

Dissertation

submitted to the
Combined Faculties for the Natural Sciences
and for Mathematics
of the Ruprecht Karl University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

M.Sc. Tim Vöth

Born in: Dachau, Germany

Oral examination: 26.07.2023

**3D Reconstruction of Interventional Material from Very
Few X-Ray Projections for Interventional Image Guidance**

Referees: Prof. Dr. Peter Bachert
Prof. Dr. Marc Kachelrieß

3D Reconstruction of Interventional Material from Very Few X-Ray Projections for Interventional Image Guidance

Today, minimally invasive endovascular interventions are usually guided by 2D fluoroscopy, i.e. a live 2D X-ray image. However, 3D fluoroscopy, i.e. a live 3D image reconstructed from a stream of 2D X-ray images, could improve spatial awareness. 3D fluoroscopy is, however, not used today, since no appropriate 3D reconstruction algorithm is known. Existing algorithms for the real-time reconstruction of interventional material (guidewires, stents, catheters, etc.) are either only capable of reconstructing a single guidewire or catheter, or use too many X-ray images and therefore too much dose per 3D reconstruction. The goal of this thesis was to reconstruct complex arrangements of interventional material from as few X-ray images as possible. To this end, a previously proposed algorithm for the reconstruction of interventional material from four X-ray images was adapted. Five key improvements allowed to reduce the number of X-ray images per 3D reconstruction from four to two: a) use of temporal information in a rotating imaging setup, b) separate reconstruction of different types of interventional material enabled by the computation of semantic interventional material extraction images, c) compensation of stent motion by spatial transformer networks, d) per-projection backprojection and e) binarization of the guidewire extraction images. While previously only single curves could be reconstructed from two newly acquired X-ray images, the proposed pipeline can reconstruct stents and even stent-guidewire combinations. Submillimeter reconstruction accuracy was demonstrated on measured X-ray images of interventional material inside an anthropomorphic phantom with simulated respiratory motion. Measurements of the dose area product rate of the proposed 3D reconstruction pipeline indicate a dose burden roughly similar to that of 2D fluoroscopy.

3D Rekonstruktion von Interventionsmaterial aus sehr wenigen Röntgenprojektionen für die Interventionsbildführung

Heutzutage werden minimalinvasive endovaskuläre Eingriffe in der Regel durch 2D-Fluoroskopie, d.h. ein Live-2D-Röntgenbild, geführt. Jedoch könnte 3D-Fluoroskopie, d.h. ein Live-3D-Bild, das aus einer Serie von 2D-Röntgenbildern rekonstruiert wird, die räumliche Vorstellung verbessern. Mangels geeigneter 3D-Rekonstruktionsalgorithmen wird 3D-Fluoroskopie heute jedoch nicht eingesetzt. Die bisher veröffentlichten Algorithmen für die Echtzeit-Rekonstruktion von Interventionsmaterial (Führungsdrähte, Stents, Katheter usw.) können entweder nur einen einzelnen Führungsdraht oder Katheter rekonstruieren, oder sie verwenden zu viele Röntgenbilder und damit zu viel Dosis pro 3D-Rekonstruktion. Das Ziel dieser Arbeit war es, komplexe Anordnungen von Interventionsmaterial aus möglichst wenigen Röntgenbildern zu rekonstruieren. Hierzu wurde ein Algorithmus für die Rekonstruktion von Interventionsmaterial aus vier Röntgenbildern angepasst. Durch fünf wesentliche Verbesserungen konnte die Anzahl der Röntgenbilder pro 3D-Rekonstruktion von vier auf zwei reduziert werden: a) Nutzung zeitlicher Informationen in einem rotierenden Bildgebungssetup, b) getrennte Rekonstruktion verschiedener Arten von Interventionsmaterial durch die Berechnung semantischer Interventionsmaterial-Extraktionsbilder, c) Kompensation von Stentbewegungen durch Spatial Transformer Networks, d) projektionsweise Rückprojektion und e) Binarisierung der Führungsdraht-Extraktionsbilder. Während bisher nur einzelne Kurven aus zwei neu aufgenommenen Röntgenbildern rekonstruiert werden konnten, kann die vorgeschlagene Pipeline Stents und sogar Stent-Führungsdraht-Kombinationen rekonstruieren. Auf gemessenen Röntgenbildern von Interventionsmaterial in einem anthropomorphen Phantom mit simulierter Atembewegung wurde eine Rekonstruktionsgenauigkeit im Submillimeterbereich erreicht. Die gemessene Dosisflächenprodukt rate der vorgeschlagenen 3D-Rekonstruktionspipeline entspricht in etwa derjenigen der 2D-Fluoroskopie.

Contents

List of Acronyms	IX
List of Tables	XI
List of Figures	XIII
1 Introduction	1
2 Fundamentals	5
2.1 Interaction of X-Rays With Matter	5
2.1.1 Beer-Lambert Law	5
2.1.2 Photoelectric Effect	5
2.1.3 Compton Scattering	6
2.1.4 Rayleigh Scattering	6
2.2 Fundamentals of X-Ray CT	8
2.3 Health Effects of X-Rays in Fluoroscopy	10
2.4 Artificial Neural Networks for Image Processing	11
2.4.1 Artificial Neurons	11
2.4.2 Layers	11
2.4.3 Supervised Training Process	12
3 Materials & Methods	15
3.1 Reconstruction Pipelines - Design Ideas	15
3.2 Reconstruction Pipelines - Details	18
3.3 Projection Geometry	19
3.4 Simulation of 3D Models of Interventional Material	20
3.4.1 Guidewires	20
3.4.2 Stents	24
3.5 Deep Tool Extraction (DTE)	31
3.5.1 Training Data	31
3.5.2 Architecture	39
3.5.3 Training Process	39
3.6 Deep Tool Reconstruction (DTR)	39
3.6.1 Training Data	40

3.6.2	Architecture	43
3.6.3	Training Process	43
3.7	Stent Motion Compensation	43
3.7.1	Training Data	45
3.7.2	Architecture	45
3.7.3	Training Process	45
3.8	Phantom Measurements	46
3.9	Evaluation on Measured Data	50
3.10	Dose Area Product Measurements	51
4	Results	53
4.1	Pipeline Evolution	53
4.2	Robustness Analysis	66
4.2.1	Tube Current Time Product	66
4.2.2	Angular Increment	69
4.2.3	Motion Amplitude	72
4.3	Dose Area Product Rate	74
4.4	Real-Time Capability	74
5	Summary & Discussion	77
6	Conclusions	81
	Bibliography	83

List of Acronyms

R_{FD}	Source-detector distance
R_{F}	Source-isocenter distance
2D	Two-dimensional
3D	Three-dimensional
a.p.	Anterior-posterior
CBCT	Cone-beam computed tomography
CNN	Convolutional neural network
CNR	Contrast-to-noise ratio
CT	Computed tomography
D1	Average Euclidean distance between a skeleton of the ground truth and a skeleton of the reconstruction
D2	Average Euclidean distance between a skeleton of the reconstruction and a skeleton of the ground truth, under the exculsion of distances greater than 50 mm
DAP	Dose area product
DTE	Deep tool extraction
DTR	Deep tool reconstruction
FEVAR	Fenestrated endovascular aneurysm repair
fps	Frames per second
HP	High-pass filter
LP	Low-pass filter
MAE	Mean absolute error

MSE	Mean squared error
NN	Neural network
p.a.	Posterior-anterior
PPBP	Per-projection backprojection
PSF	Point spread function
px	Picture element or pixel
ReLU	Rectified linear unit
rpm	Respirations per minute
SCBP	Single-channel backprojection
ST	Spatial transformer network
vps	Volumes per second

List of Tables

3.1	Reconstruction pipelines investigated in this thesis.	19
3.2	Values of guidewire model parameters or distribution from which values were sampled.	22
3.3	Values of hoop model parameters or distribution from which values were sampled.	26
3.4	Values of B-spline curve parameters or distribution from which values were sampled.	28
3.5	Arrangements of interventional material measured in stop motion scans.	47
3.6	Interventional material used for the stop motion scans.	48
4.1	Reconstruction quality of different reconstruction pipelines.	54
4.2	Reconstruction quality at different tube current time products.	67
4.3	Reconstruction quality at different angular increments.	70
4.4	Reconstruction quality at different motion amplitudes.	72
4.5	Average dose area product (DAP) rates of some abdominal procedures guided by 2D fluoroscopy.	74
4.6	Computation times on an NVIDIA RTX 4090 GPU, using TensorFlow's mixed precision and graph execution.	75
4.7	Reconstruction times of the pipelines investigated in this thesis.	76

List of Figures

1.1	Sketch of a FEVAR	1
1.2	Two different arrangements of two guidewires result in the same pair of X-ray images	4
2.1	Attenuation coefficients of water, cortical bone, iron and Nitinol	7
2.2	Nomenclature used to describe a line in 2D	8
2.3	Acquisition of projection data in 2D parallel beam geometry and fan beam geometry	10
3.1	Sketch of the four-threaded baseline reconstruction pipeline 4-1-SCBP	15
3.2	Sketch of the three-threaded reconstruction pipeline 3-1	16
3.3	Illustration of backprojections in different imaging setups and the corresponding ground truth	17
3.4	Sketch of the best two-threaded reconstruction pipeline 2-4-sep-bin-STs	18
3.5	Sketch of the generation of spline control points	20
3.6	Volume rendered guidewire with control points	21
3.7	Sample of 12 volume rendered guidewires	23
3.8	Stent-related nomenclature	24
3.9	Plots of the three functions used to generate crowns of stent hoops	25
3.10	Visualization of the steps used to create a 3D model of a stent	27
3.11	Sample of 12 volume rendered stents	30
3.12	Visualization of the inputs and outputs of the DTE	31
3.13	Four samples of input and target patches generated by the DTE data loader	38
3.14	Visualization of the inputs and outputs of a stent DTR	40
3.15	Visualization of the generation of training data for a guidewire DTR	42
3.16	Visualization of inputs, intermediate volumes and outputs of a spatial transformer network	44
3.17	The anthropomorphic phantom used for the measurements in this thesis	46
3.18	Plot of the superior-inferior position of the phantom and guidewire feed over the time steps of the stop motion scan	49
3.19	Angles visited by reconstruction pipelines with different angular increments	49
3.20	Volume rendered 3D reconstruction of a stent and the associated skeleton	50

4.1	3D reconstructions of different reconstruction pipelines of interventional material arrangement a	57
4.2	3D reconstructions of different reconstruction pipelines of interventional material arrangement b	59
4.3	3D reconstructions of different reconstruction pipelines of interventional material arrangement c	61
4.4	3D reconstructions of different reconstruction pipelines of interventional material arrangement d	63
4.5	3D reconstructions of different reconstruction pipelines of interventional material arrangement e	65
4.6	3D reconstructions of pipeline 2-4-sep-bin-STs with different tube current time products	68
4.7	3D reconstructions of pipeline 2-4-sep-bin-STs with different angular increments	71
4.8	3D reconstructions of pipeline 2-4-sep-bin-STs of stop motion scans with different motion amplitudes	73

1 | Introduction

Today, endovascular interventions are an important treatment modality for many common diseases [1], e.g. coronary artery disease [2], cerebrovascular disease [3], peripheral artery disease [2], [4], aneurysms [5], [6], portal hypertension [7] and hepatocellular carcinoma [8]. These endovascular interventions typically start by navigating a guidewire through a small incision into a blood vessel to the target site (e.g. stenosis, aneurysm, tumor). Other interventional material, like tools (e.g. balloon catheters) or implants (e.g. stents), are then passed over the guidewire to the target site. As an example of an endovascular intervention, figure 1.1 shows a sketch of a fenestrated endovascular aneurysm repair (FEVAR), i.e. the treatment of an aneurysm using stent grafts. Since access to the patient's interior is gained percutaneously (i.e. only through a small incision

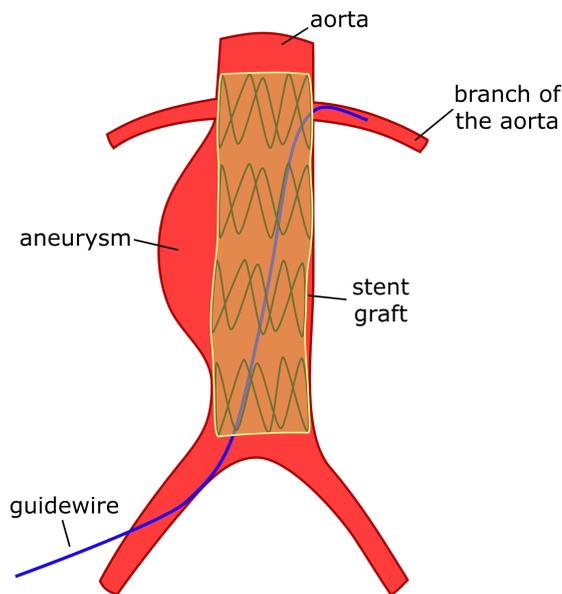


Figure 1.1: Sketch of an FEVAR. First, a guidewire is navigated through an incision into a femoral artery to the aneurysm. The stent graft is then passed over the guidewire to the aneurysm, where it is expanded. Afterwards, a guidewire is navigated into branches of the aorta to allow the placement of smaller stents there.

into the skin), these interventions are far less invasive (often called “minimally invasive”) than traditional open surgery. The percutaneous nature of endovascular interventions however also implies that the physician cannot see the interventional material and target site directly and therefore must rely on some form of image guidance. Today, this image guidance is usually provided by X-ray fluoroscopy, i.e. a live X-ray image. While X-ray fluoroscopy has high spatial and temporal resolution, it lacks depth information, since an X-ray image is a two-dimensional (2D) projection of the three-dimensional (3D) patient. This drawback is only partially mitigated by biplane X-ray fluoroscopy, i.e. X-ray fluoroscopy using two, non-parallel imaging threads (source-detector-pairs). Therefore, 3D X-ray fluoroscopy, i.e. a live 3D image reconstructed from a stream of 2D X-ray images, could improve the spatial awareness of the physician. This improved spatial awareness could result in safer and shorter procedures and enable new interventions [9]. Today, however, X-ray-based 3D fluoroscopy is not used for guiding endovascular interventions. One reason for this is the lack of a suitable algorithm for the reconstruction of 3D images from 2D X-ray images.

While there are techniques called “3D fluoroscopy” [10] or “live 3D guidance” [11], they only overlay information (e.g. a planned needle path or a volume rendered view) from a static, 3D computed tomography (CT) reconstruction with a live 2D fluoroscopy image. Because these techniques do not involve a live 3D image, they do not constitute 3D fluoroscopy, as understood in this thesis and defined above. While CT fluoroscopy, i.e. the continuous acquisition of CT scans, does provide a live 3D image reconstructed from X-ray images, its high dose rate limits its use to quick interventions, like biopsy or drainage with typical fluoroscopy times of a few minutes [12]–[14]. Many common endovascular interventions however have average (X-ray) fluoroscopy times above 20 min [15], [16]. The high dose rate of CT fluoroscopy is due to its use of conventional CT reconstruction algorithms, like filtered backprojection [17], [18], which require hundreds of X-ray images per 3D reconstruction [19], [20].

Therefore, several 3D-fluoroscopy-specific reconstruction approaches using much fewer X-ray images per 3D reconstruction have been proposed. These approaches separate the reconstruction of interventional material and patient and assume that the patient is well represented by a static or slowly updated 3D image [21]–[26]. A static 3D image of the patient can be computed easily by using conventional CT reconstruction algorithms to reconstruct a set of X-ray images acquired once prior to the intervention [21]–[23], [26]. A slowly updated 3D image of the patient can be reconstructed from a running set of X-ray images acquired slowly during the intervention using conventional CT reconstruction and registration algorithms [24], [25]. In both cases, the dose rate required for the patient reconstruction is much lower than in CT fluoroscopy, since the necessary X-ray images are acquired only once, or at a low temporal rate.

Outside of anatomical regions with strong patient motion, the remaining problem is therefore the 3D reconstruction of interventional material from few X-ray images. For this task, different algorithms have been proposed. One group of algorithms specializes

in the reconstruction of curvilinear structures, i.e. single guidewires or single catheters, from one [22], [23] or two [21], [26]–[28] X-ray images. Since these algorithms use that the object to be reconstructed is a single curve, they can neither be readily adapted to reconstruct other types of interventional material (e.g. stents) nor to reconstruct multiple guidewires or catheters, that are simultaneously present in the patient (a common occurrence). Furthermore, an algorithm for the retrieval of a 3D reconstruction of stents from a single X-ray image has been proposed [29]. Since it assumes a 3D model of the stent to be known a priori, it is more similar to a registration than a reconstruction algorithm.

Besides the aforementioned algorithms, which specialize in the reconstruction of a single instance of a single type of interventional material, two more general algorithms dedicated to reconstructing interventional material have been proposed. Kuntz et al. [24], [25] proposed a compressed-sensing-based algorithm [30], [31] for the reconstruction of interventional material, like stents and guidewires, from 16 to 20 X-ray images. This algorithm was later adapted to deal with patient motion by Flach et al. [32], [33]. However, the adoption of 3D fluoroscopy would likely require a dose rate similar to that of conventional biplane X-ray fluoroscopy. Since the dose rate is approximately proportional to the number of acquired X-ray images per update of the displayed image(s) (X-ray fluoroscopy: one or two X-ray images; 3D X-ray fluoroscopy: 3D reconstruction), it is likely necessary to acquire a number of X-ray images per 3D reconstruction that is roughly similar to two (biplane fluoroscopy). Therefore, while the 16 to 20 X-ray images of Kuntz et al.’s algorithm are much less than required by conventional CT algorithms, this number is likely still too high for use in endovascular interventions. More recently, Eulig et al. presented a deep-learning-based algorithm for the reconstruction of stents, guidewires, and coils from only four simultaneous X-ray images [34]. While this pipeline needs four times fewer X-ray images than that of Kuntz et al., it still requires four simultaneous X-ray images per 3D reconstruction, which has some drawbacks. First, acquiring four X-ray images simultaneously would require four imaging threads, which would make the X-ray system expensive and prone to hardware failure. Second, four X-ray images per image update, i.e. 3D reconstruction, would result in a dose rate that is still roughly twice as high as that of biplane X-ray fluoroscopy.

The goal of this thesis was therefore to develop an algorithm for the reconstruction of interventional material that is more capable than the one-curve-only algorithms and requires fewer imaging threads and fewer X-ray images (and therefore less dose) per 3D reconstruction than the algorithm of Eulig et al. The work described in this thesis was carried out as an employee of Ziehm Imaging GmbH, Nürnberg, Germany, a manufacturer of mobile C-arms, which are often used for 2D-fluoroscopic guidance of minimally invasive endovascular interventions. This work focuses therefore on methods that are readily transferable, e.g. in terms of hardware requirements, dose burden, and real time capability, to medical device development. This work was furthermore carried out in close cooperation with the department of X-Ray Imaging and CT at the German Cancer Research Center (DKFZ), Heidelberg, Germany. The department of X-Ray

Imaging and CT at the DKFZ performs research on algorithmic improvements of X-ray and CT imaging, in particular on 3D reconstruction algorithms and artifact reduction.

While a reduction of the number of X-ray images per 3D reconstruction from four to three could be achieved by simple retraining of the algorithm of Eulig et al. with training data with the appropriate number of imaging threads, this strategy proved insufficient for the transition to two X-ray images. While there are deep-learning-based algorithms [35]–[38] aiming to compute 3D reconstructions of complex objects like knee bones or whole CT volumes from only one or two X-ray images, reconstruction of interventional material from only two X-ray images seems indeed impossible in general, as the example of two guidewires in figure 1.2 illustrates. Therefore, the biplane X-ray system was assumed to be continuously rotating, and previously acquired X-ray images, which now contributed additional perspectives (due to the rotation) were used for the 3D reconstruction. Together with further adaptations, like the addition of a motion compensation step, this resulted in a pipeline capable of reconstructing multiple guidewires and single stents from only two newly acquired X-ray images. The effectiveness of these adaptations and the 3D reconstruction quality were investigated on X-ray images measured in an anthropomorphic phantom with respiratory motion. Measurements of the dose area product (DAP) were used to compare the dose rate with that of conventional X-ray fluoroscopy.

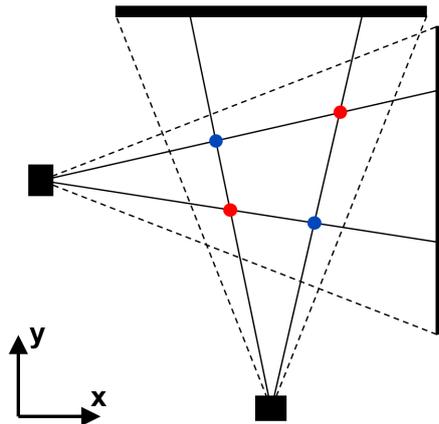


Figure 1.2: The blue and red dots indicate two different arrangements of two guidewires that result in the same pair of X-ray images. The 3D reconstruction problem has therefore no unique solution. The dotted lines indicate the edges of the X-ray beam. Only a 2D transversal (x - y) slice of the imaging setup is shown here. The red and blue dots are the locations, where the two guidewires intersect the slice.

2 | Fundamentals

2.1 Interaction of X-Rays With Matter

2.1.1 Beer-Lambert Law

When a narrow beam of photons of energy E travels through a homogeneous object of thickness d , the intensity of the incoming beam I_0 is reduced to the intensity of the outgoing beam I according to the Beer-Lambert law:

$$I = I_0 \exp(-\mu d) \quad (2.1)$$

The attenuation coefficient μ is a function of E and of the object material. For inhomogeneous objects, i.e. objects whose material changes with position \mathbf{r} , the following version of the Beer-Lambert law is used:

$$I = I_0 \exp\left(-\int_L d\mathbf{r} \mu(\mathbf{r})\right). \quad (2.2)$$

Here, L is the line describing the photon beam. Since the X-ray sources used in medical X-ray imaging are typically not monochromatic, a version of the Beer-Lambert law that also takes the energy dependence of the attenuation coefficient into account, is useful:

$$I = \int dE w(E) \exp\left(-\int_L d\mathbf{r} \mu(E, \mathbf{r})\right). \quad (2.3)$$

Here, $w(E)$ describes the spectrum of the incoming X-ray beam. Since the intensity in a narrow beam is considered here, all interactions of a photon, also scatter interactions, remove the photon from the beam and contribute to the (total) attenuation coefficient μ . In the range of photon energies produced by medical X-ray imaging devices, i.e. 0 keV to about 150 keV, there are three relevant interaction mechanisms. These three interactions, namely the photoelectric effect, Compton scattering and Rayleigh scattering are introduced in the following.

2.1.2 Photoelectric Effect

The photoelectric effect is the absorption of a photon by an electron of an atom, in which the electron is ejected from the atom. The kinetic energy of the ejected electron equals

the difference between the photon energy E and the binding energy of the electron. A rule of thumb for the dependence of the attenuation coefficient of the photoelectric effect on E and atomic number Z of the atom is given by $\mu_{\text{photo}} \propto k \cdot Z^4 \cdot E^{-3}$ [39], where k depends on the shell of the electron. Due to the strong E^{-3} dependence, the photoelectric effect leads to an absorption of most low energy photons and is the dominant interaction effect at low energies (see figure 2.1). The strong Z^4 dependence implies a higher contrast of high- Z objects than low- Z objects in X-ray images (at equal intersection length). This is one of the reasons, why interventional material, like stents and guidewires, are often made out of materials like Nitinol or stainless steel [40], [41]. Their attenuation coefficients are much higher than those of soft tissue, whose attenuation is very similar to that of water, and cortical bone (see figure 2.1).

2.1.3 Compton Scattering

Compton scattering is the inelastic scattering of a photon at an electron. It becomes the dominant interaction effect at energies above about 28 keV for water, 55 keV for cortical bone, 117 keV for iron and 115 keV for Nitinol (see figure 2.1).

2.1.4 Rayleigh Scattering

Rayleigh scattering is the elastic scattering of a photon at an atom or molecule. Its contribution to the total attenuation coefficient is typically low for materials investigated in medical X-ray imaging. For water, bone, iron and Nitinol it never exceeds 13% (compare figure 2.1).

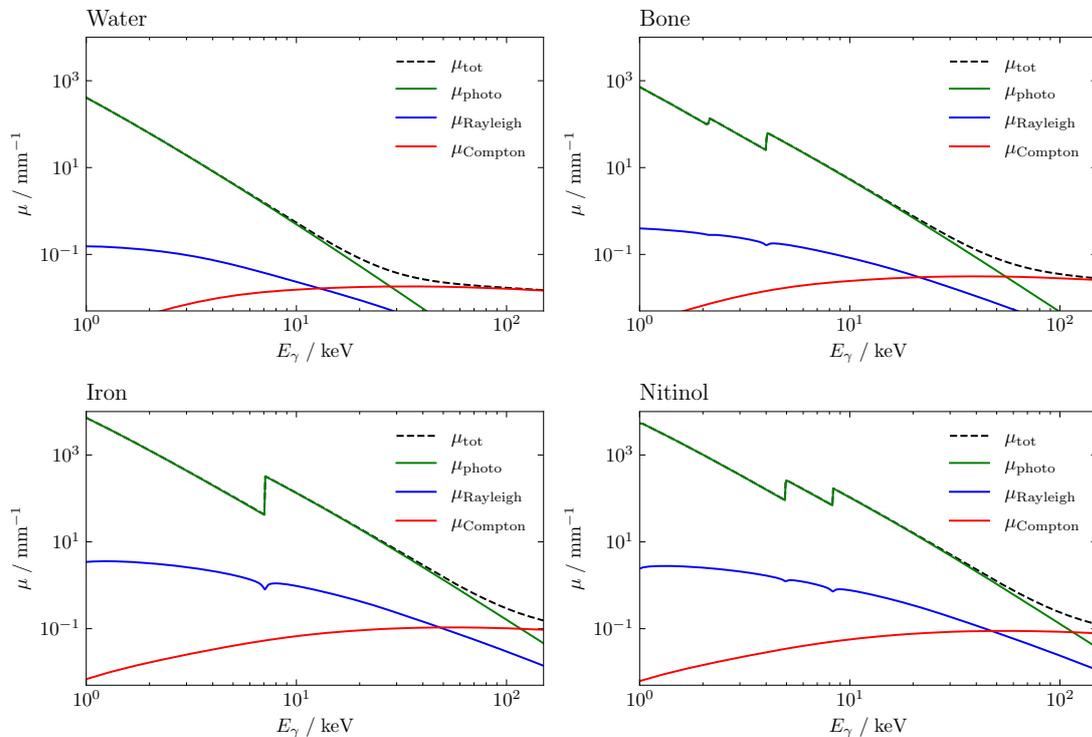


Figure 2.1: Attenuation coefficients of water, cortical bone, iron and nickel titanium naval ordnance laboratory (Nitinol) for photon energies from 1 keV to 150 keV. Cortical bone was assumed to be 39.2% hydrogen ($Z = 1$), 31.6% oxygen ($Z = 8$), 15.0% carbon ($Z = 6$), 6.5% calcium ($Z = 20$), 3.9% phosphorus ($Z = 15$), 3.5% nitrogen ($Z = 7$), 0.11% sulfur ($Z = 16$), 0.10% magnesium ($Z = 12$), 0.051% sodium ($Z = 11$) according to [42] (ratios are number-of-atoms ratios). Nitinol was assumed to be 50% nickel ($Z = 28$) and 50% titanium ($Z = 22$) (ratios are number-of-atoms ratios). The total attenuation coefficient (black dotted) is the sum of the contributions of the photoelectric effect (green), of Rayleigh scattering (blue) and of Compton scattering (red).

2.2 Fundamentals of X-Ray CT

X-ray CT is an imaging technique in which images of the internal structure of objects are computed from X-ray images (projections). The process of computing the image as well as the image itself are called reconstruction. In general, the acquisition of many X-ray projections from different angles is required for the reconstruction of complex objects. In the following, the reconstruction of 2D images from 1D parallel beam X-ray projections (2D parallel beam geometry) will be considered first. Afterwards, the more complicated reconstruction of 3D images from 2D projections will be discussed.

It is well known [43] that a real-valued function on the Euclidean plane $\mu(\mathbf{r}) = \mu(x, y)$, that meets certain regularity conditions, can be uniquely determined from the set of all line integrals $p(\vartheta, \xi)$, $\vartheta \in [0, 2\pi)$, $\xi \in \mathbb{R}$:

$$\begin{aligned} p(\vartheta, \xi) &= \int \mu(\mathbf{r}) \delta(\mathbf{r} \cdot \mathbf{n} - \xi) dx dy \\ &= \int \mu(x, y) \delta(x \cos \vartheta + y \sin \vartheta - \xi) dx dy. \end{aligned} \quad (2.4)$$

As can be seen from figure 2.2, $p(\vartheta, \xi)$ describes the integral of $\mu(x, y)$ along a line with unit normal vector $\mathbf{n} = (\cos \vartheta, \sin \vartheta)$ with distance ξ from the origin.

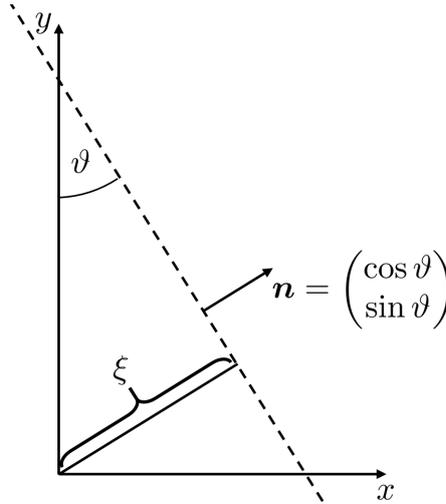


Figure 2.2: Nomenclature used to describe a line in 2D. Figure adapted from [44].

According to the monoenergetic Beer-Lambert law (2.2), such a line integral p can be determined from measuring the unattenuated and attenuated X-ray intensities I_0 and I :

$$p = -\ln(I/I_0). \quad (2.5)$$

An exact, analytic formula for the reconstruction of $\mu(x, y)$ can be found by noticing that the 1D Fourier transform of $p(\vartheta, \xi)$ with respect to ξ resembles the 2D Fourier

transform of $\mu(x, y)$:

$$\begin{aligned}
 P(\vartheta, u) &= \int p(\vartheta, \xi) e^{-2\pi i u \xi} d\xi \\
 &= \int \mu(x, y) e^{-2\pi i u (x \cos \vartheta + y \sin \vartheta)} dx dy \\
 &= (\mathcal{F}\mu)(u \cos \vartheta, u \sin \vartheta).
 \end{aligned} \tag{2.6}$$

The attenuation coefficient $\mu(x, y)$ can then be determined applying the inverse Fourier transform:

$$\begin{aligned}
 \mu(x, y) &= \int (\mathcal{F}\mu)(u \cos \vartheta, u \sin \vartheta) e^{2\pi i (xu \cos \vartheta + yu \sin \vartheta)} d(u \cos \vartheta) d(u \sin \vartheta) \\
 &= \int |u| P(\vartheta, u) e^{2\pi i u (x \cos \vartheta + y \sin \vartheta)} d\vartheta du \\
 &= \int (\mathcal{F}^{-1} | \cdot | P)(\vartheta, \xi) d\vartheta \\
 &= \int (\mathcal{F}^{-1} | \cdot |)(\xi) * p(\vartheta, \xi) d\vartheta.
 \end{aligned} \tag{2.7}$$

Here, the convolution theorem $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$, was used in the last step. The reconstruction formula (2.7) is often called filtered backprojection, since it can be implemented as the convolution of each projection $p(\vartheta, \xi)$, $\vartheta \in [0, 2\pi)$ with a filter kernel $(\mathcal{F}^{-1} | \cdot |)(\xi)$ followed by smearing all filtered projections back along their respective integration lines (backprojection).

While the acquisition of a set of parallel beam X-ray projections $p(\vartheta, \xi)$ is possible in practice, e.g. by repeatedly translating a pencil beam X-ray source in the direction of \mathbf{n} for each projection angle ϑ (figure 2.3, left), it is also very slow. Therefore, for 2D imaging, X-ray projections are usually acquired in a fan-beam geometry, i.e. each projection is acquired at once by simultaneously illuminating all pixels of a linear detector with an X-ray source (figure 2.3, right). For the reconstruction of 2D images from fan-beam projections, an exact, analytic reconstruction formula, similar to equation (2.7), exists as well [39]. For 3D imaging, often a cone-beam computed tomography (CBCT) geometry, i.e. an X-ray source illuminating a rectangular grid of detector pixels, i.e. a 3D version of the fan-beam geometry, is used. In CBCT, the X-ray source usually travels along a circular trajectory during the acquisition of the projections. In this setup, only the 2D slice, in which the source trajectory lies can be reconstructed exactly [45]. For the reconstruction of a 3D volume, approximate filtered-backprojection-type reconstruction algorithms exist [18].

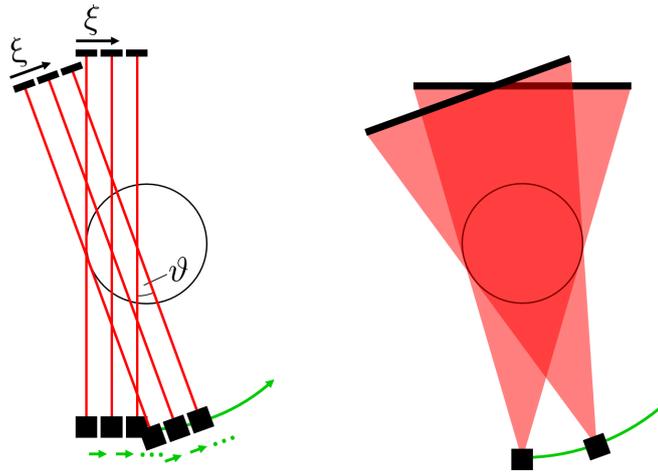


Figure 2.3: Acquisition of projection data in 2D parallel beam geometry (left) and fan-beam geometry (right). In 2D parallel beam geometry, a pencil beam X-ray source is translated in the ξ -direction to acquire a projection. This is repeated for many projections at different angles ϑ . In fan-beam geometry, whole projections are acquired at once. Red indicates X-rays. Green indicates motion of the X-ray system (shift and rotation). Black squares indicate X-ray sources. The circle indicates the reconstructable area.

2.3 Health Effects of X-Rays in Fluoroscopy

Photons in the X-ray spectrum have sufficient energy to damage biological tissue by ionizing atoms or molecules. The effects of exposure to radiation are typically categorized as being deterministic or stochastic [46]. Deterministic effects occur reliably above a certain threshold dose and their severity increases with dose. In X-ray fluoroscopy, deterministic effects are usually skin injuries [47]–[51]. They occur when peak skin dose exceeds certain thresholds, which is rare, but can result in severe injuries [47]–[51]. Peak skin dose is the maximum local absorbed dose on the skin of an individual [48]. Absorbed dose is the energy deposited by ionizing radiation in a volume divided by the mass of the volume. For stochastic effects, the severity is independent of dose, but the probability of their occurrence increases with dose. In X-ray fluoroscopy, the formation of cancer is the primary stochastic effect of concern [47]. The dose quantity most widely used to estimate the risk of stochastic effects, mainly cancer, is effective dose [52], [53]. It is the absorbed dose of different organs, weighted by factors specific to the kind of radiation (this factor is 1 for all X-rays) and the radiosensitivity of the organs. Whether the formation of cataracts, which is sometimes observed after X-ray fluoroscopy, is a deterministic or stochastic effect is a point of content in the literature [47], [54].

2.4 Artificial Neural Networks for Image Processing

Artificial neural networks, often just called neural networks (NNs), are a type of machine learning algorithm. In the development of machine learning algorithms, the transformation of inputs into outputs is not manually coded into the algorithms, but learned from data. For many tasks in image processing, NNs are state of the art [55]–[57].

2.4.1 Artificial Neurons

The smallest computational unit in a NN is the artificial neuron. It takes a tuple of real numbers $a_k \in \mathbb{R}$ as inputs and outputs a real number a'

$$a' = \sigma\left(\sum_k a_k \cdot w_k + b\right). \quad (2.8)$$

Here, $w_k \in \mathbb{R}$ are called weights, $b \in \mathbb{R}$ is called bias and σ is a non-linear real function, called activation function. The weights and the bias are usually learnable parameters. The activation function must be non-linear in order for the network to learn non-linear transformations. One of the most commonly used activation functions is the rectified linear unit (ReLU), which maps inputs x to outputs $f(x)$ via $f(x) = \max(0, x)$.

2.4.2 Layers

The inputs of the artificial neuron can be the inputs of the NN, e.g. brightness values of pixels in an image, or the outputs of other artificial neurons. Usually, the artificial neurons in a NN are arranged in layers, with the inputs and outputs being layers themselves and with the artificial neurons of one layer usually getting their inputs from the same previous layer. Some of the layers most commonly used in image processing are introduced in the following.

i.) Fully-Connected Layers

In fully-connected layers, each neuron is connected to each neuron of the previous layer with weights and biases being different for each neuron. The output a_i^j of the i th neuron of the j th fully-connected layer in a NN is

$$a_i^j = \sigma\left(\sum_k a_k^{j-1} \cdot w_{ki}^j + b_i^j\right), \quad (2.9)$$

with weights w_{ki}^j and biases b_i^j and the sum running over all neurons of the previous layer. Fully-connected layers are rarely used to process high-resolution images, since the number of operations and the number of learnable parameters grow rapidly with the square of the number of input pixels. Therefore, fully-connected layers usually get low-resolution, downsampled images as input, as they often occur deep, i.e. after a few layers, in NNs. In image processing, common uses of fully-connected layers are

classification tasks, like classifying images as showing cats or dogs [58], and regression tasks like the estimation of the parameters of spatial transformations [59].

ii.) Convolutional Layers

In a convolutional neural network (CNN), i.e. a NN with convolutional layers, the output a_i^j of the i th neuron of the j th convolutional layer is

$$a_i^j = \sigma\left(\sum_{k=-N, \dots, N} a_{i+k}^{j-1} \cdot w_k^j + b^j\right), \quad (2.10)$$

with bias b^j and convolution kernel w_k^j of width $2N + 1$ ($N \in \mathbb{N}$). This local and translation invariant pattern of connections between previous and current layer lends itself to the efficient processing of images. Equation (2.10) describes a convolutional layer that processes 1D images, i.e. the output a_i^j of each layer j and the convolutions performed in the layers are 1D. The processing of n -dimensional images is described by analogous, n -dimensional versions of equation (2.10), i.e. the output of each layer and the convolutions performed in the layers are n -dimensional. Usually, a convolutional layer does not only consist of an image with one channel, a_i^j , computed by one convolution kernel w_k^j , but of an image with N_{channels}^j channels, a_{ic}^j , computed by $N_{\text{channels}}^{j-1} \cdot N_{\text{channels}}^j$ convolution kernels w_{klc}^j :

$$a_{ic}^j = \sigma\left(\sum_{k=-N, \dots, N} \sum_{l=1, \dots, N_{\text{channels}}^{j-1}} a_{i+k, l}^{j-1} \cdot w_{klc}^j + b_c^j\right). \quad (2.11)$$

Here, j is the index of the current layer, $k \in \{-N, \dots, N\}$ is the kernel dimension, $l \in \{1, \dots, N_{\text{channels}}^{j-1}\}$ is the channel index of the previous layer and $c \in \{1, \dots, N_{\text{channels}}^j\}$ is the channel index of the current layer.

iii.) Downsampling and Upsampling Layers

Downsampling and upsampling layers are often used in CNNs to enable the processing of data at different resolutions. These layers have typically no learnable parameters. Downsampling is often performed by max pooling, i.e. by replacing patches of the channels of the input layer by the maximum value in the respective patch. Upsampling is often performed by nearest-neighbor interpolation.

2.4.3 Supervised Training Process

NNs often have millions or billions of free parameters, i.e. weights or kernel entries and biases. These parameters are not set manually, but learned from data in a process called training. The data used for training are called training data. A common type of training is supervised training. In supervised training, the training data set consists of input-target-pairs (also called samples), i.e. pairs of possible inputs to the NN and the respective desired outputs (targets) of the NN. Supervised training is usually performed

by a) feeding an input from the training data set into the network and computing the output, b) computing the value of the loss function, that is a real-valued function that measures the difference between the respective target from the training data set and the actual output of the NN, c) computing the gradient of the loss function with respect to all free parameters of the NN using the backpropagation algorithm [60] and d) using a gradient-descent-type optimization algorithm like vanilla gradient descent or Adam [61] to adapt the free parameters accordingly. Steps a) to d) are then repeated until each sample in the training data set has been used exactly once. Each use of the whole training data set for training is called an epoch. Usually, NNs are trained for many epochs. The training process described in steps a) to d), where the free parameters of the NN are adapted after computing the gradient of a single sample, is called stochastic gradient descent. However, in practice, usually, the gradients of multiple samples, called a mini-batch, are averaged before adapting the free parameters. This approach is called mini-batch gradient descent and often leads to a more stable training with convergence to lower values of the loss function. Common choices for a loss function used to compare images are the mean absolute error (MAE) and the mean squared error (MSE).

3 | Materials & Methods

3.1 Reconstruction Pipelines - Design Ideas

In this thesis, pipelines for the reconstruction of interventional material from four, three and two simultaneous X-ray images, i.e. using X-ray systems with $G = 4$, $G = 3$ and $G = 2$ imaging threads, were investigated. The reconstruction pipeline of Eulig et al. [34], which requires four X-ray images acquired simultaneously at 45° increments served as a baseline (figure 3.1; pipeline 4-1-SCBP in table 3.1). The pipeline consists of two main steps. First, the deep tool extraction (DTE), a CNN, is applied to each of the four X-ray images. The DTE is trained to output an image that is only non-zero in those pixels that contain interventional material in the input X-ray image. A second CNN, the deep tool reconstruction (DTR), then transforms the backprojection (BP) of the four DTE output images into a 3D reconstruction of the interventional material.

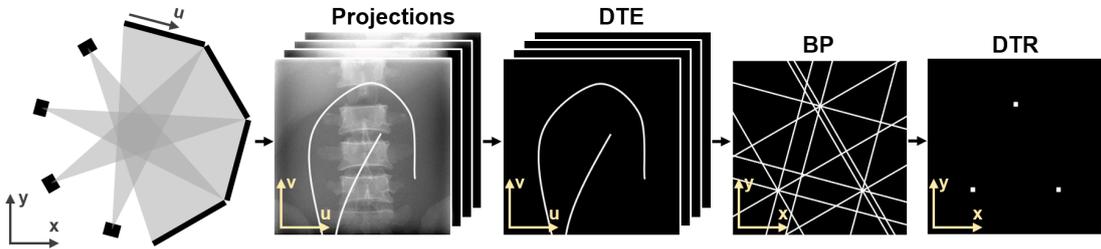


Figure 3.1: The pipeline of Eulig et al., which reconstructs interventional material (in this sketch: two guidewires) from four simultaneous X-ray images (pipeline 4-1-SCBP in table 3.1). Note that the backprojection (BP) and DTR output are 3D volumes of which only a transversal (x - y) slice is shown here.

For the reconstruction of interventional material from three X-ray images, a simple retraining of the baseline pipeline with three-threaded training data was sufficient (pipeline 3-1-SCBP in table 3.1). Additionally, a minor adaptation of the baseline pipeline was empirically found to improve reconstruction quality (figure 3.2; pipeline 3-1 in table 3.1). Instead of backprojecting all three DTE output images into a single volume (single-channel backprojection (SCBP)), each DTE output image was backprojected into a separate volume (per-projection backprojection (PPBP)). These three backprojection volumes were then passed to the DTR in separate input channels (illustrated

in figure 3.2 by using different colors). Adaptation of the pipeline to two (orthogonal)

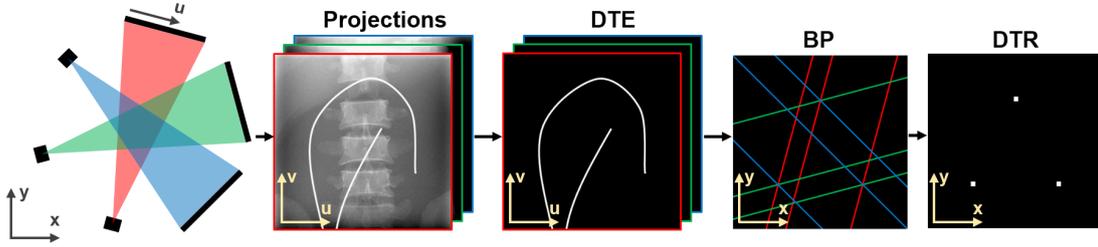


Figure 3.2: The pipeline used for computing 3D reconstructions from three X-ray images (pipeline 3-1 in table 3.1). Compared to the baseline pipeline (figure 3.1), the number of X-ray images (four \rightarrow three) and the angle between the X-ray images ($45^\circ \rightarrow 60^\circ$) changed. Furthermore, instead of SCBP, PPBP was used.

X-ray images posed a significantly greater challenge than the adaptation to three X-ray images. This challenge is illustrated in figure 3.3, where transversal slices through input volumes (backprojections) of different DTRs and the corresponding slice of the ground truth reconstruction are shown. Points where G backprojected rays intersect and where interventional material intersects the slice are marked green. Points where G backprojected rays intersect incidentally, i.e. where no interventional material is present, are marked red. When four or three X-ray images are available, i.e. $G = 4$ or $G = 3$, incidental intersections (red) are rare. The task of the DTR is then relatively simple, since the set of points where G rays intersect is a good approximation of the reconstruction. This is no longer true when reducing the number of imaging threads to $G = 2$. Most intersections of G rays are then incidental (more red than green points). Accordingly, simply training the reconstruction pipeline shown in figure 3.2 with two instead of three X-ray images, yields much lower reconstruction quality (pipeline 2-1 in table 3.1).

Therefore, for the reconstruction from two imaging threads, a continuously rotating X-ray system was assumed. In this rotating biplane setup, taking the backprojection $BP(t)$ of the current time step t and of previous time steps, e.g. $BP(t - 1)$, together (figure 3.3, second from the right), leads to a situation, which is similar to the easier case with four imaging threads (figure 3.3, left), if the motion between time steps is small. Accordingly, using the backprojections of previous time steps in the rotating biplane setup as additional inputs for the DTR, improves the reconstruction quality significantly (pipeline 2-4 in table 3.1).

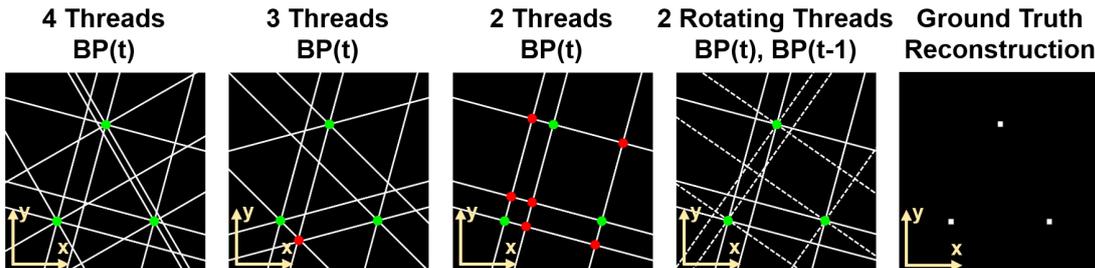


Figure 3.3: Illustration of a transversal (x - y) slice of backprojections (BPs) and the corresponding ground truth. The interventional material being imaged here could for example be three guidewires, each intersecting the depicted slice exactly once. The four images on the left show backprojections, as they are available as inputs for the DTR in different image acquisition setups. Points where G backprojected rays intersect are marked green, if interventional material intersects the slice there. Incidental intersections of G rays are marked red.

Since the reconstruction quality of this pipeline is still significantly worse than the reconstruction quality in the four- and three-threaded setups, three more improvements to the reconstruction pipeline were made. This reconstruction pipeline is shown in figure 3.4 (pipeline 2-4-sep-bin-STs in table 3.1). The first improvement is the separation of the 3D reconstruction of different types of interventional material, which is enabled by training the DTE to output different types of interventional material in separate output channels (pipeline 2-4-sep in table 3.1). In this work, DTE was trained to output two images (channels), one showing stents and one showing guidewires. The motivation for splitting the 3D reconstruction task into subtasks (guidewire reconstruction and stent reconstruction) was twofold. First, the subtasks have a higher ratio of real to incidental intersections in the backprojections (higher sparsity), which makes reconstruction easier. Second, the subtask-specific DTRs have a narrower, and therefore easier task than a DTR for the reconstruction of the full task. The second improvement is the addition of spatial transformer networks (STs) ST_1 , ST_2 and ST_3 to compensate the motion between the backprojections of the current time step and the previous steps $t - 1$, $t - 2$ and $t - 3$. This motion compensation was motivated by the fact that in the absence of motion, the combination of $BP(t)$ and $BP(t - 1)$ in the rotating biplane setup is very similar to $BP(t)$ in the much easier $G = 4$ setup (figure 3.3). The motion compensation by STs was only found to be necessary for the reconstruction of stents. Presumably, this is due to stents having finer spatial detail than guidewires. The third improvement is the binarization of the output images of the guidewire channel of the DTE prior to backprojection, which was incidentally found to improve reconstruction quality. Without binarization, the DTE output images contain in each pixel the intersection length in mm (as estimated by the DTE) between the interventional material and the ray traveling from the source to that pixel. The binarization was performed by thresholding with a value close to zero. Binarization was not tested for the stent output images of the DTE,

due to the greater amount of self-overlap of stents in projection images, which makes distinguishing individual stent struts difficult in thresholded images.

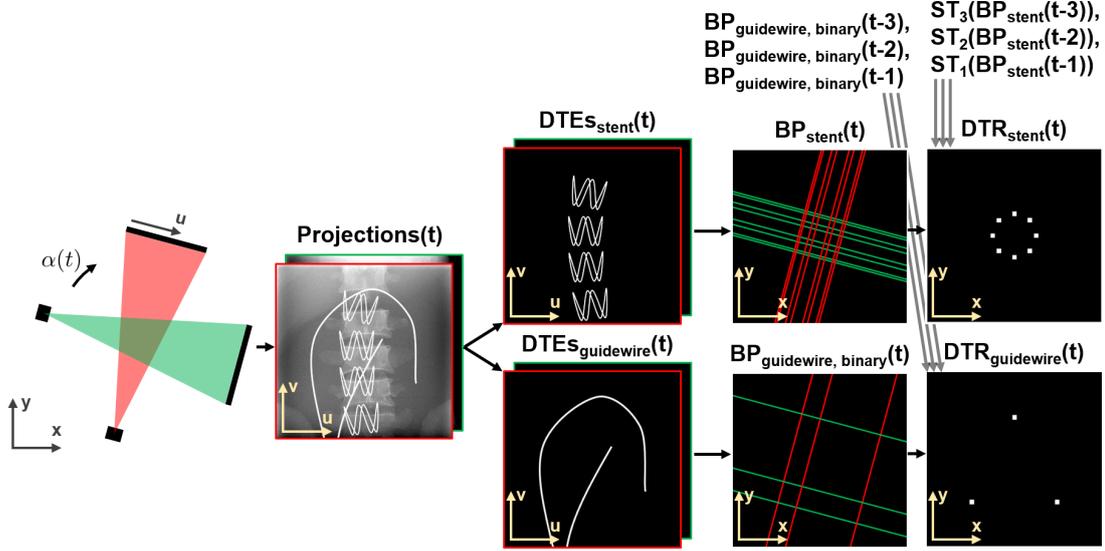


Figure 3.4: Best two-threaded reconstruction pipeline (pipeline 2-4-sep-bin-STs in table 3.1). The biplane X-ray system rotates continuously. The DTE outputs stents and guidewires in separate output channels. Stents and guidewires are then reconstructed separately, by feeding PPBPs (red and green) of the current and of previous time steps into a DTR. Binarizing the DTE guidewire outputs prior to backprojection and compensating motion between the stent backprojections using STs, further improve the reconstruction quality.

3.2 Reconstruction Pipelines - Details

In this section, the parameters of the 3D reconstruction pipelines, which were investigated in this thesis, are described in more detail. The first seven pipelines in table 3.1 are the stages in the evolution from the four-threaded baseline pipeline (4-1-SCBP) into the best two-threaded pipeline (2-4-bin-STs), which were already mentioned in the previous section. The final two pipelines in table 3.1 are used to investigate the robustness of pipeline 2-4-bin-STs with respect to changes in the angular increment $\Delta\alpha$.

For each pipeline, the angle $\alpha(t, j)$ at which the imaging thread $j \in \{0, \dots, G-1\}$ of the X-ray system acquires an X-ray image at time step $t \in \mathbb{N}_0$ is determined by the number of imaging threads G , the initial angle of thread $j = 0$, α_0 and the angular increment $\Delta\alpha$:

$$\alpha(t, j) = \alpha_0 + j \cdot 180^\circ / G + t \cdot \Delta\alpha. \quad (3.1)$$

An angle of $\alpha(t, j) = 0^\circ$ corresponds to an anterior-posterior (a.p.) X-ray image. The

initial angle α_0 was set to 0° for all reconstruction pipelines. Throughout this thesis, in particular for the simulation of training data (sections 3.6.1 and 3.7.1) and for the selection of X-ray images from stop motion measurements (section 3.8), the length of a time step, Δt , was assumed to be 0.2s. Although those pipelines, that use only a single time step as input ($T_{\text{input}} = 1$), do not profit from a rotation-induced change in perspective, it was assumed that the respective X-ray system rotates, since this rotation might enable a conventional CT reconstruction of the patient anatomy, which requires X-ray images from all directions.

Table 3.1: Reconstruction pipelines investigated in this thesis.

ID	Geometry Parameters		DTR Parameters				
	G	$\Delta\alpha$	T_{input}	k	Bin	SCBP	STs
4-1-SCBP	4	19°	1	total	no	yes	no
3-1-SCBP	3	19°	1	total	no	yes	no
3-1	3	19°	1	total	no	no	no
2-1	2	19°	1	total	no	no	no
2-4	2	19°	4	total	no	no	no
2-4-sep	2	19°	4	guidewire	no	no	no
			4	stent	no	no	no
2-4-sep-bin-STs	2	19°	4	guidewire	yes	no	no
			4	stent	no	no	yes
2-4-sep-bin-STs-less $\Delta\alpha$	2	9.5°	4	guidewire	yes	no	no
			4	stent	no	no	yes
2-4-sep-bin-STs-more $\Delta\alpha$	2	28.5°	4	guidewire	yes	no	no
			4	stent	no	no	yes

sep: separate reconstruction of stents and guidewires

Bin: binarization of DTE output images prior to backprojection

k : type of interventional material reconstructed by the DTR(s) ($k = \text{total}$ means, that one DTR reconstructs stents and guidewires; $k = \text{guidewire}$, stent means that one DTR reconstructs guidewires and a second DTR reconstructs stents)

T_{input} : number of input time steps

3.3 Projection Geometry

The experimental X-ray system used in this work has a flat detector (1024^2 pixels of area $(0.3 \text{ mm})^2$, source-detector distance $R_{\text{FD}} = 1053 \text{ mm}$, source-isocenter distance $R_{\text{F}} = 616 \text{ mm}$), which is mounted on a rotating gantry. Using a stop motion technique (section 3.8), this X-ray system was used to measure data that resembles the data, that a system with $G \in \{2, 3, 4\}$ imaging threads would measure. For the simulation of training and validation data, the same geometry was used.

3.4 Simulation of 3D Models of Interventional Material

The generation of training data for the DTE and DTR was based on a large number of simulated, randomly varying 3D models of guidewires and stents. The simulation of these 3D models is described in this section.

3.4.1 Guidewires

Each guidewire was modeled as a curved cylinder with diameter $D_{\text{guidewire}}$. The cylinder was modeled as a triangle mesh approximating the cylinder surface. The centerline of this cylinder was given by a clamped B-spline curve of degree 3, with N_{control} control points. Here, the B-spline curve implementation of NURBS-Python (class `geomdl.BSpline.Curve`, `geomdl` version 5.3.1) was used [62]. The control points $\{\mathbf{p}_i \mid i = 1, \dots, N_{\text{control}}\}$ were sampled in a random-walk-like fashion. For each step of the random walk, the step length l_{step} , polar angle θ and azimuthal angle ϕ were sampled randomly. The polar and azimuthal angles were defined with respect to local Cartesian coordinate systems, given by orthonormal vectors $\mathbf{e}_{i,x}$, $\mathbf{e}_{i,y}$ and $\mathbf{e}_{i,z}$ (see figure 3.5). Intuitively, θ corresponds to how much the direction of the current step deviates from the direction of the previous step and ϕ corresponds to how far the current step is outside the plane given by the previous two steps. The curved cylinder around the spline defined by these control points will be referred to as the base cylinder.

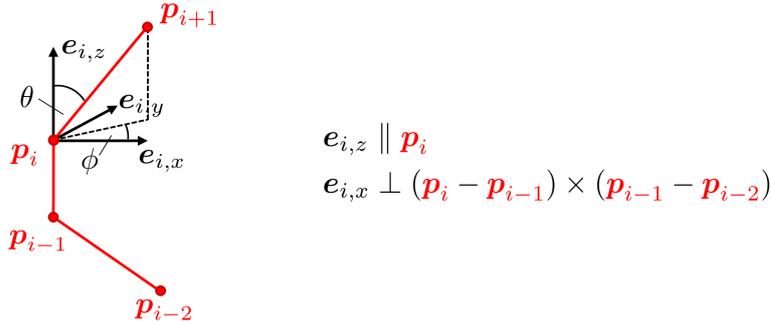


Figure 3.5: Sketch of a local coordinate system $\{\mathbf{e}_{i,x}, \mathbf{e}_{i,y}, \mathbf{e}_{i,z}\}$, which was used to define the polar angle θ and the azimuthal angle ϕ in the process of sampling control point \mathbf{p}_{i+1} .

Additionally, to simulate guidewires with curved tips, a “tip” cylinder with much shorter step lengths was sampled. The base cylinder and the tip cylinder were joined, so that their centerlines were parallel at the joint. Figure 3.6 shows one randomly generated guidewire (pink). The joint between the base and the tip cylinder is indicated by a green arrow. The control polygons of the base and the tip are shown in cyan. Real guidewires can look hollow in X-ray images, probably because their core is made from a different material than their shell. To achieve a similar effect in simulated X-ray images, the guidewires were modeled as hollow curved cylinders with inner diameter $D_{\text{guidewire,inner}}$.

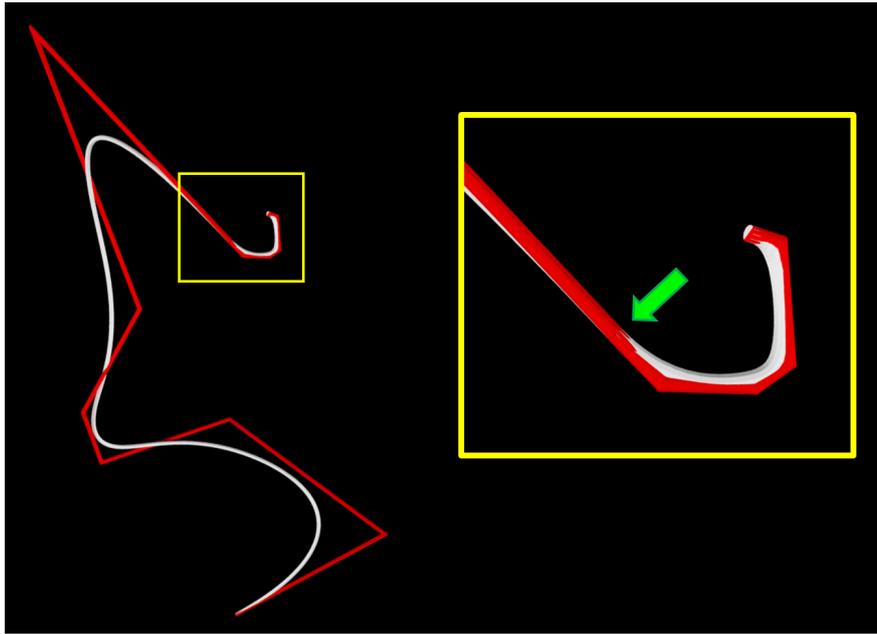


Figure 3.6: A randomly generated guidewire (white) with control polygons of the base and the tip splines (red). The green arrow indicates the point where the two control polygons meet.

The values of the parameters N_{control} , l_{step} , θ , ϕ , $D_{\text{guidewire}}$ and $D_{\text{guidewire,inner}}$ or the distributions from which these values were sampled are given in table 3.2. To give an impression of the distribution of the guidewires, 12 randomly generated guidewires are shown in figure 3.7.

Table 3.2: Values of guidewire model parameters or distribution from which values were sampled.

Parameter	Value or Distribution	Sampled for Each
Base Cylinder		
N_{control}	8	-
l_{step}	$\mathcal{U}(20 \text{ mm}, 80 \text{ mm})$	random walk step
θ	$\mathcal{U}(0^\circ, 150^\circ)$	random walk step
ϕ	$\mathcal{U}(0^\circ, 360^\circ)$	random walk step
Tip Cylinder		
N_{control}	$\mathcal{U}(2, 7)$	guidewire
l_{step}	$\mathcal{U}(1.5 \text{ mm}, 7 \text{ mm})$	random walk step
θ	$\mathcal{U}(20^\circ, 70^\circ)$	random walk step
ϕ	$\mathcal{U}(5^\circ, 25^\circ)$	random walk step
Common for Base and Tip Cylinder		
$D_{\text{guidewire}}$	$\mathcal{U}(0.25 \text{ mm}, 1 \text{ mm})$	guidewire
$D_{\text{guidewire,inner}}$	$\mathcal{U}(0, 0.8 \cdot D_{\text{guidewire}})$	guidewire
$\mathcal{U}(a, b)$ is the uniform distribution on the interval $[a, b] \subset \mathbb{R}$		

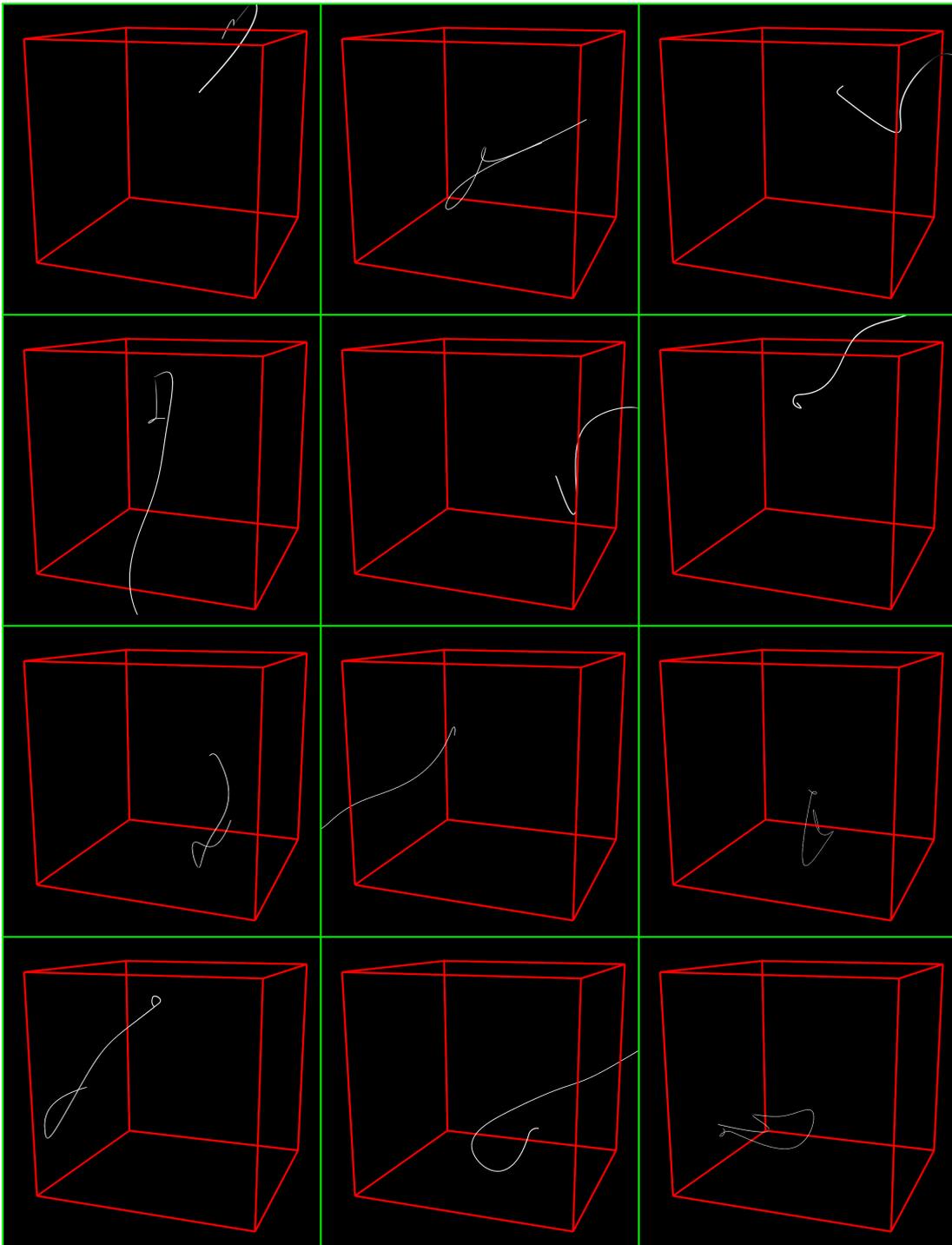


Figure 3.7: 12 randomly generated guidewires. The red cube has a side length of 153.6 mm, which is the same size as the volume reconstructed by the DTR.

3.4.2 Stents

Each stent was modeled as a stack of oscillating curved cylinder rings. Each curved cylinder was modeled as triangle meshes approximating the cylinder surface. The oscillating rings are often called hoops. The individual wire segments making up each hoop are called struts. Two adjacent struts are called a crown. The meaning of these terms is visualized in figure 3.8. The remainder of this section describes the creation of the stent models in detail.

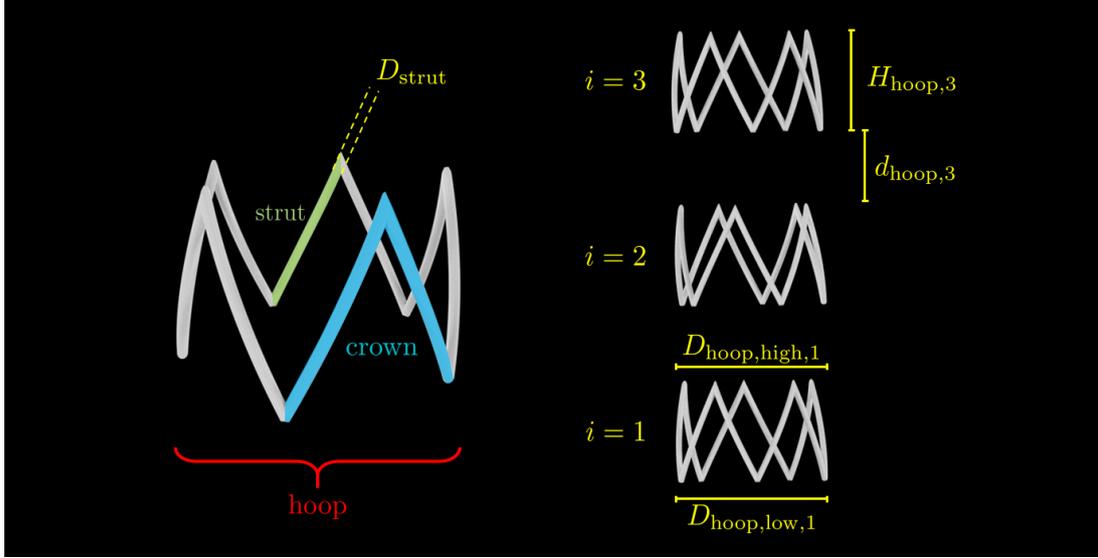


Figure 3.8: Stent-related nomenclature: a stent is a stack of hoops, wherein each hoop consists of struts. Two adjacent struts form a crown. Values belonging to the different hoops are marked with the corresponding index.

i.) Creation of Hoops

Each hoop was modeled as a curved cylinder with diameter D_{strut} . The centerline of the curved cylinder was given by a closed curve oscillating on the surface of a conical frustum:

$$\mathbf{r}(\lambda) = \begin{pmatrix} x(\lambda) \\ y(\lambda) \\ z(\lambda) \end{pmatrix} = \begin{pmatrix} r(\lambda) \sin(2\pi\lambda + \phi) \\ r(\lambda) \cos(2\pi\lambda + \phi) \\ H_{\text{hoop}} \cdot g_k(N_{\text{crown}}\lambda - \lfloor N_{\text{crown}}\lambda \rfloor) \end{pmatrix}, \text{ with}$$

$$r(\lambda) = 0.5 \cdot (D_{\text{hoop,low}} + (D_{\text{hoop,high}} - D_{\text{hoop,low}}) \cdot (z(\lambda)/H_{\text{hoop}} + 0.5)) \quad (3.2)$$

and with the curve parameter $\lambda \in [0, 1)$, the floor function $\lfloor \cdot \rfloor$, the lower and upper hoop diameters $D_{\text{hoop,low}}$, $D_{\text{hoop,high}}$, the hoop height H_{hoop} , the number of crowns N_{crown} and phase angle ϕ . The functions g_k describe the shape of the crowns. On the domain $[0, 1]$ they are symmetric with respect to $x = 0.5$ and take values in the range $[-0.5, 0.5]$.

Here, three different functions g_k , describing sinusoidal, zig-zag and sigmoidal crown shapes were used:

$$g_{\text{sinusoid}}(x) = 0.5 \cdot \sin(2\pi x), \quad (3.3)$$

$$g_{\text{zig-zag}}(x) = 2|x - 0.5| - 0.5, \quad (3.4)$$

$$g_{\text{sigmoid}}(x) = \frac{L(|24x - 12| - 6) - L(-6)}{L(6) - L(-6)} - 0.5, \quad (3.5)$$

with the logistic function $L(x) = 1/(1 + \exp(-x))$. Plots of g_{sinusoid} , $g_{\text{zig-zag}}$ and g_{sigmoid} are shown in figure 3.9.

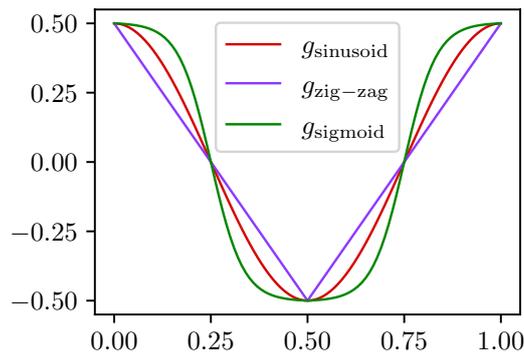


Figure 3.9: Plots of the three functions g_{sinusoid} , $g_{\text{zig-zag}}$ and g_{sigmoid} , which were used to generate the crowns of the hoops.

The hoop parameters were sampled randomly. Details on this sampling are given in table 3.3. Note that some parameters were the same for all hoops in a stent, while some parameters were sampled for each hoop. The latter were explicitly labeled with the hoop index $i = 1, 2, \dots, N_{\text{hoop}}$ in the table (unlike in equation (3.2), where the hoop index was omitted for clarity). Here N_{hoop} is the number of hoops in the stent which was sampled uniformly from the integer interval $[5, 10]$. Also note that some of the parameters depended on each other: the lower hoop diameter $D_{\text{hoop,low}}$ was always equal to the upper hoop diameter $D_{\text{hoop,high}}$ of the previous hoop and the ratio between upper and lower diameter of each hoop was sampled uniformly from the real interval $[0.85, 1.15]$.

A hoop with $g_k = g_{\text{sigmoid}}$ is shown in figure 3.10 a.

ii.) Hoop Deformation by Thin Plate Splines

Equation (3.2) produces very regular hoop centerlines. Real stents however often deviate slightly from their regular, ideal shape, e.g. due to uneven expansion and/or the not perfectly cylindrical shape of the surrounding vessel wall. Therefore, the centerlines of each hoop produced by equation (3.2) are deformed randomly by randomly generated 3D thin plate splines. Here the thin plate spline implementation of Tensor-

Table 3.3: Values of hoop model parameters or distribution from which values were sampled.

Parameter	Value or Distribution	Sampled for Each
N_{crown}	$\mathcal{U}_I(5, 10)$	stent
D_{strut}	$\mathcal{U}(0.2 \text{ mm}, 0.6 \text{ mm})$	stent
g_k	$\mathcal{U}(\{g_{\text{sinusoid}}, g_{\text{zig-zag}}, g_{\text{sigmoid}}\})$	stent
ϕ_i	$\mathcal{U}(0^\circ, 360^\circ)$	$i = 1, 2, \dots, N_{\text{hoop}}$
$D_{\text{hoop,low},1}$	$\mathcal{U}(12 \text{ mm}, 31 \text{ mm})$	stent
$D_{\text{hoop,low},i}$	$D_{\text{hoop,high},i-1}$	$i = 2, \dots, N_{\text{hoop}}$
$D_{\text{hoop,high},i}$	$D_{\text{hoop,low},i} \cdot \mathcal{U}(0.85, 1.15)$	$i = 1, 2, \dots, N_{\text{hoop}}$
$H_{\text{hoop},i}$	$\mathcal{U}(6 \text{ mm}, 16 \text{ mm})$	$i = 1, 2, \dots, N_{\text{hoop}}$

$\mathcal{U}(a, b)$ is the uniform distribution over the interval $[a, b] \subset \mathbb{R}$
 $\mathcal{U}_I(a, b)$ is the discrete uniform distribution over the interval $[a, b] \subset \mathbb{Z}$
 $\mathcal{U}(\{a, b, c\})$ is the discrete uniform distribution over the set $\{a, b, c\}$

Flow Addons [63] (function `tensorflow-addons.image.interpolate_spline`, in `tensorflow-addons`, version 0.14.0) was used. Each thin plate spline is generated such that it interpolates a first set of 3D points A and a second set of 3D points B . Set $A = \{\mathbf{a}_{ijk} \mid i, j = 0, 1, 2, 3, 4; k = 0, 1, 2\}$ is a regular rectilinear $5 \times 5 \times 3$ grid aligned with, but 20% larger than, the axis-aligned bounding box of the hoop:

$$\mathbf{a}_{ijk} = \mathbf{a}_0 + \begin{pmatrix} i \\ j \\ k \end{pmatrix} \cdot \mathbf{d}\mathbf{a}, \text{ with}$$

$$\mathbf{a}_0 = \begin{pmatrix} x_{\min} - 0.1 \cdot (x_{\max} - x_{\min}) \\ y_{\min} - 0.1 \cdot (y_{\max} - y_{\min}) \\ z_{\min} - 0.1 \cdot (z_{\max} - z_{\min}) \end{pmatrix} \text{ and } \mathbf{d}\mathbf{a} = \begin{pmatrix} 1.2 \cdot (x_{\max} - x_{\min})/4 \\ 1.2 \cdot (y_{\max} - y_{\min})/4 \\ 1.2 \cdot (z_{\max} - z_{\min})/2 \end{pmatrix}. \quad (3.6)$$

and with $x_{\min}, x_{\max}, y_{\min}, y_{\max}, z_{\min}, z_{\max}$ being the minimum and maximum x -, y - and z -coordinates of the hoop centerline. Grid A is made larger than the bounding box to avoid extrapolation artifacts. Set $B = \{\mathbf{b}_{ijk} \mid i, j = 0, 1, 2, 3, 4; k = 0, 1, 2\}$ is generated from set A by translating each point in set A into an isotropically sampled 3D direction by a distance sampled uniformly from $[0, d_{\max}]$. Here d_{\max} is a third of the smallest step size in grid A , i.e. a third of the smallest component of $\mathbf{d}\mathbf{a}$ (see equation (3.6)). The effect of the thin plate spline deformation is shown in figure 3.10. Subfigure a) shows a regular hoop created by equation (3.2). Subfigure b) shows the hoop after the application of the thin plate spline deformation.

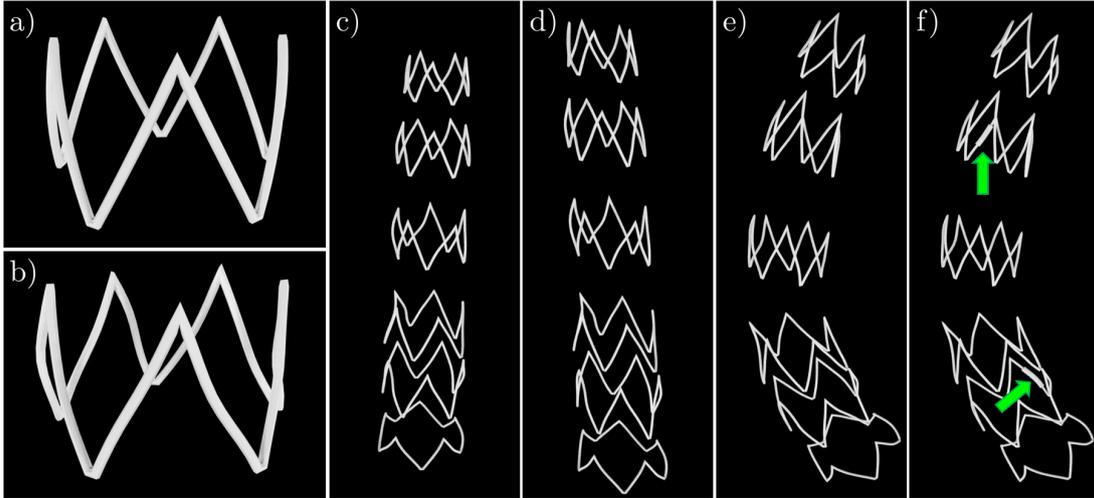


Figure 3.10: Visualization of the steps used to create a 3D model of a stent: a) each hoop is generated as a curved cylinder around a closed curve oscillating on the surface of a conical frustum, b) each hoop is deformed randomly by a thin plate spline, c) multiple stacked hoops form the stent, d) the stent is scaled along an isotropically sampled 3D direction, e) the stent is bent along a randomly sampled spline, f) sleeves (see green arrows) are added to some of the hoops.

iii.) Stacking of Hoops

To create the stent, N_{hoop} hoops were sampled and translated along the z -axis (same z -axis as in equation (3.2)) to form a stack of hoops. The distance $d_{\text{hoop},i}$ between hoops $i - 1$ and i was sampled uniformly between -2% and 120% of the local average hoop height $(H_{\text{hoop},i-1} + H_{\text{hoop},i})/2$. See figure 3.8 for a visual definition of $H_{\text{hoop},i}$ and $d_{\text{hoop},i}$. A stack of hoops is shown in figure 3.10 c).

iv.) Scaling

To include deformations that change the cross section of the stent from circular to elliptical, a scaling along a 3D direction $\mathbf{d} \in \mathbb{R}^3$ was simulated. The direction \mathbf{d} was sampled isotropically. The scaling factor s was sampled uniformly from the real interval $[0.6, 1.5]$. Scaling was implemented by multiplication of the coordinate vectors of the hoop centerline points with the following matrix:

$$A = \begin{pmatrix} e'_{1x} & e'_{2x} & e'_{3x} \\ e'_{1y} & e'_{2y} & e'_{3y} \\ e'_{1z} & e'_{2z} & e'_{3z} \end{pmatrix} \begin{pmatrix} s & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e'_{1x} & e'_{1y} & e'_{1z} \\ e'_{2x} & e'_{2y} & e'_{2z} \\ e'_{3x} & e'_{3y} & e'_{3z} \end{pmatrix}. \quad (3.7)$$

Here, the vectors $\mathbf{e}'_i = (e'_{ix}, e'_{iy}, e'_{iz})^\top$, with $i \in \{1, 2, 3\}$ form a system of orthonormal

basis vectors with $\mathbf{e}'_1 \parallel \mathbf{d}$. The effect of scaling can be observed in figure 3.10, by comparing subfigure c) (the stent before scaling) with subfigure d) (the stent after scaling).

v.) Bending

Since stents follow the form of the surrounding vessel, stents are often not straight, but curved. To simulate this, the stent models are bent so that their centerline is no longer the z -axis but follows a randomly sampled B-spline curve $\gamma(t)$. This bending was implemented by transforming each point $\mathbf{r} = (x, y, z)^\top = \mathbf{e}_x \cdot x + \mathbf{e}_y \cdot y + \mathbf{e}_z \cdot z$ into the point $\mathbf{B}(x, y, z) = \mathbf{e}_a(z) \cdot x + \mathbf{e}_b(z) \cdot y + \mathbf{e}_c(z) \cdot z$. Here, $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$ is the standard basis, which was (implicitly) used before, e.g. in equation (3.2), while $\{\mathbf{e}_a(z), \mathbf{e}_b(z), \mathbf{e}_c(z)\}$ is a set of local, orthonormal basis vectors parameterized by the z -coordinate of \mathbf{r} :

$$\mathbf{e}_c(z) = \frac{\gamma'(z - z_{\min})}{\|\gamma'(z - z_{\min})\|} \quad (\text{normalized tangent vector}) \quad (3.8)$$

$$\mathbf{e}_b(z) = \frac{\gamma''(z - z_{\min})}{\|\gamma''(z - z_{\min})\|} \quad (\text{normalized derivative of tangent vector}) \quad (3.9)$$

$$\mathbf{e}_a(z) = \mathbf{e}_b(z) \times \mathbf{e}_c(z). \quad (3.10)$$

Here the spline curve $\gamma: [0, L] \rightarrow \mathbb{R}^3$ is assumed to be parameterized by its arc-length (L is the length of the curve) and z_{\min} is the minimum of the z -coordinates of all points of all hoops of the stent. The effect of bending can be observed in figure 3.10, by comparing subfigure d) (the stent before bending) with subfigure e) (the stent after bending). The spline curve γ was generated as in section 3.4.1, but with different parameters of the control point-generating random walk (see table 3.4).

Table 3.4: Values of B-spline curve parameters or distribution from which values were sampled.

Parameter	Value or Distribution	Sampled for Each
N_{control}	20	-
l_{step}	$\mathcal{U}(30 \text{ mm}, 50 \text{ mm})$	random walk step
θ	$\mathcal{U}(0^\circ, 30^\circ)$	random walk step
ϕ	$\mathcal{U}(0^\circ, 30^\circ)$	random walk step

$\mathcal{U}(a, b)$ is the uniform distribution on the interval $[a, b] \subset \mathbb{R}$

vi.) Addition of Sleeves

Hoops of real stents sometimes have a metal sleeve around a short section of the hoop. Presumably, this sleeve serves as a connection between the two ends of the wire from which the hoop is manufactured. These sleeves were simulated as short, curved, hollow cylinders with the centerline being a section of the hoop centerline. The

sleeve length was sampled from $\mathcal{U}(0.3 \cdot H_{\text{hoop}}, 0.7 \cdot H_{\text{hoop}})$. The outer diameter D_{outer} was sampled from $\mathcal{U}(1.8 \cdot D_{\text{strut}}, 2.5 \cdot D_{\text{strut}})$. The inner diameter D_{inner} was sampled by first sampling a sleeve thickness t_{sleeve} from $\mathcal{U}(0.07 \text{ mm}, 0.12 \text{ mm})$ and then using $D_{\text{inner}} = \max(D_{\text{outer}} - t_{\text{sleeve}}, D_{\text{strut}})$. The sleeve was placed at a random location along the hoop centerline. To obtain hoops with and without sleeve, the probability for a given hoop to have a sleeve was set to 50%. A stent with sleeves is shown in figure 3.10 f) (green arrows indicate sleeves).

To give an impression of the distribution of the stents, 12 randomly generated stents are shown in figure 3.11.

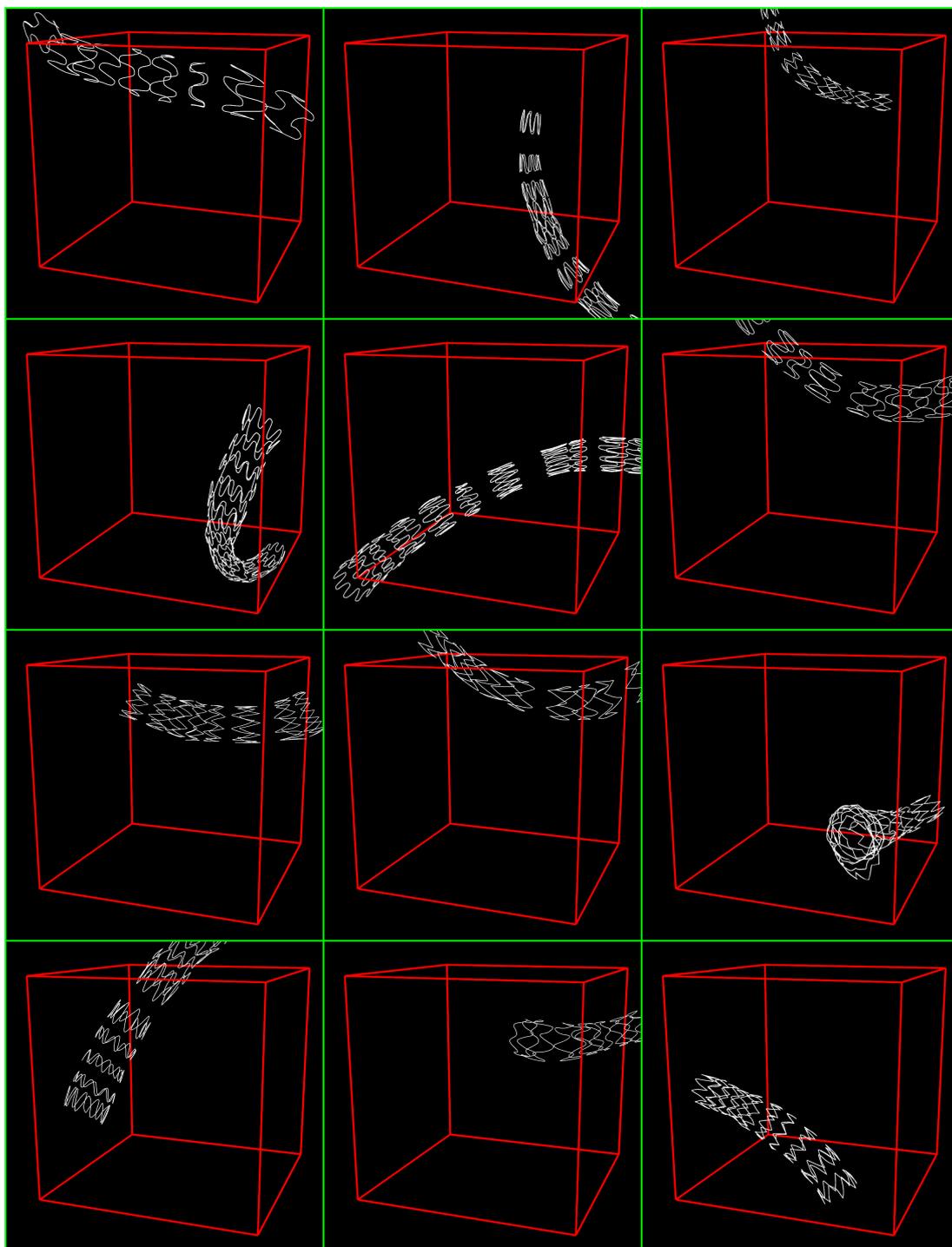


Figure 3.11: 12 randomly generated stents. The red cube has a side length of 153.6 mm, which is the same size as the volume reconstructed by the DTR.

3.5 Deep Tool Extraction (DTE)

The task of the DTE, as proposed by Eulig et al. [34], is to transform an X-ray image of a patient into an image, which is zero everywhere, except in those pixels that contain interventional material. More specifically, the DTE was trained to output in each pixel the intersection length (in mm) of the interventional material with the ray traveling from the X-ray source to that pixel. In this work, DTE was adapted to output the intersection lengths with stents and guidewires in two separate channels of the output image. This separation is advantageous for the subsequent 3D reconstruction (see section 3.1). The application of the DTE to an exemplary X-ray image is shown in figure 3.12.

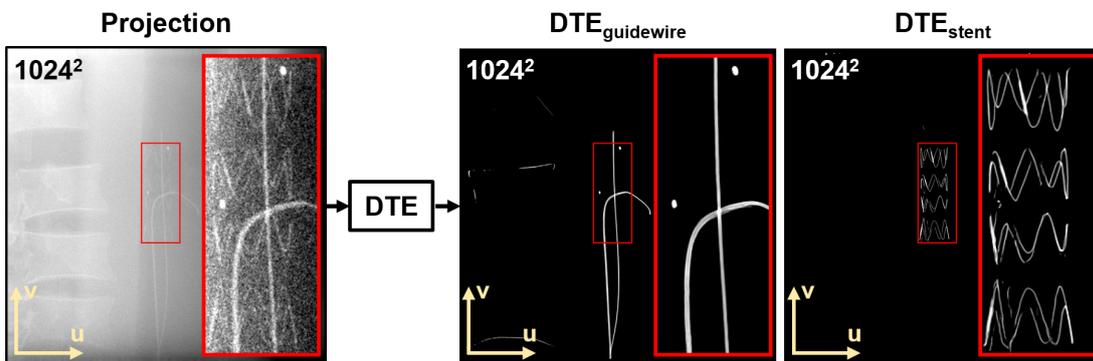


Figure 3.12: The DTE applied to an X-ray image (projection) of an anthropomorphic phantom with two guidewires and a stent. The input and output images have 1024^2 pixels of size 0.3 mm.

The remainder of this section on the DTE details the generation of training data, the CNN architecture and the training process.

3.5.1 Training Data

The training data for the DTE consist of input-target-pairs. Each input is an X-ray image of a patient with stents and guidewires. Each target is an image pair, consisting of the image of the stents and the image of the guidewires. The training data were generated online, i.e. during training, by combining real patient X-ray images and simulated X-ray images of stents and guidewires. This generation of training data is detailed in the following. First, an overview is given by describing the data loader, i.e. the function that generates the random input-target-pairs during training. Then, some details on the components of the data loader are given in section i.) to vi.). Finally, some samples (input-target pairs) generated by the data loader are shown in section vii.).

i.) Overview

The data loader is described in algorithm 1. First, random patient, stent, and guidewire patches are drawn (patches instead of full X-ray images were used to accelerate training). To simulate the blurriness of real X-ray images, the simulated stent and guidewire patches were convolved with an isotropic 2D Gaussian filter. The spatial variations of the contrast of interventional material, that is induced by scattered radiation, was simulated using a small-angle-scatter model (section v.)). To reduce the number of input-target-pairs, in which the contrast-to-noise ratio (CNR) of the interventional material is so low, that the interventional material is invisible or so high, that the task of the DTE becomes very easy, the contrast of the interventional material is adjusted, so that an estimate of the average CNR lies between one and eight. Finally, the sum of the patient, stent and guidewire patches (input patch) and the intersection-length stent and guidewire patches (target patches) are returned.

ii.) Generation of Patient Patches

The generation of random patient patches was based on X-ray images of nine CBCT scans of patients without interventional material. For training, 1793 images from seven scans (1 pelvis, 5 thorax, 1 head) were used. For validation, 958 images from two scans (1 abdomen, 1 thorax) were used. All images had 1024^2 pixels of size 0.3 mm. The generation of a random patient patch from these data is described in algorithm 2.

iii.) Generation of Stent Patches

The generation of random stent patches was based on 160 000 simulated intersection length images of 40 000 3D stent models (four images at 0° , 45° , 90° and 135° per 3D model). These images were split 70–30 between training and validation. The generation of a random stent patch from these images is described in algorithm 3. To increase the ratio of foreground-to-background pixels, the patch was drawn such that the center of mass of the image was inside the patch.

iv.) Generation of Guidewire Patches

The generation of random guidewire patches was based on 160 000 simulated intersection length images of 40 000 3D guidewire models (four images at 0° , 45° , 90° and 135° per 3D model). These images were split 70–30 between training and validation. The generation of a random guidewire patch from these images is described in algorithm 4. To increase the ratio between foreground and background pixels, each patch was generated by adding five patches of five different guidewire images. Since the tips of guidewires are of particular interest to the interventional radiologist, each of these five patches was drawn such that it contains the guidewire tip.

v.) Simulation of Scatter-Induced Contrast Variations

The simulation of the contrast variations induced by scattered radiation is described in algorithm 5. A simple scatter model [64] was used to keep computation fast enough for the online generation of DTE training data.

vi.) Estimation of Contrast and Noise

The estimation of the average CNR in algorithm 1, was based on estimations of the average contrast and the average noise standard deviation. The average contrast was approximated as the median of the heights of the local maxima of 20 line profiles through the interventional material patch (algorithm 6). The estimation of the average noise standard deviation is described in algorithm 7.

vii.) Visualization of Input-Target Pairs

Four exemplary input-target pairs generated by the data loader (algorithm 1) are shown in figure 3.13.

Algorithm 1 DTE Data Loader

```

# Sample a 384 px × 384 px patient, stent and guidewire patch
1:  $p_{\text{patient}} = \text{RANDOMPATIENTPATCH}()$ 
2:  $x_{\text{stent}} = \text{RANDOMSTENTPATCH}()$ 
3:  $x_{\text{guidewire}} = \text{RANDOMGUIDEWIREPATCH}()$ 
# Sample parameters related to blurring, scatter and contrast-to-noise ratio (CNR)
4: Sample  $\sigma_{\text{PSF}}$  from  $\mathcal{U}(0.5 \text{ px}, 1.8 \text{ px})$ 
5: Sample  $k_{\text{pep}}$  from  $\mathcal{U}(0.014, 0.017)$ 
6: Sample  $\text{CNR}_{\text{target,stent}}$  and  $\text{CNR}_{\text{target,guidewire}}$  independently from the piecewise
uniform distribution with probability density function  $f(t) = \begin{cases} 0.28, & \text{if } 1 \leq t < 2 \\ 0.12, & \text{if } 2 \leq t < 8 \\ 0, & \text{otherwise} \end{cases}$ 
7: for type in {guidewire, stent} do
# 2D Gaussian as point spread function (PSF)
8:  $p_{\text{type}} = G(\sigma_x = \sigma_y = \sigma_{\text{PSF}}) * x_{\text{type}}$ 
# Scatter induces variations of tool contrast
9:  $p_{\text{type}} = \text{TOOLCONTRASTWITHSCATTER}(p_{\text{patient}}, p_{\text{type}}, k_{\text{pep}})$ 
# Estimate contrast-to-noise ratio
10:  $C = \text{MEDIANLOCALMAXIMUMESTIMATION}(p_{\text{type}})$ 
11:  $N = \text{AVERAGENOISESTANDARDDEVIATIONESTIMATION}(p_{\text{patient}})$ 
12:  $\text{CNR} = C/N$ 
# Adjust contrast-to-noise ratio
13:  $p_{\text{type}} = \text{CNR}_{\text{target,type}}/\text{CNR} \cdot p_{\text{type}}$ 
# Pixel values are the negative logarithm of normalized intensities
14:  $\text{inputPatch} = p_{\text{patient}} + p_{\text{stent}} + p_{\text{guidewire}}$ 
# Pixel values are intersection lengths (in mm) between the X-rays and the tools
15:  $\text{targetPatches} = (x_{\text{stent}}, x_{\text{guidewire}})$ 
16: return (inputPatch, targetPatches)

```

$G(\sigma_x = \sigma_y = x)$ is an isotropic 2D Gaussian filter kernel

$*$ is a 2D convolution

$\mathcal{U}(a, b)$ is the uniform distribution on the interval $[a, b] \subset \mathbb{R}$

$(., .)$ is a pair

Algorithm 2 Sampling of Random Patient Patch

```

1: function RANDOMPATIENTPATCH()
    # Pixel values are the negative logarithm of normalized intensities
2:   Choose a patient X-ray image  $q_{\text{patient}}$  at random.
3:   Choose a random  $384 \text{ px} \times 384 \text{ px}$  patch  $p_{\text{patient}}$  from  $q_{\text{patient}}$ .
    # Data augmentation
4:   With a probability of 0.5, flip  $p_{\text{patient}}$  laterally (the patient's right side becomes
    the left side and vice versa; only lateral flipping, since humans are roughly
    symmetric with respect to the median plane).
5:   return  $p_{\text{patient}}$ 

```

Algorithm 3 Sampling of Random Stent Patch

```

1: function RANDOMSTENTPATCH()
    # Pixel values are intersection lengths (in mm) between the X-rays and the
    stent
2:   Choose a stent image  $q_{\text{stent}}$  at random.
3:   Choose a  $384 \text{ px} \times 384 \text{ px}$  patch  $x_{\text{stent}}$  from  $q_{\text{stent}}$  randomly, so that the center
    of mass of  $q_{\text{stent}}$  is contained in  $x_{\text{stent}}$ .
    # Data augmentation
4:   With a probability of 0.5, flip  $x_{\text{stent}}$  about the  $v$ -axis.
5:   With a probability of 0.5, flip  $x_{\text{stent}}$  about the  $u$ -axis.
6:   return  $x_{\text{stent}}$ 

```

Algorithm 4 Sampling of Random Guidewire Patch

```

1: function RANDOMGUIDEWIREPATCH()
    # Pixel values are intersection lengths (in mm) between the X-rays and the
    # guidewire
2:    $x_{\text{guidewire}}$  is a  $384 \text{ px} \times 384 \text{ px}$  array initialized with zeros.
3:   for  $n = 0, \dots, 4$  do
4:     Choose a guidewire image  $q_{\text{guidewire},n}$  at random.
5:     Choose a  $384 \text{ px} \times 384 \text{ px}$  patch  $x_{\text{guidewire},n}$  from  $q_{\text{guidewire},n}$  randomly, so that
    the guidewire tip is contained in  $x_{\text{guidewire},n}$ .
    # Data augmentation
6:     With a probability of 0.5, flip  $x_{\text{guidewire},n}$  about the  $v$ -axis.
7:     With a probability of 0.5, flip  $x_{\text{guidewire},n}$  about the  $u$ -axis.
8:      $x_{\text{guidewire}} = x_{\text{guidewire}} + x_{\text{guidewire},n}$ 
9:   return  $x_{\text{guidewire}}$ 

```

Algorithm 5 Scatter-Induced Changes in Tool Contrast

```
1: function TOOLCONTRASTWITHSCATTER( $p_{\text{patient}}, p_{\text{tool}}, k_{\text{pep}}$ )  
   # Scatter can be approximated as a smoothed, pep( $p \cdot e^{-p}$ )-like image [64]  
2:  $I_{\text{scatter}} = G(\sigma_x = \sigma_y = 167 \text{ px}) * (k_{\text{pep}} \cdot p_{\text{patient}}^{1.2} \cdot \exp(-p_{\text{patient}})^{0.8})$   
   # low-pass filter (LP) to eliminate noise  
3:  $I_{\text{patient,LP}} = G(\sigma_x = \sigma_y = 5 \text{ px}) * \exp(-p_{\text{patient}})$   
   # Avoid extreme scatter-to-primary ratios  
4:  $I_{\text{scatter}} = \min(I_{\text{scatter}}, 0.8 \cdot I_{\text{patient,LP}})$   
   # Compute tool contrast image  $p_{\text{tool}}$  assuming  $I_{\text{patient,LP}}$  already contains  
   scatter  
5:  $I_{\text{total}} = (I_{\text{patient,LP}} - I_{\text{scatter}}) \cdot \exp(-p_{\text{tool}}) + I_{\text{scatter}}$   
6:  $p_{\text{tool}} = -\log(I_{\text{total}}) + \log(I_{\text{patient,LP}})$   
7: return  $p_{\text{tool}}$ 
```

$G(\sigma_x = \sigma_y = x)$ is an isotropic 2D Gaussian filter kernel

* is a 2D convolution

Algorithm 6 Median Local Maximum Estimation

```
1: function MEDIANLOCALMAXIMUMESTIMATION( $p$ )  
2:   lineProfiles is a list containing 10, randomly chosen, non-empty rows of the  
   image  $p$  and 10, randomly chosen, non-empty columns of the image  $p$ .  
3:   localMaxima is an empty list.  
4:   for each lineProfile in lineProfiles do  
5:     Append the heights of the local maxima of lineProfile to localMaxima.  
6:   return median(localMaxima)
```

Algorithm 7 Average Noise Standard Deviation Estimation

```

1: function AVERAGENOISESTANDARDDEVIATIONESTIMATION( $p$ )
   # Low-pass (LP) and high-pass (HP) filtering using a 2D Gaussian
2:   LP =  $G(\sigma_x = \sigma_y = 3 \text{ px}) * p$ 
3:   HP =  $p - \text{LP}$ 
   # Approximate relationship between gray value and noise standard deviation
4:    $x$  is an array of 50 floating-point numbers.
5:    $y$  is an array of 50 floating-point numbers.
6:   for  $k = 0, \dots, 49$  do
7:      $t_{\text{low}} = \min(\text{LP}) + k \cdot (\max(\text{LP}) - \min(\text{LP}))/50$ 
8:      $t_{\text{high}} = \min(\text{LP}) + (k + 1) \cdot (\max(\text{LP}) - \min(\text{LP}))/50$ 
9:     indices =  $\{(i, j) \mid t_{\text{low}} < \text{LP}_{i,j} \leq t_{\text{high}} \wedge i \in \{1, \dots, N_u\} \wedge j \in \{1, \dots, N_v\}\}$ 
10:     $x_k = \text{mean}(\{\text{LP}_{i,j} \mid (i, j) \in \text{indices}\})$ 
11:     $y_k = \text{std}(\{\text{HP}_{i,j} \mid (i, j) \in \text{indices}\})$ 
12:     $f$  is the function that interpolates the set of data points  $\{(x_1, y_1), \dots, (x_{50}, y_{50})\}$ 
       linearly.
       # Apply  $f$  pixelwise to the low-pass filtered image
13:    stdImage =  $f \odot \text{LP}$ 
14:   return mean(stdImage)

```

 $G(\sigma_x = \sigma_y = x)$ is an isotropic 2D Gaussian filter kernel* is a 2D convolution

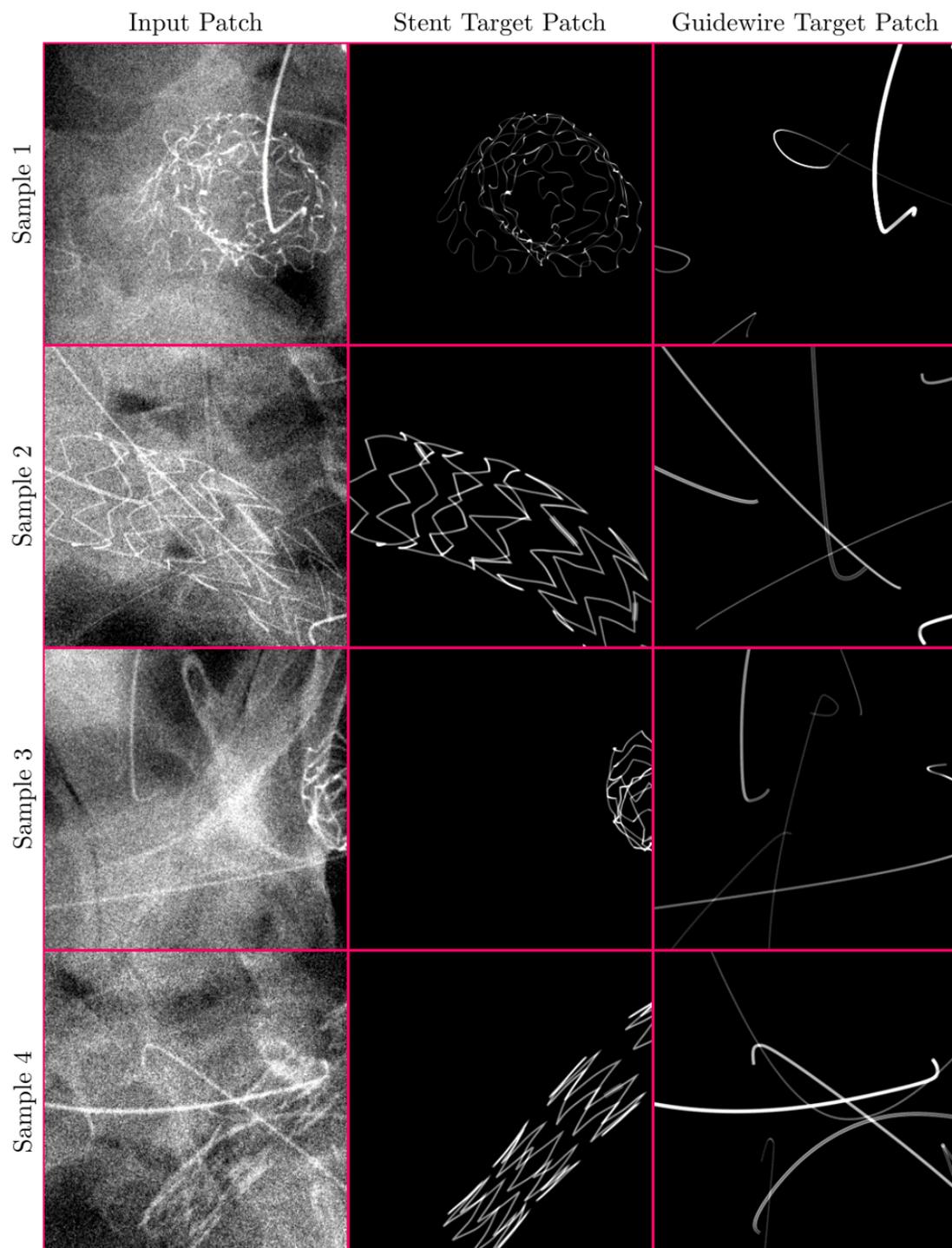


Figure 3.13: Four samples of input and target patches generated by the DTE data loader. During training, X-ray images of patients are used by the data loader to generate the input patches. To make this figure, X-ray images of an anthropomorphic phantom were used. The input patches were windowed to the range [1st percentile, 99th percentile]. The target patches were windowed to the range [0 mm, 1 mm]. All patches have $384 \text{ px} \times 384 \text{ px}$.

3.5.2 Architecture

The CNN architecture used to implement the DTE is similar to the U-Net [65]: two (3×3 convolution + batch normalization + ReLU)-blocks per resolution stage in the encoder and decoder, downsampling by 2×2 max-pooling, upsampling by nearest-neighbor interpolation followed by 3×3 convolution, number of feature channels of the convolution layers = $64 \cdot 2^n$, with the resolution-stage-index $n = 0, \dots, 6$. The final 64-channel layer of the decoder is the input of two separate 1×1 convolution layers, which output the stent and guidewire images respectively.

3.5.3 Training Process

Training was performed for 200 epochs, with each epoch consisting of 16 000 training and 4000 validation samples. As the loss function, the mean absolute error was used. Optimization was performed by the Adam optimizer [61] (learning rate = 1×10^{-4} , $\beta_1 = 0.9$, $\beta_2 = 0.999$) with a mini-batch size of 8. Training took about 78 h on an NVIDIA RTX 6000 GPU.

3.6 Deep Tool Reconstruction (DTR)

In this thesis, a DTR is a CNN trained to transform backprojection volumes $\text{BP}_k(t_{\text{input}})$ of T_{input} time steps $t_{\text{input}} \in \{t - T_{\text{input}} + 1, \dots, t\}$ into a 3D segmentation $\text{DTR}_k(t)$ of the interventional material at the current time step t . Each backprojection $\text{BP}_k(t_{\text{input}})$ is computed by applying the DTE to the G X-ray images acquired in time step t_{input} and backprojecting the G DTE output images into a volume. Some of the DTRs in this thesis are specific to a single type $k \in \{\text{guidewire}, \text{stent}\}$ of interventional material. In this case, the DTE output channel of the respective type k is used to compute the backprojections $\text{BP}_k(t_{\text{input}})$. Other DTRs are trained to reconstruct stents and guidewires simultaneously ($k = \text{total}$). In this case, the sum of the stent and guidewire DTE channels was used to compute $\text{BP}_k(t_{\text{input}})$. Besides the usual SCBP, that backprojects the G DTE output images into the same volume, the PPBP, that backprojects the G DTE output images into separate volumes (equivalently: different channels of the same volume) was used. All but one of the DTRs are used to compute high resolution reconstructions for display to the interventional radiologist (input and output volumes have 256^3 voxels of size 0.6 mm). The exception is an auxiliary DTR, which is used in the motion compensation step (section 3.7). This auxiliary DTR processes volumes with lower resolution (64^3 voxels of size 2.4 mm) to save computation time. The inputs and the output of the stent DTR of a two-threaded pipeline ($G = 2$, $T_{\text{input}} = 4$, $k = \text{stent}$, PPBP) are visualized in figure 3.14.

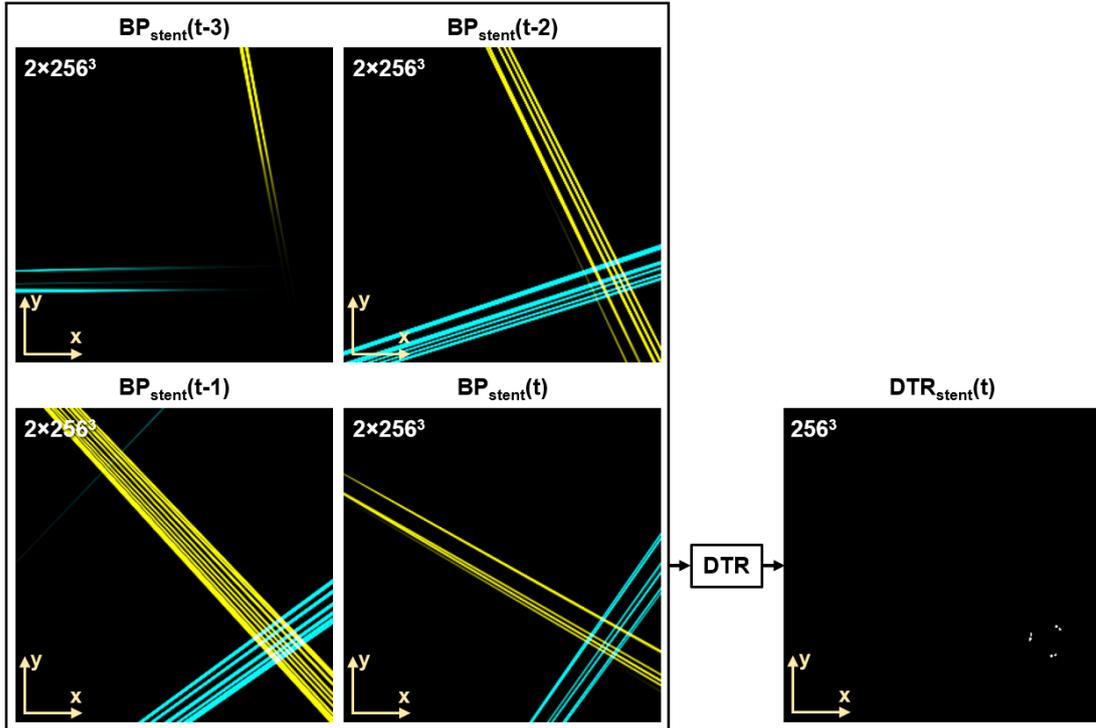


Figure 3.14: Input volumes and output volume of a stent DTR ($G = 2$, $T_{input} = 4$, $k = \text{stent}$, PPBP), when applied to measured data of a stent in an anthropomorphic phantom. Note the changing direction of the backprojected rays, which is induced by the rotation of the X-ray system (19° between time steps). Only a transversal (x - y) slice of each volume is shown. All volumes have 256^3 voxels of size 0.6 mm. The channels of the backprojection volumes are indicated by colors (yellow, cyan).

3.6.1 Training Data

The training data for the DTRs consist of input-target-pairs. Each input is a tuple of backprojection volumes of different time steps. Each target is a segmentation volume (background = 0, interventional material = 1). The input volumes were generated by backprojecting forward projections of simulated 3D models of stents and/or guidewires. To make the DTR robust against errors of the DTE, simulated false positives and false negatives were added to the forward projections. False positives were simulated by adding forward projections of randomly generated, curved cylinders (similar to short guidewires or stent struts) to the forward projections. False negatives were simulated by setting the pixel values inside random rectangles to zero. The false positives and false negatives were sampled independently for each projection. The target volumes were computed from the triangle mesh models of the interventional material by voxelization. Voxelization refers to the process of generating a volume, that is zero everywhere, except

in those voxels, that are fully inside the triangle mesh or that are intersected by a triangle of the mesh. The different steps of generating training data for a guidewire DTR are illustrated in figure 3.15.

For DTRs, that take backprojections of multiple time steps as input, some motion of the interventional material between the time steps had to be simulated. To simulate respiratory motion, the interventional material at time step t_{input} was generated by translating the triangle mesh by the vector $\mathbf{x}(t_{\text{input}}) = \mathbf{d} \cdot \sin(t_{\text{input}} \cdot \Delta t \cdot f + \phi)$. Here, the length of a time step Δt was set to 0.2 s, the 3D unit vector \mathbf{d} was sampled isotropically, the amplitude A was sampled uniformly from [15 mm, 25 mm], the frequency f was sampled uniformly from [6 rpm, 30 rpm] and the phase ϕ was sampled uniformly from $[0, 2\pi)$.

The dataset for the low-resolution auxiliary DTR contained 200 000 samples, with each sample containing the full 64^3 volumes. The datasets for all other DTRs contained 40 000 samples, with each sample containing only randomly drawn 128^3 patches of the full 256^3 volumes. All datasets were split 70–30 between training and validation.

Each sample in the datasets of the stent DTRs contained one stent. Each sample in the datasets for the guidewire DTRs contained a random number of guidewires, which was drawn uniformly from [1, 5]. Each sample in the datasets for the total DTRs contained one stent and a random number of guidewires, which was drawn uniformly from [1, 5]. For each sample of each dataset, new random 3D models of the interventional material were sampled.

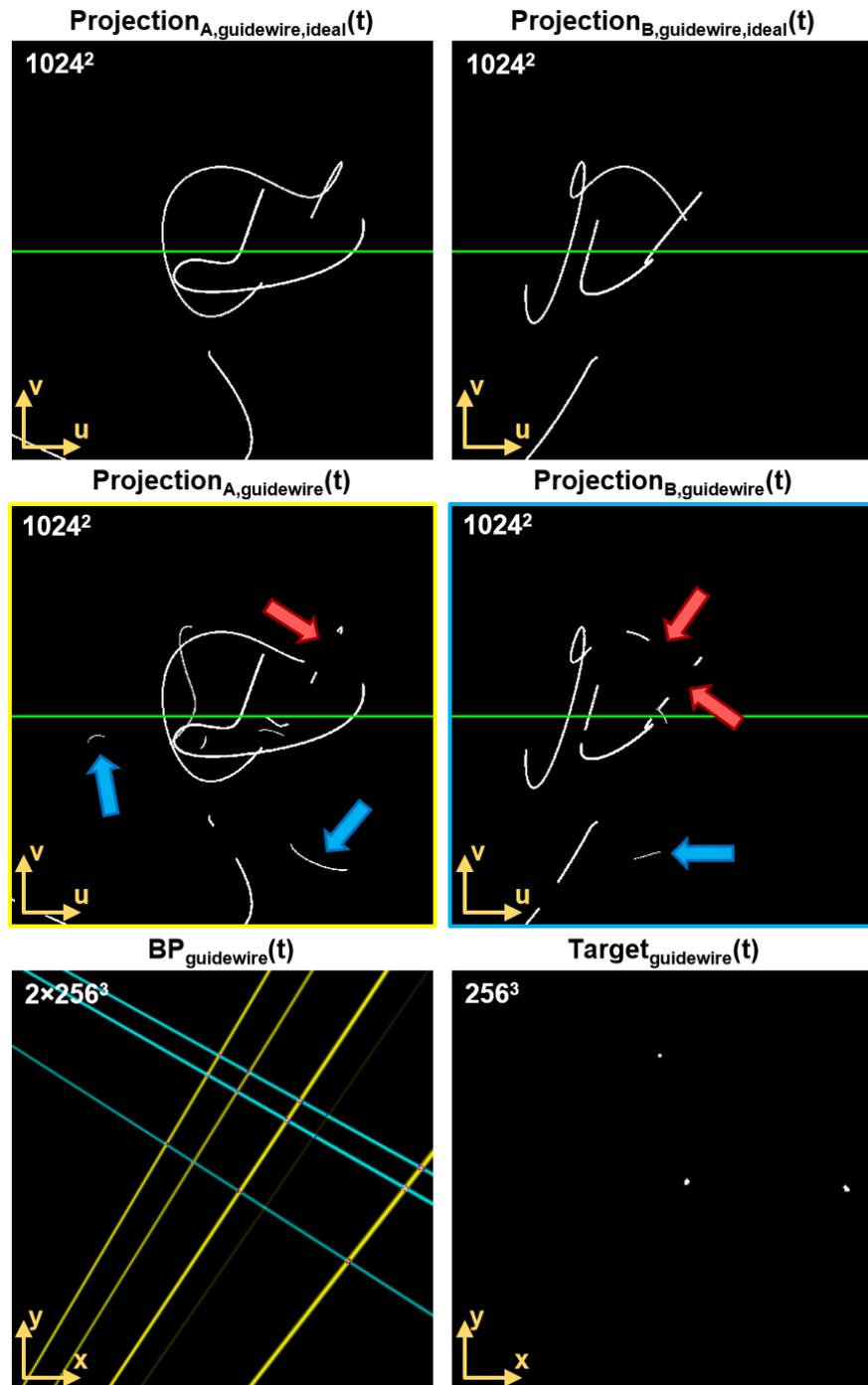


Figure 3.15: Visualization of the generation of training data for a guidewire DTR with two orthogonal imaging threads (A,B). Top row: (ideal) forward projections of three randomly generated guidewires. Center row: the forward projections with simulated false positives (blue arrows) and false negatives (red arrows). Bottom row, left: transversal (x - y) slice through the PPBP of the central row. Bottom row, right: transversal slice through a segmentation volume of the guidewires. The transversal slices are the central transversal slices of the respective volumes. Therefore only the central rows (green) of the projections contribute to the backprojection.

3.6.2 Architecture

The CNN architecture used to implement the DTRs is similar to the 3D U-Net [66]: two ($3 \times 3 \times 3$ convolution + batch normalization + ReLU)-blocks per resolution stage in the encoder and decoder, downsampling by $2 \times 2 \times 2$ max-pooling, upsampling by nearest-neighbor interpolation followed by $3 \times 3 \times 3$ convolution, number of feature channels of the convolution layers = $16 \cdot 2^n$, with the resolution-stage-index n ranging from 0 to 3 (auxiliary DTR) or 5 (all other DTRs). The input volumes were passed as separate input channels. The final layer was a sigmoid activation layer.

3.6.3 Training Process

Training was performed for 200 epochs, with each epoch consisting of 16 000 training and 4000 validation samples. As the loss function, the soft Dice loss with Laplacian smoothing L was used:

$$L(X, Y) = 1 - \frac{2 \cdot (\sum_i x_i \cdot y_i) + 1}{(\sum_i x_i + y_i) + 1}. \quad (3.11)$$

Here, X is the output volume of the DTR (voxel values lie between 0 and 1), Y is the target segmentation volume (voxel values are either 0 or 1) and x_i and y_i are the values of the i th voxels of X and Y . In the sums, the index i runs over all voxels of the volume (not just over one dimension). Optimization was performed by the Adam optimizer (learning rate = 1×10^{-4} , $\beta_1 = 0.9$, $\beta_2 = 0.999$) with a mini-batch size of 8. Training of the auxiliary DTR took about 4 days on an NVIDIA RTX 6000 GPU. Training of all other DTRs took about 18 days each on an NVIDIA RTX 6000 GPU.

3.7 Stent Motion Compensation

Some of the reconstruction pipelines investigated in this thesis perform a motion compensation between the (per-projection) stent backprojections $\text{BP}_{\text{stent}}(t_{\text{input}})$ of different time steps $t_{\text{input}} \in \{t - T_{\text{input}} + 1, \dots, t\}$, which are the input to the stent DTR at time step t . The backprojections $\text{BP}_{\text{stent}}(t_{\text{input}})$ differ not only in the depicted time step, but also in the angle of the backprojected rays (the X-ray system rotates by some angle $\Delta\alpha$ between adjacent time steps). The motion compensation was performed by STs [59] ST_i , $i \in \{1, \dots, T_{\text{input}} - 1\}$. A ST is a NN, that estimates and applies a spatial transformation. The part of a ST, that estimates the transformation parameters is called localization network (LocNet). Here, each ST_i takes $\text{BP}_{\text{stent}}(t - i)$ and $\text{BP}_{\text{stent}}(t)$ as input, estimates the transformation that describes the motion of the interventional material between time steps $t - i$ and t , and applies this transformation to $\text{BP}_{\text{stent}}(t - i)$. ST_1 trained for a pipeline with $\Delta\alpha = 19^\circ$ is visualized in figure 3.16. In this thesis, the transformation performed by the STs was chosen to be a global translation, i.e. the localization networks estimate a 3D translation vector $(\Delta x, \Delta y, \Delta z)$. More complex transformations (e.g. affine or thin plate spline) could be learned [59], should this prove

to be advantageous.

To keep computation times low, downsampled versions $\text{BP}_{\text{stent,lr}}(t_{\text{input}})$ (64^3 voxels of size 2.4 mm) of the backprojections $\text{BP}_{\text{stent}}(t_{\text{input}})$ (256^3 voxels of size 0.6 mm) were passed to the localization networks. Since it was found to improve the accuracy of the translation estimation significantly, low resolution, auxiliary stent reconstructions $\text{DTR}_{\text{stent,lr,aux}}(t-i)$ and $\text{DTR}_{\text{stent,lr,aux}}(t)$ were passed as additional inputs to LocNet_i (see figure 3.16). These auxiliary stent reconstructions were computed by an auxiliary, low-resolution (64^3 voxels of size 2.4 mm) stent DTR, which was trained to reconstruct stents from the PPBP of two orthogonal stent DTE output images.

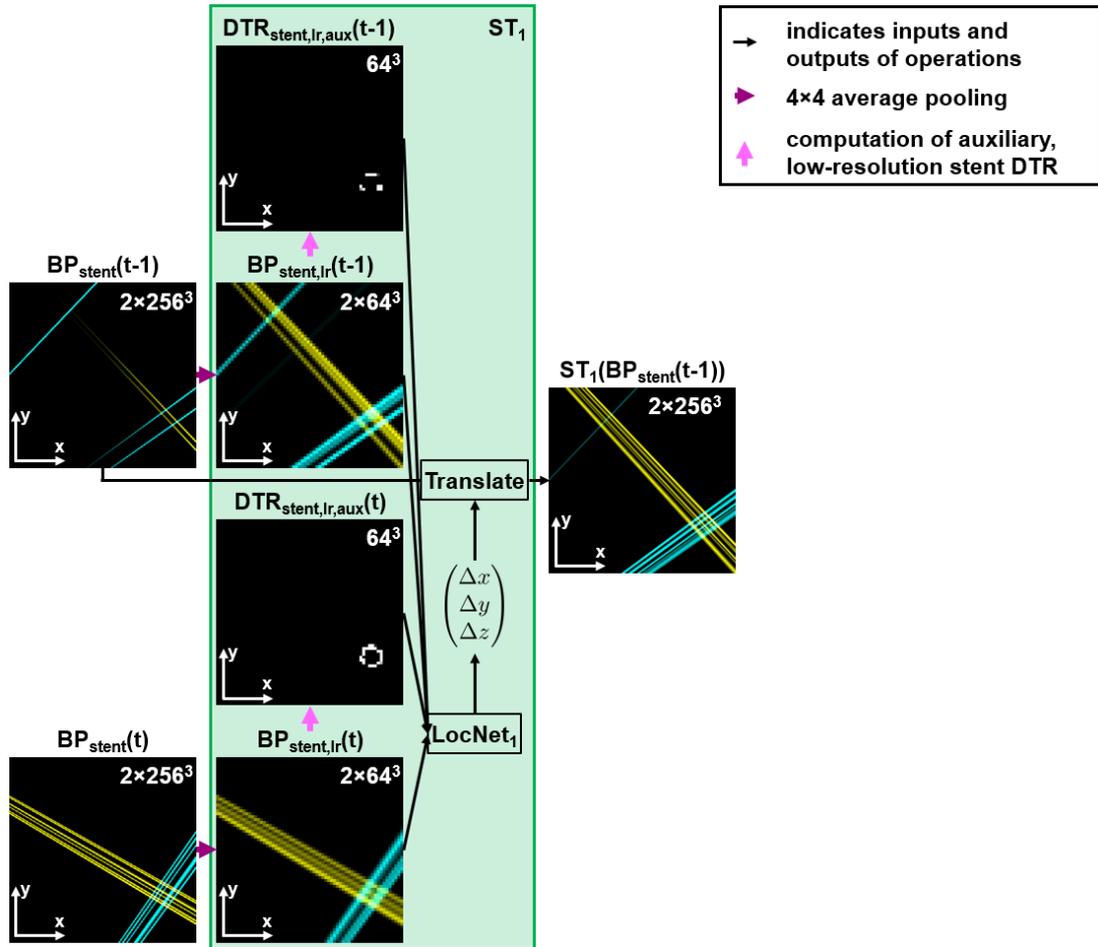


Figure 3.16: Visualization of a spatial transformer network ST_1 , that estimates the translation vector $(\Delta x, \Delta y, \Delta z)$ between time steps $t-1$ and t from the backprojection volumes $\text{BP}_{\text{stent}}(t-1)$ and $\text{BP}_{\text{stent}}(t)$ and applies this translation to $\text{BP}_{\text{stent}}(t-1)$. In this case, the rotation of the X-ray system between adjacent time steps was $\Delta\alpha = 19^\circ$ (compare angles of backprojected rays).

3.7.1 Training Data

The training data for the localization networks LocNet_i consist of input-target-pairs. Each input is a pair of backprojection volumes ($\text{BP}_{\text{stent}}(t-i), \text{BP}_{\text{stent}}(t)$). Each target is a translation vector $(\Delta x, \Delta y, \Delta z)$. These data were simulated just as the training data for the DTRs (section 3.6.1), with two differences: a) the backprojections had lower resolution (64^3 voxels of size 2.4 mm) and b) targets were not voxelization volumes, but the translation vectors of the simulated patient motion (section 3.6.1). For each localization network LocNet_i , a dataset of 200 000 samples (70–30 training–validation split) was simulated. Each sample contained one stent.

3.7.2 Architecture

The localization networks were implemented by an architecture similar to a 3D version of VGG13 [58]: two ($3 \times 3 \times 3$ convolution + batch normalization + ReLU)-blocks per resolution stage, downsampling by $2 \times 2 \times 2$ max-pooling, number of feature channels of the convolution layers = $16 \cdot 2^n$, with the resolution-stage-index $n = 0, \dots, 4$, three fully-connected layers with sizes 2048, 1024 and 3. The input volumes were passed as separate input channels.

3.7.3 Training Process

Training was performed for 200 epochs, with each epoch consisting of 16 000 training and 4000 validation samples. As the loss function, the mean absolute error was used. Optimization was performed by the Adam optimizer (learning rate = 1×10^{-4} , $\beta_1 = 0.9$, $\beta_2 = 0.999$) with a mini-batch size of 8. Training of a localization network took about 2.75 days on an NVIDIA RTX 6000 GPU.

3.8 Phantom Measurements

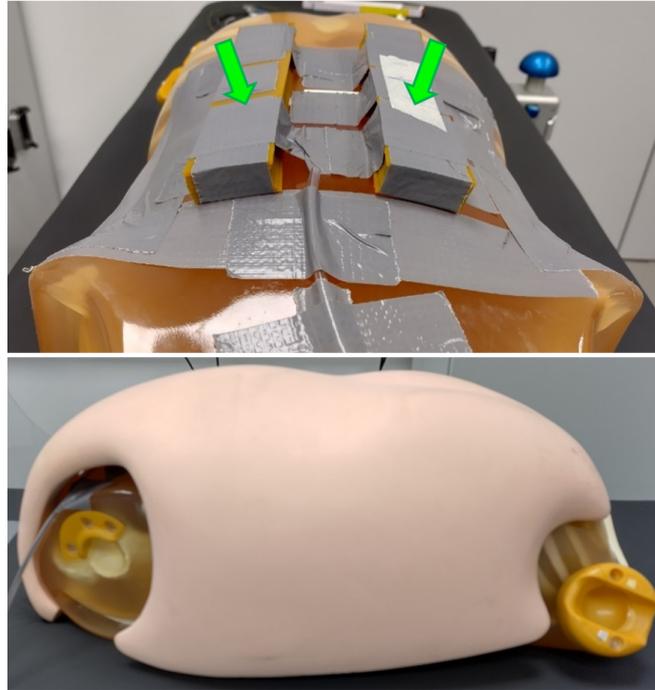


Figure 3.17: The anthropomorphic phantom (PBU-50, Kyoto Kagaku Co. Ltd., Japan) used for the measurements in this thesis. Top: inferior-superior view. The interventional material (not in the picture) was placed over the abdomen of the phantom, between the two polyurethane rubber spacers (green arrows). When the interventional material scene did not contain a stent, the spacers were removed. Bottom: lateral view of the phantom with the soft tissue-equivalent extension on the abdomen. Unlike in this picture, the phantom was in a prone position during all measurements.

The performance of the reconstruction pipelines was evaluated on measured data of an anthropomorphic phantom (PBU-50, Kyoto Kagaku Co. Ltd., Japan; figure 3.17) with respiratory motion. The interventional material was placed on the abdomen, between the phantom and a soft tissue-equivalent extension (figure 3.17).

The data were measured using the experimental X-ray system described in section 3.3 in a stop motion fashion, i.e. a series of N_t 360° scans with fine angular sampling was acquired with motion happening not during, but between scans. Retrospectively, G -tuples of X-ray images were extracted from each 360° scan according the respective scan sequence (equation (3.1) and table 3.1). Compared to acquiring these G -tuples in real time, this stop motion approach has the advantages that a) ground truth reconstructions can be reconstructed from conventional CT reconstructions of the 360°

Table 3.5: Arrangements of interventional material measured in stop motion scans.

ID	Description	PU-spacers
a	gw1	no
b	gw1, gw2	no
c	st	yes
d	st, gw1	yes
e	st, gw1, gw2, ca	yes

For stent measurements, polyurethane (PU) spacers were placed between phantom and soft tissue-equivalent extension.

scans and b) a one-threaded X-ray system can be used to simulate X-ray systems with multiple imaging threads.

Stop motion scans of interventional material arrangements of different complexity were acquired (table 3.5). The pieces of interventional material are described in table 3.6. All 360° scans were acquired with a tube voltage of 100 kV and a pulse duration of $T_{\text{pulse}} = 20$ ms. Each arrangement was measured with different tube currents $I \in \{7.5 \text{ mA}, 30 \text{ mA}, 120 \text{ mA}\}$, i.e. different tube current time products $Q = I \cdot T_{\text{pulse}} \in \{0.15 \text{ mA s}, 0.60 \text{ mA s}, 2.40 \text{ mA s}\}$ to investigate the influence of the tube current time product on reconstruction quality. Two kinds of motion were simulated. Guidewire feed was simulated by pushing one of the guidewires (gw1, see table 3.6) into the phantom by about 4 mm between 360° scans. Respiratory motion was simulated by translating the phantom in the superior-inferior (z) direction (using the longitudinal slide of the patient table), so that the z -coordinate of the t th 360° scan was given by:

$$z(t) = A \cdot \cos(2\pi \cdot 0.25 \text{ Hz} \cdot t \cdot 0.2 \text{ s}), \quad (3.12)$$

with $t \in \{0, \dots, 10\}$. The arrangements c, d and e were acquired with two different respiration amplitudes $A = 11$ mm and $A = 22$ mm. The extent of the simulated respiratory motion with $A = 11$ mm (extent = twice the amplitude, i.e. 22 mm) is roughly similar to the mean extent of respiratory motion of upper abdominal arteries [67], of the kidneys [68] and of the pancreas [69]. The extent of the simulated respiratory motion with $A = 22$ mm (extent = twice the amplitude, i.e. 44 mm) is roughly similar to the maximum extent of respiratory motion of the above mentioned organs [67]–[69]. The frequency of the simulated respiratory motion ($0.25 \text{ Hz} = 15 \text{ min}^{-1}$) is typical of adults at rest [70]. Arrangements a and b were only acquired with $A = 11$ mm due to technical difficulties. All stop motion scans consisted of $N_t = 11$ 360° scans. Each 360° scan consisted of about 208 X-ray images with an angular increment of 1.73° (rotation time $T_{\text{rot}} = 25$ s, acquisition of images at 8.3 fps). The rotational blur of all acquired X-ray images was 0.29° ($360^\circ \cdot T_{\text{pulse}}/T_{\text{rot}}$).

Although the stop motion scans consisted of only 11 360° scans, i.e. 11 time steps t , they

were extended to 60 time steps by temporal mirroring and repetition of the measured sequence of 360° scans. The reconstruction pipelines were evaluated on these extended scans for better statistics. The graphs of the resulting z -position of the phantom and the guidewire feed are depicted in figure 3.18.

For application of the pipelines to these extended stop motion scans, the angular increments $\Delta\alpha$, which were used for selecting X-ray images from the 360° scans, were chosen slightly different from the values $\Delta\alpha \in \{9.5^\circ, 19^\circ, 28.5^\circ\}$ used during the training of the DTRs, to make the distribution of the angles visited by each imaging thread during the 60 time steps more uniform. In particular 9.31° was used instead of 9.50° , 18.30° was used instead of 19.00° and 27.93° was used instead of 28.50° . These angles lead to a much more uniform distribution of the angles visited by each imaging thread during the 60 time steps (figure 3.19). This uniformity has two advantages. First, it makes the evaluation of pipelines less dependent on the starting angle α_0 (equation (3.1)). Second, it makes the evaluations of pipelines with different angular increments more comparable.

Table 3.6: Interventional material used for the stop motion scans.

ID	Type	Make and Model	Description
gw1	guidewire	Terumo Radiofocus TM Guide Wire M	$D = 0.70$ mm, angled tip
gw2	guidewire	unknown	$D = 0.83$ mm, straight tip
ca	catheter	Terumo Radiofocus TM Glidecath vertebral	$D = 1.40$ mm, angled tip
st	stent	contralateral leg of the Jotec [®] E-tegra stent graft system	$D = 16$ mm, $D_{\text{strut}} = 0.36$ mm

D : outer diameter. For guidewires, it is measured at the tip.

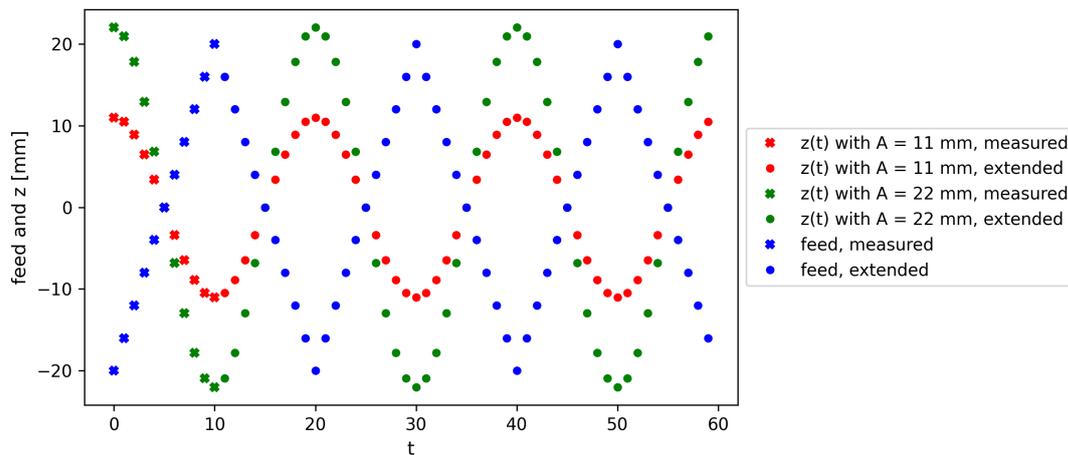


Figure 3.18: Superior-inferior (z) position of the phantom and guidewire feed over the time step t of the stop motion scan. Only the first 11 360° scans (time steps) were measured. The other time steps were generated by temporal mirroring and repetition.

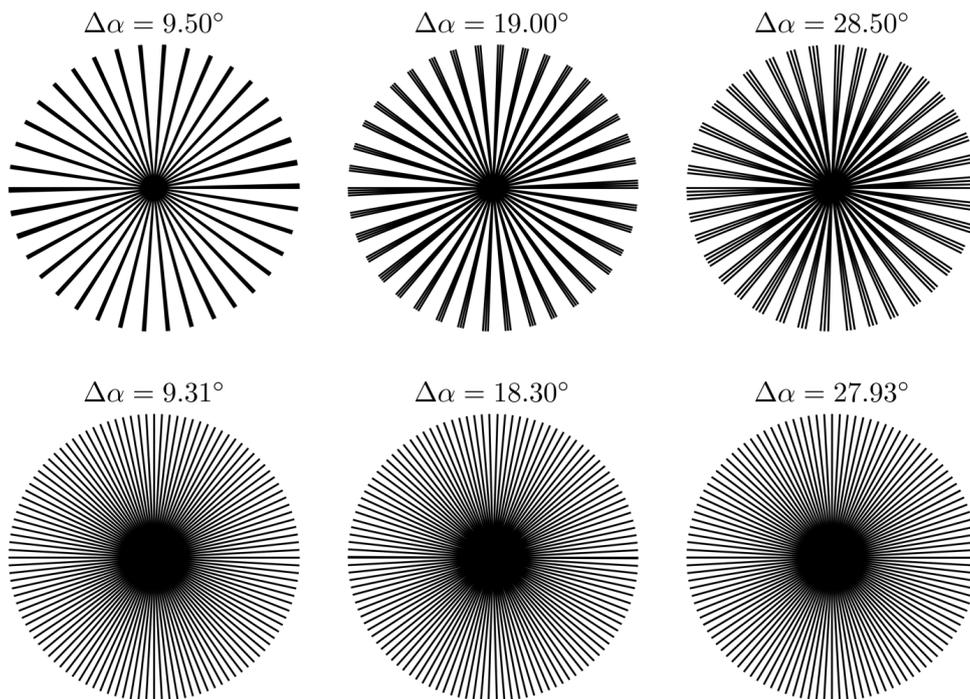


Figure 3.19: Angles $\alpha \in \{\Delta\alpha \cdot t \mid t = 1, \dots, 59\}$ visited by an imaging thread of reconstruction pipelines with different angular increments $\Delta\alpha$ during 60 time steps. Top row: angular increments used during training. Bottom row: angular increments used for application of trained pipelines to stop motion scans (of length 60 time steps). Each angle is represented by a diameter, instead of a radius, since X-ray images differing by 180° are similar to each other.

3.9 Evaluation on Measured Data

To evaluate the quality of 3D reconstructions of the investigated reconstruction pipelines, ground truth 3D reconstructions were used. The ground truths were generated by thresholding conventional CT reconstructions, which were computed from all of the X-ray images of the respective 360° scans of the stop motion scan. The CT reconstructions were computed by the algorithm of Feldkamp, Davis and Kress [18].

Since the task of the DTRs is formulated as a segmentation, using the Sørensen–Dice coefficient to quantify the similarity of a binary 3D reconstruction Y and a binary ground truth 3D reconstruction GT , would be an obvious choice. A voxelwise comparison of Y and GT is however not useful here, since GT only contains thin structures (guidewires and stent struts), whose diameter is uncertain, since small changes of the above-mentioned threshold can lead to large changes in the diameter (e.g. with a voxel size of 0.6 mm, is the cross-section of a 0.7 mm guidewire 1×1 voxels or 2×2 voxels?). Therefore, custom metrics based on skeletons $S(Y)$ and $S(GT)$ were used. Skeleton here refers to the set of voxels in a binary voxel volume, that form a one-voxel-thick representations of the centerline of the non-zero voxels in the volume. Skeletons were computed by the algorithm of Lee et al. [71], implemented in the `skeletonize_3d` function of `scikit-image` 0.17.2 [72]. A skeleton of an exemplary 3D reconstruction of a stent is shown in figure 3.20.

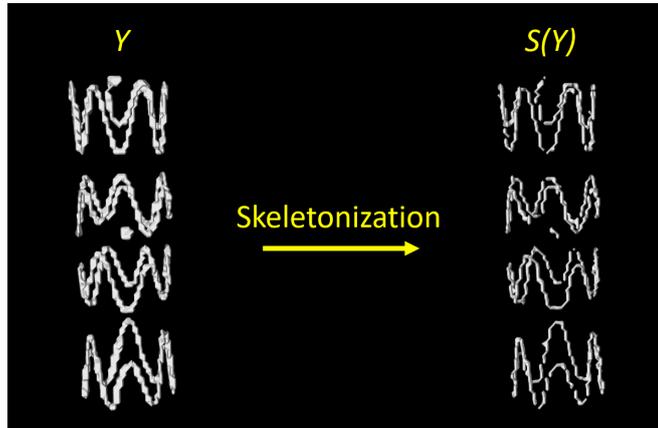


Figure 3.20: Volume rendered 3D reconstruction of a stent Y and the associated skeleton $S(Y)$.

Two custom metrics were used. D1 is the average distance of voxels in $S(GT)$ to voxels in $S(Y)$:

$$D1 = \text{mean}(\{d(\mathbf{r}, S(Y)) \mid \mathbf{r} \in S(GT)\}). \quad (3.13)$$

The average includes all voxels $\mathbf{r} = (x, y, z)$ in $S(GT)$ and $d(\mathbf{a}, B)$ is the Euclidean

distance between a point \mathbf{a} and a set of points B :

$$d(\mathbf{a}, B) = \min_{\mathbf{b} \in B} \sqrt{\sum_{i \in \{x,y,z\}} (a_i - b_i)^2}. \quad (3.14)$$

D2 is the average distance of voxels in $S(Y)$ to voxels in $S(GT)$, where all distances greater than 50 mm are excluded from the average:

$$D2 = \text{mean}(\{d(\mathbf{r}, S(GT)) \mid \mathbf{r} \in S(Y) \wedge d(\mathbf{r}, S(GT)) < 50 \text{ mm}\}). \quad (3.15)$$

D1 and D2 are therefore complementary metrics, which are sensitive to false-negatives (gaps) and false-positives (phantoms), respectively. In D2, distances greater than 50 mm are excluded, since false-positives, which are far away from the true-positives, are generally easy to distinguish visually from the true positives. This is especially true when watching sequences of reconstructions (as is the case in fluoroscopy), since the false-positives are not temporally consistent. In the following, subscripts “guidewire”, “stent”, “total” will be used to indicate whether Y and GT are reconstructions and ground truths of guidewires, stents or both.

3.10 Dose Area Product Measurements

The dose burden of the investigated 3D reconstruction pipelines was compared to the dose burden of conventional 2D fluoroscopy based on their dose area product rates \dot{DAP} . \dot{DAP} was chosen over other dose quantities because of its ease of measurement, its good correlation with effective dose [73] and the availability of studies on average \dot{DAP} values of conventional 2D fluoroscopy. The \dot{DAP} rates of the reconstruction pipelines during the stop motion scans were determined from a single measurement $\dot{DAP}_{0.04} = 0.00132 \text{ Gy cm}^2$ of the \dot{DAP} of an X-ray image at $U = 100 \text{ kV}$ and 0.04 mA s ($T_{\text{pulse}} = 8 \text{ ms}$ and $I = 5 \text{ mA}$):

$$\dot{DAP}(Q, G, r) = \frac{Q}{0.04 \text{ mA s}} \cdot 0.00132 \text{ Gy cm}^2 \cdot G \cdot r. \quad (3.16)$$

Here, Q is the tube current time product of one X-ray image of the stop motion scan, G is the number of imaging threads of the reconstruction pipeline (table 3.1) and r is the rate at which 3D reconstructions are computed. $\dot{DAP}_{0.04}$ was determined by measuring the dose in the unattenuated X-ray beam at a distance $r = 59.3 \text{ cm}$ from the X-ray focus using the RaySafeTM X2 R/F sensor and multiplying this value by the scaled-down area of the flat detector of the experimental X-ray system $((30.7 \text{ cm} \cdot r / R_{\text{FD}})^2)$.

4 | Results

4.1 Pipeline Evolution

Starting from a four-threaded imaging setup, several changes to the reconstruction pipeline were necessary to be able to reconstruct interventional material from two threads. The influence of each of these changes is investigated in this section, by comparing the image quality of each pipeline on five stop motion scans of the five different arrangements (a to e) of interventional material. The five stop motion scans were performed with a tube current time product of $Q = 0.6 \text{ mAs}$ and a motion amplitude of $A = 11 \text{ mm}$ (see section 3.8). Image quality will be judged quantitatively, by considering median $D1_{\text{total}}$ and $D2_{\text{total}}$ values, which are given in table 4.1. Greater importance will be given to the D1 values, since the human brain seems to be more capable to ignore false positives in the DTR reconstructions, than to fill false negatives. To get a visual impression, for each pipeline i and each stop motion scan j , the reconstruction of the most representative time step $t_{\text{rep}}(i, j)$ out of the $60 - T_{\text{input}} + 1$ total reconstructed time steps of the respective stop motion scan is shown in figures 4.1 to 4.5. For pipelines, that reconstruct guidewires and stents simultaneously, the most representative time step was defined to be:

$$t_{\text{rep}}(i, j) = \underset{T_{\text{input}} - 1 \leq t < 60}{\operatorname{argmin}} 3 \cdot |D1_{\text{total}}(t, i, j) - \widetilde{D1}_{\text{total}}(i, j)| + |D2_{\text{total}}(t, i, j) - \widetilde{D2}_{\text{total}}(i, j)|, \quad (4.1)$$

with the tilde referring to the median over all time steps of the stop motion scan. For pipelines, that reconstruct guidewires and stents separately, the most representative time step was defined to be:

$$t_{\text{rep}}(i, j) = \underset{T_{\text{input}} - 1 \leq t < 60}{\operatorname{argmin}} \sum_k 3 \cdot |D1_k(t, i, j) - \widetilde{D1}_k(i, j)| + |D2_k(t, i, j) - \widetilde{D2}_k(i, j)|, \quad (4.2)$$

with the sum over the interventional material type k running over $\{\text{guidewire, stent}\}$, $\{\text{guidewire}\}$, or $\{\text{stent}\}$ depending on whether both stents and guidewires or whether only one of the two types was present in the stop motion scan.

Starting from the four-threaded pipeline 4-1-SCBP (four threads, 1 input time step, single-channel backprojection), reducing the number of threads to three (pipeline 3-1-SCBP) leads, as expected, to worse reconstructions for arrangements c and d (see

Pipeline	a		b		c		d		e		Average	
4-1-SCBP	0.30	0.31	0.31	0.33	0.32	0.35	0.52	0.44	1.71	0.54	0.63	0.39
3-1-SCBP	0.31	0.31	0.30	0.31	0.43	0.69	0.65	0.72	1.17	0.68	0.57	0.54
3-1	0.32	0.32	0.33	0.35	0.37	0.55	0.56	0.63	0.79	0.71	0.47	0.51
2-1	0.35	1.83	4.05	5.76	0.76	0.67	0.98	0.76	2.00	1.14	1.63	2.03
2-4	0.29	0.35	0.75	1.33	0.54	0.68	0.79	0.80	1.26	1.30	0.73	0.89
2-4-sep	0.34	0.35	0.70	0.73	0.53	0.70	0.51	0.65	0.72	0.74	0.56	0.63
2-4-sep-bin-STs	0.33	0.39	0.34	0.35	0.39	0.43	0.39	0.43	0.58	0.54	0.40	0.43

(a) Left (right) part of each double-column a to e: median of $D1_{\text{total}}$ ($D2_{\text{total}}$) over the $60 - T_{\text{input}} + 1$ reconstructed time steps of the respective stop motion scan. Average: median values averaged over the arrangements a to e. All values in mm.

Pipeline	a		b		c		d		e		Average	
4-1-SCBP	N/A	N/A										
3-1-SCBP	0.01	-0.02	-0.04	-0.04	0.33	0.95	0.24	0.64	-0.32	0.27	-0.10	0.38
3-1	0.05	0.06	0.09	0.10	-0.14	-0.20	-0.13	-0.12	-0.32	0.05	-0.17	-0.05
2-1	0.09	4.64	11.42	15.62	1.05	0.22	0.74	0.20	1.53	0.61	2.43	2.96
2-4	-0.17	-0.81	-0.81	-0.77	-0.29	0.02	-0.20	0.05	-0.37	0.14	-0.55	-0.56
2-4-sep	0.16	0.01	-0.07	-0.45	-0.01	0.02	-0.35	-0.18	-0.43	-0.43	-0.23	-0.29
2-4-sep-bin-STs	-0.04	0.11	-0.52	-0.52	-0.27	-0.39	-0.23	-0.34	-0.20	-0.27	-0.28	-0.32

(b) Relative change between row $i - 1$ and i in sub table a.

Table 4.1: Reconstruction quality of different reconstruction pipelines on stop motion scans of the interventional material arrangements a to e with $Q = 0.6$ mA s and $A = 11$ mm (see section 3.8).

table 4.1a). While both pipelines are only able to reconstruct the guidewires (not the stent) in the most complex arrangement e convincingly, surprisingly, pipeline 3-1-SCBP performs better in terms of D1 (see table 4.1a and figure 4.5). The guidewire-only arrangements a and b are reconstructed without visible deviations from the ground truth by both pipelines (see figures 4.1 and 4.2).

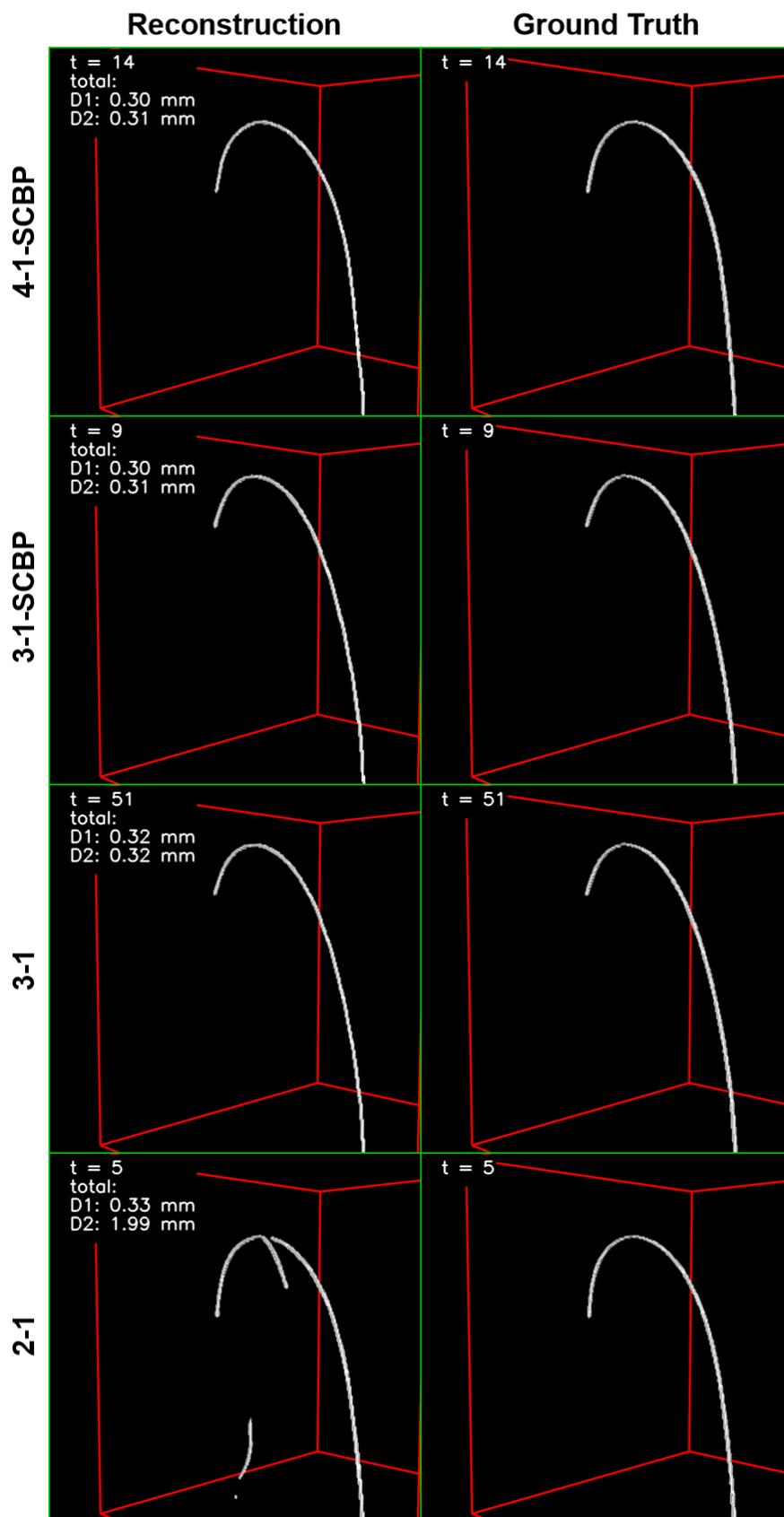
Going from single-channel backprojection (3-1-SCBP) to per-projection backprojection (3-1), unexpectedly leads to a small deterioration in reconstruction quality of the guidewire-only arrangements a and b, which is only noticeable as an inconspicuous false positive in the reconstruction of arrangement b (figure 4.2). Reconstruction quality is however visibly improved for the more complex, stent-containing arrangements c to e (figures 4.3 to 4.5 and table 4.1a).

Going from three threads (3-1) to two threads (2-1), leads to a large drop in image quality, as was expected from theoretical considerations (figure 3.3). While each of the two previous changes, changed the average median D1 and D2 values by no more than 38% (table 4.1b), average median D1 and D2 values increase by 243% and 296% respectively, when going from 3-1 to 2-1. Median D1 and D2 values increase for all five

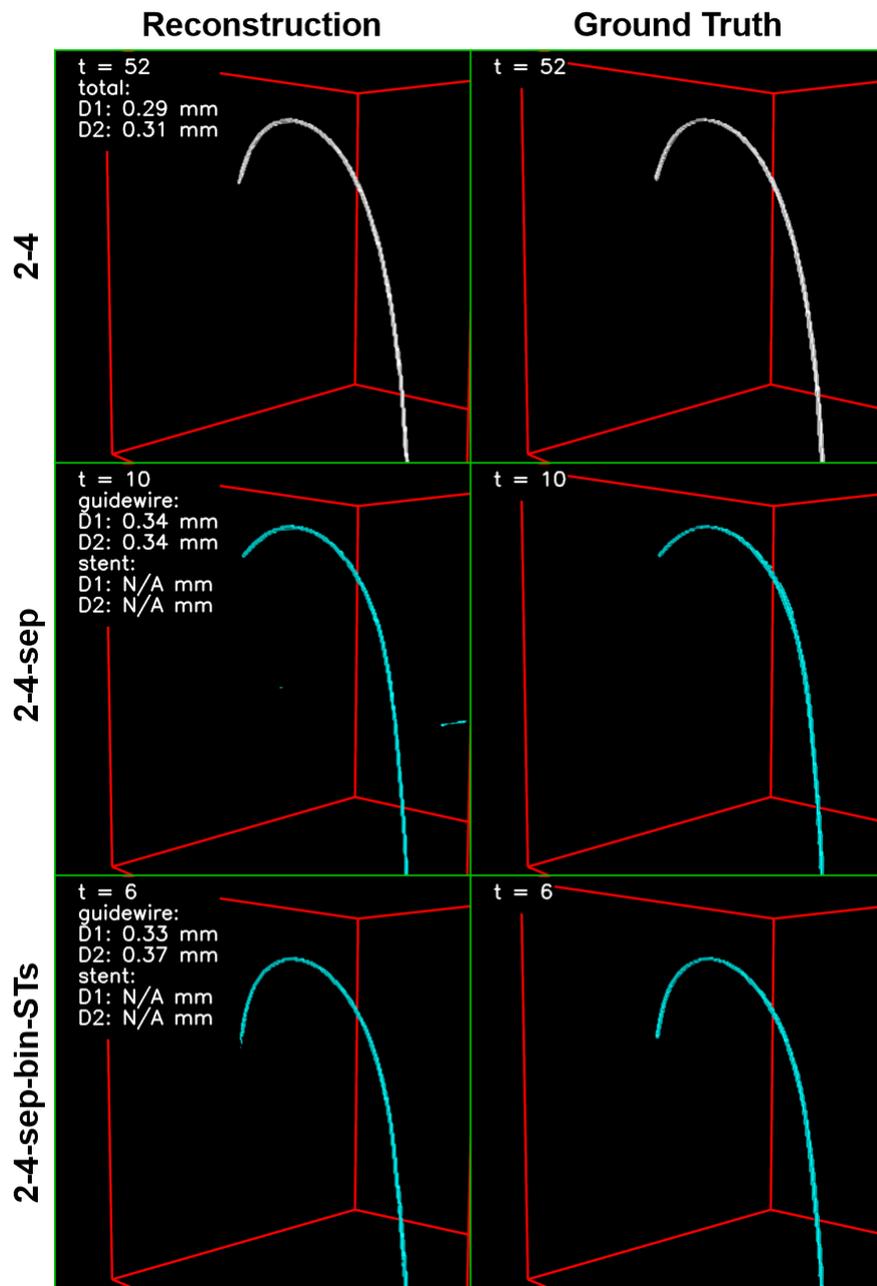
arrangements of interventional material. Visually, for guidewires, this corresponds to few but large false negatives and false positives, while for stents it corresponds to many small false negatives and false positives.

Going from only using the backprojections of the current time step as input (2-1) to using the four latest backprojections (2-4), lowers the average median D1 and D2 value by 55 % and 56 % respectively. Improved median D1 values are observed for all five interventional material arrangements. D2 values improve by 81 % and 77 % for the guidewire-only arrangements a and b, while deteriorating by 2 % to 14 % for arrangements c to e (table 4.1b). Going from one DTR reconstructing stents and guidewires (2-4) to separate DTRs for the reconstruction of stents and guidewires (2-4-sep) improves the average median D1 and D2 value by 23 % and 29 % respectively. As expected, improvement in terms of D1 is greatest for arrangements containing stents and guidewires, i.e. e and d (table 4.1b).

Binarizing the guidewire DTE output images prior to backprojection and using STs to compensate the motion between backprojections (2-4-sep-bin-STs), improves the average median D1 and D2 value further by 28 % and 32 % respectively (table 4.1a). Improvements are visible for the most representative reconstruction of all arrangements (figures 4.1 to 4.5). The most representative reconstruction of all arrangements is visually only barely distinguishable from the respective ground truth (figures 4.1 to 4.5).

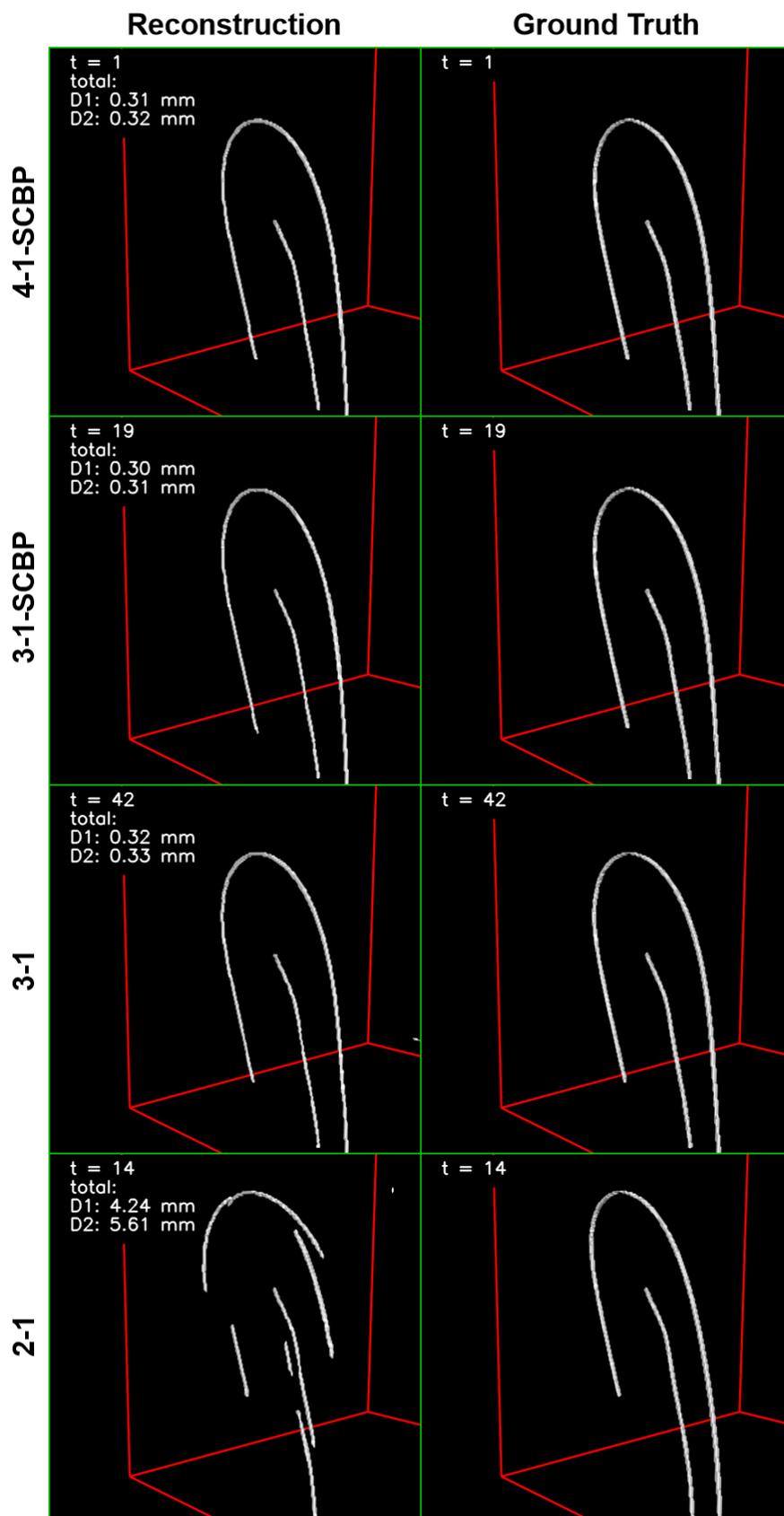


(a) Reconstructions of pipelines 4-1-SCBP, 3-1-SCBP, 3-1 and 2-1.

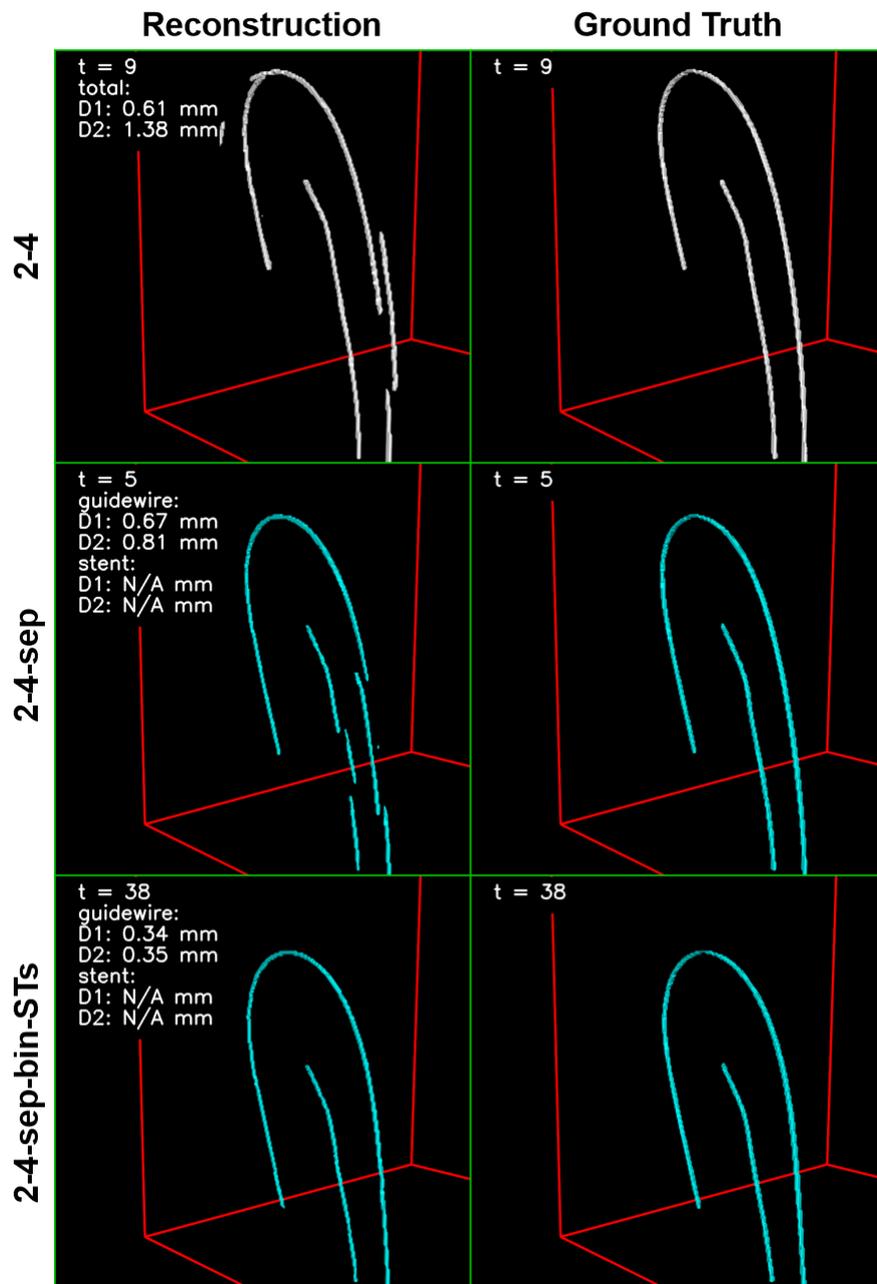


(b) Reconstructions of pipelines 2-4, 2-4-sep and 2-4-sep-bin-STs.

Figure 4.1: 3D reconstructions computed from a stop motion scan of interventional material arrangement a, using different reconstruction pipelines. Total, stent and guidewire reconstructions are white, yellow and cyan, respectively. Arrangement a contains a single guidewire. For each pipeline, the time step most representative (see equations (4.1) and (4.2)) of the average image quality of the stop motion scan is shown. The red cube (153.6 mm side length) indicates the reconstructed volume.

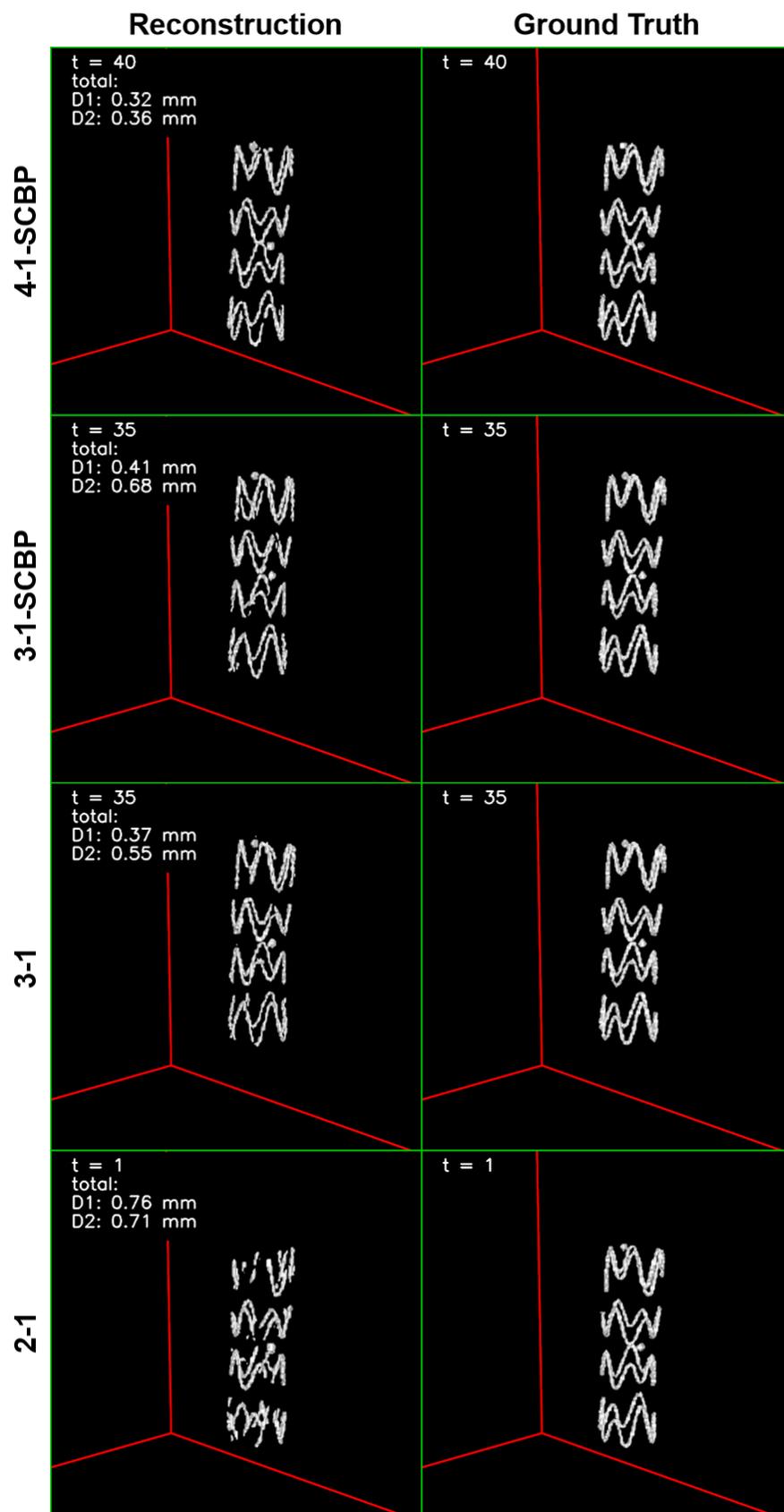


(a) Reconstructions of pipelines 4-1-SCBP, 3-1-SCBP, 3-1 and 2-1.

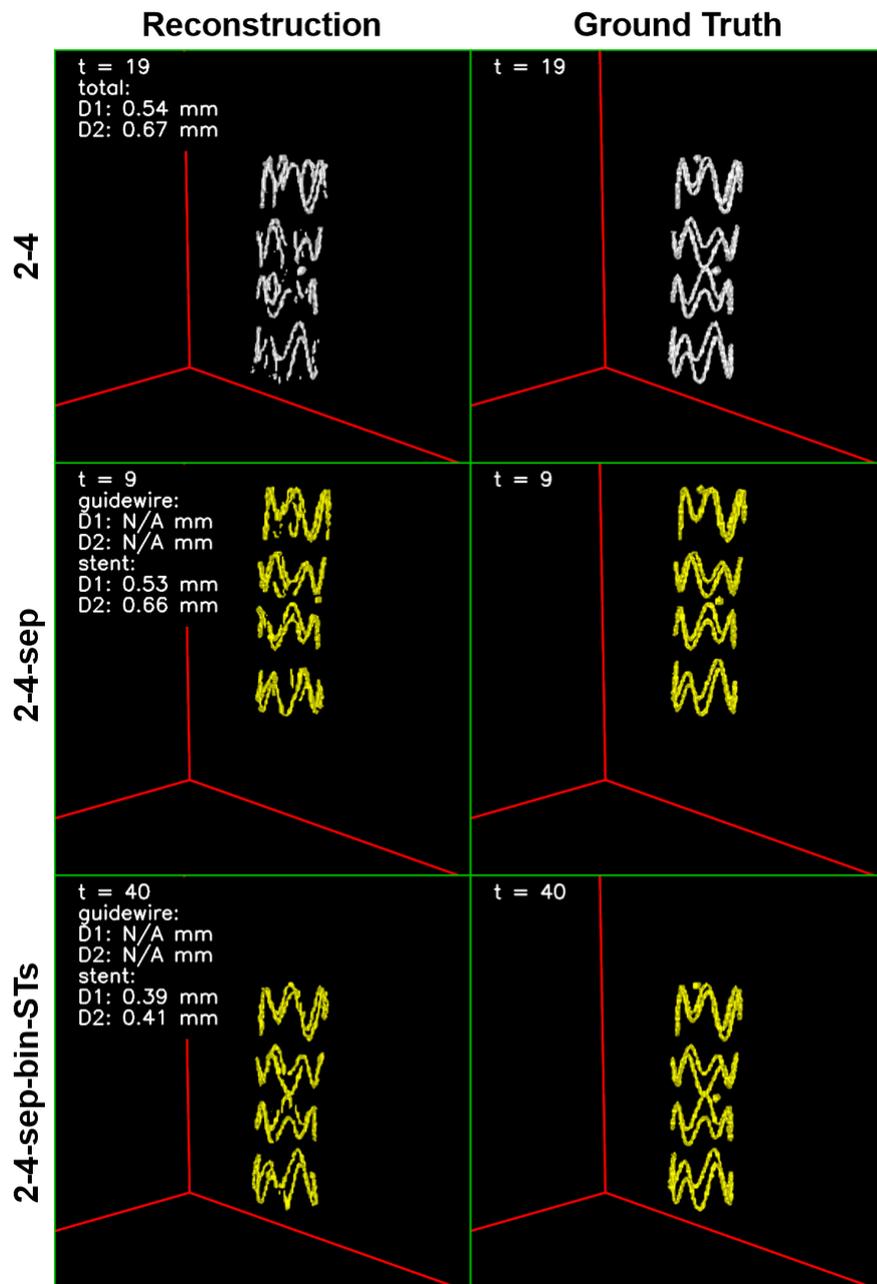


(b) Reconstructions of pipelines 2-4, 2-4-sep and 2-4-sep-bin-STs.

Figure 4.2: 3D reconstructions computed from a stop motion scan of interventional material arrangement b, using different reconstruction pipelines. Total, stent and guidewire reconstructions are white, yellow and cyan, respectively. Arrangement b contains two guidewires. For each pipeline, the time step most representative (see equations (4.1) and (4.2)) of the average image quality of the stop motion scan is shown. The red cube (153.6 mm side length) indicates the reconstructed volume.

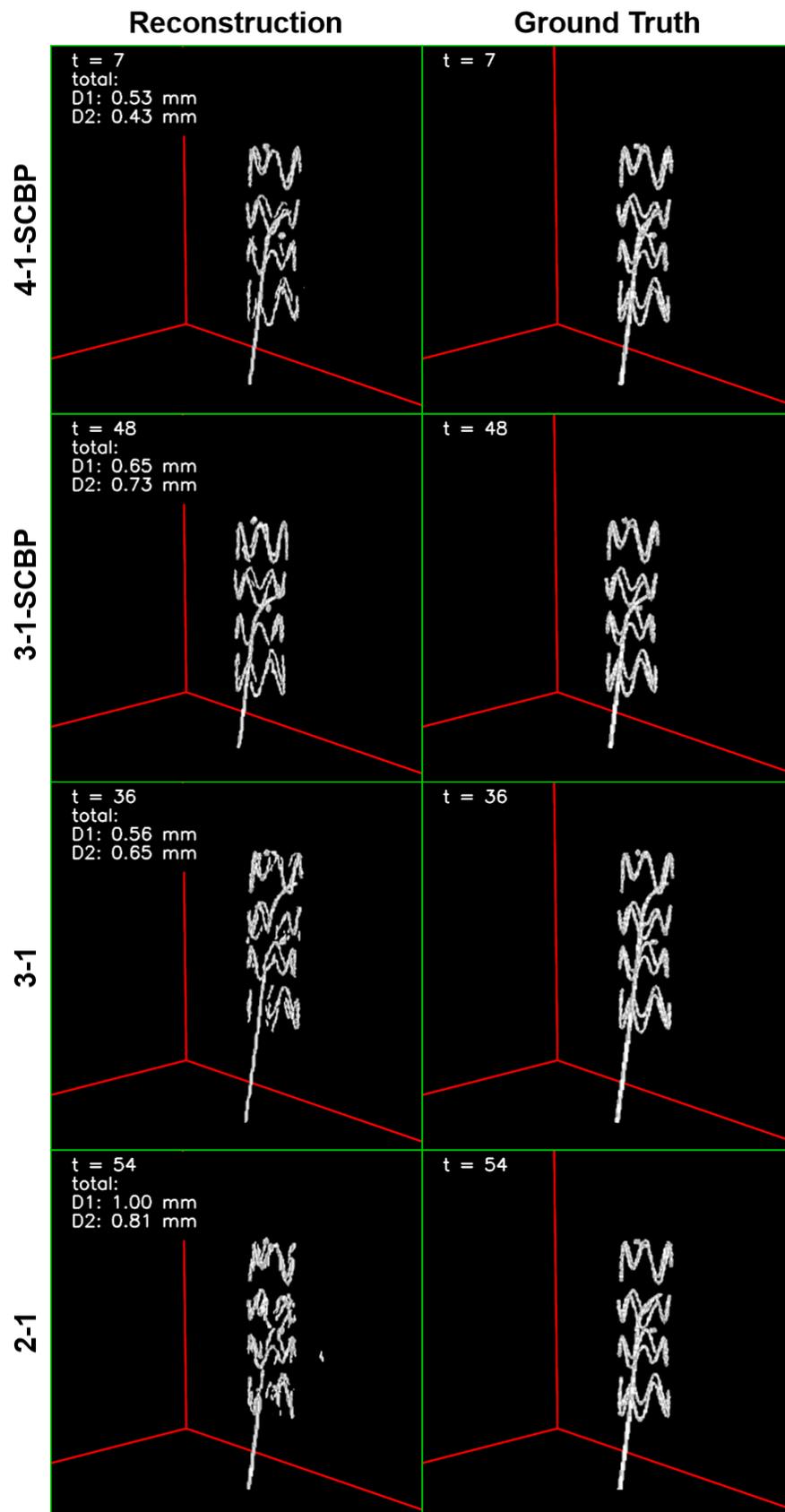


(a) Reconstructions of pipelines 4-1-SCBP, 3-1-SCBP, 3-1 and 2-1.

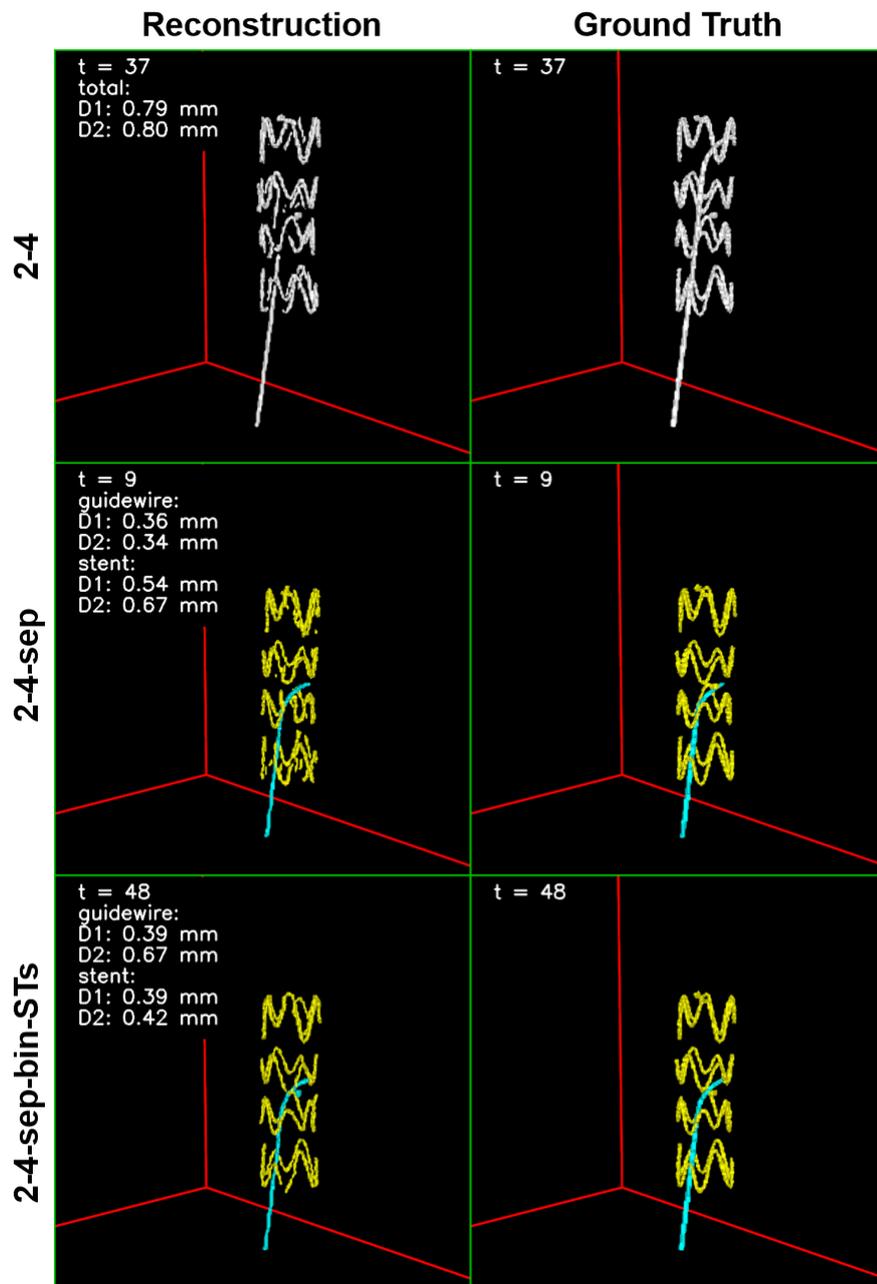


(b) Reconstructions of pipelines 2-4, 2-4-sep and 2-4-sep-bin-STs.

Figure 4.3: 3D reconstructions computed from a stop motion scan of interventional material arrangement c , using different reconstruction pipelines. Total, stent and guidewire reconstructions are white, yellow and cyan, respectively. Arrangement c contains a single stent. For each pipeline, the time step most representative (see equations (4.1) and (4.2)) of the average image quality of the stop motion scan is shown. The red cube (153.6 mm side length) indicates the reconstructed volume.

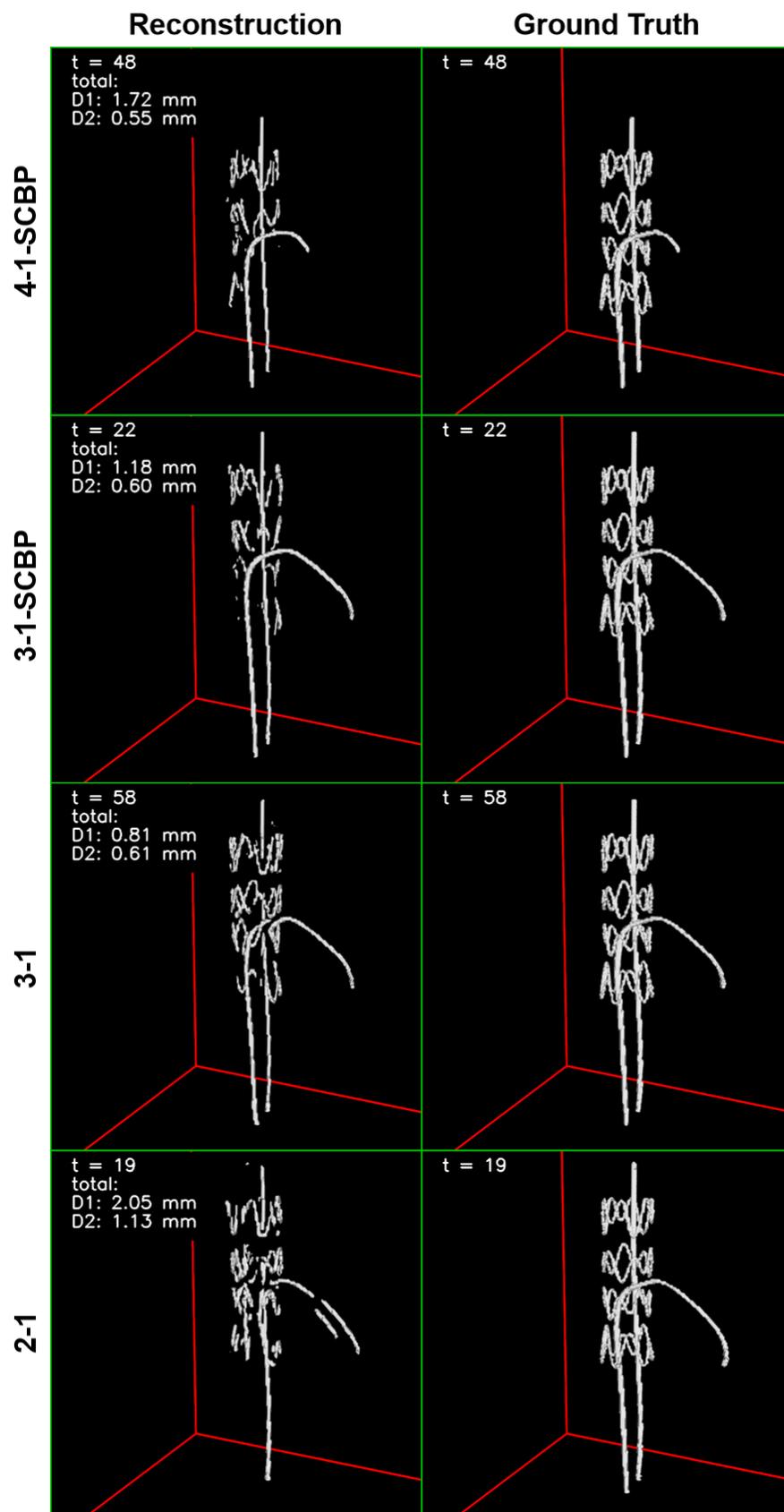


(a) Reconstructions of pipelines 4-1-SCBP, 3-1-SCBP, 3-1 and 2-1.

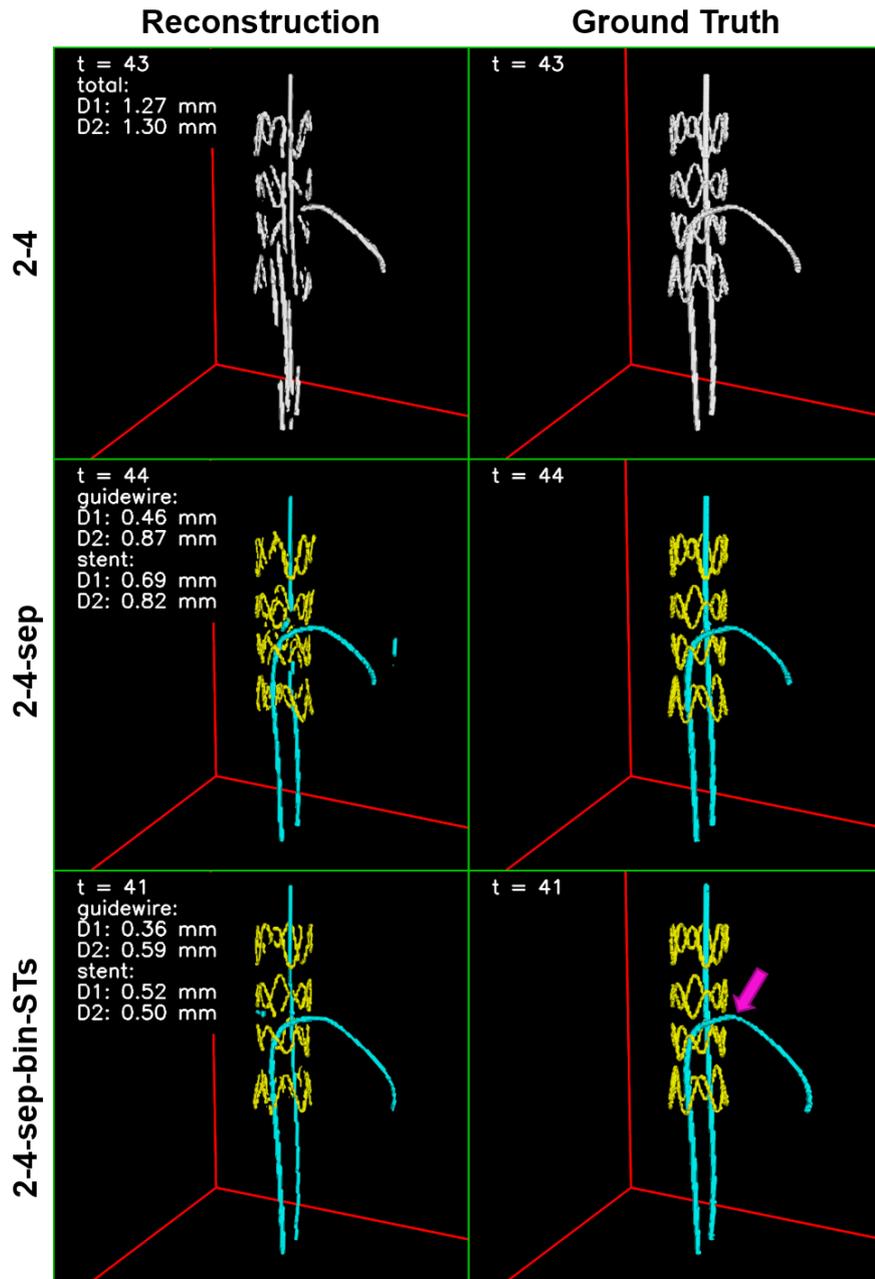


(b) Reconstructions of pipelines 2-4, 2-4-sep and 2-4-sep-bin-STs.

Figure 4.4: 3D reconstructions computed from a stop motion scan of interventional material arrangement d , using different reconstruction pipelines. Total, stent and guidewire reconstructions are white, yellow and cyan, respectively. Arrangement d contains a stent and a guidewire. For each pipeline, the time step most representative (see equations (4.1) and (4.2)) of the average image quality of the stop motion scan is shown. The red cube (153.6 mm side length) indicates the reconstructed volume.



(a) Reconstructions of pipelines 4-1-SCBP, 3-1-SCBP, 3-1 and 2-1.



(b) Reconstructions of pipelines 2-4, 2-4-sep and 2-4-sep-bin-STs.

Figure 4.5: 3D reconstructions computed from a stop motion scan of interventional material arrangement e , using different reconstruction pipelines. Total, stent and guidewire reconstructions are white, yellow and cyan, respectively. Arrangement e contains a stent and two guidewires, with one of the guidewires being partially inside a catheter. The point where the guidewire exits the catheter is marked with a pink arrow. For each pipeline, the time step most representative (see equations (4.1) and (4.2)) of the average image quality of the stop motion scan is shown. The red cube (153.6 mm side length) indicates the reconstructed volume.

4.2 Robustness Analysis

In this section, the robustness of the best two-threaded reconstruction pipeline, 2-4-sep-bin-STs, with respect to changes in the tube current time product per X-ray image Q , the angular increment between X-ray image pairs $\Delta\alpha$ and the amplitude A of the respiration-induced motion of the interventional material is investigated. Four- and three-threaded pipelines are not investigated, since they require more X-ray images and therefore dose, per 3D reconstruction and require more expensive imaging hardware.

4.2.1 Tube Current Time Product

In this section, the reconstruction quality of pipeline 2-4-sep-bin-STs is evaluated on stop motion scans with different tube current time products $Q \in \{0.15 \text{ mA s}, 0.60 \text{ mA s}, 2.40 \text{ mA s}\}$. Table 4.2 shows the median D1 and D2 values for both channels (guidewire, stent) of the reconstruction pipeline for interventional material arrangements a to e. Since arrangements a and b contain no stent and arrangement c contains no guidewire, the respective ground truths are empty and D1 and D2 values cannot be computed (nan-entries in table 4.2). Figure 4.6 shows a 3D reconstruction of arrangement e at different tube current time products. Details of the two input X-ray images of the current time step and the respective details of the DTE outputs are also shown.

Going from 0.60 mA s to 2.40 mA s improves the median $D1_{\text{stent}}$ value of arrangement e by 31 % (0.48 mm \rightarrow 0.33 mm). The $D1_{\text{stent}}$ values of arrangements c and d and the $D1_{\text{guidewire}}$ values of scenes a, b and d improve by less than 13%. Unexpectedly, the $D1_{\text{guidewire}}$ value of arrangement e gets worse (0.37 mm \rightarrow 0.53 mm). These changes can also be observed visually in figure 4.6. The stent reconstruction at 2.40 mA s has slightly fewer false positives and false negatives, while the guidewire reconstruction surprisingly has more false negatives. These false negatives stem from false negatives in the DTE output (pink arrows), which is surprising, since the guidewire is perfectly visible there. A possible explanation is, that the noise level in the training data of the DTE could be more similar to the noise level in 0.60 mA s X-ray images than to that in 2.40 mA s X-ray images.

Going from 0.60 mA s to 0.15 mA s strongly deteriorates the median $D1_{\text{stent}}$ value of all stent arrangements c to e by 323 %, 598 % and 1813 %. $D1_{\text{guidewire}}$ is less affected: for arrangements a, d and e it is increased by less than 25 % and for scene b it is increased by 400 %. These changes can also be observed visually in figure 4.6. While at 0.15 mA s, the guidewire reconstruction is only missing a part towards the tip, the stent reconstruction is almost entirely empty. The empty stent reconstruction is due to the stent struts being invisible in X-ray image A at 0.15 mA s due to the low contrast-to-noise ratio (figure 4.6).

Q [mA s]	Channel	a		b		c		d		e		Average	
0.15	guidewire	0.37	0.43	1.71	0.61	nan	nan	0.39	0.40	0.46	0.46	0.73	0.47
0.60	guidewire	0.33	0.37	0.34	0.35	nan	nan	0.37	0.67	0.37	0.63	0.35	0.51
2.40	guidewire	0.33	0.36	0.35	0.43	nan	nan	0.37	0.79	0.53	0.71	0.40	0.57
0.15	stent	nan	nan	nan	nan	1.65	0.97	2.79	0.96	9.18	1.16	4.54	1.03
0.60	stent	nan	nan	nan	nan	0.39	0.43	0.40	0.43	0.48	0.52	0.42	0.46
2.40	stent	nan	nan	nan	nan	0.34	0.34	0.35	0.34	0.33	0.36	0.34	0.35

Table 4.2: Reconstruction quality of pipeline 2-4-sep-bin-STs on stop motion scans of the interventional material arrangements a to e with $A = 11$ mm and tube current time products $Q \in \{0.15 \text{ mA s}, 0.60 \text{ mA s}, 2.40 \text{ mA s}\}$ (see section 3.8). Left (right) part of each double-column a to e: median of $D1_{\text{Channel}}$ ($D2_{\text{Channel}}$) over the 57 reconstructed time steps of the respective stop motion scan. “Channel” $\in \{\text{guidewire}, \text{stent}\}$ here describes the channel of the reconstruction. All values in mm.

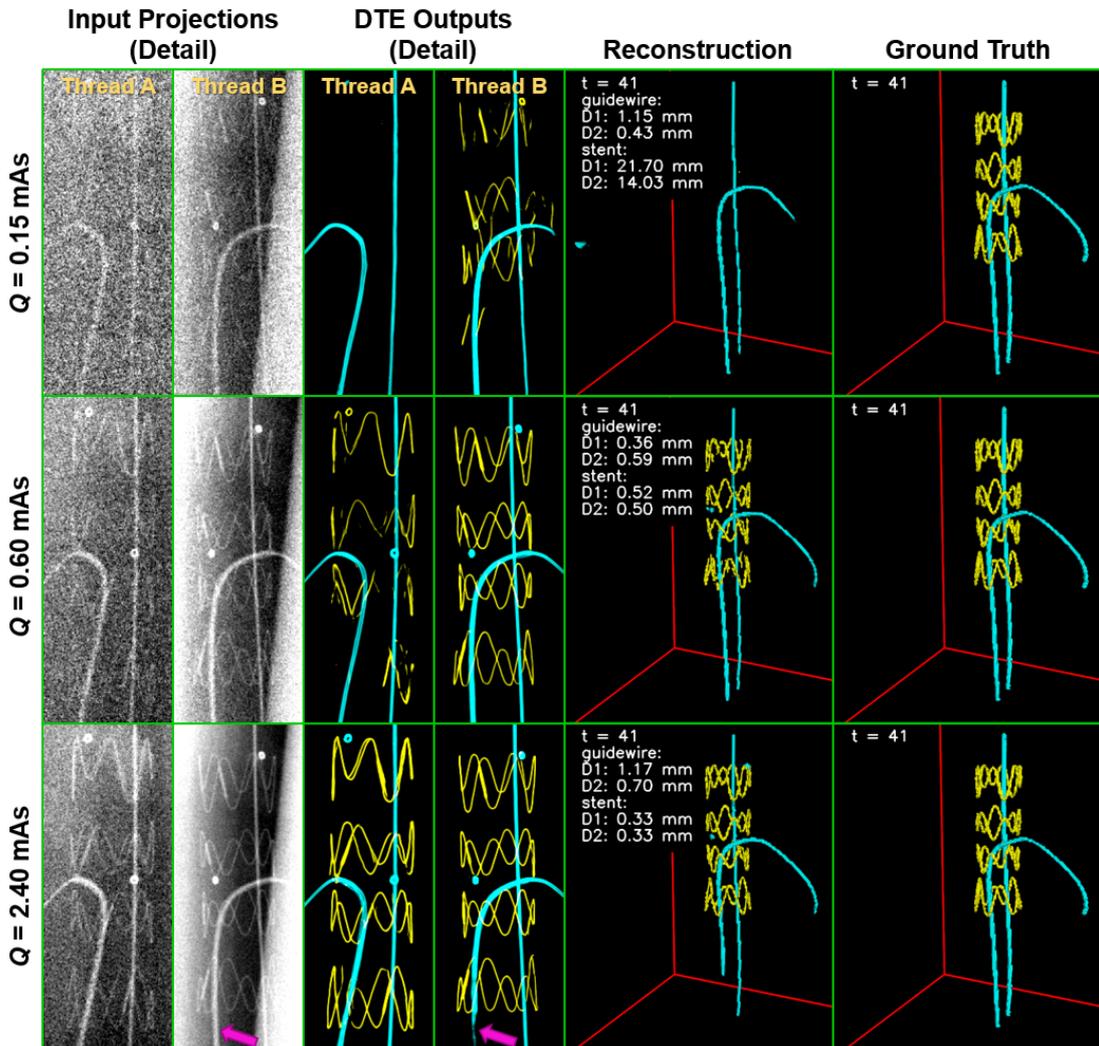


Figure 4.6: 3D reconstructions of the best two-threaded pipeline, 2-4-sep-bin-STs, on stop motion scans of the interventional material arrangement e with different tube current time products $Q \in \{0.15 \text{ mAs}, 0.60 \text{ mAs}, 2.40 \text{ mAs}\}$ and with $A = 11 \text{ mm}$ (see section 3.8). The left columns show details of the input X-ray images and DTE outputs of the current time step. In the DTE outputs, the stent and guidewire channels of the DTE are overlaid (yellow, cyan). Time step 41 of the stop motion scan is shown here, since it is the most representative (see equation (4.2)) for $Q = 0.60 \text{ mAs}$. The red cube (153.6 mm side length) indicates the reconstructed volume.

4.2.2 Angular Increment

In this section, the influence of the angular increment $\Delta\alpha$ between X-ray image pairs on reconstruction quality is investigated. To this end, pipeline 2-4-sep-bin-STs, which was originally trained with $\Delta\alpha = 19^\circ$, was trained again with $\Delta\alpha = 9.5^\circ$ and $\Delta\alpha = 28.5^\circ$. Median D1 and D2 values of these variants on the interventional material arrangements a to e are given in table 4.3. Reconstructions of arrangement e computed by these three variants are shown in figure 4.7.

Table 4.3 reveals no strong dependence of the stent reconstruction quality on $\Delta\alpha$ in the range $[9.5^\circ, 28.5^\circ]$. $D1_{\text{stent}}$ values change by no more than 17% when going from 19° to 9.5° or 28.5° . Accordingly, the stent reconstructions at different angular increments in figure 4.7 look very similar.

A stronger dependence of reconstruction quality on $\Delta\alpha$ can be observed for guidewires. When going from $\Delta\alpha = 19.0^\circ$ to $\Delta\alpha = 9.5^\circ$, median $D1_{\text{guidewire}}$ values increase for all guidewire-containing arrangements a, b, d and e by 36%, 735%, 3% and 400% respectively. The increase seems to be stronger for arrangements of guidewires, that have a higher maximum number of intersections $N_{\text{int,max}}$ of the guidewires with a transversal (x - y) plane:

$$N_{\text{int,max}} = \max_{j \in \{0, \dots, J-1\}} N_{\text{int},j}, \quad (4.3)$$

with $N_{\text{int},j}$ being the number of intersections of guidewires with the transversal plane with the z -index j and the number of voxels in the z -direction J . Arrangement d has $N_{\text{int,max}} = 1$, a has $N_{\text{int,max}} = 2$, e and b have $N_{\text{int,max}} = 3$ (see the ground truths in figures 4.4, 4.1, 4.5 and 4.2). Going from $\Delta\alpha = 19.0^\circ$ to $\Delta\alpha = 28.5^\circ$ does not decrease median $D1_{\text{guidewire}}$ values. Surprisingly, the median $D1_{\text{guidewire}}$ value even increases by 106% for arrangement b. In accordance with these observations on the median $D1_{\text{guidewire}}$ values, figure 4.7, shows guidewire reconstructions, that are nearly indistinguishable from the ground truths for $\Delta\alpha = 19.0^\circ$ and $\Delta\alpha = 28.5^\circ$. For $\Delta\alpha = 9.5^\circ$ however, a part of one of the guidewires is missing.

$\Delta\alpha$	Channel	a		b		c		d		e		Average	
9.5°	guidewire	0.45	0.38	2.84	1.82	nan	nan	0.38	0.60	1.85	0.61	1.38	0.85
19°	guidewire	0.33	0.37	0.34	0.35	nan	nan	0.37	0.67	0.37	0.63	0.35	0.51
28.5°	guidewire	0.35	0.37	0.70	0.75	nan	nan	0.37	0.54	0.40	0.63	0.46	0.57
9.5°	stent	nan	nan	nan	nan	0.43	0.46	0.45	0.49	0.56	0.57	0.48	0.51
19°	stent	nan	nan	nan	nan	0.39	0.43	0.40	0.43	0.48	0.52	0.42	0.46
28.5°	stent	nan	nan	nan	nan	0.43	0.42	0.43	0.41	0.55	0.53	0.47	0.45

Table 4.3: Reconstruction quality of the best two-threaded pipeline, 2-4-sep-bin-STs, trained and applied with different angular increments $\Delta\alpha$ on stop motion scans of the interventional material arrangements a to e with $A = 11$ mm and $I = 30$ mA (see section 3.8). Left (right) part of each double-column a to e: median of $D1_{\text{Channel}}$ ($D2_{\text{Channel}}$) over the 57 reconstructed time steps of the respective stop motion scan. “Channel” \in {guidewire, stent} here describes the channel of the reconstruction. All values in mm.

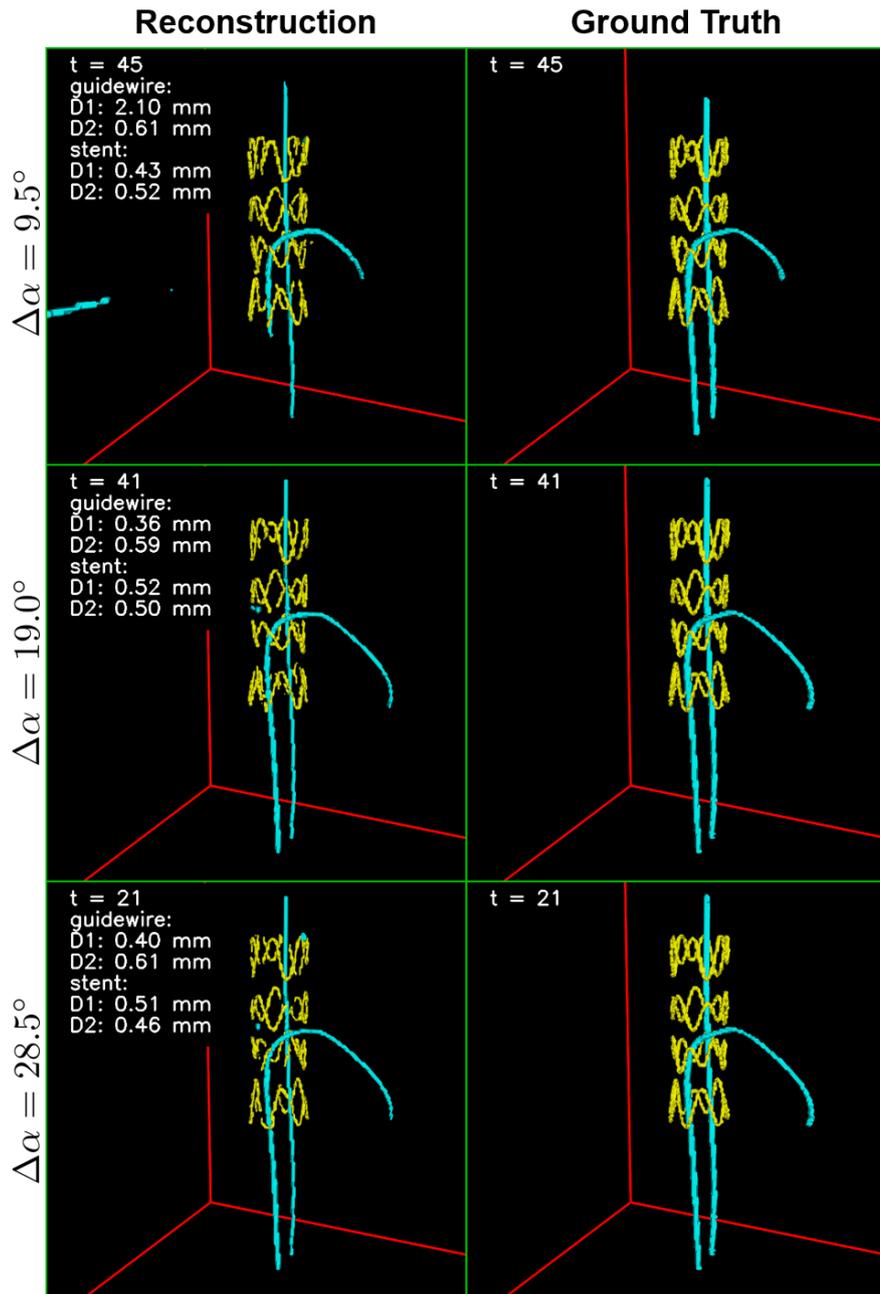


Figure 4.7: 3D reconstructions of three variants of the best two-threaded pipeline, 2-4-sep-bin-STs. The variants differ in the angular increment $\Delta\alpha$ between X-ray image pairs, that was used during training and during application to stop motion scans. Shown here, is the application to a stop motion scan of interventional material arrangement e with $Q = 0.60$ mAs and with $A = 11$ mm (see section 3.8). For each variant, the time step most representative (see equation (4.2)) of the average image quality of the stop motion scan is shown. The red cube (153.6 mm side length) indicates the reconstructed volume.

4.2.3 Motion Amplitude

The amplitude A of the simulated respiratory motion (equation (3.12)) was $A = 11$ mm in the previous sections. In this section, the change in reconstruction quality of pipeline 2-4-sep-bin-STs, when A is increased to 22 mm, is investigated. The respective stop motion scans with $A = 22$ mm were only acquired for interventional material arrangements c to e, due to technical difficulties.

Table 4.4 shows no strong influence of the motion amplitude A on reconstruction quality in the range [11 mm, 22 mm]. Median $D1_{\text{stent}}$ and $D1_{\text{guidewire}}$ values change by no more than 13 %, when going from 11 mm to 22 mm. Accordingly, the reconstructions of arrangement e at $A = 11$ mm and $A = 22$ mm look very similar (figure 4.8).

$A[\text{mm}]$	Channel	c		d		e		Average	
11	guidewire	nan	nan	0.37	0.67	0.37	0.63	0.37	0.65
22	guidewire	nan	nan	0.34	0.69	0.37	0.64	0.36	0.66
11	stent	0.39	0.43	0.40	0.43	0.48	0.52	0.42	0.46
22	stent	0.39	0.43	0.42	0.43	0.54	0.56	0.45	0.47

Table 4.4: Reconstruction quality of pipeline 2-4-sep-bin-STs on stop motion scans of the interventional material arrangements c to e with $Q = 0.60$ mA s and different motion amplitudes $A \in \{11 \text{ mm}, 22 \text{ mm}\}$ (see section 3.8). Left (right) part of each double-column c to e: median of $D1_{\text{Channel}}$ ($D2_{\text{Channel}}$) over the 57 reconstructed time steps of the respective stop motion scan. “Channel” $\in \{\text{guidewire}, \text{stent}\}$ here describes the channel of the reconstruction. All values in mm.

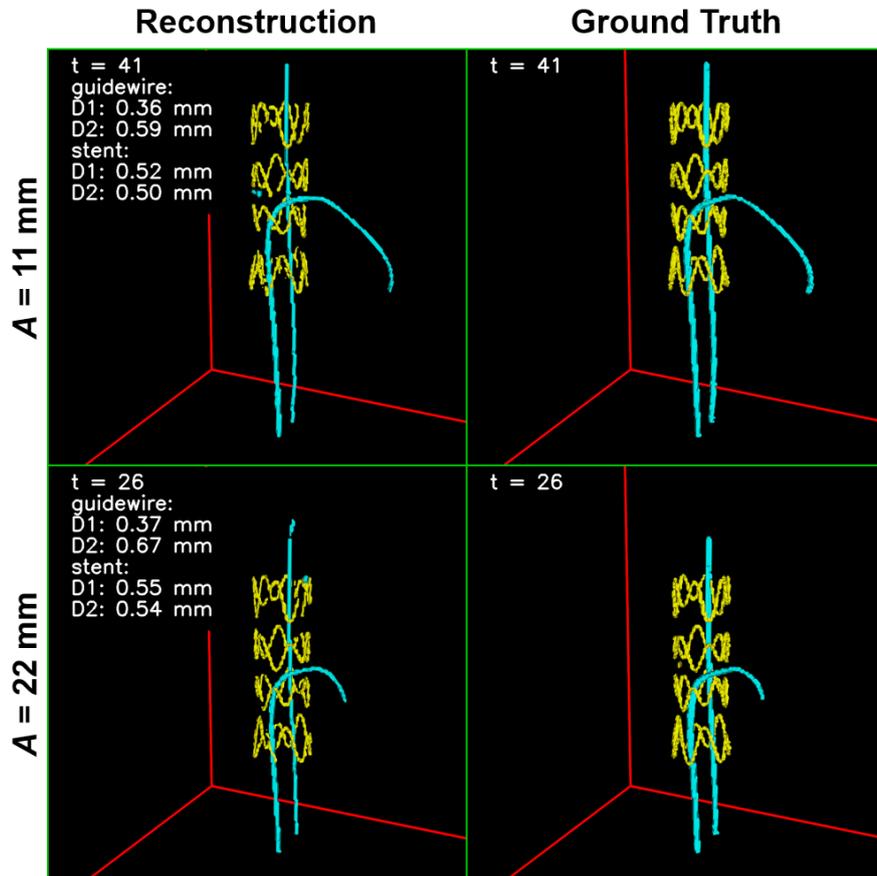


Figure 4.8: 3D reconstructions of pipeline 2-4-sep-bin-STs of stop motion scans of interventional material arrangement e with $Q = 0.60 \text{ mA s}$ and with different motion amplitudes $A \in \{11 \text{ mm}, 22 \text{ mm}\}$ (see section 3.8). For each amplitude, the time step most representative (see equation (4.2)) of the average image quality of the stop motion scan is shown. The red cube (153.6 mm side length) indicates the reconstructed volume.

4.3 Dose Area Product Rate

According to section 4.2.1, for the phantom used in this work, the lowest investigated tube current time product per X-ray image Q , at which the best two-threaded reconstruction pipeline, 2-4-sep-bin-STs, still yields good reconstruction quality, is $Q = 0.60$ mAs. According to equation (3.16), the DAP rate of a two-threaded pipeline at $Q = 0.60$ mAs and the intended 3D reconstruction rate of $r = 5$ vps is $\text{DAP} = 0.198 \text{ Gy cm}^2 \text{ s}^{-1} = 11.9 \text{ Gy cm}^2 \text{ min}^{-1}$.

Since the interventional material was placed on the abdomen of the phantom during the stop motion scans, it makes sense to compare this DAP rate to the DAP rate of conventional fluoroscopy in abdominal interventions. Average values for DAP and fluoroscopy time of four common abdominal interventions guided by 2D fluoroscopy (table 4.5) were taken from retrospective, clinical studies [15], [16]. These averages include, among other things, different patient sizes, single-plane and biplane imaging systems, different sites and operators, different imaging parameters (frame rate, tube voltage, etc.). DAP rate was computed by dividing DAP through fluoroscopy time (table 4.5).

Table 4.5: Average dose area product (DAP) rates of some abdominal procedures guided by 2D fluoroscopy.

Interventional Procedure	DAP [Gy cm ²]	Fluoroscopy Time [min]	DAP Rate [Gy cm ² min ⁻¹]	Image Type
TIPS creation ^a	335	38.7	8.7	2D
Hepatic chemoembolization ^a	282	16.8	16.8	2D
EVAR, infrarenal abdominal aorta ^b	108	21	5.1	2D
TACE ^b	121	17	7.1	2D
2-4-sep-bin-STs at $Q = 0.60$ mAs and $r = 5$ vps	N/A	N/A	11.9	3D

EVAR: endovascular aneurysm repair, TACE: transarterial chemoembolization, TIPS: transjugular intrahepatic portosystemic shunt.

^a DAP and fluoroscopy time are mean values from a 2003 study [16].

^b DAP and fluoroscopy time are median values from a 2019 study [15].

DAP rate was computed by dividing DAP through fluoroscopy time.

4.4 Real-Time Capability

Computation times of the NNs in the reconstruction pipelines on an NVIDIA RTX 4090 GPU are given in table 4.6. The computation times for backprojection, 4×4 average pooling (in the STs) and translation (in the STs) are negligible compared to the NN

computation times.

Table 4.6: Computation times on an NVIDIA RTX 4090 GPU, using TensorFlow’s mixed precision and graph execution.

Algorithm	Input Size	Output Size	Computation Time
DTE	1024^2	2×1024^2	30 ms
DTR	$N_{\text{channels}} \times 256^3$	256^3	165 ms
auxiliary DTR	2×64^3	64^3	7 ms
localization network	6×64^3	3	3 ms

N_{channels} is $G \cdot T_{\text{input}}$ (T_{input}) for DTRs with (without) per-projection backprojection.

The 3D reconstruction of each time step involves the application of the DTE to all G X-ray images acquired in the time step and the application of N_{DTR} DTRs ($N_{\text{DTR}} = 1$ for all pipelines, except for pipelines 10 to 13, where $N_{\text{DTR}} = 2$ (see table 3.1)). The pipelines 10, 12 and 13 employ motion compensation and therefore additionally involve the application of the low-resolution auxiliary DTR to the current time step (the low-resolution auxiliary stent reconstructions of previous time steps are reused) and the application of $T_{\text{input}} - 1$ STs (one per previous time step). Therefore, when using a single NVIDIA RTX 4090 GPU, the 3D reconstruction of each time step takes:

$$T_{\text{reco},4090} = G \cdot 30 \text{ ms} + N_{\text{DTR}} \cdot 165 \text{ ms} + \begin{cases} 7 \text{ ms} + (T_{\text{input}} - 1) \cdot 3 \text{ ms}, & \text{if motion compensation} \\ 0, & \text{else} \end{cases} \quad (4.4)$$

When using G NVIDIA RTX 4090 GPUs to parallelize the applications of the DTE, the DTRs and the STs, the computation times become:

$$T_{\text{reco},G4090s} = 30 \text{ ms} + \lceil N_{\text{DTR}}/G \rceil \cdot 165 \text{ ms} + \begin{cases} 7 \text{ ms} + \lceil (T_{\text{input}} - 1)/G \rceil \cdot 3 \text{ ms}, & \text{if motion compensation} \\ 0, & \text{else} \end{cases} \quad (4.5)$$

The reconstruction times of all reconstruction pipelines (table 3.1) using one GPU or G GPUs are given in table 4.7.

Table 4.7: Reconstruction times of the pipelines investigated in this thesis.

ID	$T_{\text{reco},4090}[\text{ms}]$	$T_{\text{reco},G4090s}[\text{ms}]$
4-1-SCBP	285	195
3-1-SCBP	255	195
3-1	255	195
2-1	225	195
2-4	225	195
2-4-sep	390	195
2-4-sep-bin-STs	406	208
2-4-sep-bin-STs-less $\Delta\alpha$	406	208
2-4-sep-bin-STs-more $\Delta\alpha$	406	208

5 | Summary & Discussion

The goal of this thesis was the development of an algorithm for the 3D reconstruction of interventional material, which requires as few newly acquired X-ray images per 3D reconstruction (for dose efficiency) as possible and as few imaging threads as possible (those two quantities are the same for all reconstruction pipelines tested in this thesis and are therefore used synonymously in the following). To this end, the four-threaded pipeline 4-1-SCBP proposed by Eulig et al. was adapted to work with fewer imaging threads. The transition from four threads to three threads (4-1-SCBP \rightarrow 3-1-SCBP) did not result in a visually noticeable decrease in reconstruction quality (figures 4.1 to 4.5), which was expected from the consideration illustrated in figure 3.3. The transition from three threads to two threads (3-1 \rightarrow 2-1) however did: obvious deviations from the respective ground truths are present in the reconstructions of all interventional material arrangements (figures 4.1 to 4.5), which was expected from the considerations illustrated in figures 1.2 and 3.3. Each of the four improvements subsequently made to this pipeline, i.e. i) use of previous time steps in a rotating imaging setup, ii) separate reconstruction of stents and guidewires, iii) binarization of DTE output images, and iv) stent motion compensation resulted in improved reconstruction quality, as measured by the figures of merit D1 and D2 (table 4.1, rightmost double-column). The combined improvement (2-1 \rightarrow 2-4-sep-bin-STs) is 75 % for the average D1 value (1.63 mm to 0.40 mm) and 79 % for the average D2 value (2.03 mm to 0.43 mm). The reconstructions of 2-4-sep-bin-STs of all tested interventional material arrangements are visually nearly indistinguishable from the respective ground truths (figures 4.1 to 4.5). Therefore, to the author's best knowledge, this pipeline is the first algorithm capable of reconstructing i) a stent, ii) multiple guidewires, and iii) a stent and multiple guidewires from only two newly acquired X-ray images. Each of i), ii), and iii) is a first. The reconstruction quality of this best two-threaded reconstruction pipeline (2-4-sep-bin-STs) is even higher than that of the four-threaded baseline (4-1-SCBP): interventional arrangements a to d are reconstructed well by both pipelines (figures 4.1 to 4.4), but only pipeline 2-4-sep-bin-STs reconstructs the most complex arrangement e well (figure 4.5). 2-4-sep-bin-STs-reconstructions are also more useful than 4-1-SCBP-reconstructions in a way not measured by D1 and D2: the separate reconstruction of stents and guidewires performed by 2-4-sep-bin-STs makes it possible to display stents and guidewires in different colors, which makes it much easier to distinguish them visually (see figures 4.4 and 4.5). Since two-threaded pipelines are presumably most promising for use in a 3D fluoroscopy application due to their dose efficiency and relatively low hardware

demands, the following robustness analysis was performed for 2-4-sep-bin-STs.

The investigation of different tube current time products $Q \in \{0.15 \text{ mA s}, 0.60 \text{ mA s}, 2.40 \text{ mA s}\}$ showed that $Q \in \{0.60 \text{ mA s}, 2.40 \text{ mA s}\}$ result in similar reconstruction quality: average D1 and D2 values of the stent and guidewire reconstructions differed by no more than 0.08 mm between $Q = 0.60 \text{ mA s}$ and $Q = 2.40 \text{ mA s}$ (table 4.2, right-most double-column). Reducing the tube current time product from $Q = 0.60 \text{ mA s}$ to $Q = 0.15 \text{ mA s}$ however decreased reconstruction quality considerably: $D1_{\text{guidewire}}$ increased by 109 % (0.35 mm to 0.73 mm) and $D1_{\text{stent}}$ increased by 981 % (0.42 mm to 4.54 mm). These observations only apply to the specific anthropomorphic phantom used in this work. In general, the necessary tube current time product will depend strongly on the size of the patient or phantom. It was expected that the stent reconstruction quality suffers more from a decrease in Q since stent struts are thinner than guidewires and therefore have a lower CNR in X-ray images.

The investigation of different angular increments $\Delta\alpha$ showed similar stent reconstruction quality for all three tested values $\Delta\alpha \in \{9.5^\circ, 19.0^\circ, 28.5^\circ\}$: average D1 values differed by no more than 0.06 mm (table 4.3). The guidewire reconstruction quality was similar for $\Delta\alpha \in \{19.0^\circ, 28.5^\circ\}$ (average D1 values differed by 0.11 mm) but decreased strongly when going from 19.0° to 9.5° (average D1 value increased from 0.35 mm to 1.38 mm). Since lower $\Delta\alpha$ implies a slower rotation of the X-ray system, which permits the use of longer and therefore more energetic X-ray pulses without increasing rotation-induced angular blur, low $\Delta\alpha$ values increase the penetration ability of the X-ray system.

During the stop motion scans, the interventional material was translated according to a sinusoid (equation (3.12)) with amplitude A and frequency $f = 0.25 \text{ Hz}$ to simulate respiratory motion. The investigation of different motion amplitudes A showed very similar reconstruction quality for $A = 11 \text{ mm}$ and $A = 22 \text{ mm}$: the respective reconstructions look very similar (figure 4.8) and average D1 and D2 values for the stent and guidewire reconstructions differ by no more than 0.03 mm (table 4.4). For the 22 mm-motion, the maximum translation between adjacent time steps is 6.8 mm (equation (3.12) with $A = 22 \text{ mm}$). While this translation could be compensated by 2-4-sep-bin-STs, a decrease in image quality with increasing translation is expected from the consideration illustrated in figure 3.3 (second image from the right) and from the fact that motion-compensating spatial transformers improve reconstruction quality (figures 4.3 to 4.5).

Since the DAP rates of conventional 2D fluoroscopy in table 4.5 and the DAP rate of the 3D reconstruction pipeline 2-4-sep-bin-STs were determined very differently (section 4.3) (the former are averages over many different clinical situations, while the latter is specific to the anthropomorphic phantom, the imaging system and the scan parameters used during the stop motion scans), a comparison must be interpreted carefully. Nevertheless, this comparison indicates that the DAP rate of the 3D reconstruction pipeline 2-4-sep-bin-STs ($11.9 \text{ Gy cm}^2 \text{ min}^{-1}$ at 5 vps and $Q = 0.60 \text{ mA s}$) is in the same order of magnitude

as that of conventional 2D fluoroscopy ($5.1 \text{ Gy cm}^2 \text{ min}^{-1}$ to $16.8 \text{ Gy cm}^2 \text{ min}^{-1}$), while providing 3D instead of 2D information (section 4.3). This similarity was expected since the X-ray image acquisition rate of 2-4-sep-bin-STs (10 fps = 2 threads \times 5 vps) is typical for 2D fluoroscopy (7.5 fps to 15 fps [74], [75]) and because, like 2D fluoroscopy, the proposed pipeline operates at a noise level that is just low enough to permit seeing the interventional material (central row in figure 4.6). Guiding an intervention with the proposed 3D reconstruction pipeline instead of conventional 2D fluoroscopy could even lower the cumulative DAP, since the improved spatial awareness provided by 3D instead of 2D images could lead to shorter intervention times. Shorter intervention times could furthermore reduce the use of anaesthetics, the use of toxic iodinated contrast agents [76], [77], and the cost of the intervention. A limitation of the experiments in this work is the unrealistic placement of the interventional material between the abdomen of the phantom and the soft tissue-equivalent extension. In a patient, the interventional material would be further away from the patient surface and therefore noise in lateral X-ray images would be higher. This effect leads to an underestimation of the required DAP. A more accurate DAP estimation might require a custom phantom.

Compared to previous work on interventional material reconstruction from very few X-ray images, the proposed reconstruction pipeline 2-4-sep-bin-STs has some obvious advantages. Compared to previous two-threaded approaches [21], [26]–[28], which are only able to reconstruct single curves (e.g. a single guidewire or a single catheter), the proposed pipeline is significantly more capable since it can reconstruct a stent and multiple guidewires being simultaneously present in the scene (see figure 4.5). Furthermore, the proposed pipeline is readily extendable to a) further types of interventional material (e.g. vena cava filter or radiopaque markers) by adding further output channels to the DTE or b) more crowded scenes by training the DTE to distinguish instances of the same type of interventional material (similar to an instance segmentation [78]). Compared to the previously proposed four-threaded reconstruction pipeline [34], which was the starting point of this thesis, the number of X-ray images per 3D reconstruction and the number of required imaging threads could be reduced from four to two while even increasing reconstruction quality (table 4.1 and figures 4.1 to 4.5).

Compared to those previously proposed approaches, the presented pipeline also has some disadvantages which stem from the required rotation of the X-ray system. First, the rotation induces angular blur in the X-ray images. To achieve an angular increment of $\Delta\alpha = 19^\circ$ at a 3D reconstruction rate of 5 vps, a rotation time of 3.79 s is necessary. Therefore, to match the angular blur of the stop motion scans of 0.29° (section 3.8), a relatively short pulse duration of $T_{\text{pulse}} = 3.1 \text{ ms}$ would be necessary, which limits the penetration ability of the X-ray system. Future work should investigate the maximum tolerable angular blur, which will depend on the size of the structures being imaged (e.g. struts of 4 mm diameter stents are much thinner and more densely packed than those of 30 mm stents) and on the distance to the center of rotation. Second, the rotation makes the use of simple but effective temporal filtering techniques, which are commonly used for noise reduction in 2D fluoroscopy, impossible. Third, while in conventional 2D

fluoroscopy, X-ray images are usually acquired in a posterior-anterior (p.a.) fashion, the 3D reconstruction pipeline acquires from all angles. Since the effective dose of a.p. X-ray image is about twice as high as that of p.a. X-ray images [73], [79] at constant DAP, the effective dose of the 3D reconstruction pipeline is expected to impart a slightly higher effective dose, even when DAP values are comparable.

The rotation of the X-ray system however also implies a dose-related advantage: peak skin dose is much lower since the radiation is distributed over a larger skin area. In 2D fluoroscopy, excessive peak skin dose occasionally causes severe damage to the patient's skin and underlying tissue [49], [50].

Reconstructions of pipeline 2-4-sep-bin-STs of average quality, like the ones shown in figures 4.1 to 4.5, show only minor deviations from the respective ground truths, e.g. small false negatives and false positives of the stent. These reconstruction errors are often associated with high-noise-induced false negatives in one of the two DTE output images of the current time step (see figure 4.6). Since these errors are not temporally consistent, better use of temporal information, e.g. by using the reconstructions of previous time steps as additional inputs or by using recurrent neural networks (e.g. convolutional long short-term memory (ConvLSTM) networks [80]), could be advantageous. Reconstructions of below-average quality sometimes show major deviations, e.g. larger false negatives in the stents and guidewires. The most severe deviations occur when one of the two input X-ray images is approximately lateral since noise is highest there. To remedy this issue, an automatic tube current modulation [81] might be necessary in addition to the above-mentioned algorithmic improvements.

Using two NVIDIA RTX 4090 GPUs, TensorFlow's mixed precision and graph execution, the computation time of pipeline 2-4-sep-bin-STs is 208 ms (table 4.6). This allows the reconstruction of 4.8 vps. While this makes the pipeline already real-time capable, an increase in computational efficiency would be advantageous to reduce image latency, to allow for higher spatial resolution 3D reconstructions, or to add further reconstruction channels (other than the stent and guidewire channels). Since the computation time is dominated by the DTR computation time (table 4.6), accelerating the DTR is most important. Such an acceleration could for example be achieved by integer quantization [82] or by neural network pruning [83].

6 | Conclusions

To enable guidance of interventional procedures by 3D instead of 2D fluoroscopy, the computation of a 3D reconstruction of interventional material must require as few newly acquired X-ray images as possible. Too many new X-ray images per 3D reconstruction would result in an excessive radiation dose rate and would require expensive imaging hardware. In this work, multiple improvements to an existing pipeline for the 3D reconstruction of interventional material were proposed, which allowed reducing the number of X-ray images per 3D reconstruction from four to two, while not only maintaining but even improving reconstruction quality. While previously proposed algorithms for the 3D reconstruction of interventional material from two X-ray images are only able to reconstruct a single guidewire or a single catheter, the proposed pipeline can reconstruct significantly more complex scenes, e.g. an arrangement of a stent and two guidewires. Submillimeter reconstruction accuracy was demonstrated on measured X-ray images of interventional material inside an anthropomorphic phantom with simulated respiratory motion. The measured DAP rate of the proposed 3D reconstruction pipeline is roughly similar to that of conventional single-plane (at half the frame rate) and biplane (at the same frame rate) 2D fluoroscopy. Its capability to reconstruct complex arrangements of interventional material, its straightforward extensibility to further types of interventional material, its real-time capability, its relatively low hardware demands, and its dose efficiency make the proposed pipeline a promising candidate to enable 3D fluoroscopy.

Bibliography

- [1] S. Rudin, D. R. Bednarek, and K. R. Hoffmann, “Endovascular Image-Guided Interventions (EIGIs),” *Medical Physics*, volume 35, number 1, page 11, 2008.
- [2] S. B. King and B. Meier, “Interventional Treatment of Coronary Heart Disease and Peripheral Vascular Disease,” *Circulation*, volume 102, number 4, 2000.
- [3] A. Ciccone, L. Valvassori, M. Nichelatti, A. Sgoifo, M. Ponzio, R. Sterzi, and E. Boccardi, “Endovascular Treatment for Acute Ischemic Stroke,” *New England Journal of Medicine*, volume 368, number 10, pages 904–913, 2013.
- [4] S. Allaqaband, R. Kirvaitis, F. Jan, and T. Bajwa, “Endovascular Treatment of Peripheral Vascular Disease,” *Current Problems in Cardiology*, volume 34, number 9, pages 359–476, 2009.
- [5] J. A. Ellis, E. Nossek, A. Kronenburg, D. J. Langer, and R. A. Ortiz, “Intracranial Aneurysm: Diagnostic Monitoring, Current Interventional Practices, and Advances,” *Current Treatment Options in Cardiovascular Medicine*, volume 20, number 12, page 94, 2018.
- [6] M. L. Schermerhorn and P. Cotterill, “Endovascular vs. Open Repair of Abdominal Aortic Aneurysms in the Medicare Population,” *New England Journal of Medicine*, 2008.
- [7] N. Fidelman, S. W. Kwan, J. M. LaBerge, R. L. Gordon, E. J. Ring, and R. K. Kerlan, “The Transjugular Intrahepatic Portosystemic Shunt: An Update,” *American Journal of Roentgenology*, volume 199, number 4, pages 746–755, 2012.
- [8] J.-L. Raoul, A. Forner, L. Bolondi, T. T. Cheung, R. Kloeckner, and T. de Baere, “Updated Use of TACE for Hepatocellular Carcinoma Treatment: How and When to Use It Based on Clinical Evidence,” *Cancer Treatment Reviews*, volume 72, pages 28–36, 2019.
- [9] R. Gupta, C. Walsh, I. S. Wang, M. Kachelrieß, J. Kuntz, and S. Bartling, “CT-Guided Interventions: Current Practice and Future Directions,” in *Intraoperative Imaging and Image-Guided Therapy*, New York, NY: Springer New York, 2014, pages 173–191.

- [10] S. J. Braak, M. J. L. van Strijen, M. van Leersum, H. W. van Es, and J. P. M. van Heesewijk, "Real-Time 3D Fluoroscopy Guidance During Needle Interventions: Technique, Accuracy, and Feasibility," *American Journal of Roentgenology*, volume 194, number 5, W445–W451, 2010.
- [11] J. M. Racadio, D. Babic, R. Homan, J. W. Rampton, M. N. Patel, J. M. Racadio, and N. D. Johnson, "Live 3D Guidance in the Interventional Radiology Suite," *American Journal of Roentgenology*, volume 189, number 6, W357–W364, 2007.
- [12] N Keat, "Real-Time CT and CT Fluoroscopy," *The British Journal of Radiology*, volume 74, number 888, pages 1088–1090, 2001.
- [13] J. J. Froelich and H.-J. Wagner, "CT-Fluoroscopy: Tool or Gimmick?" *CardioVascular and Interventional Radiology*, volume 24, number 5, pages 297–305, 2001.
- [14] S. G. Silverman, K. Tuncali, D. F. Adams, R. D. Nawfel, K. H. Zou, and P. F. Judy, "CT Fluoroscopy-guided Abdominal Interventions: Techniques, Results, and Radiation Exposure," *Radiology*, volume 212, number 3, pages 673–681, 1999.
- [15] A. Schegerer, R. Loose, L. J. Heuser, and G. Brix, "Diagnostic Reference Levels for Diagnostic and Interventional X-Ray Procedures in Germany: Update and Handling," *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 191, number 08, pages 739–751, 2019.
- [16] D. L. Miller, S. Balter, P. E. Cole, H. T. Lu, B. A. Schueler, M. Geisinger, A. Berenstein, R. Albert, J. D. Georgia, P. T. Noonan, J. F. Cardella, J. S. George, E. J. Russell, T. W. Malisch, R. L. Vogelzang, G. L. Miller, and J. Anderson, "Radiation Doses in Interventional Radiology Procedures: The RAD-IR Study Part I: Overall Measures of Dose," *Journal of Vascular and Interventional Radiology*, volume 14, number 6, pages 711–727, 2003.
- [17] G. N. Ramachandran and A. V. Lakshminarayanan, "Three-dimensional Reconstruction from Radiographs and Electron Micrographs: Application of Convolutions instead of Fourier Transforms," *Proceedings of the National Academy of Sciences*, volume 68, number 9, pages 2236–2240, 1971.
- [18] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical Cone-Beam Algorithm," *Journal of the Optical Society of America A*, volume 1, number 6, pages 612–619, 1984.
- [19] P. M. Joseph and R. A. Schulz, "View Sampling Requirements in Fan Beam Computed Tomography," *Medical Physics*, volume 7, number 6, pages 692–702, 1980.
- [20] Z. Zhao, G. J. Gang, and J. H. Siewerdsen, "Noise, Sampling, and the Number of Projections in Cone-Beam CT with a Flat-Panel Detector: Noise and Sampling in Cone-Beam CT," *Medical Physics*, volume 41, number 6, page 061909, 2014.

-
- [21] S. Baert, E. van de Kraats, T. van Walsum, M. Viergever, and W. Niessen, “Three-Dimensional Guide-Wire Reconstruction from Biplane Image Sequences for Integrated Display in 3D Vasculature,” *IEEE Transactions on Medical Imaging*, volume 22, number 10, pages 1252–1258, 2003.
- [22] T. van Walsum, S. A. M. Baert, and W. J. Niessen, “Guide Wire Reconstruction and Visualization in 3DRA Using Monoplane Fluoroscopic Imaging,” *IEEE Transactions on Medical Imaging*, volume 24, number 5, pages 612–623, 2005.
- [23] M. Brückner, F. Deinzer, and J. Denzler, “Temporal Estimation of the 3D Guide-Wire Position Using 2D X-Ray Images,” *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, Eds., pages 386–393, 2009.
- [24] J. Kuntz, B. Flach, R. Kueres, W. Semmler, M. Kachelrieß, and S. Bartling, “Constrained Reconstructions for 4D Intervention Guidance,” *Physics in Medicine and Biology*, volume 58, number 10, pages 3283–3300, 2013.
- [25] J. Kuntz, R. Gupta, S. O. Schönberg, W. Semmler, M. Kachelrieß, and S. Bartling, “Real-Time X-Ray-Based 4D Image Guidance of Minimally Invasive Interventions,” *European Radiology*, volume 23, number 6, pages 1669–1677, 2013.
- [26] M. Wagner, S. Schafer, C. Strother, and C. Mistretta, “4D Interventional Device Reconstruction from Biplane Fluoroscopy,” *Medical Physics*, volume 43, number 3, pages 1324–1334, 2016.
- [27] M. Hoffmann, A. Brost, C. Jakob, F. Bourier, M. Koch, K. Kurzdin, J. Hornegger, and N. Strobel, “Semi-automatic Catheter Reconstruction from Two Views,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, 2012, pages 584–591.
- [28] M. Hoffmann, A. Brost, C. Jakob, M. Koch, F. Bourier, K. Kurzdin, J. Hornegger, and N. Strobel, “Reconstruction Method for Curvilinear Structures from Two Views,” *SPIE Medical Imaging*, D. R. Holmes and Z. R. Yaniv, Eds., 86712F, 2013.
- [29] S. Demirci, A. Bigdelou, L. Wang, C. Wachinger, M. Baust, R. Tibrewal, R. Ghotbi, H.-H. Eckstein, and N. Navab, “3D Stent Recovery from One X-Ray Projection,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, volume 6891, 2011, pages 178–185.
- [30] E. Candes, J. Romberg, and T. Tao, “Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information,” *IEEE Transactions on Information Theory*, volume 52, number 2, pages 489–509, 2006.
- [31] D. Donoho, “Compressed Sensing,” *IEEE Transactions on Information Theory*, volume 52, number 4, pages 1289–1306, 2006.
- [32] B. Flach, J. Kuntz, M. Brehm, R. Kueres, S. Bartling, and M. Kachelrieß, “Low Dose Tomographic Fluoroscopy: 4D Intervention Guidance with Running Prior,” *Medical Physics*, volume 40, number 10, page 101 909, 2013.

- [33] B. Flach, M. Brehm, S. Sawall, and M. Kachelrieß, “Deformable 3D–2D Registration for CT and Its Application to Low Dose Tomographic Fluoroscopy,” *Physics in Medicine and Biology*, volume 59, number 24, pages 7865–7887, 2014.
- [34] E. Eulig, J. Maier, M. Knaup, N. R. Bennett, K. Hörndler, A. S. Wang, and M. Kachelrieß, “Deep Learning-Based Reconstruction of Interventional Tools and Devices from Four X-Ray Projections for Tomographic Interventional Guidance,” *Medical Physics*, volume 48, number 10, pages 5837–5850, 2021.
- [35] P. Henzler, V. Rasche, T. Ropinski, and T. Ritschel, “Single-image Tomography: 3D Volumes from 2D Cranial X-Rays,” *Computer Graphics Forum*, volume 37, number 2, pages 377–388, 2018.
- [36] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, “X2CT-GAN: Reconstructing CT From Biplanar X-Rays With Generative Adversarial Networks,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA: IEEE, 2019, pages 10 611–10 620.
- [37] Y. Kasten, D. Doktovsky, and I. Kovler, “End-To-End Convolutional Neural Network for 3D Reconstruction of Knee Bones from Bi-planar X-Ray Images,” in *Machine Learning for Medical Image Reconstruction*, Cham: Springer International Publishing, 2020, pages 123–133.
- [38] Y. Lei, Z. Tian, T. Wang, K. Higgins, J. D. Bradley, W. J. Curran, T. Liu, and X. Yang, “Deep Learning-Based Real-Time Volumetric Imaging for Lung Stereotactic Body Radiation Therapy: A Proof of Concept Study,” *Physics in Medicine & Biology*, volume 65, number 23, page 235 003, 2020.
- [39] T. M. Buzug, “Computed Tomography,” in *Springer Handbook of Medical Technology*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pages 311–342.
- [40] D. Stoeckel, C. Bonsignore, and S. Duda, “A Survey of Stent Designs,” *Minimally Invasive Therapy & Allied Technologies*, volume 11, number 4, pages 137–147, 2002.
- [41] G. G. Tóth, M. Yamane, and G. R. Heyndrickx, “How to Select a Guidewire: Technical Features and Key Characteristics,” *Heart*, volume 101, number 8, pages 645–652, 2015.
- [42] “Tissue Substitutes in Radiation Dosimetry and Measurement,” International Commission on Radiation Units and Measurements, Bethesda, MD (United States), Technical Report ICRU-44, 1989.
- [43] J. Radon, “Über die Bestimmung von Funktionen längs gewisser Mannigfaltigkeiten,” *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse*, volume 69, pages 262–277, 1917.
- [44] T. Vöth, “Cross-Scatter Correction in Dual Source CT Using the Deep Scatter Estimation,” M.S. thesis, Heidelberg University, Heidelberg, 2019.

-
- [45] H. K. Tuy, “An Inversion Formula for Cone-Beam Reconstruction,” *SIAM Journal on Applied Mathematics*, volume 43, number 3, pages 546–552, 1983.
- [46] D. M. Christensen, C. J. Iddins, and S. L. Sugarman, “Ionizing Radiation Injuries and Illnesses,” *Emergency Medicine Clinics of North America*, volume 32, number 1, pages 245–265, 2014.
- [47] L. K. Wagner, “Potential Biological Effects Following High X-ray Dose Interventional Procedures,” *Journal of Vascular and Interventional Radiology*, volume 5, number 1, 1994.
- [48] D. L. Miller, S. Balter, P. E. Cole, H. T. Lu, A. Berenstein, R. Albert, B. A. Schueler, J. D. Georgia, P. T. Noonan, E. J. Russell, T. W. Malisch, R. L. Vogelzang, M. Geisinger, J. F. Cardella, J. S. George, G. L. Miller, and J. Anderson, “Radiation Doses in Interventional Radiology Procedures: The RAD-IR Study Part II: Skin Dose,” *Journal of Vascular and Interventional Radiology*, volume 14, number 8, pages 977–990, 2003.
- [49] S. Balter, J. W. Hopewell, D. L. Miller, L. K. Wagner, and M. J. Zelefsky, “Fluoroscopically Guided Interventional Procedures: A Review of Radiation Effects on Patients’ Skin and Hair,” *Radiology*, volume 254, number 2, pages 326–341, 2010.
- [50] S. Balter and D. L. Miller, “Patient Skin Reactions From Interventional Fluoroscopy Procedures,” *American Journal of Roentgenology*, volume 202, number 4, W335–W342, 2014.
- [51] W. Jaschke, M. Schmuth, A. Trianni, and G. Bartal, “Radiation-Induced Skin Injuries to Patients: What the Interventional Radiologist Needs to Know,” *CardioVascular and Interventional Radiology*, volume 40, number 8, pages 1131–1140, 2017.
- [52] “ICRP Publication 26,” International Commission on Radiological Protection, Technical Report, 1977.
- [53] C. J. Martin, J. D. Harrison, and M. M. Rehani, “Effective Dose from Radiation Exposure in Medicine: Past, Present, and Future,” *Physica Medica*, volume 79, pages 87–92, 2020.
- [54] K. F. Seals, E. W. Lee, C. H. Cagnon, R. A. Al-Hakim, and S. T. Kee, “Radiation-Induced Cataractogenesis: A Critical Literature Review for the Interventional Radiologist,” *CardioVascular and Interventional Radiology*, volume 39, number 2, pages 151–160, 2016.
- [55] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep Learning for Visual Understanding: A Review,” *Neurocomputing*, volume 187, pages 27–48, 2016.
- [56] D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” 2017.

- [57] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, “A Survey on Deep Learning in Medical Image Analysis,” *Medical Image Analysis*, volume 42, pages 60–88, 2017.
- [58] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv, Tech. Rep. arXiv:1409.1556, 2015, arXiv:1409.1556 [cs].
- [59] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial Transformer Networks,” in *Advances in Neural Information Processing Systems*, volume 28, Curran Associates, Inc., 2015.
- [60] R. Rojas, “The Backpropagation Algorithm,” in *Neural Networks: A Systematic Introduction*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pages 149–182.
- [61] D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization,” *3rd International Conference for Learning Representations*, 2015.
- [62] O. R. Bingol and A. Krishnamurthy, “NURBS-Python: An Open-Source Object-Oriented NURBS Modeling Framework in Python,” *SoftwareX*, volume 9, pages 85–94, 2019, Publisher: Elsevier.
- [63] Martín Abadi, Ashish Agarwal, Paul Barham, *et al.*, *Tensorflow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015.
- [64] B. Ohnesorge, T. Flohr, and K. Klingenberg-Regn, “Efficient Object Scatter Correction Algorithm for Third and Fourth Generation CT Scanners,” *European Radiology*, volume 9, number 3, pages 563–569, 1999.
- [65] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015.
- [66] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432, 2016.
- [67] A. Takemura, M. Suzuki, K. Sakamoto, T. Kitada, H. Iida, Y. Okumura, and H. Harauchi, “Analysis of Respiration-Related Movement of Upper Abdominal Arteries: Preliminary Measurement for the Development of a Respiratory Motion Compensation Technique of Roadmap Navigation,” *Radiological Physics and Technology*, volume 1, number 2, pages 178–182, 2008.
- [68] M. T. Draney, C. K. Zarins, and C. A. Taylor, “Three-Dimensional Analysis of Renal Artery Bending Motion During Respiration,” *Journal of Endovascular Therapy*, volume 12, number 3, pages 380–386, 2005.
- [69] “Tumor Motion Ranges Due to Respiration and Respiratory Motion Characteristics,” in *Treating Tumors That Move with Respiration*, series Robotic Radiosurgery, H. C. Urschel, J. J. Kresl, J. D. Luketich, L. Papiez, and E. Thomson, Eds., Berlin Heidelberg New York: Springer, 2007, page 10.

-
- [70] K. Barrett, S. Barman, S. Boitano, and H. Brooks, *Ganong's Review of Medical Physiology*, 24th. McGraw Hill Professional, 2012.
- [71] T. C. Lee, R. L. Kashyap, and C. N. Chu, "Building Skeleton Models Via 3-D Medial Surface Axis Thinning Algorithms," *CVGIP: Graphical Models and Image Processing*, volume 56, number 6, pages 462–478, 1994.
- [72] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Goullart, and T. Yu, "Scikit-Image: Image Processing in Python," *PeerJ*, volume 2, e453, 2014.
- [73] J. C. L. Heron, "Estimation of Effective Dose to the Patient During Medical X-Ray Examinations from Measurements of the Dose-Area Product," *Physics in Medicine and Biology*, volume 37, number 11, pages 2117–2126, 1992.
- [74] N. I. Akkus, G. S. Mina, A. Abdalbaki, F. Shafiei, and N. Tandon, "Using 7.5 Frames Per Second Reduces Radiation Exposure in Lower Extremity Peripheral Vascular Interventions.," *Vascular*, volume 23, number 3, pages 240–244, 2015, Place: England.
- [75] K. Sadamatsu and Y. Nakano, "The Effect of Low Frame Rate Fluoroscopy on the X-ray Dose during Coronary Intervention," *Internal Medicine*, volume 55, number 15, pages 1943–1946, 2016.
- [76] K. M. Hasebroock and N. J. Serkova, "Toxicity of MRI and CT contrast agents," *Expert Opinion on Drug Metabolism & Toxicology*, volume 5, number 4, pages 403–416, 2009.
- [77] H. Lusic and M. W. Grinsta, "X-ray-Computed Tomography Contrast Agents," *Chem. Rev.*, 2013.
- [78] A. M. Hafiz and G. M. Bhat, "A Survey on Instance Segmentation: State of the Art," *International Journal of Multimedia Information Retrieval*, volume 9, number 3, pages 171–189, 2020.
- [79] D. Gosch, K. Gosch, and T. Kahn, "Konversionsfaktoren zur Ermittlung der effektiven Dosis für Patienten aus dem Dosisflächenprodukt bei Röntgendurchleuchtungsuntersuchungen," *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 179, number 10, pages 1035–1042, 2007.
- [80] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., volume 28, Curran Associates, Inc., 2015.
- [81] M. K. Kalra, M. M. Maher, T. L. Toth, B. Schmidt, B. L. Westerman, H. T. Morgan, and S. Saini, "Techniques and Applications of Automatic Tube Current Modulation for CT," *Radiology*, volume 233, number 3, pages 649–657, 2004.
- [82] H. Wu, P. Judd, X. Zhang, M. Isaev, and P. Micikevicius, "Integer Quantization for Deep Learning Inference: Principles and Empirical Evaluation," arXiv, Tech. Rep. arXiv:2004.09602, 2020, arXiv:2004.09602 [cs, stat].

- [83] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Gutttag, “What is the State of Neural Network Pruning?” In *Proceedings of the 3rd MLSys Conference*, Austin, TX, USA, 2020.

List of Publications

Journal Articles

- [1] T. Vöth, T. Koenig, E. Eulig, M. Knaup, V. Wiesmann, K. Hörndler, and M. Kachelrieß, “Real-Time 3D Reconstruction of Guidewires and Stents Using Two Update X-Ray Projections,” *submitted to Medical Physics*, 2023.
- [2] J. Erath, T. Vöth, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, “Deep Learning-Based Forward and Cross-Scatter Correction in Dual-Source CT,” *Medical Physics*, volume 48, number 9, pages 4824–4842, 2021.

Conference Contributions

- [1] T. Vöth, T. König, E. Eulig, M. Knaup, K. Hörndler, and M. Kachelrieß, “Real-Time 3D Reconstruction of Multiple Guidewires at Dose Values of Conventional 2D Fluoroscopy,” in *Medical Imaging 2022: Physics of Medical Imaging*, San Diego, United States: SPIE, 2022, page 75.
- [2] T. Vöth, T. König, E. Eulig, M. Knaup, V. Wiesmann, K. Hörndler, and M. Kachelrieß, “3D Reconstruction of Stents and Guidewires in an Anthropomorphic Phantom From Three X-Ray Projections,” in *Proceedings of the 7th International Conference on Image Formation in X-Ray Computed Tomography*, 2022, pages 468–471.
- [3] P. Trapp, T. Vöth, C. Amato, S. Sawall, and M. Kachelriess, “DeepRAR: A CNN-based Approach for CT and CBCT Ring Artifact Reduction,” in *Medical Imaging 2022: Physics of Medical Imaging*, W. Zhao and L. Yu, Eds., San Diego, United States: SPIE, 2022, page 70.

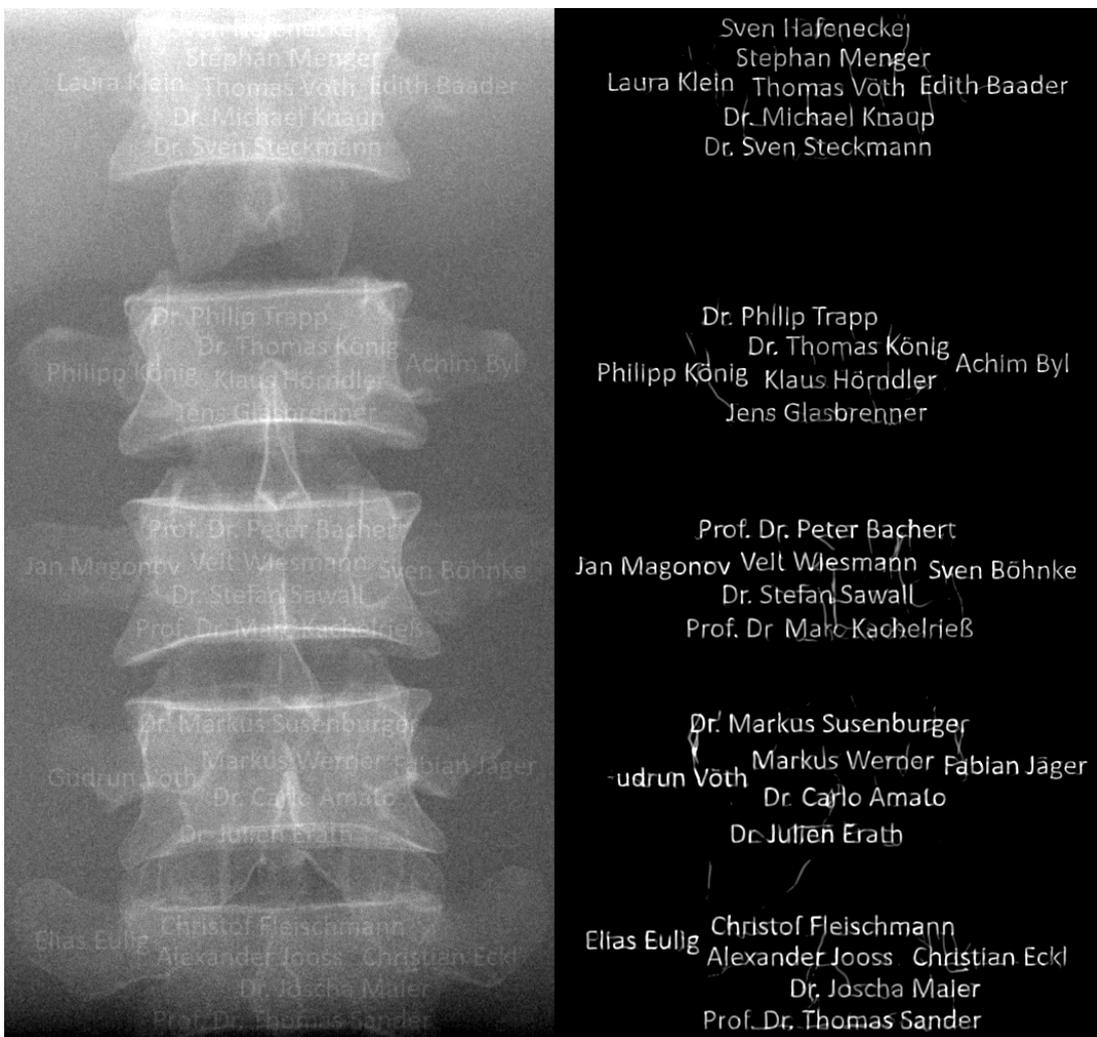
Patent Applications

- [1] T. Vöth, T. Koenig, K. Hörndler, M. Knaup, and M. Kachelrieß, “Verfahren zum Betreiben eines medizinischen Bildgebungsgeräts zur lagerichtigen Darstellung von nicht-anatomischen Strukturen während einer bildgebenden Untersuchung und Vorrichtung hierfür,” patent DE102021003956A1, 2023.

-
- [2] T. Vöth, T. Koenig, K. Hörndler, M. Knaup, and M. Kachelrieß, “Method and Device for Operating a Medical Imaging Device for the Positionally Correct Representation of Non-Anatomical Structures During an Imaging Examination,” patent US20230032731A1, 2023.
 - [3] T. Vöth, T. Koenig, K. Hörndler, M. Knaup, and M. Kachelrieß, “Method for Operating a Medical Imaging Device for the Correct Presentation of Non-Anatomical Structures During an Imaging Examination and Device Therefore,” patent EP4129189A1, 2023.

Acknowledgments

I want to thank everyone who enabled this thesis. Without your supervision, advice, help and support, this thesis would not have been possible.



I hereby assure, that I composed this work by myself and did not use any other than the listed resources.

Heidelberg, May 15, 2023

Tim Vollm

