyPAR+, InpaintingNets and ModelingNets rely on the outputs of the SegmentationNets, i.e. they have to deal with false positive and false negative metal shadow segmentations. For each data split in the bagging approach, the model selected in step 5. is applied on the test cases. Binary segmentation masks resulting by thresholding with $T = 0.15$ are stored in P_{segm} . As described in the following sections, InpaintingNets and ModelingNets are trained on P_{segm} , i.e. the testing output masks with slight inaccuracies, rather than on the ground truth masks P_{mask} .

6.3.3. INPAINTINGNETS

InpaintingNets are trained to map metal-afflicted projection data P_{input} masked by $M = \neg P_{\text{segm}}$ to corresponding metal-free line integrals from the original projection data P_{org} . Line integrals with a mask value of zero are treated as missing data.

Patch sampling During training, the network takes patches of size $2 \times 100 \times 100 \times 1$ as input, whereby channel information are sampled from the projection data P_{input} (first channel) and the corresponding mask M (second channel). Online data augmentation is performed by randomly treating up to ten additional line integrals in every tenth input patch as unknown. The sampling process is controlled in such a way that each patch contains at least one missing line integral.

Learning framework Several DL-based approaches dealing with free-form masks and non-blind image inpainting (i.e. regions to be inpainted are known a priori) have been presented in the last years [85, 121]. The network architecture illustrated in Figure 6.4 is

adapted for the inpainting task by replacing all convolutional layers with partial convolutions as suggested by Liu et al. [64]. In a partial convolution only valid pixels are taken into account and the layers output is re-normalized according to the ratio of kernel size and masking area. The mask is updated after every layer and passed as additional single-channel feature map through the network. Within the skip connections, merely common feature maps are copied and concatenated, i.e. upsampled masks from deeper layers are utilized in the expanding path. After the last convolution layer with linear activation function, inpainted areas of the networks output $\text{NN}_{\text{inpaint}}$ are combined with the original input by $P_{\text{clean}} = M \odot P_{\text{input}} + (1 - M) \odot \text{NN}_{\text{inpaint}}$. The network has an increased number of feature maps compared to the SegmentationNets and 3 805 313 learned parameters in total. Training is performed over 60 epochs using a mini-batch size of 64. The learning is driven by the combined loss function:

$$L_{\text{inpaint}} = \frac{1}{2N^{\boxplus}} (\|P_{\text{clean}}^{\boxplus} - P_{\text{org}}^{\boxplus}\|_1 + \|S_x * P_{\text{clean}}^{\boxplus} - S_x * P_{\text{org}}^{\boxplus}\|_1 + \|S_y * P_{\text{clean}}^{\boxplus} - S_y * P_{\text{org}}^{\boxplus}\|_1)$$

The loss function is defined on patch-level, i.e. $P_{\text{clean}}^{\boxplus}$ and $P_{\text{org}}^{\boxplus}$ represent patches sampled from P_{clean} and P_{org} , respectively. $N^{\boxplus} = 100^2$ is defined as the number of line integrals in a training patch and $S_{x/y}$ denote Sobel convolution kernels for vertical and horizontal derivative approximation. The loss function is a combination of the MAE on the projection data and the MAE on the corresponding gradients. It therefore penalizes differences in edge information. The validation metric utilized in step 4. of the bagging approach is replaced by a per-pixel reconstruction accuracy

$$\text{nMAE} = \frac{1}{\|P_{\text{segm}}\|_1} \|P_{\text{clean}} - P_{\text{org}}\|_1 \quad (6.1)$$

normalized by the number of metal-affected line integrals.

Application in DyPAR+ InpaintingNets take the raw projection data P_{input} and the corresponding metal shadow mask P_{segm} predicted by the SegmentationNets as input. View-wise processing and subsequent ensemble averaging yields the inpainted projection data P_{clean} . The metal-free image volume is obtained by reconstruction of P_{clean} using either FBP for ungated cases or AWCR when concomitant ECG data is available.

6.3.4. MODELINGNETS

ModelingNets are trained to predict metal positions in the reconstructed image volume based on segmented binary metal shadow masks. Since the SegmentationNets might produce slight inaccuracies, the networks have to be robust regarding false positives and false negatives. Furthermore, cardiac motion needs to be compensated by the network to produce metal masks without blurring artifacts. The metal reinsertion method is inspired by existing motion compensation approaches which exploit the increased temporal resolution of partial angle reconstructions (PARs) [38, 40, 55].

Partial angle reconstruction Each projection view P_j is associated with a specific gantry rotation angle γ_j and acquisition time point $t_{c,j} \in [0 \% \text{ R-R}, 100 \% \text{ R-R}]$ within the

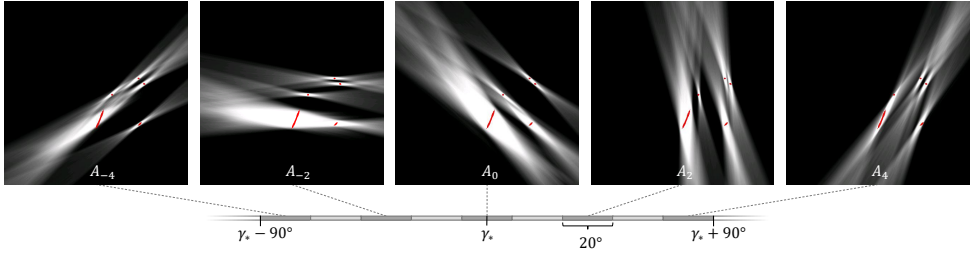


Figure 6.5: The ModelingNets take a stack of partial angle reconstructions $A_{-4}, A_{-3}, \dots, A_4$ obtained from the segmented metal shadow mask as input and deliver the corresponding metal mask in the image domain $I_{\text{metal}}(t_{cc,c})$ as output (highlighted in red). Projection beams in A_0 are closest in time to the target motion state determined by $t_{cc,c}$. With increasing temporal distance, slight shifts between back-projected beams and target metal mask can be observed.

cardiac cycle. Given a center projection view P_c , the back-projection operator \mathcal{B} without high-pass filtering is used to reconstruct nine partial angle volumes

$$A_i = \sum_{j \in \Gamma_i} \mathcal{B}(P_{\text{segm},j}), \quad \text{with } i \in \{-4, -3, \dots, 4\}, \quad \Gamma_i = \{j : |\gamma_j - 20^\circ i - \gamma_c| < 10^\circ\} \quad (6.2)$$

of disjoint 20° angle segments. By this procedure 180° gantry rotation are covered in total as illustrated in Figure 6.5. The target metal mask $I_{\text{metal}}(t_{cc,c})$ which corresponds to the motion state at acquisition time point $t_{cc,c}$ is highlighted in red. Depending on the center index c , a partial field of view (pFOV) is defined as

$$\Omega_c = \begin{cases} 1, & \text{if voxel is part of the FOV over the full angle range of } 180^\circ. \\ 0, & \text{otherwise.} \end{cases} \quad (6.3)$$

For each reference case, four stacks of PARs $S_c = \{A_{-4}, A_{-3}, \dots, A_4\}$ with varying center index c and corresponding target metal masks $I_{\text{metal}}(t_{cc,c})$ are created. Solely image regions which are part of the pFOV are included in the learning data. By this procedure, a total amount of $14 \cdot 512 \cdot 512 \cdot 4 \cdot 170 \approx 2.5 \cdot 10^9$ labeled voxels is collected whereby on average 0.024% of the image data contains object voxels. In order to compensate this class imbalance, a patch-based approach is applied again.

Patch sampling During training, the ModelingNets take patches of size $9 \times 60 \times 60 \times 7$ as input and deliver patches of size 60×60 as output. The PARs A_{-4}, \dots, A_4 which belong to different angular segments are provided as channel information. The second and third dimension of the networks input contain the information of the axial plane. Seven neighboring axial slices are included, whereby the ModelingNets target the segmentation mask $I_{\text{metal}}(t_{cc,c})$ of the middle fourth view. The pacemaker leads can be arbitrarily oriented in the image volume. By including neighboring axial views, the networks get additional information on the lead pathways. Online data augmentation is performed by random axial rotation of the input and target patches by $0^\circ, 90^\circ, 180^\circ$ or 270° . Furthermore, mirroring along the scanners x- and z-axis increases the training database. To enforce a clockwise system rotation direction, the channel order is inverted in case of mirroring along the

x-axis. The sampling process is controlled in such a way that 90% of the target output patches contain at least one object voxel while the remaining 10% are randomly sampled.

Learning framework The number of feature maps is doubled compared to the SegmentationNets, resulting in a total number of 1 398 114 learned parameters. The learning setup including loss function, validation metric and hyper-parameter settings remains unchanged compared to Section 6.3.2. Case sampling is performed with regard to the template cases, i.e. 8 testing, 16 validation and 32 training volumes are selected for each split in the bagging approach.

Application in DyPAR+ Gated CT scans are reconstructed by AWCR whereas simple FBP is used for ungated test cases. Therefore, application of the ModelingNets is adapted depending on the availability of ECG data.

ECG-gated test case: A specific heart phase $t_{cc,j}$ is associated with each projection views P_j by means of the ECG data. During AWCR a reference heart phase r needs to be specified which determines the center of the cardiac gating window. For each recorded cardiac cycle \mathcal{C} a corresponding subset of views is assigned by $\Lambda(\mathcal{C}) = \{j|P_j \text{ is aquired within } \mathcal{C}\}$. The projection view $P_{c(\mathcal{C})}$ with heart phase closest to the reference phase is identified by $c(\mathcal{C}) = \underset{j \in \Lambda(\mathcal{C})}{\operatorname{argmin}} |t_{cc,j} - r|$. The stack of corresponding

partial angle volumes $S_{c(\mathcal{C})}$ is generated according to equation (6.2) and fed into the ModelingNets. The networks output NN_{model} includes averaging across the ensemble. Under consideration of the pFOV $\Omega_{c(\mathcal{C})}$, the output probability map is calculated by

$$I_{\text{metal}} = \frac{\sum_{\mathcal{C}} \text{NN}_{\text{model}}(S_{c(\mathcal{C})}) \odot \Omega_{c(\mathcal{C})}}{\sum_{\mathcal{C}} \Omega_{c(\mathcal{C})}}. \quad (6.4)$$

Thresholding with 0.5 finally delivers the binary metal mask in image domain.

Ungated test case: In case of ungated CT data, projection views P_j can not be associated with a specific time point $t_{cc,j}$. But, for each axial image slice I_z , a corresponding nearest projection view P_j can be calculated by

$$j(z) = \frac{\text{NVPT}}{\text{Pitch}} (\text{res}_z \cdot z + T_z - \text{Source}_z), \quad (6.5)$$

whereby NVPT denotes the number of views per gantry turn and res_z is the image resolution in z . The expression $T_z - \text{Source}_z$ specifies the distance in z of the first axial slice in the image FOV to the center of rotation for the first projection view. In order to avoid blending of different motion phases in the metal image, the metal probability map is block-wise filled, i.e. $I_{\text{metal},Z} = \text{NN}_{\text{model}}(S_{c(Z)})$, whereby Z denotes a subset of axial slice indices. The center index $c(Z) = j(z^*)$ is calculated according to equation (6.5) whereby z^* is defined as the center slice of Z . The axial block size $|Z|$ is an adjustable parameter. A small block size leads to smoother output probability maps, but requires a longer runtime. It has to be noted that the metal reinsertion process for ungated cases can be significantly accelerated by reusing partial image volume under consideration of the table movement during 20° gantry rotation. Furthermore, PARs should only be back-projected to the relevant image area defined by Z . Thresholding of $I_{\text{metal},Z}$ with 0.5 finally delivers the binary metal mask in image domain.

Table 6.2: Test results on the synthetic learning data including mean and standard deviation of selected performance metrics for segmentation and inpainting tasks. The threshold T defines the metal-background class separation, whereby $T = 0$ would correspond to classifying all pixels/voxels as metal-affected. Except for the mean absolute error (MAE) in the projection and in the image domain, all scores are expressed in percent.

Approach	Metric	Score
SegmentationNets (threshold $T = 0.5$)	Dice coefficient	94.16±1.49
	Sensitivity	93.88±1.37
	Precision	94.57±2.34
SegmentationNets (threshold $T = 0.15$)	Dice coefficient	88.05±2.17
	Dice coefficient (D1)	97.88±2.42
	Sensitivity	99.02±0.79
	Sensitivity (D1)	99.53±0.53
	Precision	79.44±2.94
	Precision (D1)	96.51±4.03
InpaintingNets	nMAE (projection)	6.040±0.88
	MAE (image) [HU]	11.54±2.49
Inverse Distance Weighting	nMAE (projection)	6.337±0.87
	MAE (image) [HU]	12.17±2.57
PatchMatch	nMAE (projection)	6.912±1.16
	MAE (image) [HU]	12.72±2.71
ModelingNets (threshold $T = 0.5$)	Dice coefficient	55.60±4.79
	Dice coefficient (D1)	76.02±6.98
	Sensitivity	53.35±5.07
	Sensitivity (D1)	73.24±7.34
	Precision	59.15±5.03
	Precision (D1)	80.96±6.70

*D1-corrected: false positives and false negatives within the dilated true positive area (using a $3 \times 3 \times 3$ structure element) are ignored

6.4. EXPERIMENTS AND RESULTS

6

For all experiments, the Microsoft Cognitive Toolkit (CNTK v2.5+, Microsoft Research, Redmond, WA, USA) is used as deep learning framework. Section 6.4.1 deals with quantitative and qualitative evaluation of the network’s performance on the synthetic learning data. The network’s generalization capabilities to clinical test data with real pacemakers are examined in 6.4.2.

6.4.1. EVALUATION ON SYNTHETIC LEARNING DATA

Performance measurements of the network ensembles achieved on the testing subsets are summarized in Table 6.2. As most segmentation errors produced by the SegmentationNets and the ModelingNets occur at the boundaries of the metal mask, also D1-corrected performance measures are considered. Despite thin line-shaped object masks, remarkably high dice coefficients are achieved by SegmentationNets. The ModelingNets have to deal with a more extreme class imbalance and segmentation errors of the SegmentationNets. Furthermore, predicting the exact position and diameter of the pacemaker leads, based on the PARs is indeed a difficult task. The image volumes in the learning data are reconstructed with a voxel resolution between 0.4 mm and 0.5 mm. Therefore, lead pathways shifted by few voxels are tolerable.

The InpaintingNets are compared with the rule-based approaches PatchMatch [4] and inverse distance weighting [101]. For PatchMatch, a third-party implementation¹ is used. The IDW is performed in the following experiments by interpolating metal-affected line integrals based on border pixels (defined in a 8-neighborhood around the segmentation mask) using the L_∞ metric as distance function. All inpainting approaches are tested by view-wise processing using the projection data P_{input} and the ground-truth metal shadow masks P_{mask} as input. Besides the normalized mean absolute error in the projection do-

¹<https://github.com/younesse-cv/patchmatch>

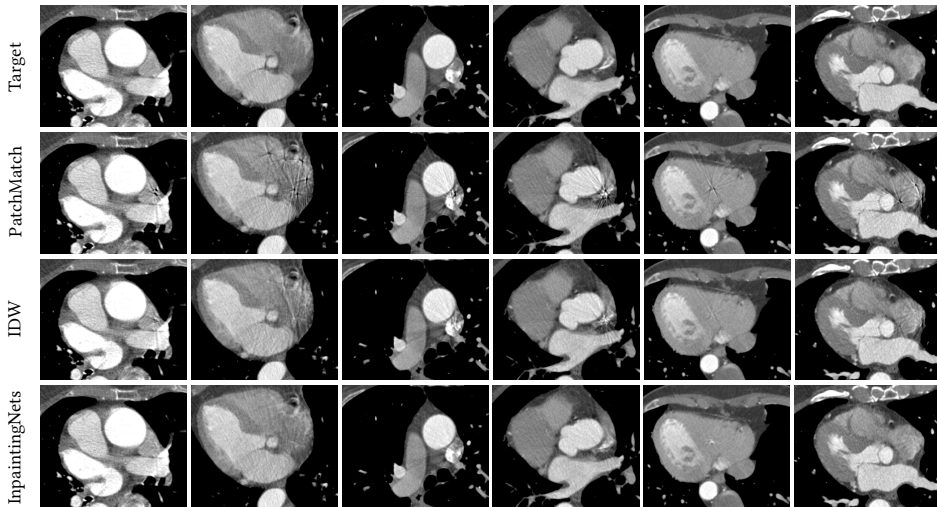


Figure 6.6: Comparison of reconstructed axial image slices after inpainting of synthetic metal shadows using PatchMatch, inverse distance weighting (IDW) and the InpaintingNets with the original reference image volume without synthetic leads.

main (nMAE) introduced in Equation (6.1), the mean absolute deviation from I_{org} is regarded, as the image quality after reconstruction is most crucial. In both domains, the DL-based approach outperforms the rule-based ones. In Figure 6.6, example reconstruction results after inpainting are compared. The visual impression coincides with the performance scores. The InpaintingNets induced least streak-shaped artifacts and seem to fill the projection data with higher consistency across the projection views.

6.4.2. EVALUATION ON CLINICAL TEST DATA

DyPAR+ is applied to 9 clinical test cases with real pacemakers described in Section 6.1.2. Figure 6.7 provides qualitative evaluation results of the networks outputs in the projection domain. As already stated in [70], the SegmentationNets object-background separation also generalizes to electrodes and defibrillators despite the lag of dedicated learning data. False negatives occur especially at the pacemaker leads due to a low deviation of metal shadow and background line integrals (see Figure 6.7d,f). ECG-leads and pacemaker leads are visually hard to distinguish based on a single projection view. Apart from few exceptions (see Figure 6.7h), the SegmentationNets are remarkably successful in their separation and seem also to consider rotation velocities (see Figure 6.7b,c,e). The InpaintingNets are able to fill metal-affected line integrals. However, inpainted areas exhibit removed noise patterns and reduced edge preservation, e.g. in the case of interrupted ECG-leads (see Figure 6.7b,e). This is a known effect of many data-driven inpainting approaches that use the MAE as loss function.

Maximum intensity projections visualized in Figure 6.8 compare between real and predicted lead pathways. The output of the ModelingNets highly depends on the quality of the previous metal shadow segmentation step. In all test cases, hardly false positive ac-

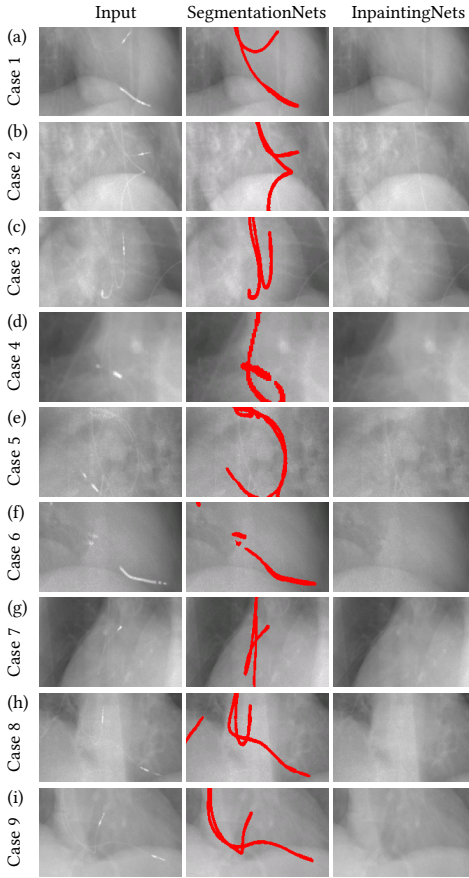


Figure 6.7: For each clinical test case, one example view of the input projection data P_{input} with corresponding outputs of the SegmentationNets P_{segm} and InpaintingNets P_{clean} is depicted.

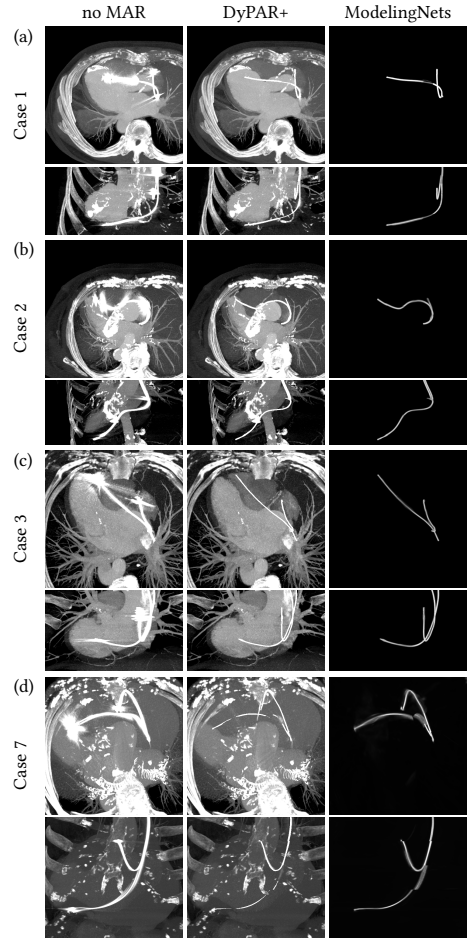


Figure 6.8: Maximum intensity projections of axial (top) and sagittal (bottom) slices are compared for image volumes reconstructed without MAR, image volumes after DyPAR+ and the output probability maps of the ModelingNets I_{metal} .

tivations are present. In the gated test cases, edges at the boundaries of the partial FOVs $\Omega_{c(\mathcal{C})}$ are visible in the probability maps (see Figure 6.8a-c). However, extracted binary metal masks after thresholding are coherent and inserted at the correct positions. The redundancy in the projection data due to the low pitches and averaging across different cardiac cycles increases the robustness of metal modeling for gated cases. In contrast, the predicted metal probability map of the ungated case in Figure 6.8d exhibits stack transition artifacts and increased blurring. One fifth of the axial slices is selected as block size $|Z|$. Interrupted leads occur in the extracted binary metal mask after thresholding. In this case it might be more sensible to use the probability map without binarization as heatmap



Figure 6.9: A selection of axial image slices without MAR, after 2nd pass MAR, after DyPAR and after DyPAR+ are visualized using a level/window setting of 150/750 HU. Reinserted metal is highlighted in red. Observer rankings of the MAR approaches are provided in the upper left corner of each slice whereby a score of 1 corresponds to least artifacts.

overlay. It has to be mentioned that especially in the ungated test cases, contrast enhancement and acquisition settings like pitch and gantry rotation speed vary from the learning data (see Table 6.1). Nevertheless, a significant metal artifact reduction can be observed in the DyPAR+ output.

The proposed pacemaker artifact removal method is compared with a common second pass approach and the previous DyPAR pipeline. In the second pass approach, metal

masks are segmented in the image domain using 3D hysteresis thresholds of 1000 HU and 1500 HU. The metal shadow areas are yielded by forward projection and thresholding with zero. IDW is applied as inpainting strategy. The DyPAR pipeline proposed in [70] comprises DL-based metal shadow segmentation and IDW. For each clinical test case, two axial slices are depicted in Figure 6.9 which exhibit severe metal artifacts after conventional reconstruction. Image slices after MAR are ranked by visual comparison of blurring and streak-shaped artifact levels, whereby a ranking of 1 corresponds to the highest image quality. In case of similar artifact levels, MAR approaches yield the same score.

Mean observer rankings of 2.5 and 1.94 are achieved by the second pass and the DyPAR approach. The second pass approach is not robust regarding cardiac motion and leads to increased blurring in the neighboring anatomy, incomplete metal removal and introduction of new severe artifacts in several slices (see Figure 6.9b,c,g,h,i). For visual inspection of motion perturbations in metal shadow masks resulting from image-based segmentation, please refer to Figure 2.7 and [70]. In [70], two additional test data sets without pacemakers are considered for false positives quantification. Incorrect metal shadow segmentations are caused by ECG leads and sternal steel wires. In contrast to the image-based approach which exhibits higher false positive rates, severe calcifications and stents are not misinterpreted by the networks. However, metal shadow masks are post-processed in DyPAR by largest connected component extraction in order to reduce the number of false positives. This post-processing step is not applied in DyPAR+, as it might lead to incorrect removal of true positives in case of gaps in the metal shadow masks (see Figure 6.9f). Furthermore, no metal reinsertion is performed in [70] in contrast to second pass MAR and DyPAR+.

With a mean observer ranking of 1.0, axial image slices after DyPAR+ exhibit least artifacts, i.e. it benefits from the DL-based metal shadow segmentation and inpainting. However, partial angle artifacts due to incomplete metal shadow segmentations (see Figure 6.9d,g) and introduction of streak artifacts due to inconsistencies among the 2D projections occurred after the inpainting step (see Figure 6.9a) are still the main sources of image quality degradation. Nevertheless, metal artifacts are successfully reduced by DyPAR+ and the evaluation of neighboring anatomy is facilitated in most cases. In Figure 6.9h number and position of pacemaker leads can be identified without cardiac motion blur and in Figure 6.9a-c, metal artifacts are removed which hamper evaluation of portions of the coronary arteries. It shows a high robustness to different noise levels, contrast agent densities and motion velocities. The experiments demonstrate the feasibility of pacemaker artifact removal without the need of an initial image reconstruction and the transferability from synthetic leads to real pacemakers.

6.5. DISCUSSION

The proposed DyPAR+ pipeline can facilitate evaluation of neighboring anatomy, for instance, with regard to inflammations or calcification. Especially, the procedure planning from cardiac CT data for minimal invasive pacemaker lead extraction might be improved by means of DyPAR+. Pacemaker leads need to be removed in case of infection, damage or needlessness. Furthermore, a multitude of existing model-based and DL-based approaches for risk assessment, anatomical segmentation etc. fail in the presence of pacemaker metal artifacts. Figure 6.10 shows an example of perturbed coronary vessel segmentation by the Comprehensive Cardiac Analysis Software (IntelliSpace Portal 9.0, Philips Healthcare,

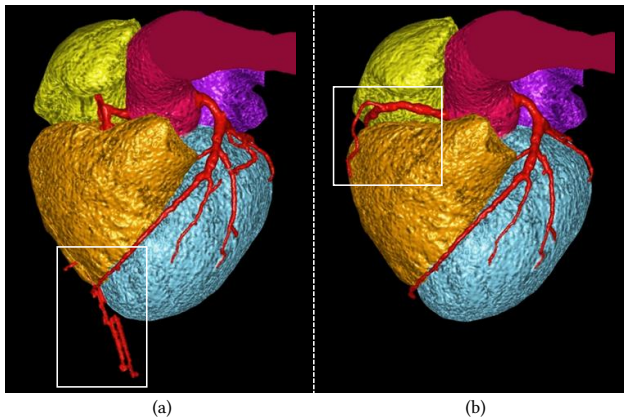


Figure 6.10: Coronary lumen segmentation in case 3 of the clinical test data without (a) and with (b) DyPAR+. False positives caused by streak-shaped motion artifacts and false negatives at the right coronary artery are highlighted in white boxes.

Cleveland, OH, USA). DyPAR+ might thus also be used as pre-processing step for subsequent examination of other anatomies and pathologies.

The proposed approach for dynamic MAR offers a lot of potential for further research and exhibits several tunable parameters. By reducing the angular range of the PARs in Section 6.3.4 to less than 20° , the temporal resolution could be further increased. During parameter tuning the observation was made that especially the number of input slices k in the SegmentationNets and the ModelingNets significantly influences the network performances. Multi-slice inputs are preferable to the single-slice inputs. Except for simple thresholding, no post-processing of the network outputs is performed so far. Application of morphological operations and frequency splitting [80] might further increase the resulting image quality.

Establishment and maintenance of consistent projection data is most crucial for artifact removal and avoidance, but the current view-wise processing precludes the consideration of long time dependencies. Approaches for spatial propagation, e.g. by means of recurrent neural networks [19] or spatio-temporal adversarial objectives [114], could provide solutions here. The extension of the dynamic forward model for introduction of synthetic electrodes and defibrillators with additional simulation of beam hardening and Poisson noise as performed by Zhang and Yu [125] might enable the combination of the segmentation and the inpainting step by directly learning to predict residual metal shadows. By this, also information behind the metal shadow are exploitable.

The methodology of first transforming template device models into clinical reference cases for subsequent device-specific network training is, in principle, not restricted to pacemakers. Transferability to projection-based detection and removal of other metal implants like artificial valves, electrodes, or left ventricular assist devices is part of future research. In contrast to second pass MAR approaches, DyPAR+ furthermore enables suppression of artifacts caused by high-density object outside the scan FOV. Besides the device-specific learning, also protocol-specific MAR is feasible. Cardiac CT images are acquired with a wide variety of imaging protocols. By providing a set of sinograms without metal implants, DyPAR+ is on-site trainable on data of arbitrary scanner type, acquisition mode and contrast protocol.

7

DISCUSSION, SYNTHESIS AND CONCLUSIONS

COMPUTED tomography image quality enhancement and artifact reduction is an excellent area to develop learning based methods to improve the imaging capabilities of modern scanners. This is due to the in depth knowledge of the underlying imaging physics and the associated ability to derive efficient and accurate forward models. In the previous chapters, the feasibility of coronary motion artifact detection, quantification and reduction as well as pacemaker artifact removal and metal reinsertion by means of DL-models trained on synthetic data is demonstrated. The working hypothesis was thus confirmed.

The proposed forward model CoMoFACT enables introduction of local motion in adjustable complexity at arbitrary positions of the vessel tree. The forward model for pacemaker lead synthesis is designed in such a way that sensible insertion positions, pathways and motion trajectories are ensured. An essential step for subsequent supervised learning is the sampling of *rich* features, e.g. multi-slice instead of single-slice inputs or patches rotated w.r.t. the angular reconstruction range. Careful forward model design, input feature sampling and supervised learning are thus the methodological building blocks of this work. Figure 7.1 summarizes the pursued strategy of developing knowledge-driven forward models to learn the AI-enabled backward models.

As task-related discussions, i.e. method limitations and potential extensions, are already provided in the corresponding Chapters 4, 5 and 6, the following sections give an interpretation of the methodology and the results with a more generic view. Section 7.1 compares risks and advantages of classical manual annotation and the conducted labeled data synthesis. An outlook of potential further applications in medical imaging is provided in Section 7.2.

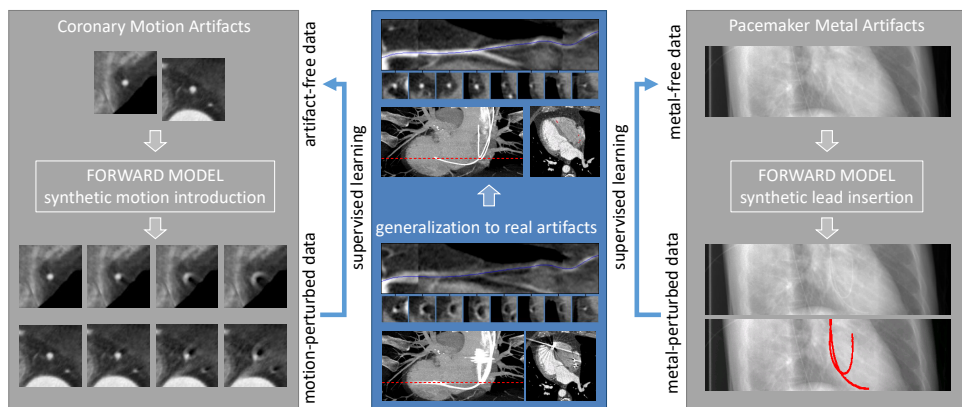


Figure 7.1: System knowledge is exploited to create synthetic artifact-perturbed data from artifact-free clinical cases. Subsequently, supervised learning is applied to develop predictive models which enable reversion and also generalize to real coronary motion and pacemaker metal artifacts.

7.1. LABELED DATA SYNTHESIS VS. MANUAL ANNOTATION

In the following, all challenges and strengths of labeled data synthesis identified during this work are listed, whereby - denotes a challenge, + indicates a strength and +/- is neural.

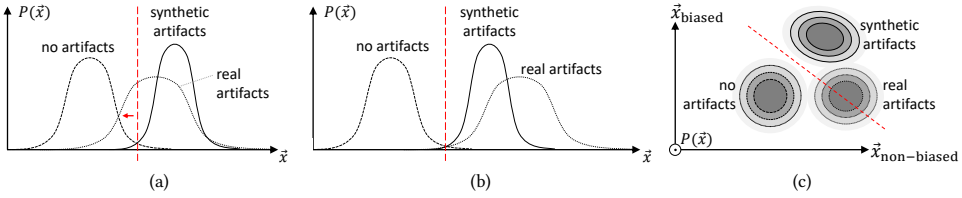


Figure 7.2: Examples of a classification networks trained on artifact-free data \bar{x}_{clean} and data with synthetic artifacts \bar{x}_{syn} which generalize badly (a) or well (b) on real artifact-perturbed data \bar{x}_{real} . The decision boundary is illustrated as dashed red line. (c) One potential cause of bad generalization is the occurrence of an unexpected bias. The input features \bar{x} are separated into biased \bar{x}_{biased} and non-biased components $\bar{x}_{\text{non-biased}}$. The predictive model may learn to separate between artifact-free and synthetic data based on feature components \bar{x}_{biased} . As the bias is not present in real data, the model will not generalize.

- **Challenging forward model development** The complexity of the forward model is similarly critical as the hypothesis space complexity. A forward model of increased complexity is prone to bad label calibration. One example is the required limitation of the CoMoFACT to constant linear motion in order to define accurate labels for coronary motion artifact quantification. Furthermore, the information content in the synthetic data \bar{x}_{syn} might be a limiting factor in the prediction of underlying artifact parameters \bar{y}_{syn} if the model is too complex. For this reason, CoMPACT covers solely linear trajectories in the axial plane despite potential extension to arbitrary non-linear 3D motion trajectories.

The input representation can be crucial as well. In Figure 7.2a, the classes *no artifact* and *real artifact* are not perfectly separable with given feature representation. One example are the similar appearances of bifurcations and blurring artifacts in single view data. In comparison, multi-slice input data led to increased classification accuracies.

A forward model of insufficient complexity, on the other hand, is related to the out-of-distribution problem. The larger the deviation of real and synthetic data, the higher is the risk of lacking generalization capabilities. However, Figure 7.2b shows the ideal case of a classification network trained on artifact-free data \bar{x}_{clean} and data with synthetic artifacts \bar{x}_{syn} which also generalizes to real artifact-perturbed data \bar{x}_{real} despite mismatching data distributions. One example is here the learned metal shadow segmentation which also holds for electrodes and defibrillators despite lack of dedicated learning data.

- **Overfitting to synthesized data** Figure 7.2a shows the contrary case of poor generalization due to an overfitting to the synthesized data. Another potential cause, besides the generation of insufficiently realistic data \bar{x}_{syn} due to the limited forward model complexity, is the occurrence of an unexpected bias as visualized in Figure 7.2c. The classification of coronary motion artifacts is, for instance, prone to such a bias. Centerlines are extracted based on the artifact-free reference cases and motion introduction may lead to slight vessel shifts. Therefore, image translation by cropping is necessary to avoid supplemental information embedded in the in-plane coronary position during supervised learning.

- **Forward model requirements** Collection of the required input data for the forward models can be challenging. The CoMoFACT database, for instance, is limited in scope, since clinical cases which meet the requirements on the reference data are hardly available. CCTA image volumes which are completely free of motion artifacts at the given resolution level are rare. Furthermore, corresponding raw projection data is most often unavailable. Typically, these are erased from the systems after a finite time period. In addition, raw data download from the system requires interruption of the clinical work flow and is therefore also difficult. In contrast, the input data of the pacemaker forward model is easier to get as solely the raw projection data of the pacemaker-free reference cases is required.
- +/- **Label noise** Labeled data synthesis and manual annotation may both suffer from label noise due to missing calibrateability. In Chapter 4, the fact is discussed that the artifact levels are not equivalent to the motion magnitudes as artifact appearances additionally depend on other influencing factors like the tube positions during acquisition, the orientation of the coronary arteries and the attenuation coefficient of the neighboring anatomy. Therefore, accurate label assignment is challenging. In case of manual annotation, influencing factors can, for instance, be subjective perceptions of different observers or annotation duration and time.
- + **Data efficiency** Several approaches for data augmentation, data generation and transfer learning have been developed in the last years in order to deal with the problem of machine learning on data sets of limited scope. Forward models allow one to generate for each artifact-free image patch multiple artifact-perturbed variants. Besides proposed applications in artifact detection and removal, forward models can also be used to augment the database for other learning or testing tasks like, for instance, DL-based centerline extraction and robustness evaluation with respect to increasing motion artifact levels. In Section 4.4.2 and Figure 5.12, predictive models trained for artifact level regression are evaluated on clinical data which is synthetically perturbed using a more complex motion model. By such a procedure, the forward model can also be used to generate test data for the investigation of generalization capabilities.
- + **Time efficiency** Manual annotation processes can be very time-consuming and, thus, often represent a bottleneck in predictive model development. Synthetic label generation, in contrast, requires the first time to set the forward model up, but afterwards works without user-interaction and furthermore enables on-site training.
- + **Increasing system knowledge** Development and application of the forward model can provide additional knowledge about the system itself. In the phantom studies of [25] and [68], the CoMoFACT is used to identify additional factors which influence the appearance of coronary motion artifacts.
- + **Controllable data distribution** Several approaches have been developed which deal with the problem of supervised learning on imbalanced data sets [52, 63]. Label and data distributions are, to a certain extend, steerable by the forward model. Uniformly distributed coronary motion magnitudes are, for instance, enforced in Chapter 4 and Chapter 5.

- + **Reproducible data** As mentioned before, intra- and interobserver variabilities occur during hand-labeling as results are influenced by data-independent factors like the annotating person or the annotation time (see Figure 4.10). In several cases like coronary motion artifact quantification, no accepted standard exists. Forward models offer the advantage of a deterministic feature-label-assignment on a quantitative scale.
- + **High-level label generation** Some types of ground truth labels like the underlying motion trajectories in [68] or the inpainting targets in [71] are not manually annotatable. They can only be approximated, e.g. by motion estimation using existing model-based approaches and subsequent manual selection of clinical cases with significantly improved image quality. In contrast, target outputs utilized to train the AI-enabled backward models are exact, i.e. also do not exhibit label noise.

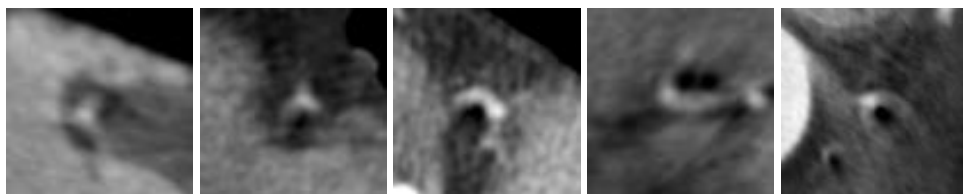
In total, three negative, one neutral and six positive arguments for the synthesis of labeled data are collected. However, the proposed procedure relies on the assumptions that

- (1) the information content in the input data is sufficient for target prediction,
- (2) the chosen training loss (e.g. the euclidean distance of predicted and real motion vector) is a good surrogate of the real target metric (e.g. the resulting image quality after artifact reduction),
- (3) the complexity of the hypothesis space \mathcal{H} is sufficient,
- (4) the synthesized data is sufficiently realistic in such a way that predictive models also generalize to real data.

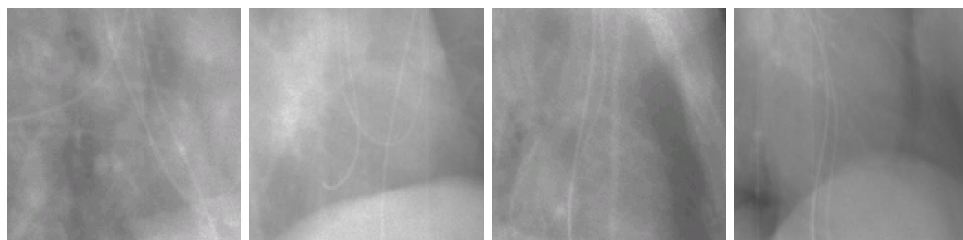
The assumptions (1) to (3) are generally made in supervised learning. As none of them can be assured, multiple iterations of forward model tuning, input sampling and supervised learning might be necessary to yield the desired results. It is worthwhile noting that manual annotation is still performed in both applications. During reference data collection, visual inspection is done to identify CCTA images which exhibit no coronary motion artifacts. The forward model for synthetic lead insertion includes the manual selection of B-spline knots along the pacemaker leads. So, it should not necessarily be an *either-or* question, but a task-specific consideration of how to combine both methods, labeled data synthesis and manual selection, in a clever way.

7.2. OUTLOOK AND CONCLUSION

Visual inspection is the first step in identifying to which extent the forward model is able to approximate the real data distribution $P(\vec{x}_{\text{real}})$. In Figure 7.3, real and synthetic data is anonymized and compared. Are you as a human observer able to separate both classes? The corresponding solution which patches obtain real coronary motion artifacts or pacemaker metal shadows is provided in Figure 7.4. The training of a discriminator network for the task of real and fake data separation should be considered as well. Correctly classified patches with synthetic artifacts can subsequently be analyzed, e.g. by means of Grad-CAM [99], in regard to unexpected biases. In case of insufficient generalization to real data, synthetic data might be useful for pre-training only. Besides the option of model fine-tuning on real data, direct mixture of manually annotated and synthetic data might further increase the models accuracy on real data.



(a) Example image patches with either real or synthetic motion artifacts.



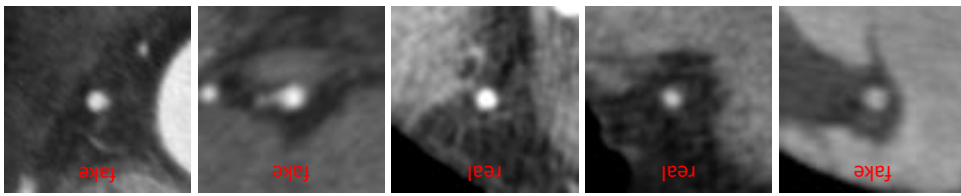
(b) Example projection patches with either real or synthetic pacemaker leads.

Figure 7.3: Is it possible to fool human observers with the data generated by the proposed forward models? Try it out! Which patches are real and which are fake?

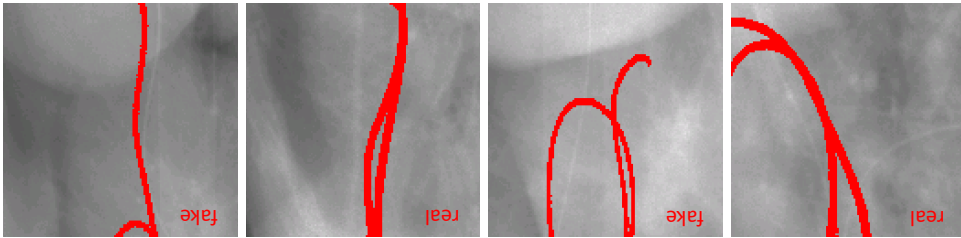
7

Further potential applications A multitude of existing model-based and DL-based approaches for risk assessment, anatomical segmentation etc. fail in the presence of imaging artifacts. Data with synthetically introduced artifacts, in general, could also be used to **evaluate the robustness** of such methods. Furthermore, forward artifact models can be integrated as **data augmentation strategy** into data-driven learning approaches in order to build invariants into the resulting predictive models. The **extraction of coronary centerlines** in CCTA is one potential application. In the first step, an existing extraction tool like the Comprehensive Cardiac Analysis Software (IntelliSpace Portal 9.0, Philips Healthcare, Cleveland, OH, USA) can be applied to extract centerlines in artifact-free reference cases. Afterwards, artifact introduction, label propagation and supervised learning is performed to train, for instance, an iterative tracking method as proposed in [119] which does not break in the presence of motion artifacts. However, one has to be aware that the label propagation is associated with label noise as motion introduction may lead to slight shifts of the centerline position.

An image-related task which suffers from missing ground truth, in general, is registration. Weak supervision and unsupervised learning are usually performed instead [43]. A forward model for motion introduction using MC-FBP could provide a solution for the **generation of ground truth data in mono-modal image registration**. However, reference cases which exhibit no motion-artifacts during 360 degrees gantry rotation are required to also take different angular reconstruction ranges into account. Forward models which create perturbed or incomplete CT projection data for subsequent DL-based **detector element recalibration, partial angle reconstruction, denoising or FOV extension** [27] are less complex. According to the motto *“it is easier to make the mess than to clean it up”*, supervised learning is applied here to solve the more difficult inverse problem of artifact removal either directly in the projection domain or via retrospective



(a) Corresponding artifact-free image patches with fake/real solution.



(b) Corresponding pacemaker metal shadow masks with fake/real solution.

Figure 7.4: The solution of the fake/real test from Figure 7.3 is rotated by 180 degrees. (a) Artifact-free patches are sampled from the reference data for fake examples or from the CoMPACT MC output for real examples. (b) The pacemaker metal shadow masks (highlighted as red overlay) are obtained by the forward model for fake examples or by the SegmentationNets of DyPAR+ for real examples. ECG leads are not masked.

correction of the reconstructed image data. Besides the simulation of imaging artifacts, **synthetic anomalies** may be inserted into clinical data. Corresponding forward models will take data of healthy patients as input and deliver image data with artificial pathologies plus accurate annotation for **anomaly detection, classification or segmentation** as output. Potential examples are, for instance, tumor or coronary plaque synthesis. The described potential applications are not limited to cardiac CT, in principle. **Other parts of the human anatomy and imaging modalities** should be considered as well.

Conclusion In this dissertation, the feasibility of learning backward models based on synthesized data is demonstrated by two clinical relevant examples of artifact reduction. Key contributions are the

- development of sufficiently complex forward models for realistic coronary motion and pacemaker metal introduction,
- adaptation of existing DL approaches for prediction of underlying artifact parameters,
- integration of the learned predictive models into end-to-end processing pipelines, e.g. for artifact detection and removal,
- assessment of the transferability to clinical practice based on data with real artifacts,
- final analysis of challenges and benefits of labeled data synthesis and identification of potential further applications.

To sum up, many steps in application of machine learning in cardiac CT imaging, or more general in medical imaging, have already been taken (see Section 3.2) and there is still a lot

of potential for further research. However, despite the recent success of DL approaches, model-based methods should not be neglected. One way to continue to benefit from them is the integration into forward models for labeled data synthesis. In this dissertation, two examples of successful implementation in cardiac CT reconstruction are presented. Software solutions for artifact removal on clinical cases with real coronary motion artifacts and pacemaker metal artifacts are developed based on synthesized data. Improved CT imaging quality and diagnostic value enhancement are achieved. Furthermore, the feasibility of coronary motion artifact recognition, quantification and metal reinsertion is demonstrated.

REFERENCES

- [1] C. K. Ahn, H. Jin, C. Heo, and J. H. Kim, “Combined low-dose simulation and deep learning for CT denoising: application of ultra-low-dose cardiac cta,” in *Medical Imaging 2019: Physics of Medical Imaging*, vol. 10948. International Society for Optics and Photonics, 2019, p. 1094846.
- [2] S. J. Al’Aref and J. K. Min, “Cardiac ct: current practice and emerging applications,” *Heart*, vol. 105(20), pp. 1597–1605, 2019.
- [3] F. Bamberg, A. Dierks, K. Nikolaou, M. F. Reiser, C. R. Becker, and T. R. Johnson, “Metal artifact reduction by dual energy computed tomography using monoenergetic extrapolation,” *European Radiology*, vol. 21, no. 7, pp. 1424–1429, 2011.
- [4] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “Patchmatch: A randomized correspondence algorithm for structural image editing,” in *ACM Transactions on Graphics (ToG)*, vol. 28, no. 3. ACM, 2009, p. 24.
- [5] J. F. Barrett and N. Keat, “Artifacts in ct: recognition and avoidance,” *Radiographics*, vol. 24, no. 6, pp. 1679–1691, 2004.
- [6] R. Bhargalia, J. D. Pack, J. V. Miller, and M. Iatrou, “Nonrigid registration-based coronary artery motion correction for cardiac computed tomography,” *Medical Physics*, vol. 39, no. 7, pp. 4245–4254, 2012.
- [7] R.-D. Bippus, T. Koehler, F. Bergner, B. Brendel, E. Hansis, and R. Proksa, “Projector and backprojector for iterative CT reconstruction with blobs using CUDA,” in *Fully 3D 2011: 11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Potsdam, Germany, 11-15 July 2011*, 2011.
- [8] Y.-L. Boureau, J. Ponce, and Y. LeCun, “A theoretical analysis of feature pooling in visual recognition,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 111–118.
- [9] O. Bousquet, S. Boucheron, and G. Lugosi, “Introduction to statistical learning theory,” in *Summer School on Machine Learning*. Springer, 2003, pp. 169–207.
- [10] B. Brendel, R. Bippus, S. Kabus, and M. Grass, “Motion compensated backprojection versus backproject-then-warp for motion compensated reconstruction,” in *The Third International Conference on Image Formation in X-ray Computed Tomography, Salt Lake City, USA*, 2014, pp. 169–172.

- [11] S. Bruns, J. M. Wolterink, R. W. van Hamersvelt, T. Leiner, and I. Išgum, “Cnn-based segmentation of the cardiac chambers and great vessels in non-contrast-enhanced cardiac CT,” *Medical Imaging with Deep Learning (MIDL), arXiv preprint arXiv:1908.07727*, 2019.
- [12] M. J. Budoff, D. Li, E. A. Kazerooni, G. S. Thomas, J. H. Mieres, and L. J. Shaw, “Diagnostic accuracy of noninvasive 64-row computed tomographic coronary angiography (CCTA) compared with myocardial perfusion imaging (MPI): the PICTURE study, a prospective multicenter trial,” *Academic Radiology*, vol. 24, no. 1, pp. 22–29, 2017.
- [13] A. Busse, D. Cantré, E. Beller, F. Streckenbach, A. Öner, H. Ince, M.-A. Weber, and F. G. Meinel, “Cardiac ct: why, when, and how,” *Der Radiologe*, doi: 10.1007/s00117-019-0530-9, 2019.
- [14] T. M. Buzug, “Computed tomography,” in *Springer Handbook of Medical Technology*. Springer, 2011, pp. 311–342.
- [15] G. C. Camargo, T. Peclat, A. C. Souza, R. d. S. L. Lima, and I. Gottlieb, “Prognostic performance of coronary computed tomography angiography in asymptomatic individuals as compared to symptomatic patients with an appropriate indication,” *Journal of Cardiovascular Computed Tomography*, vol. 11, no. 2, pp. 148–152, 2017.
- [16] C. Cano-Espinosa, G. González, G. R. Washko, M. Cazorla, and R. S. J. Estépar, “Automated agatston score computation in non-ECG gated CT scans using deep learning,” in *Medical Imaging 2018: Image Processing*, vol. 10574. International Society for Optics and Photonics, 2018, p. 105742K.
- [17] A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, “Adversarial image synthesis for unpaired multi-modal cardiac data,” in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2017, pp. 3–13.
- [18] F. Chen, Y. Li, T. Tian, F. Cao, and J. Liang, “Automatic coronary artery lumen segmentation in computed tomography angiography using paired multi-scale 3d CNN,” in *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10578. International Society for Optics and Photonics, 2018, p. 105782R.
- [19] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, “Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation,” in *Advances in Neural Information Processing Systems*, 2016, pp. 3036–3044.
- [20] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, “Metal streak artifacts in X-ray computed tomography: a simulation study,” *IEEE Transactions on Nuclear Science*, vol. 46, no. 3, pp. 691–696, 1999.
- [21] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

- [22] T. Durand, T. Mordan, N. Thome, and M. Cord, "Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 642–651.
- [23] O. Ecabert, J. Peters, H. Schramm, C. Lorenz, J. von Berg, M. J. Walker, M. Vembar, M. E. Olszewski, K. Subramanyan, G. Lavi *et al.*, "Automatic model-based segmentation of the heart in CT images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1189–1201, 2008.
- [24] T. Elss, R. Bippus, H. Schmitt, T. Ivanc, M. Morlock, and M. Grass, "Motion compensated reconstruction of the aortic valve for computed tomography," in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573. International Society for Optics and Photonics, 2018, p. 1057331.
- [25] T. Elss, H. Nickisch, T. Wissel, R. Bippus, M. Morlock, and M. Grass, "Motion estimation in coronary CT angiography images using convolutional neural networks," *Medical Imaging with Deep Learning (MIDL)*, <https://openreview.net/forum?id=HkBtaBjz>, 2018.
- [26] T. Elss, H. Nickisch, T. Wissel, H. Schmitt, M. Vembar, M. Morlock, and M. Grass, "Deep-learning-based CT motion artifact recognition in coronary arteries," in *Medical Imaging 2018: Image Processing*, vol. 10574. International Society for Optics and Photonics, 2018, p. 1057416.
- [27] É. Fournié, M. Baer-Beck, and K. Stierstorfer, "CT field of view extension using combined channels extension and deep learning methods," *Medical Imaging with Deep Learning (MIDL)*, *arXiv preprint arXiv:1908.09529*, 2019.
- [28] A. J. Foy, S. S. Dhruva, B. Peterson, J. M. Mandrola, D. J. Morgan, and R. F. Redberg, "Coronary computed tomography angiography vs functional stress testing for patients with suspected coronary artery disease: A systematic review and meta-analysis," *JAMA Internal Medicine*, vol. 177, no. 11, pp. 1623–1631, 2017.
- [29] M. Freiman, R. Manjeshwar, and L. Goshen, "Unsupervised abnormality detection through mixed structure regularization (MSR) in deep sparse autoencoders," vol. 46, no. 5. Wiley Online Library, 2019, pp. 2223–2231.
- [30] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.
- [31] O. Ghekiere, R. Salgado, N. Buls, T. Leiner, I. Mancini, P. Vanhoenacker, P. Dendale, and A. Nchimi, "Image quality in coronary CT angiography: challenges and technical solutions," *The British Journal of Radiology*, vol. 90, no. 1072, p. 20160567, 2017.
- [32] L. Gjestebj, B. De Man, Y. Jin, H. Paganetti, J. Verburg, D. Giantsoudi, and G. Wang, "Metal artifact reduction in CT: where are we after four decades?" *IEEE Access*, vol. 4, pp. 5826–5849, 2016.

- [33] L. Gjestebj, Q. Yang, Y. Xi, H. Shan, B. Claus, Y. Jin, B. De Man, and G. Wang, “Deep learning methods for CT image-domain metal artifact reduction,” in *Developments in X-ray Tomography XI*, vol. 10391. International Society for Optics and Photonics, 2017, p. 103910W.
- [34] L. Gjestebj, Q. Yang, Y. Xi, Y. Zhou, J. Zhang, and G. Wang, “Deep learning methods to guide CT image reconstruction and reduce metal artifacts,” in *Medical Imaging 2017: Physics of Medical Imaging*, vol. 10132. International Society for Optics and Photonics, 2017, p. 101322W.
- [35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, ISBN-13: 978-0262035613, 2016.
- [36] M. Grass, R. Bippus, A. Thran, D. Schäfer, S. Kabus, and K. Brown, “Motion compensation for non-gated helical ct: application to lung imaging,” in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573. International Society for Optics and Photonics, 2018, p. 105733H.
- [37] M. Grass, R. Manzke, T. Nielsen, P. Koken, R. Proksa, M. Natanzon, and G. Shechter, “Helical cardiac cone beam reconstruction using retrospective ECG gating,” *Physics in Medicine & Biology*, vol. 48, no. 18, p. 3069, 2003.
- [38] M. Grass, A. Thran, R. Bippus, S. Kabus, R. Wiemker, M. Vembar, and H. Schmitt, “Fully automatic cardiac motion compensation using vessel enhancement,” in *Abstracts of the 11th Annual Scientific Meeting of the Society of Cardiovascular Computed Tomography*. JCCT, 2016.
- [39] A. Hahn, M. Knaup, M. Brehm, S. Sauppe, and M. Kachelrieß, “Two methods for reducing moving metal artifacts in cone-beam CT,” *Medical Physics*, vol. 45, no. 8, pp. 3671–3680, 2018.
- [40] J. Hahn, H. Bruder, C. Rohkohl, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelrieß, “Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short-scan CT data,” *Medical Physics*, vol. 44, no. 11, pp. 5795–5813, 2017.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.
- [42] —, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [43] A. Hering, S. Kuckertz, S. Heldmann, and M. P. Heinrich, “Memory-efficient 2.5D convolutional transformer networks for multi-modal deformable registration with weak label supervision applied to whole-heart CT and MRI scans,” *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–12, 2019.
- [44] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

- [45] J. Hsieh, J. Londt, M. Vass, J. Li, X. Tang, and D. Okerlund, "Step-and-shoot data acquisition and reconstruction for cardiac X-ray computed tomography," *Medical Physics*, vol. 33, no. 11, pp. 4236–4248, 2006.
- [46] X. Huang, J. Wang, F. Tang, T. Zhong, and Y. Zhang, "Metal artifact reduction on cervical CT images by deep residual learning," *Biomedical Engineering Online*, vol. 17, no. 1, p. 175, 2018.
- [47] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [48] A. A. Isola, M. Grass, and W. J. Niessen, "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction," *Medical Physics*, vol. 37, no. 3, pp. 1093–1109, 2010.
- [49] T. Joyce, A. Chartsias, and S. A. Tsaftaris, "Deep multi-class segmentation without ground-truth labels," *Medical Imaging with Deep Learning (MIDL)*, <https://openreview.net/forum?id=S11Xr-3iM>, 2018.
- [50] S. Jung, S. Lee, B. Jeon, Y. Jang, and H.-J. Chang, "Deep learning based coronary artery motion artifact compensation using style-transfer synthesis in CT images," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2018, pp. 100–110.
- [51] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants." *Radiology*, vol. 164, no. 2, pp. 576–577, 1987.
- [52] H. Kervadec, J. Bouchtiba, C. Desrosiers, É. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," *arXiv preprint arXiv:1812.07032*, 2018.
- [53] T. Köhler, R. Proksa, and M. Grass, "A fast and efficient method for sequential cone-beam tomography," *Medical Physics*, vol. 28, no. 11, pp. 2318–2327, 2001.
- [54] T. Köhler, T. Klinder, U. van Stevendaal, C. Lorenz, and P. Forthmann, "Correction of breathing motion in the thorax for helical ct," *Tsinghua Science and Technology*, vol. 15, no. 1, pp. 87–95, 2010.
- [55] S. Kim, Y. Chang, and J. B. Ra, "Cardiac motion correction based on partial angle reconstructed images in X-ray CT," *Medical Physics*, vol. 42, no. 5, pp. 2560–2571, 2015.
- [56] —, "Cardiac motion correction for helical CT scan with an ordinary pitch," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1587–1596, 2018.
- [57] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2015.
- [58] P. Koken and M. Grass, "Aperture weighted cardiac reconstruction for cone-beam CT," *Physics in Medicine and Biology*, vol. 51, no. 14, p. 3433, 2006.

- [59] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.
- [60] K. Lange and J. A. Fessler, “Globally convergent algorithms for maximum a posteriori transmission tomography,” *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1430–1438, 1995.
- [61] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [62] M. Lell, M. Marwan, T. Schepis, T. Pflederer, K. Anders, T. Flohr, T. Allmendinger, W. Kalender, D. Ertel, C. Thierfelder *et al.*, “Prospectively ECG-triggered high-pitch spiral acquisition for coronary CT angiography using dual source CT: technique and initial experience,” *European Radiology*, vol. 19, no. 11, p. 2576, 2009.
- [63] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2018, pp. 2980–2988.
- [64] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, “Image inpainting for irregular holes using partial convolutions,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 85–100.
- [65] G. Liu, K. J. Shih, F. A. Wang, Ting-Chun and Reda, K. Sapra, Z. Yu, A. Tao, and B. Catanzaro, “Partial convolution based padding,” *arXiv preprint arXiv:1811.11718*, 2018.
- [66] T. Liu, P. Maurovich-Horvat, T. Mayrhofer, S. B. Puchner, M. T. Lu, K. Ghemigian, P. H. Kitslaar, A. Broersen, A. Pursnani, U. Hoffmann *et al.*, “Quantitative coronary plaque analysis predicts high-risk plaque morphology on coronary computed tomography angiography: results from the ROMICAT II trial,” *The International Journal of Cardiovascular Imaging*, pp. 1–9, 2017.
- [67] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [68] T. Lossau, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. Grass, “Motion estimation and correction in cardiac CT angiography images using convolutional neural networks,” *Computerized Medical Imaging and Graphics*, vol. 76, p. 101640, 2019.
- [69] —, “Motion artifact recognition and quantification in coronary CT angiography using convolutional neural networks,” *Medical Image Analysis*, vol. 52, pp. 68–79, 2019.

- [70] T. Lossau, H. Nickisch, T. Wissel, S. Hakmi, C. Spink, M. Morlock, and M. Grass, “Dynamic pacemaker artifact removal (DyPAR) from CT data using CNNs,” *Medical Imaging with Deep Learning (MIDL)*, <https://openreview.net/forum?id=rkx5InjA1N>, 2019.
- [71] T. Lossau, H. Nickisch, T. Wissel, M. Morlock, and M. Grass, “Learning metal artifact reduction in cardiac CT images with moving pacemakers,” *Medical Image Analysis*, vol. 61, p. 101655, 2020.
- [72] H. Ma, E. Gros, S. G. Baginski, Z. R. Laste, N. M. Kulkarni, D. Okerlund, and T. G. Schmidt, “Automated quantification and evaluation of motion artifact on coronary CT angiography images,” *Medical Physics*, vol. 45, no. 12, pp. 5494–5508, 2018.
- [73] H. Ma, E. Gros, A. Szabo, S. G. Baginski, Z. R. Laste, N. M. Kulkarni, D. Okerlund, and T. G. Schmidt, “Evaluation of motion artifact metrics for coronary CT angiography,” *Medical Physics*, vol. 45, no. 2, pp. 687–702, 2018.
- [74] G. S. Mak and Q. A. Truong, “Cardiac CT: imaging of and through cardiac devices,” *Current Cardiovascular Imaging Reports*, vol. 5, no. 5, pp. 328–336, 2012.
- [75] N. Mayer, E. Ilg, P. Fischer, C. Hazirbas, D. Cremers, A. Dosovitskiy, and T. Brox, “What makes good synthetic training data for learning disparity and optical flow estimation?” *International Journal of Computer Vision*, vol. 126, no. 9, pp. 942–960, 2018.
- [76] K. P. McGee, A. Manduca, J. P. Felmlee, S. J. Riederer, and R. L. Ehman, “Image metric-based correction (autocorrection) of motion effects: analysis of image metrics,” *Journal of Magnetic Resonance Imaging*, vol. 11, no. 2, pp. 174–181, 2000.
- [77] A. Mehranian, M. R. Ay, A. Rahmim, and H. Zaidi, “3D prior image constrained projection completion for X-ray CT metal artifact reduction,” *IEEE Transactions on Nuclear Science*, vol. 60, no. 5, pp. 3318–3332, 2013.
- [78] J. Menke, C. Unterberg-Buchwald, W. Staab, J. M. Sohns, A. S. A. Hosseini, and A. Schwarz, “Head-to-head comparison of prospectively triggered vs retrospectively gated coronary computed tomography angiography: meta-analysis of diagnostic accuracy, image quality, and radiation dose,” *American Heart Journal*, vol. 165, no. 2, pp. 154–163, 2013.
- [79] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, “Normalized metal artifact reduction (NMAR) in computed tomography,” *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, 2010.
- [80] —, “Frequency split metal artifact reduction (FSMAR) in computed tomography,” *Medical Physics*, vol. 39, no. 4, pp. 1904–1916, 2012.
- [81] M. Motwani, D. Dey, D. S. Berman, G. Germano, S. Achenbach, M. H. Al-Mallah, D. Andreini, M. J. Budoff, F. Cademartiri, T. Q. Callister *et al.*, “Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease:

- a 5-year multicentre prospective registry analysis,” *European Heart Journal*, vol. 38, no. 7, pp. 500–507, 2016.
- [82] A. Mouton, N. Megherbi, K. van Slambrouck, J. Nuyts, and T. P. Breckon, “An experimental survey of metal artefact reduction in computed tomography,” *Journal of X-ray Science and Technology*, vol. 21, no. 2, pp. 193–226, 2013.
- [83] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT Press, ISBN-13: 978-0262018029, 2012.
- [84] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.
- [85] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, “Edgeconnect: Generative image inpainting with adversarial edge learning,” *arXiv preprint arXiv:1901.00212*, 2019.
- [86] H. Nickisch, Y. Lamash, S. Prevrhal, M. Freiman, M. Vembar, L. Goshen, and H. Schmitt, “Learning patient-specific lumped models for interactive coronary blood flow simulations,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 433–441.
- [87] T. Nielsen, R. Manzke, R. Proksa, and M. Grass, “Cardiac cone-beam ct volume reconstruction using art,” *Medical Physics*, vol. 32, no. 4, pp. 851–860, 2005.
- [88] J. M. Noothout, B. D. de Vos, J. M. Wolterink, T. Leiner, and I. Išgum, “CNN-based landmark detection in cardiac CTA scans,” *arXiv preprint arXiv:1804.04963*, 2018.
- [89] M. Oehler and T. M. Buzug, “The λ -mlem algorithm: An iterative reconstruction technique for metal artifact reduction in CT images,” in *Advances in Medical Engineering*. Springer, 2007, pp. 42–47.
- [90] H. S. Park, S. M. Lee, H. P. Kim, J. K. Seo, and Y. E. Chung, “CT sinogram-consistency learning for metal-induced beam hardening correction,” *Medical Physics*, vol. 45, no. 12, pp. 5376–5384, 2018.
- [91] M. Perez-Ortiz, P. Tino, R. Mantiuk, and C. Hervas-Martinez, “Exploiting synthetically generated data with semi-supervised learning for small and imbalanced datasets,” *arXiv preprint arXiv:1903.10022*, 2019.
- [92] M. Petersilka, H. Bruder, B. Krauss, K. Stierstorfer, and T. G. Flohr, “Technical principles of dual source CT,” *European Journal of Radiology*, vol. 68, no. 3, pp. 362–368, 2008.
- [93] C. Rohkohl, H. Bruder, K. Stierstorfer, and T. Flohr, “Improving best-phase image quality in cardiac CT by motion correction with MAM optimization,” *Medical Physics*, vol. 40, no. 3, 2013.

- [94] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015, pp. 234–241.
- [95] S. Roux, L. Desbat, A. Koenig, and P. Grangeat, "Exact reconstruction in 2D dynamic CT: compensation of time-dependent affine deformations," *Physics in Medicine & Biology*, vol. 49, no. 11, p. 2169, 2004.
- [96] D. Schäfer, J. Borgert, V. Rasche, and M. Grass, "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 7, pp. 898–906, 2006.
- [97] U. J. Schoepf, *CT of the Heart*. Humana Press, ISBN 978-1-59259-818-2, 2018.
- [98] H. Schöndube, T. Allmendinger, K. Stierstorfer, H. Bruder, and T. Flohr, "Evaluation of a novel CT image reconstruction algorithm with enhanced temporal resolution," in *Medical Imaging 2011: Physics of Medical Imaging*, vol. 7961. International Society for Optics and Photonics, 2011, p. 79611N.
- [99] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [100] Y. Shen, Z. Fang, Y. Gao, N. Xiong, C. Zhong, and X. Tang, "Coronary arteries segmentation based on 3D FCN with attention gate and level set function," *IEEE Access*, vol. 7, pp. 42 826–42 835, 2019.
- [101] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM National Conference*. ACM, 1968, pp. 517–524.
- [102] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, *arXiv preprint arXiv:1409.1556*, 2014.
- [103] G. Singh, S. J. Al'Aref, M. Van Assen, T. S. Kim, A. van Rosendael, K. K. Kolli, A. Dwivedi, G. Maliakal, M. Pandey, J. Wang *et al.*, "Machine learning in cardiac CT: basic concepts and contemporary data," *Journal of Cardiovascular Computed Tomography*, vol. 12, no. 3, pp. 192–201, 2018.
- [104] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [105] K. Taguchi and H. Kudo, "Motion compensated fan-beam reconstruction for non-rigid transformation," *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 907–917, 2008.
- [106] Q. Tang, J. Cammin, S. Srivastava, and K. Taguchi, "A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT," *Medical Physics*, vol. 39, no. 7Part1, pp. 4291–4305, 2012.

- [107] J. Toftegaard, W. Fledelius, D. Seghers, M. Huber, M. Brehm, E. S. Worm, U. V. Elstrøm, and P. R. Poulsen, "Moving metal artifact reduction in cone-beam CT scans with implanted cylindrical gold markers," *Medical Physics*, vol. 41, no. 12, p. 121710, 2014.
- [108] K. van Slambrouck and J. Nuyts, "Metal artifact reduction in computed tomography using local models in an image block-iterative scheme," *Medical Physics*, vol. 39, no. 11, pp. 7080–7093, 2012.
- [109] U. van Stevendaal, J. von Berg, C. Lorenz, and M. Grass, "A motion-compensated scheme for helical cone-beam reconstruction in cardiac CT angiography," *Medical Physics*, vol. 35, no. 7, pp. 3239–3251, 2008.
- [110] U. van Stevendaal, P. Koken, P. G. Begemann, R. Koester, G. Adam, and M. Grass, "ECG gated circular cone-beam multi-cycle short-scan reconstruction algorithm," in *Medical Imaging 2007: Physics of Medical Imaging*, vol. 6510. International Society for Optics and Photonics, 2007, p. 65105P.
- [111] M. Vembar, M. Garcia, D. Heuscher, R. Haberl, D. Matthews, G. Böhme, and N. Greenberg, "A dynamic approach to identifying desired physiological phases for cardiac imaging using multislice spiral CT," *Medical Physics*, vol. 30, no. 7, pp. 1683–1693, 2003.
- [112] J. Šprem, B. D. de Vos, P. A. de Jong, M. A. Viergever, and I. Išgum, "Classification of coronary artery calcifications according to motion artifacts in chest CT using a convolutional neural network," in *Medical Imaging 2018: Image Processing*, M. A. Styner and E. D. Angelini, Eds. International Society for Optics and Photonics, 2017, p. 101330R.
- [113] G. Wang, D. L. Snyder, J. A. O'Sullivan, and M. W. Vannier, "Iterative deblurring for CT metal artifact reduction," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 657–664, 1996.
- [114] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, and B. Catanzaro, "Video-to-video synthesis," *arXiv preprint arXiv:1808.06601*, 2018.
- [115] Y. Wang, E. Vidan, and G. W. Bergman, "Cardiac motion of coronary arteries: variability in the rest period and implications for coronary MR angiography," *Radiology*, vol. 213, no. 3, pp. 751–758, 1999.
- [116] R. Wiemker, T. Klinder, M. Bergtholdt, K. Meetz, I. C. Carlsen, and T. Bulow, "A radial structure tensor and its use for shape-encoding medical visualization of tubular and nodular structures," *IEEE Transactions on visualization and computer graphics*, vol. 19, no. 3, pp. 353–366, 2013.
- [117] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.

- [118] J. M. Wolterink, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever, and I. Išgum, “Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks,” *Medical Image Analysis*, vol. 34, pp. 123–136, 2016.
- [119] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum, “Coronary artery centerline extraction in cardiac CT angiography using a cnn-based orientation classifier,” *Medical Image Analysis*, vol. 51, pp. 46–60, 2019.
- [120] S. Xu and H. Dang, “Deep residual learning enabled metal artifact reduction in ct,” in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573. International Society for Optics and Photonics, 2018, p. 105733O.
- [121] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Free-form image inpainting with gated convolution,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 4471–4480.
- [122] M. D. Zeiler, “Adadelata: an adaptive learning rate method,” *arXiv preprint arXiv:1212.5701*, 2012.
- [123] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *European conference on computer vision*. Springer, 2014, pp. 818–833.
- [124] X. Zhang, J. Wang, and L. Xing, “Metal artifact reduction in X-ray computed tomography (CT) by constrained optimization,” *Medical Physics*, vol. 38, no. 2, pp. 701–711, 2011.
- [125] Y. Zhang and H. Yu, “Convolutional neural network based metal artifact reduction in X-ray computed tomography,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1370–1381, 2018.
- [126] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [127] M. Zreik, N. Lessmann, R. W. van Hamersvelt, J. M. Wolterink, M. Voskuil, M. A. Viergever, T. Leiner, and I. Išgum, “Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis,” *Medical Image Analysis*, vol. 44, pp. 72–85, 2018.

LIST OF ABBREVIATIONS
LIST OF SYMBOLS
LIST OF FIGURES
LIST OF TABLES
LIST OF PUBLICATIONS
ACKNOWLEDGEMENTS
CURRICULUM VITAE

LIST OF ABBREVIATIONS

ACM	All-cause Mortality
AI	Artificial Intelligence
ANN	Artificial Neural Network
ART	Algebraic Reconstruction Technique
AWCR	Aperture Weighted Cardiac Reconstruction
AWR	Aperture Weighted Reconstruction
BHC	Beam Hardening Correction
BP	Backprojection
BPW	Backproject-than-warp
CAC	Coronary Artery Calcium
CAD	Coronary Artery Disease
CAE	Convolutional Autoencoder
CCT	Cardiac Computed Tomography
CCTA	Coronary Computed Tomography Angiography
CNN	Convolutional Neural Network
CoMoFACT	Coronary Motion Forward Artifact model for CT data
CoMPACT	Coronary Motion Estimation by Patch Analysis in CT data
CSCT	Calcium Scoring Computed Tomography
CT	Computed Tomography
DFT	Discrete Fourier Transform
DL	Deep Learning
DLB	Deep-learning-based
DyPAR	Dynamic Pacemaker Artifact Removal
ECG	Electrocardiogram
EM	Expectation Maximization
ERM	Empirical Risk Minimization
FBP	Filtered Back-projection
FC	Fully Connected
FCN	Fully Convolutional Network
FFR	Fractional Flow Reserve
FFT	Fast Fourier Transform
FOV	Field of View
FT	Fine-tuning
GAN	Generativ Adversarial Network
GPU	Graphical Processing Unit
HU	Hounsfield Units
IDW	Inverse Distance Weighting
LAD	Left Anterior Descending

LCX	L eft C ircum f lex artery
LI	L inear I nterpolation
LV	L eft V entricle
MAE	M ean A bsolute E rror
MAM	M otion A rtifact M easure
MAR	M etal A rtifact R eduction
MC	M otion C ompensation
MC-FBP	M otion C ompensated F iltered B ack- p rojection
MC-IR	M otion C ompensated I terative R econstruction
MCR	M otion C ompensated R econstruction
ME	M otion E stimation
ML	M aximum L ikelihood
MLP	M ultilayer P erceptron
MMAR	M oving M etal A rtifact R eduction
MR	M agnetic R esonance
MV	M otion V ector
MVF	M otion V ector F ield
NCCT	N on-contrast C omputed T omography
NGF	N ormalized G radient F ield
nMAE	N ormalized M ean A bsolute E rror
NMAR	N ormalized M etal A rtifact R eduction
NNLS	N on- n egative L east- s quares
PAR	P artial A ngle R econstruction
pFOV	P artial F ield of V iew
RCA	R ight C oronary A rtery
ReLU	R ectified L inear U nit
RMSE	R oot M ean S quared E rror
SGD	S tochastic G radient D escent
SNR	S ignal N oise R atio
SVM	S upport V ector M achine
TRIM	T emporal R esolution I mprovement M ethod
VNC	V irtual N on-contrast

LIST OF SYMBOLS

α	angle between the mean reconstruction direction and the target motion direction (control parameter of the CoMoFACT)
β	cone-angle of an X-ray beam
$\gamma(t)$	gantry rotation angle at time point t
γ_j	gantry rotation angle associated with projection view P_j
γ_{mean}	gantry rotation angle at the reference heart phase r
Γ	angular scan range required for a motion compensation algorithm
$\vec{\delta}_{\vec{c}}$	defines motion trajectory in the CoMoFACT
$\vec{\delta}_i$	sample vectors for piecewise linear motion
ϵ	learning rate for iterative neural network optimization
η	thickness of the irradiated object
θ	specific parameter setting of an input-output mapping function
$\theta_{\vec{z}}$	optimal parameter setting in empirical risk minimization
Θ	parameter space
Θ_{τ}	bounding volume with radius τ
$\Lambda(\mathcal{C})$	subset of projection views acquired during the cardiac cycle \mathcal{C}
$\mu(\vec{v})$	attenuation coefficient at voxel position \vec{v}
\vec{v}	voxel position within the reconstruction field of view
$\Pi(t)$	set of pi-partner acquisition times associated with t
\vec{p}	defines motion direction in CoMoFACT
\vec{p}_i	sample points for piecewise linear motion
\vec{p}_{orth}	orthogonal motion direction (cross product of the normal vector $\vec{n}_{\vec{c}}$ and the mean reconstruction direction)
σ	standard deviation of the Gaussian kernel
τ	bounding volume radius
ϕ	non-linear activation function
$\Phi(\vec{x}_{\text{clean}})$	forward model for labeled data synthesis
Ω	reconstruction field of view
$\Omega(\vec{c})$	restricted field of view for sub-volume reconstruction
Ω_c	partial field of view (pFOV) indicating angular completeness
A_i	partial angle volume (input of the ModelingNets)
b	number of channels included in the input data for supervised learning
B	back-projection matrix
\mathcal{B}	back-projection operator
c	center projection index for partial angle volume calculation
\vec{c}	centerline point within the set C
C	set of centerline points
\mathcal{C}	cardiac cycle

$\vec{d}_{\vec{c}}$	synthetic motion vector field of the CoMoFACT
$\hat{d}_{\vec{c}}$	predicted motion vector at centerline point \vec{c}
\hat{d}	predicted dense motion vector field
$\vec{D}(t, u, v)$	position of detector element (u, v) at time point t
$\epsilon_{x,y}$	euclidean error of the motion vector estimation
ϵ_o	component-wise absolute error of the motion vector estimation
ϵ_α	angle error of the motion vector estimation
$E(f)$	generalization error of function f
$E_{\vec{z}}(f)$	empirical risk of function f
f	input-output mapping function
f_θ	parametric input-output mapping function
F_{1D}	1D discrete Fourier transform matrix
\mathcal{F}_{1D}	1D Fourier transform operator
g	band limited ramp filter
G	ramp filter matrix
h	number of slices included in the input data for supervised learning
H	high-pass filtering matrix
\mathcal{H}	hypothesis space
HR_{mean}	mean heart rate
i	general index
$I_{\Omega}^{\vec{c},s}$	output image volume of CoMoFACT for artifact recognition and quantification
$I_{100}^{\vec{c},s}$	image patch sampled from $I_{\Omega}^{\vec{c},s}$
$I_{60}^{\vec{c},s}$	image patch cropped from $I_{100}^{\vec{c},s}$
$I_{\Omega}^{\vec{c},s,\alpha}$	output image volume of CoMoFACT for motion estimation
$I_{2.5D}^{\vec{c},s,\alpha}$	sampled patch consisting of three orthogonal image slices
$I_{\text{metal}}(t)$	time-dependent metal mask in the image domain
I_D	radiation intensity at the detector
I_S	radiation intensity at the X-ray source
j	general index
k	iteration index of the CoMPACT MC pipeline
k_{max}	total number of CoMPACT MC iteration steps
L	loss function for supervised learning
L_{leaky}	piecewise L1 loss function
L_{inpaint}	inpainting loss function
m	number of data pairs for supervised learning
m_{batch}	minibatch size during stochastic gradient descent
$m_{\vec{c}}$	spatial weighting function of the CoMoFACT
m_C	spatial weighting function of the CoMPACT MC pipeline
M	binary mask of valid line integrals
n	hyper-parameter determining the plane size $n \times n$ of an input patch for supervised learning
$\vec{n}_{\vec{c}}$	normal vector associated with centerline point \vec{c}
N	number of layers of an artificial neural network

N^{\boxplus}	number of line integrals in a training patch
NN_{class}	artifact classification network
NN_{regr}	artifact level regression network
NN_{inpaint}	sinogram inpainting network
NN_{model}	pacemaker modeling network
o	hyper-parameter determining the number of feature maps at each network layer
p	original raw projection data
p_{wedge}	rebinned projection data
p_{fit}	high-pass filtered projection data
P_j	projection view at index j
P_{org}	original metal-free projection data
P_{mask}	binary mask of metal-affected line integrals
P_{metal}	synthetic metal shadow
P_{input}	projection data with synthetic leads
P_{segm}	predicted binary mask of metal-affected line integrals
P_{clean}	inpainted projection data
P	probability function
P_{clean}	distribution function of \vec{x}_{clean}
P_{syn}	distribution function of \vec{x}_{syn}
P_{real}	distribution function of \vec{x}_{real}
q	3D Gaussian kernel
r	reference heart phase for AWCR
R	Radon transform matrix
\mathcal{R}	Radon transform operator
s	target motion strength (control parameter of the CoMoFACT)
$S_{x/y}$	Sobel convolution kernels for vertical and horizontal derivative approximation
S_c	stack of partial angle volumes (input of the ModelingNets)
$\vec{S}(t)$	position of the X-ray source at time point t
t	acquisition time point
t_j	time point associated with projection view P_j
t_{cc}	time in percent cardiac cycle
t_{start}	scan start time
t_{end}	scan end time
T	threshold of the pixel-wise or voxel-wise classification tasks
u	index of the detector row
v	index of the detector column
w_{cos}	cosine weighting function
w_c	cardiac weighting function
w_{ap}	aperture weighting function
w_{AWCR}	aperture and cardiac weighting function with pi-partner normalization
$W^{(l)}$	weight matrix at layer l of an artificial neural network
x	Cartesian coordinate
x_d	displacement in x-direction (target value for CoMPACT)

$\vec{x} \in X$	independent random variable (input for supervised learning)
\vec{x}_i	drawn data samples
\vec{x}_{clean}	clinical reference data without artifacts
\vec{x}_{syn}	clinical data with synthetic artifacts
\vec{x}_{real}	clinical data with real artifacts
$\vec{x}^{(i)}$	output feature map of layer i of an artificial neural network
y	Cartesian coordinate
y_d	displacement in y-direction (target value for CoMPACT)
y_{class}	artifact classification target
y_{regr}	artifact level regression target
$\vec{y} \in Y$	independent random variable (target output for supervised learning)
\vec{y}_i	drawn data samples associated with \vec{x}_i
\vec{y}_{syn}	known underlying artifact parameters of \vec{x}_{syn}
\vec{y}_{real}	unknown underlying artifact parameters of \vec{x}_{real}
\hat{y}	predicted output
z	Cartesian coordinate
\vec{z}_i	data pair for supervised learning

LIST OF FIGURES

A symbol at the end of the caption indicates that the corresponding Figure is taken from published literature. (Legend: † [25]; ◇ [69]; ‡ [68]; △ [71])

1.1	Examples of coronary motion and pacemaker metal artifacts	2
1.2	Manual annotation vs. labeled data synthesis	3
1.3	Flowchart of the proposed methodology to build knowledge-driven forward artifact models to learn AI-enabled backward models	4
1.4	Flowchart depicting the structure of the work	5
2.1	Schematic drawings of the 3D cone-beam projection geometry	11
2.2	Scan information plot of a clinical case with retrospectively ECG-gated spiral acquisition	12
2.3	Scan information plot of a clinical case with prospectively ECG-triggered sequential acquisition	12
2.4	Schematic drawing of voxel-driven back-projection without and with motion compensation ◇	16
2.5	Comparison of coronary motion artifacts in step-and-shoot and helical data ◇	17
2.6	Flowcharts of sinogram completion without and with motion perturbation	21
2.7	Mixture of motion states during reconstruction	22
3.1	Overview on the different categories of artificial intelligence	26
3.2	Relation between hypothesis space complexity and the problems of under- or overfitting	28
3.3	Exemplary visualization of a convolutional neural network	29
3.4	Statistics of publications listed in Web of Science Core Collection database	31
3.5	Procedure of labeled data synthesis	34
4.1	Introducing motion by means of the MC-FBP algorithm	38
4.2	Graphical abstract for the recognition and quantification of coronary motion artifacts ◇	40
4.3	Schematic drawing of the time-dependent motion trajectories ◇	41
4.4	Example image patches generated by the CoMoFACT ◇	43
4.5	The relation of artifact appearance and motion direction investigated on a phantom vessel ◇	44
4.6	Schematic drawing of the network architecture ◇	45
4.7	Confusion matrix of the regression network for multi-slice inputs ◇	47
4.8	Relative artifact measurements on clinical test data with real artifacts ◇	48

4.9	Absolute artifact measurements on clinical data with synthetic artifacts	◇	50
4.10	Absolute artifact measurements on clinical test data with real artifacts	◇	51
5.1	Paired data generated by the CoMoFACT enables two options for supervised learning		56
5.2	Different types of constant linear motion introduced to a phantom vessel tree using CoMoFACT	‡	58
5.3	Example centerline point in phantom and clinical mode for varying parameter settings	‡	61
5.4	Schematic drawing of the network architecture	‡	62
5.5	Motion vector extrapolation	‡	63
5.6	Bar plot of mean and median angle error evaluated for subsets determined by the selected s ranges.		65
5.7	Confusion matrix of the target motion strength	‡	65
5.8	Local CoMPACT MC on clinical cases with real artifacts	‡	66
5.9	Global CoMPACT MC on clinical cases with real artifacts	‡	67
5.10	Cross-sectional image patches before and after MC are rated by four human observers in a five point Likert scale (from 1: excellent to 5: non-diagnostic).		68
5.11	Schematic drawing of the cascaded MC pipeline		70
5.12	DL-based motion artifact measurements on clinical validation data	†	70
6.1	Flowcharts of sinogram completion with motion perturbation and purely raw-data-based MAR		74
6.2	Forward model for synthetic lead insertion	△	76
6.3	Schematic drawing of the DyPAR+ pipeline	△	78
6.4	Schematic drawing of the shared network architecture	△	79
6.5	Partial angle reconstructions	△	82
6.6	Comparison of different inpainting approaches	△	85
6.7	Example outputs of the SegmentationNets and InpaintingNets	△	86
6.8	Example outputs of the ModelingNets	△	86
6.9	Comparison of a 2nd pass MAR, DyPAR and DyPAR+	△	87
6.10	Coronary lumen without and with DyPAR+		89
7.1	Graphical abstract of the pursued methodology		92
7.2	Examples of good and bad generalizations of predictive models trained on synthetic data		93
7.3	Comparison of real and synthetic data		96
7.4	Solution which patches obtain real or synthetic data		97

LIST OF TABLES

A symbol at the end of the caption indicates that the corresponding Table is taken from published literature. (Legend: \diamond [69]; \ddagger [68]; \triangle [71])

2.1	Weighting functions in aperture weighted cardiac reconstruction.	15
2.2	Literature survey in the research field of coronary motion artifacts in CCTA images \ddagger	18
2.3	Literature survey in the research field of CT metal artifact reduction \triangle	20
3.1	Representative selection of publications in the research field of machine learning in cardiac CT.	32
4.1	Test results of the classification and the regression networks \diamond	47
5.1	Test results of the motion estimation network \ddagger	64
5.2	Duration measurements of CoMPACT MC \ddagger	68
6.1	Comparison of clinical test database and synthetic learning database \triangle	75
6.2	Test results on the synthetic learning data \triangle	84

LIST OF PUBLICATIONS

JOURNAL PAPERS

- [A] **T. Lossau**, H. Nickisch, T. Wissel, M. Morlock, and M. Grass, “Learning Metal Artifact Reduction in Cardiac CT Images with Moving Pacemakers” submitted to *Medical Image Analysis, Special Issue on Medical Imaging with Deep Learning*, 2019.
- [B] **T. Lossau**, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. Grass, “Motion Estimation and Correction in Cardiac CT Angiography Images using Convolutional Neural Networks”, *Computerized Medical Imaging and Graphics*, vol 76, p. 1016140, 2019.
- [C] **T. Lossau**, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. Grass, “Motion Artifact Recognition and Quantification in Coronary CT Angiography using Convolutional Neural Networks”, *Medical Image Analysis*, vol. 52, pp. 68-79, 2019.

CONFERENCE PAPERS

- [a] **T. Lossau**, H. Nickisch, T. Wissel, S. Hakmi, C. Spink, M. Morlock, and M. Grass, “Dynamic Pacemaker Artifact Removal (DyPAR) from CT Scans using CNNs ”, *2nd Medical Imaging with Deep Learning (MIDL) Conference*, London, United Kingdom, 2019. (Oral; Invitation to Medical Image Analysis Special Issue; Top-5 Shortlist for the Best Paper Award)
- [b] **T. Elss**, H. Nickisch, T. Wissel, R. Bippus, M. Morlock, and M. Grass “Motion Estimation in Coronary CT Angiography Images using Convolutional Neural Networks”, *1st Medical Imaging with Deep Learning (MIDL) Conference*, Amsterdam, Netherlands, 2018. (Poster; Longlist for the Medical Image Analysis Special Issue)
- [c] **T. Elss**, H. Nickisch, T. Wissel, H. Schmitt, M. Vembar, M. Morlock, and M. Grass “Deep-Learning-Based CT Motion Artifact Recognition in Coronary Arteries ”, *Medical Imaging 2018: Image Processing*, vol. 10574, p. 1057416, International Society for Optics and Photonics, 2018. (Oral)
- [d] **T. Elss**, R. Bippus, H. Schmitt, T. Ivanc, M. Morlock and M. Grass “Motion Compensated Reconstruction of the Aortic Valve for Computed Tomography ”, *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573, p. 1057331, International Society for Optics and Photonics, 2018. (Poster; Best Paper Runner-Up)

CONFERENCE ABSTRACTS

- [α] **T. Lossau**, H. Nickisch, T. Wissel, S. Hakmi, A. Saalbach, C. Spink, M. Morlock, and M. Grass “Deep-learning-based reduction of moving CT metal artifacts”, *European Congress of Radiology (ECR)*, Vienna, Austria, 2020. (Oral)
- [β] **T. Lossau**, M. Vembar, H. Nickisch, T. Wissel, R. Bippus, M. Morlock, and M. Grass “Deep-Learning-Based Motion Artifact Measures for Coronary CT Angiography Images”, *European Congress of Radiology (ECR)*, Vienna, Austria, 2019. (E-Poster)
- [γ] C. Spink, **T. Elss**, R. Bippus, T. Ivanc, M. Morlock, M. Grass, and G. Lund “Motion compensated aortic valve CT imaging”, *Annual Meeting of the Cardiovascular and Interventional Radiological Society of Europe (CIRSE)*, Lisbon, Portugal, 2018. (Poster)
- [δ] **T. Elss**, C. Spink, R. Bippus, M. Morlock, and M. Grass “Motion compensated reconstruction of the aortic valve for non-gated CT scans”, *104th Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, Illinois, United States, 2018. (Oral)
- [ϵ] **T. Elss**, C. Spink, R. Bippus, T. Ivanc, G. Lund, M. Morlock, and M. Grass “Motion compensated CT imaging of the aortic valve”, *103rd Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, Illinois, United States, 2017. (Oral)

INVENTION DISCLOSURES

- [1] **T. Lossau**, H. Schmitt, and M. Grass “Learned Motion Estimation and Compensation in CT Images Based on Artifact Appearance Analysis”, 2019. (submitted)
- [2] **T. Lossau**, M. Grass, T. Wissel, and H. Nickisch “On-site Learning for Metal Artifact Reduction in Cardiac CT”, 2019. (submitted)
- [3] H. Schmitt, M. Grass, **T. Lossau**, and M. Vembar “AI-enabled cardiac CT rest phase determination”, 2018. (patent application filed)
- [4] **T. Lossau**, H. Schmitt, and M. Grass “Patch-based motion estimation in CCTA images”, 2018. (archived)
- [5] H. Schmitt, M. Grass, **T. Elss**, and M. Vembar “Method for optimized viewing order of arteries in coronary CTA”, 2017. (archived)
- [6] **T. Elss**, M. Grass, and T. Ivanc “Triggered Pre-TAVR CT imaging protocol for cardiac phase regulation”, 2017. (archived)
- [7] **T. Elss**, M. Grass, R. Bippus, and A. Thran “Motion compensated cardiac valve reconstruction”, 2017. (patent application published)

ACKNOWLEDGMENTS

In Section 3.1 of this dissertation, the terms supervised, unsupervised, weakly-supervised and self-supervised have been introduced in the context of statistical learning. However, multiple discussions with PhD students from different sites formed another interpretation describing the amount of support during the exiting time of PhD. I count myself among the lucky ones with a great mixture of supervision and self-supervision. Thus, I would like to acknowledge the people who encouraged, guided and inspired me while giving me the opportunity to find my own path to solutions.

First of all, I would like to thank my doctoral father Prof. Michael Morlock for his academic supervision and constant availability in spite of his busy schedule and the large (but lovely) group at the Institute of Biomechanics at the Hamburg University of Technology.

The thesis was enabled by and carried out at the Philips Research Laboratories in Hamburg. Therefore, I would especially like to thank my Philips supervisors Michael Grass, Hannes Nickisch and Tobias Wissel for their various and valuable ideas, feedback and discussions at our bi-weekly meetings and beyond. Michael Grass made this PhD possible and had the initial idea of corrupting artifact-free images for data generation which turned out to be the central theme of this work. I am grateful that he always pushed me to the right limit and let me benefit from his deep expertise in CT imaging. I would like to thank Hannes Nickisch for his eye for the big picture and for giving me important tips in paper writing, oral/poster presentation and, in general, nice research practice. In particular, he contributed to this work by his comprehensive knowledge in computer vision and his strength of thinking out of the box. I thank Tobias Wissel for his eye for details and for providing the most careful reviews of conference/journal papers and this dissertation. Furthermore, he shared diverse ML-related source code with me and helped me to optimize my implementations. Thank you for a lot of simple ideas with great impact.

Of course, multiple colleagues at Philips supported me during the last three years and I am very happy to also stay in this experienced and cheerful research group in the future. I would like to thank Rolf Bippus and Frank Bergner who helped me with all kinds of questions about CT image reconstruction and provided and introduced corresponding software to me. I thank Holger Schmitt who made me familiar with the IntelliSpace Portal and the coronary artery segmentation. And special thanks to Malte Steinhoff for the great times in our office and for several exciting kicker matches.

I am grateful for the backing of the people from my private life:

- my best friends Alessa Hering and Ramona Bei der Kellen - thank you for making studies and conferences such a joy and for all the evenings with oven cheese and wine,
- my mother Heike Elß - thank you that I can always count on you if necessary and especially for the last minute printing of this thesis,
- my husband Stefan Loßau - thank you for your love, patience and craziness. *Du bist mein Zuhause.*

CURRICULUM VITAE

Name Tanja Loßau, nee Elß
Date of birth 21 September 1992
Place of birth Lübeck, Germany

UNIVERSITY EDUCATION

Hamburg University of Technology since 2016/12 **PhD (intended doctorate: Dr. rer. nat.)**
Doctoral thesis: *Machine learning in cardiac CT image reconstruction*
(in cooperation with Philips Research Hamburg)

University of Lübeck 2014/10 – 2016/10 **M.Sc., Computational Life Science (Grade: 1.0)**
Master thesis: *Applying convolutional neural networks for structure detection in medical imaging* (in cooperation with Pattern Recognition Company; Grade: 1.0)
Particular contents: Machine Learning, Imaging Systems, Medical Image Processing
Award: Best Master Award

University of Lübeck 2011/09 – 2014/10 **B.Sc., Computational Life Science (Grade: 1.1)**
Bachelor thesis: *Numerics for differential-algebraic equations* (Grade: 1.0)
Award: Philips Best Bachelor Award

WORK EXPERIENCE

Philips Research Hamburg since 2016/12 **PhD student**
Goal: Detection and reduction of coronary motion and pacemaker metal artifacts
Methods: Combined application of existing handcrafted algorithms for synthesis of labeled data using Python, supervised learning using CNTK

Pattern Recognition Company 2016/04 – 2016/12 **Master student and development engineer**
Goal: Image-based structure detection in tissue sections and X-ray images
Methods: Data preparation using Python, supervised learning using CAFFE and DIGITS, network application and evaluation using C++

Philips Research Hamburg 2015/11 – 2016/02 **Intern**
Goal: Motion compensated reduction of metal artifacts in CT data with pacemakers
Methods: Combined application of filters, morphological operators, forward projection and (partial) reconstruction using C++, Python and Shell

University of Lübeck 2013/03 – 2013/08 **Research assistant**
Evaluation of EEG data for early detection of epileptic seizures using EEGLAB

University of Lübeck 2012/10 – 2016/08 **Student assistant**
Tutor in Numerical Methods for Image Computing, Optimisation, Numerical Analysis I and Analysis I & II

SCHOOL EDUCATION

Thomas-Mann-Schule 2002/09 – 2011/05 **Abitur (Grade: 1.2)**
Profile: Sciences
Award: DPG Book Award 2011

lütting 2009/08 – 2010/07 **Mobile Monitoring Project (in cooperation with Dräger Medical Lübeck)**
Evaluation and graphical representation of medical sensor data using Java