Learning projection matrices for marker free motion compensation in weight-bearing CT scans

Valentin Bacher[1], Christopher Syben[1], Andreas Maier[1], and Adam Wang[2]

[1]Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
[2]Departments of Radiology and Electrical Engineering, Stanford University, Stanford, California

Abstract Weight-Bearing Computed Tomography is gaining popularity due to its ability to generate a three-dimensional reconstruction of joints under weight-bearing condition. The major problem regarding scanning are motion artifacts which are inevitable, because of the standing position of the patient. State of the art methods overcome these problems using aids, such as external tracking devices, prior knowledge, or fiducial markers. Those methods require knowledge which might not be existent or are tedious or computationally expensive. Therefore, we investigate the possibility of using trainable CT operators to compensate rigid motion without any kind of aids or prior knowledge.

The motion is estimated and corrected using a TensorFlow based API (PYRO-NN) to incorporate the learning of projection matrices into an iterative comparison of the original sinogram with digitally reconstructed radiographs (DRRs) obtained from the motion-corrupted backprojected volume. The loss function used is a squared loss. This error is then used to calculate a gradient using a finite difference, which is used in an Adam optimizer to iteratively reduce the loss. This approach is able to estimate and correct the shifts, although it is unable to correct small rotations. The results of our simulation study show that this is still sufficient to obtain a very good reconstruction with a low level of motion artifacts.

1 Introduction

Weight-Bearing Computed Tomography (WBCT), based on Cone-Beam CT (CBCT) geometries, has established itself as a core modality in orthopedic imaging of the lower extremities [1]. It allows the imaging of the joints under weight-bearing conditions and thus valuable insights in the pathology of the bones, cartilage, and other soft tissue in proximity of the joints [1]. The greatest challenge associated with this imaging setup is motion artifacts. These are inevitable due to the upright standing position of the patient during scan time. Combined with relatively long scan times associated with cone beam scanners this causes severe motion artifacts.

Most state of the art methods use either physical aids or prior knowledge to reduce motion artifacts. Several approaches use fiducial markers placed on the skin as orientation [2–4]. However the placement of these markers is tedious and can cause errors due to movement of the skin relative to the bone. Thus Berger et al. proposed to use prior scans of the respective bone structure to register the projections [5]. While this is promising, it requires prior scans which are not always available. Algorithms based on sharpness metrics or image consistencies have been developed for other use cases with different approaches. One such, which originated in emission tomography, uses subdivision of the sinogram into almost motion-free subsets [6–8]. Other approaches use sharpness metrics on the motion corrupted reconstruction as a loss function [9]. Bodensteiner et al. proposed to evaluate a least squares difference between digitally reconstructed radiographs (DRRs) and the original projections to iteratively reduce motion artifacts [10].

Integrating trainable trainable layers into a fixed weight pipelines has proven it self as very beneficial [11]. We propose to use the differential CT operators introduced by PYRO-NN [12]. PYRO-NN implements forward and backward projectors as differential layers in deep learning frameworks. However, in the current version the weights of the CT operator layers are not trainable. In this proof-of-concept work we extend PYRO-NN to trainable CT operators and therefore allow to optimize for geometrical misalignment, e.g. occurring motion or miscalibration by learning compensated projection matrices specific to each scan.

2 Definition of the Model

For the proof-of-concept study we propose a model very similar to an iterative reconstruction pipeline. For this we make use of the differentiable CT layers of PYRO-NN [12]. Their theoretical description can be introduced by the system matrix $\mathbf{A}$ for the forward projection operator and by $\mathbf{A}^\top$ for the back-projection operand. As shown in the paper, the filtered back-projection (FBP) algorithm to reconstruct tomographic images $\mathbf{v}$ from sinogram data $\mathbf{p}$ can be fully described within the context of neural networks. The fundamental reconstruction problem can be described by performing a pseudo inverse of the forward model $\mathbf{p} = \mathbf{A}\mathbf{v}$, resulting in $\mathbf{v} = \mathbf{A}^\top \mathbf{F}^\mathbf{H} \mathbf{C}\mathbf{F}$, where $\mathbf{C}$ denotes the reconstruction filter in Fourier domain and $\mathbf{F}, \mathbf{F}^\mathbf{H}$ are the Fourier and inverse Fourier transform, respectively. As for realistic settings the system matrix $\mathbf{A}$ is infeasible to store in memory, the layers make use of the concept of projection matrices $\mathbf{R}^{3\times 4}$, where one projection matrix describes a certain projection and a set of projection matrices forms the scanning trajectory acquiring sinogram data for tomographic reconstruction. However, up to now these CT operator layers are non-trainable and differentiable with respect to their input. In this proof-of-concept work we investigate the possibility to extend the CT operator layers to trainable layers, where the projection matrices are the weights of the layers. This can be described by $\mathbf{A}(w_i)$ and $\mathbf{A}^\top(w_i)$, where w_i are the newly introduced trainable weights.

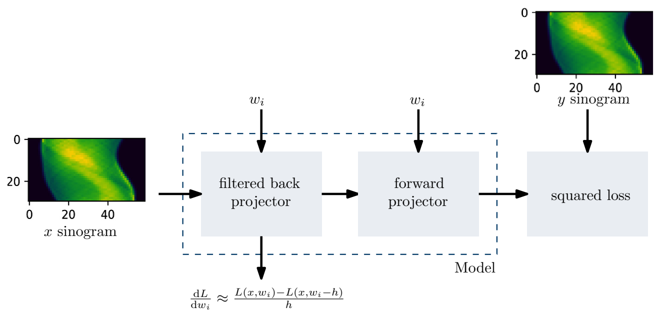


Figure 1: This model using a complete pipeline is used to train the projection matrices.

Trainable layers allow setups to update the geometric situation of the projection process and therefore, allow to setup pipelines for motion compensation. A naive implementation where we learn the projections matrices directly would lead to $N \times 3 \times 4$ trainable weights for N projections in the trajectory. To reduce the amount of trainable parameters we used a static definition of the intrinsic camera parameters and a non static description of the extrinsic parameters consisting of three rotations and three translations. With this, we define the x-axis parallel to the detector plane horizontal, the scanner rotation is about the y-axis, and the z-axis points towards the detector. Linear motion parallel to the central ray is neglected based on some initial experiments and the findings of Gullberg et al. [13]. Scaling effects caused by this kind of motion are very small due to small cone angles. Consequently our model for motion compensation takes a motion-corrupted sinogram, which is reconstructed by the first layer. The output of the first layer is directly forward projected using the same geometric parameters as the first layer. The resulting re-projected sinogram is evaluated by a loss against the original sinogram. The complete model with the loss is shown in Fig. 1. The mathematical description of the layers results in

$$\hat{\mathbf{y}} = \mathbf{A}(w_i)\mathbf{A}^\top(w_i)\mathbf{F^H}\mathbf{C}\mathbf{F}\mathbf{x} \quad . \tag{1}$$

The loss is determined using a simple squared difference $L(\mathbf{x}, w_i) = (\hat{\mathbf{y}}(\mathbf{x}, w_i) - \mathbf{y})^2$. For the proof-of-concept study, we compute the gradient with respect to the weights of the layers using a finite backward difference resulting in

$$\frac{\partial L(\mathbf{x}, w_i)}{\partial w_i} = \frac{L(\mathbf{x}, w_i) - L(\mathbf{x}, w_i - h)}{h} \quad . \tag{2}$$

The training is conducted using the Adam optimizer and the stepsize for the backward difference is set to $h = 0.002$.

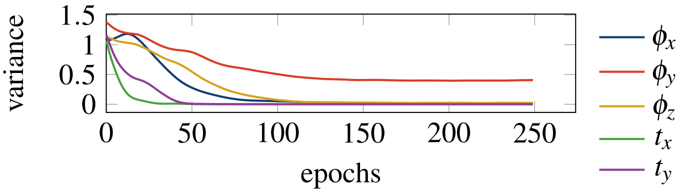
3 Experiments Description

The experiments aim to evaluate the possibility of iteratively reducing motion artifacts in cone-beam CT data using a PYRO-NN based trainable reconstruction pipeline. The first experiment investigates the possibility to achieve meaningful gradients for the layer weights. For this, the model shown in Fig. 1 is further simplified to just do a reconstruction and compute a loss with a motion-free label reconstruction. The second experiment is conducted with the presented model to be free of supervision and is inspired by iterative reconstruction methods.

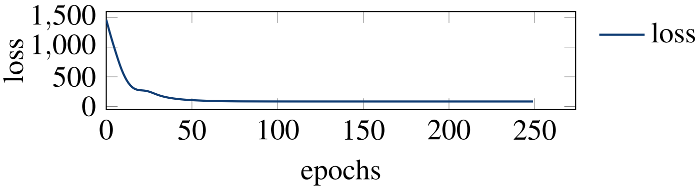
The simulated scanner setup consists of a 120×120 pixel flat-panel detector with a pixel size of 1×1 mm^2. The source is located at a distance of 1200 mm, while the source-to-isocenter distance measures 750 mm. The scanner simulated 30 projections obtained in a scan range of $180° + 2 \cdot 2.84°$. The data used in the experiments are simulated 3D Shepp-Logan phantoms. The phantom measures $60 \times 60 \times 60$ mm^3 with a pixel size of $0.5 \times 0.5 \times 0.5$ mm^3. Rigid motion is modeled by uniformly random shifts and rotations in the range of -2 to 2 [mm] or [°], respectively, for each projection with which the Shepp-Logan phantom is forward projected to obtain the motion-corrupted sinogram to reconstruct. The range of motion is oriented at the range of motion of a standing person [14]. To gain independence of the initial position of the phantom, the phantom is uniformly randomly shifted and rotated ensuring that the phantom stays within the field of view. The learning rate for the first experiment is chosen to be $\eta = 9 \cdot 10^{-2}$. The second experiment is conducted with a learning rate of $\eta = 1 \cdot 10^{-1}$. The trainable variables consisted of the five motion parameters $w_i = \{\phi_x, \phi_y, \phi_z, t_x, t_y\}$ for each of the 30 projections. Note that in the proof-of-concept study we seek to find the true weights to a specific corrupted scan and therefore only one sinogram is used for training.

4 Results

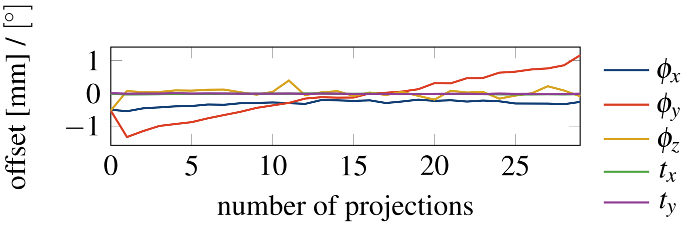
The results of the first experiment are shown in Figure 2 and show the training results for a single scan of a randomly initialized phantom, where these results are representative of other randomly initialized phantoms and motion. In Fig. 2a the variance of the difference over all 30 projections of the layer weights with respect to the ground truth parameter over the training procedure are plotted. For all parameters, except ϕ_y, which describes the axis of the scanner's rotation, the ground truth parameter can be learned. Fig. 2c indicates that the high variance is due to a point symmetrical error of the rotation angle. The loss (cf. Fig. 2b) correlates well with the variance of the shifts, while the rotations do not have a large effect on the loss.Fig. 3 shows the results of the second experiment. While for the shifts the true parameter could be found, this is not true for the rotations. Fig. 3c plots the difference between the learned parameters and the ideal ones. The shifts approach nearly zero, with a small negative constant offset for t_y. The rotations show random behaviour. Fig. 4 shows multi-planar views of the reconstruction of the phantom before and after the motion correction. While the transformation of the phantom stays the same, the sinograms used for the reconstruction differs in their type of motion corruption. A distortion just caused by rotations does not lead



(a) Variance across all 30 projections.

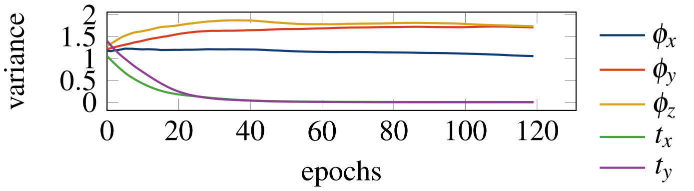


(b) The squared loss over the epochs.

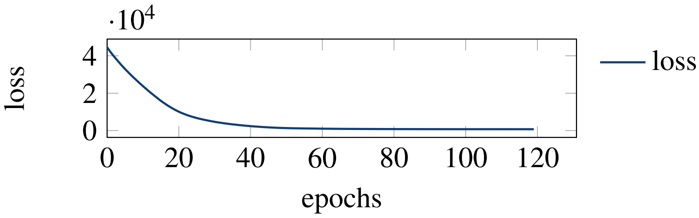


(c) Learned variables plotted over all 30 projections.

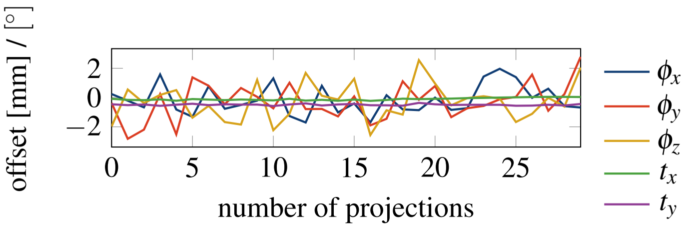
Figure 2: All results above are obtained from training on a randomly chosen single phantom, assuming a known reconstruction.



(a) Variance across all 30 projections.



(b) The squared loss over the epochs.



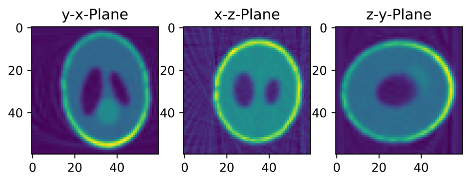
(c) Trained variables plotted over all 30 projections.

Figure 3: Results from training on a randomly chosen single phantom, without prior knowledge. For better visualization, the expected value are subtracted in sub figure (c).

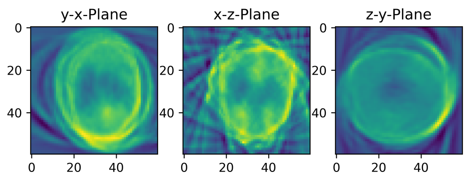
to a substantial reduction in image quality. The shifts however cause severe motion artifacts. In general, all reconstructions using the learned parameters are very well defined, and the main artifacts are caused by the relatively low number of projections used.

5 Discussion

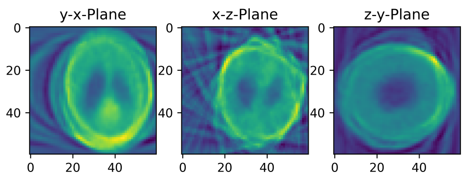
With both experiments, we were able to show that it is possible to interpret the geometric parameters of the CT operators as trainable layers and successfully update these layer weights in a training setting. With the learned parameters the image quality of the tomographic can be improved, even though the true value of all parameters could not be learned. While we were able to learn the shifts, this is not true for the rotations. In the first experiment, two out of three rotation parameter could be learned, while in the second experiment no stable solution for all three could be learned. The final learned rotation parameters do not have a substantial effect on the image quality of the reconstruction nor the loss. This indicates that the gradient response of artifacts imposed by rotation are neglected in the training process compared to the artifacts introduced by the translation parameters. This shortcoming could be overcome by introducing a weighted loss, focusing more on the rotation parameters or by choosing a composite loss function including metrics more sensitive to the artifacts introduced by the rotation parameters, e.g., incorporating sharpness metrics of the reconstructed image such as total variation. In 2017, Sisniega et al. published promising results using such a metric. They also added a term penalizing unrealistic jumps in motion parameters. Also, the object to reconstruct has an impact on the sensitivity of the loss function w.r.t changes in rotations, with the Shepp-Logan phantom particularly insensitive. Another observation to discuss is that for both experiments the reconstruction with only 30 projections leads to strong sparse sampling artifacts in the tomographic domain and therefore a substantial portion of the loss reflects artifacts which cannot be compensated by the proposed setup. This can negatively effects the training process, e.g., situations where motion artifacts can cancel out or reduce sparse sampling artifacts would lead to minima in the cost function which do not correspond with ideal layer weights. We are aware that this limitation is introduced by the limited amount of data, which is due to memory limitations in the context of 3D tomography and the use of finite difference for the gradient computation. However, the general results are promising and this does not diminish their insights as even with the limited data, the algorithm estimated the trajectory well enough to obtain good results. For this proof-of-concept study we use a finite difference, which is a computationally expensive estimation of the gradient. In the future, approximate or analytical partial gradients, which would utilize the back propagation algorithm, are desirable. The robustness with respect to a low amount of data may allow use of a Gaussian pyramid, which not only reduces spatial resolution but also the number of projections to estimate motion.



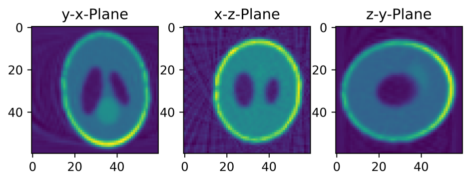
(a) Distorted reconstruction, rotations only.



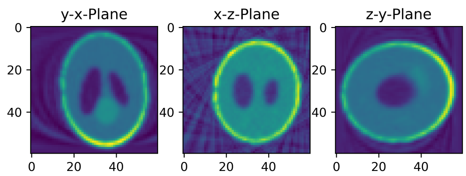
(b) distorted reconstruction, shifts only



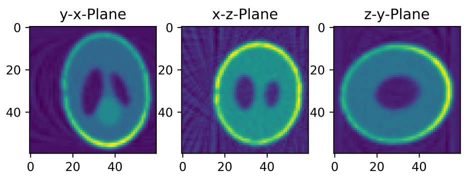
(c) Distorted reconstruction, all parameters.



(d) Learned reconstruction, rotations only.



(e) Learned reconstruction, shifts only.



(f) Learned reconstruction, all parameters.

Figure 4: Visualized are the reconstructions before (a-c) and after (d-f) the learning of the projection matrices. All graphs are the result of a training without any prior knowledge.

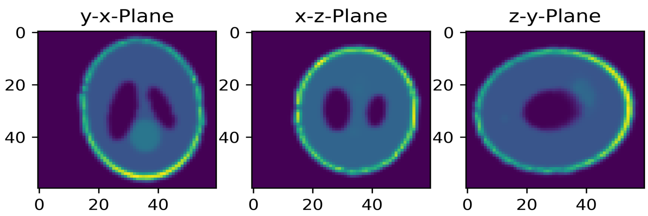


Figure 5: The randomly transformed phantom used for the results shown in Figure 4.

6 Conclusion

We were able to show that the general concept of trainable CT operators in neural networks are feasible. Further, an iterative reduction of motion artifacts by defining the trajectory parameters as trainable weights is possible. To make our approach applicable to clinical use, gradient computation needs to be improved and the loss function needs to be further investigated. The inability of the rotations to converge did not have a significant effect on the image quality. Still this approach is promising since it has proven itself as robust with respect to a small amount of data. Thus it could be combined with multi-scale algorithms increasing the performance.

7 Acknowledgements

This work was partially supported by NIH S10RR026714, NIH R01AR065248, and Siemens Healthineers.

References

[1] M. Richter, F. Lintz, C. de Cesar Netto, et al. *Weight Bearing Cone Beam Computed Tomography (WBCT) in the Foot and Ankle*. Springer, 2020. DOI: 10.1007/978-3-030-31949-6_24.

[2] M. Berger. “Motion-Corrected Reconstruction in Cone-Beam Computed Tomography for Knees under Weight-Bearing Condition”. PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg.

[3] J.-H. Choi, R. Fahrig, A. Keil, et al. “Fiducial marker-based correction for involuntary motion in weight-bearing C-arm CT scanning of knees. Part I. Numerical model-based optimization”. *Medical physics* 40.9 (2013). DOI: 10.1118/1.4817476.

[4] J.-H. Choi, A. Maier, A. Keil, et al. “Fiducial marker-based correction for involuntary motion in weight-bearing C-arm CT scanning of knees. II. Experiment”. *Medical physics* 41.6Part1 (2014). DOI: 10.1118/1.4873675.

[5] M. Berger, K Müller, A Aichert, et al. “Marker-free motion correction in weight-bearing cone-beam CT of the knee joint”. *Medical physics* 43.3 (2016). DOI: 10.1118/1.4941012.

[6] A. Z. Kyme, B. F. Hutton, R. L. Hatton, et al. “Practical aspects of a data-driven motion correction approach for brain SPECT”. *IEEE transactions on medical imaging* 22.6 (2003). DOI: 10.1109/TMI.2003.814790.

[7] M. Blume, A. Keil, N. Navab, et al. “Blind motion compensation for positron-emission-tomography”. Vol. 7258. 2009. DOI: 10.1117/12.811118.

[8] S Ens, J Ulrici, E Hell, et al. “Automatic motion correction in cone-beam computed tomography”. IEEE. 2010. DOI: 10.1109/NSSMIC.2010.5874405.

[9] A. Sisniega, J. W. Stayman, J Yorkston, et al. “Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion”. *Physics in Medicine & Biology* 62.9 (2017). DOI: 10.1088/1361-6560/aa6869.

[10] C. Bodensteiner, C. Darolti, H. Schumacher, et al. “Motion and Positional Error Correction for Cone Beam 3D-Reconstruction with Mobile C-Arms”. Springer Berlin Heidelberg, 2007. DOI: 10.1007/978-3-540-75757-3_22.

[11] A. K. Maier, C. Syben, B. Stimpel, et al. “Learning with known operators reduces maximum error bounds”. *Nature machine intelligence* 1.8 (2019). DOI: 10.1038/s42256-019-0077-5.

[12] C. Syben, M. Michen, B. Stimpel, et al. “Technical Note: PYRO-NN: Python reconstruction operators in neural networks”. *Medical Physics* (2019). DOI: 10.1002/mp.13753.

[13] G. T. Gullberg, B. M. W. Tsui, C. R. Crawford, et al. “Estimation of geometrical parameters and collimator evaluation for cone beam tomography”. *Medical Physics* 17.2 (1990). DOI: 10.1118/1.596505.

[14] A Maier, J.-H. Choi, A Keil, et al. “Analysis of vertical and horizontal circular C-arm trajectories”. Vol. 7961. International Society for Optics and Photonics. 2011. DOI: 10.1117/12.878502.