Dynamic Pacemaker Artifact Removal (DyPAR) from CT Data using CNNs

Tanja Lossau (née Elss)[1,2]　　　　　　　　　　　　　　　　TANJA.LOSSAU@PHILIPS.COM

Hannes Nickisch[1]　　　　　　　　　　　　　　　　HANNES.NICKISCH@PHILIPS.COM

Tobias Wissel[1]　　　　　　　　　　　　　　　　TOBIAS.WISSEL@PHILIPS.COM

Samer Hakmi[3]　　　　　　　　　　　　　　　　S.HAKMI@UKE.DE

Clemens Spink[4]　　　　　　　　　　　　　　　　C.SPINK@UKE.DE

Michael M. Morlock[2]　　　　　　　　　　　　　　　　MORLOCK@TUHH.DE

Michael Grass[1]　　　　　　　　　　　　　　　　MICHAEL.GRASS@PHILIPS.COM

[1] *Philips Research, Hamburg, Germany*　　[2] *Hamburg University of Technology, Germany*
[3] *University Heart Center Hamburg, Germany*　　[4] *Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Germany*

Abstract

Metal objects in the human heart like implanted pacemakers frequently occur in elderly patients. Due to cardiac motion, they are not static during the CT acquisition and lead to heavy artifacts in reconstructed CT image volumes. Furthermore, cardiac motion precludes the application of standard metal artifact reduction methods which assume that the object does not move. We propose a deep-learning-based approach for dynamic pacemaker artifact removal which deals with metal shadow segmentation directly in the projection domain. The data required for supervised learning is generated by introducing synthetic pacemaker leads into 14 clinical data sets without pacemakers. CNNs achieve a Dice coefficient of 0.913 on test data with synthetic metal leads. Application of the trained CNNs on eight data sets with real pacemakers and subsequent inpainting of the post-processed segmentation masks leads to significantly reduced metal artifacts in the reconstructed CT image volumes.

Keywords: Cardiac CT, Metal Artifact Reduction, Convolutional Neural Networks

1. Introduction

High-density objects like metallic devices lead to streak-shaped artifacts in reconstructed CT images which significantly degrade the image quality and the diagnostic value, see Figure 1. These artifacts are mainly caused by inconsistencies in the projection data due to effects like beam hardening and photon starvation [9]. Therefore, standard procedures for metal artifact reduction (MAR) are based on sinogram inpainting by treating metal-affected values as missing data. They comprise the following steps: 1) segmentation of the metal in an initially reconstructed image volume 2) forward projection of the metal mask to yield the metal shadow in the sinogram 3) inpainting of the metal shadow 4) reconstruction of the improved image volume with reinserted metal mask from step 1). A multitude of variants of this method is known [3; 8], but all of them assume that the object is static during acquisition.
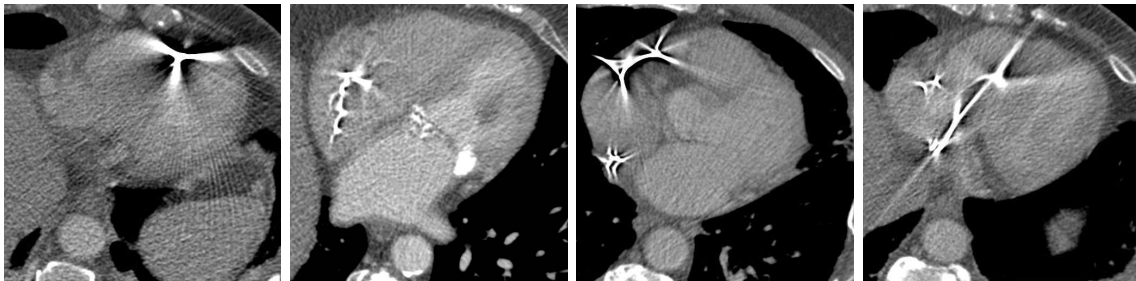
Figure 1: Pacemakers may lead to heavy streak-shaped artifacts in cardiac CT images. Especially in ungated CT data, motion aggravates artifacts caused by electrodes and leads.

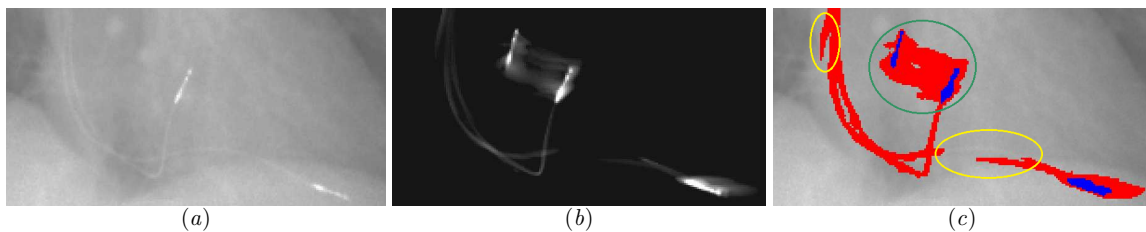(a)   (b)   (c)

Figure 2: Standard MAR procedures are not applicable for moving metal objects like pacemakers. (a) Projection view which contains the metal shadow of two pacemaker leads and electrodes. (b) The corresponding forward projected pacemaker shadow is perturbed by cardiac motion artifacts. (c) Duplicated electrode shadows (green circle) as well as shifted and interrupted lead shadows (yellow cicles) preclude conventional sinogram inpainting.

In case of implanted pacemakers, motion aggravates the metal artifacts (see Figure 1) and additionally hampers the evaluation of neighboring anatomy for example with regard to inflammation or calcification. As illustrated in Figure 2, standard MAR procedures fail in the presence of motion due to inconsistencies between the real metal shadow and the estimated one (see Figure 2b) which is obtained by step 2) of the standard MAR procedure.

In this paper, an alternative approach for dynamic MAR is investigated which deals with deep-learning-based segmentation of the metal shadow directly in projection data. Convolutional neural networks (CNNs) have been driving advances in several pattern recognition and semantic segmentation tasks [6; 10]. Also in the field of CT MAR they are increasingly used, e.g. for sinogram correction [2] and image-based artifact reduction [13; 11]. In contrast to these methods, our dynamic pacemaker artifact removal (DyPAR) approach does not rely on initially reconstructed image volumes which are potentially motion perturbed, i.e. it is applicable to static and non-static objects.

We use a forward model to synthesize pacemaker lead projection data and train CNNs to segment the metal directly in the projection domain. Inpainting of the post-processed segmentation masks and subsequent reconstruction yield CT image volumes with reduced artifacts. The generalization capabilities and transferability to MAR in clinical practice are investigated based on eight test cases with real pacemakers and two additional test cases without pacemaker.

2. Data

2.1. Synthetic learning data

Section 3.1 details the generation of projection data with synthetic pacemaker leads and corresponding segmentation masks for subsequent network training. The following described *reference data with pacemakers* and *target data without pacemakers* form the basis of this synthetic learning data. Leads are inserted into the *target data without pacemakers*, whereby reasonable lead positions and pathways are extracted from the *reference data with pacemakers*.

Reference data with pacemakers Seven reconstructed CT image volumes with pacemakers are collected for the extraction of pacemaker lead positions with respect to the cardiac anatomy. Dual as well as triple chamber pacemakers are included, i.e. synthesis of right atrial leads, right ventricular leads and coronary sinus leads is aimed for.

Target data without pacemakers Synthetic leads are inserted into 14 contrast-enhanced clinical cardiac CT data sets without pacemakers. The reconstructed image volumes as well as the corresponding raw projection data are required for the insertion process. In all target cases, acquisition was performed with a 256-slice CT scanner (Brilliance iCT, Philips Healthcare, Cleveland, Oh , USA) using a retrospective gating protocol without dose modulation. The helical trajectory exhibits a gantry rotation speed of 0.272 sec per turn. The pitch and the number of recorded projection views per turn vary between 0.16/0.18 and 1800/2400, respectively.

2.2. Clinical test data

In order to investigate generalization capabilities of the trained networks and the proposed DyPAR approach in clinical practice, ten additional cardiac CT data sets (i.e. the reconstructed image volume and the corresponding raw projection data) are collected. Eight data sets exhibit real pacemakers and allow one to evaluate the robustness of the CNNs with regard to variations in contrast-enhancement, motion levels, acquisition settings (pitch, w/wo dose modulation, w/wo gating) and lead pathways. Furthermore, the networks behavior in the presence of unseen features like electrodes or defibrillators is investigated. The remaining two test data sets without pacemakers are considered to quantify false positives. Both exhibit severe calcifications at the coronary arteries and the aortic valve. Furthermore, sternal steel wires for median sternotomy closure and a stent at the left main artery are present in one of these *no-pacemaker data sets*. The evaluation results are presented in Section 4.1.

3. Method

CNNs are trained for the task of segmenting pacemaker metal shadows in the projection domain. The required data for supervised learning is generated using a forward model for synthetic lead introduction. The Subsections 3.1 and 3.2 detail the data generation and the supervised learning processes. Finally, the trained networks are integrated into the DyPAR pipeline for application on clinical test data as described in Subsection 3.3.
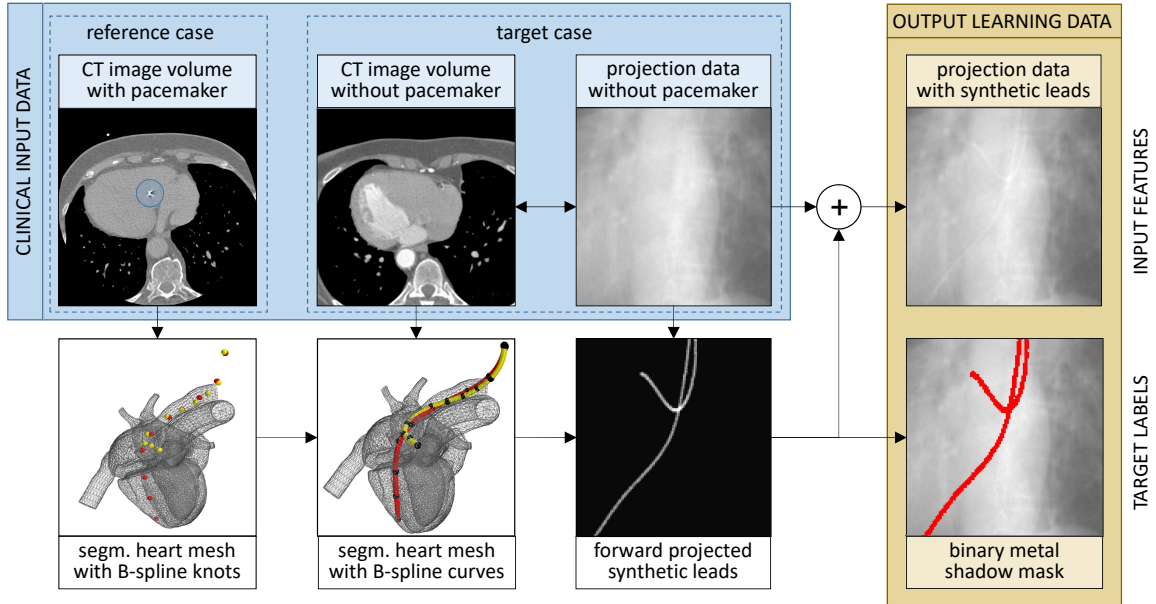
Figure 3: Synthetic leads are introduced into the projection data of clinical cases without pacemakers. Consideration of the segmented heart meshes allows sensible insertion.

3.1. Learning data generation

The data generation process is visualized in Figure 3 and requires one reference case with pacemaker and one target case without pacemaker. For both CT image volumes, corresponding heart meshes are obtained by model-based heart segmentation according to [1]. At least ten B-spline knots are manually selected along each pacemaker lead in the reference case. Based on the point-to-point correspondences in the segmented heart meshes, thin plate spline smoothing is performed to transform the B-spline knots from the reference case into the target case. The corresponding B-spline curve is determined by cubic B-spline interpolation for each pacemaker lead, separately. Dilation of the resulting lines with a chosen lead diameter of 2 millimeters and an attenuation value of 4500 HU yields a binary lead mask in the target image geometry. Subsequent forward projection delivers the corresponding metal shadow in the originally acquired projection geometry of the target case. The required input data for supervised learning is finally obtained by summation of the original projection data and the forward projected lead mask. Thresholding with zero defines the corresponding target segmentation mask.

3.2. Supervised learning

The data generation process is applied two times per reference case as twice as many target cases are available. An average number of 10 000 views per case is included in the supervised learning process. Each view contains 128×672 line integrals determined by the detector height and width of the Brilliance iCT. On average, 1.4% of the projection data contains object voxels. The data is case-wise separated into the subsets training, validation and testing with a ratio of 6:4:4, or rather 3:2:2 with respect to the corresponding reference cases.
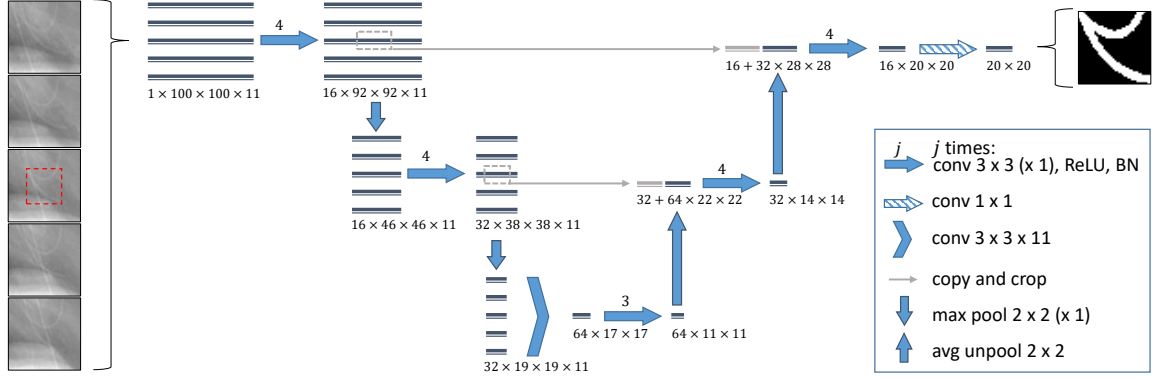
Figure 4: Adapted U-net architecture for 2D lead segmentation from multi-view input patches. Each box corresponds to a multi-channel feature map. For clarity only five instead of eleven input views are illustrated. The actual shape of each feature map is denoted at the lower left edge of the box. The arrows represent the different operations.

By this data separation strategy, it is ensured that pacemaker geometries and background line integrals are disjoint among the subsets.

Patch sampling During training, the CNN takes 3D patches of size $100 \times 100 \times 11$ voxels as input and delivers 2D patches of size 20×20 as output. The first and second dimension of the network input contain the information of the detector row and column, while the third dimension corresponds to the projection view. The third patch dimension is equidistantly sampled with respect to the number of views per gantry turn in such a way that 12 degrees gantry rotation are captured. Inclusion of neighboring views has the benefit, that the network gets additional information about the rotation velocity of supposed pacemaker leads, i.e. information about the distance to the rotation center. The target patch corresponds to the center of the sixth view. In order to compensate the extreme foreground-background class imbalance, the sampling process is controlled in such a way that 75% of the training patches contain at least one object voxel. The remaining 25% are randomly sampled.

Network architecture Figure 4 shows the utilized adapted U-Net architecture [10]. In the contracting path, 2D lead features are extracted for each view, separately. These features are joint in the bottleneck to exploit the temporal information. In the expanding path, merely location information of the center slice to be segmented are copied and cropped from the contracting path. The slim architecture with its shared weights in the contracting path has a relatively low number of 423 730 learned parameters. The networks output is reduced in the first two dimensions, as no internal padding is performed. The fully convolutional network design allows for arbitrary output shapes. During validation and testing, the full detector size of 128×672 voxels is segmented in a single step by previously enlarging the input projection data sets with respect to the networks receptive field of $11 \times 81 \times 81$ pixels using symmetric padding. During parameter tuning we observed, that in particular receptive field size and number of projection views per patch are crucial to the network performance.

Learning setup The stochastic gradient descent solver Adam [4] with an initial learning rate of 0.01, a mini-batch size of 32 and a momentum of 0.8 is used for network optimization.
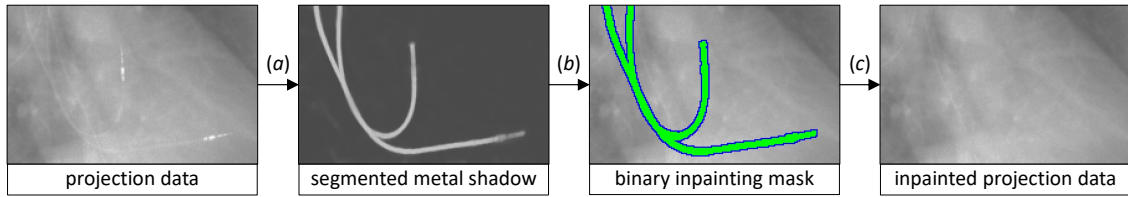
Figure 5: Schematic steps for sinogram correction. (a) The pacemaker metal shadow is segmented by CNNs. (b) The resulting probability map is transformed into a binary mask. (c) Metal-affected values are inpainted to yield pacemaker-free projection data.

Training is performed over 30 epochs while one epoch is defined by 10e5 processed samples. The learning rate decreases with a factor of two after every 10th epoch and L2 regularization with a weight of 0.0002 is used. The learning process is driven by the focal loss [7] with focusing parameter $\gamma = 2$.

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

1. Four validation cases and four test cases are randomly sampled.
2. Network training is performed based on the remaining six clinical cases.
3. After every epoch, the networks generalization capabilities are examined by calculating the Dice coefficient with a threshold of 0.5 on the validation set.
4. The model with the highest validation metric within 30 epochs of training is selected for performance evaluation and application in the DyPAR pipeline.
5. Steps 1.-4. are performed five times in total.

3.3. DyPAR pipeline

The proposed dynamic pacemaker artifact removal (DyPAR) approach deals with sinogram correction and subsequent filtered backprojection [5]. As illustrated in Figure 5, the sinogram correction comprises three steps:

(a) **Metal shadow segmentation** In order to increase the robustness of the metal shadow segmentation, the entire ensemble of five CNNs yielded by the bagging approach is applied on the previously padded input raw projection data. It has to be noted that merely projection data which is required for cardiac field-of-view reconstruction is processed. That means, the pulse generator below the clavicle, for instance, is not segmented. The output probability map is averaged across the ensemble and contains values between 0 and 1. In contrast to standard MAR approaches, the segmentation delivers probability maps which are independent of the 3D motion blur and exactly located on the real metal shadows.

(b) **Post-processing** By thresholding the probability map with 0.15 and extracting the largest connected component using a $3 \times 3 \times 3$ structure element, a binary inpainting mask is determined. The choice of the relatively low threshold 0.15 is motivated by the fact that incomplete metal masks may introduce new artifacts after inpainting and reconstruction. For the task of MAR, false positives are less severe than false negatives, i.e. the sensitivity is a more important performance measure than the specificity.

Table 1: Mean network performance measures (in percent) on training, validation and testing subsets are compared for different post-processing settings, i.e. thresholds t for separation of object and background voxels and optional largest connected component (LCC) extraction.

subset	Dice coefficient			sensitivity			specificity			AUC
t / LCC	0.5 ✗	0.15 ✗	0.15 ✓	0.5 ✗	0.15 ✗	0.15 ✓	0.15 ✗	0.5 ✗	0.15 ✓	✗
testing	91.27	79.44	84.92	89.63	98.05	97.96	99.90	99.28	99.50	99.86
validation	90.93	78.40	84.23	90.04	98.25	97.88	99.90	99.31	99.54	99.86
training	94.55	80.22	86.32	95.48	99.89	99.87	99.92	99.36	99.98	99.98

(c) **Inpainting** The predicted metal-affected values (green voxels in Figure 5) are treated as missing data. The projection data is filled by distance weighted 2D linear interpolation, based on surrounding line integrals (blue voxels in Figure 5). Of course, alternative inpainting approaches instead of view-wise linear interpolation are possible. The focus here is however on the metal shadow segmentation step. Filtered backprojection [5] of the inpainted projection data finally delivers the CT image volume without pacemaker and concomitant artifacts.

4. Experiments and results

For all experiments, the Microsoft Cognitive Toolkit (CNTK v2.5+, Microsoft Research, Redmond, WA, USA) is used as deep learning framework. Table 1 summarizes networks performance measures during the five-fold bagging approach. These quantitative results illustrate that the CNNs are able to identify synthetic pacemaker leads in clinical projection data. To further evaluate generalization capabilities of the bagging ensemble in clinical practice, DyPAR is applied on the *clinical test data* described in Section 2.2.

4.1. Evaluation on clinical test data

In order to identify strengths and weaknesses of the trained neural networks, qualitative evaluations of the segmentation masks and the output image volumes are performed by visual inspection. For comparison, image-based metal shadow extraction is considered, whereby the metal masks in the image domain are segmented using 3D hysteresis thresholds of 1000 HU and 1500 HU. The metal shadow areas are yielded by forward projection and thresholding with zero.

Both metal shadow extraction approaches, the image-based one and ours, are compared in Figure 6. For both methods, few false negatives are observable (see Figure 6c,h,j,k). In contrast to the image-based approach, our networks deliver segmentation results which are not affected by blurring between different motion states (see Figure 6a,g,h,j). Despite the lag of dedicated learning data, the networks object-background separation also holds for electrodes (see Figure 6a-k) and defibrillators (see Figure 6b,f). Misinterpretation of spine as lead is identified as one possible error source (see Figure 6f), whereas metal shadow kinks (see Figure 6d) and moderate noise do not seem to confuse the neural networks. More severe noise caused by dose modulation is indeed a limiting factor, but projection views recorded with reduced radiation doses are generally not taken into account during back projection. In cases with dose modulation, the post-processing step has to be adapted in such a way that the largest connected component is determined for each high dose segment, separately.
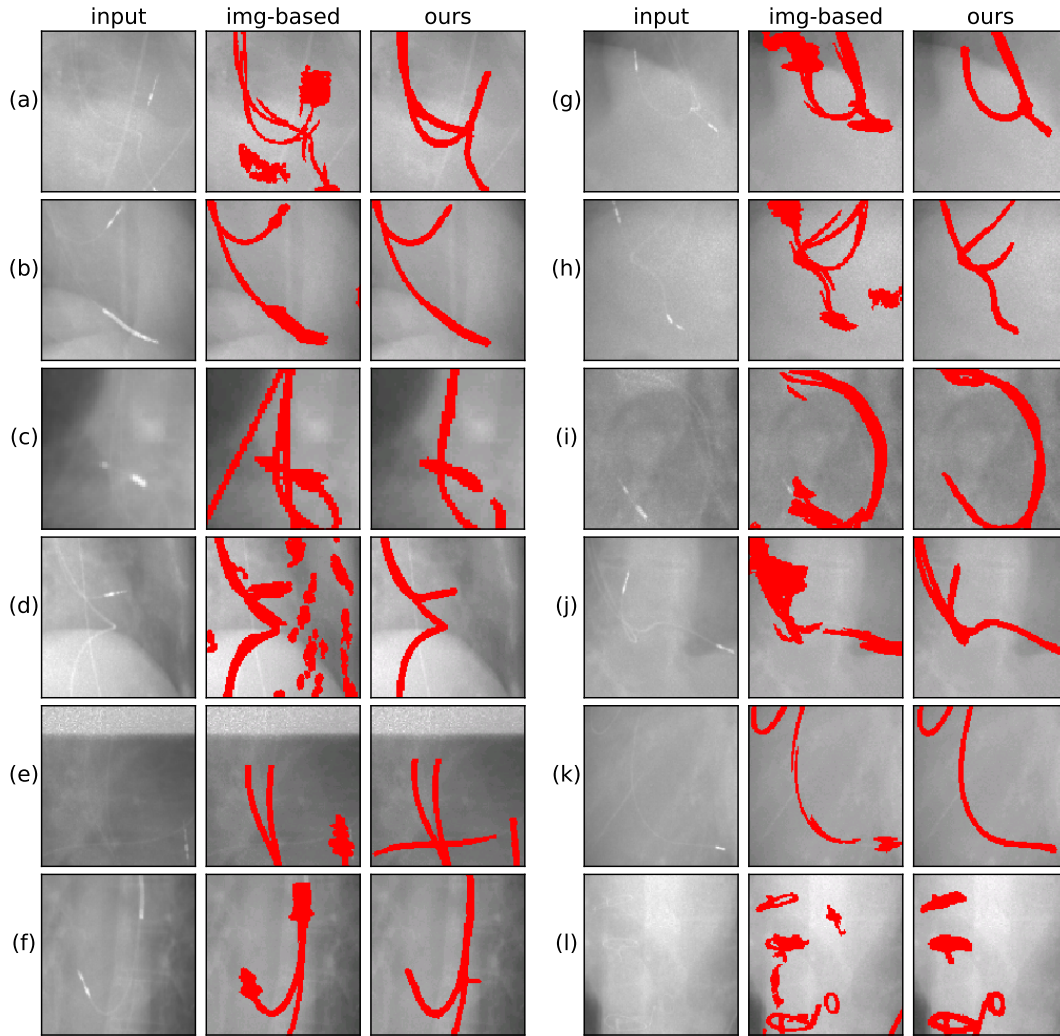
Figure 6: Example projections with real metal implants and corresponding metal shadow areas predicted by the image-based approach and during our DyPAR pipeline.

Based on a single projection view, ECG and pacemaker leads are visually hard to distinguish. The CNNs are remarkably successful in their differentiation and seem also to consider the rotation velocities (see Figure 6a,c,d). This strength is restricted in case of horizontally aligned ECG leads (see Figure 6e) as for these segments, the rotation velocity can not be measured locally. False positives are increasingly present in the image-based metal shadows, e.g. caused by severe calcifications, ECG-leads and bone (see Figure 6a-c,d,h). In the two *no-pacemaker data sets*, false positive rates of 0.1125% and 3.3297% are achieved by the networks. In both cases, false positive activations are caused by parts of the ECG leads. In the second case, the sternal steel wires are additionally misinterpreted by the networks. In contrast to the image-based approach which exhibits false positive rates of 0.8647% and 3.6866%, calcifications and stents did not lead to false positive network activations.
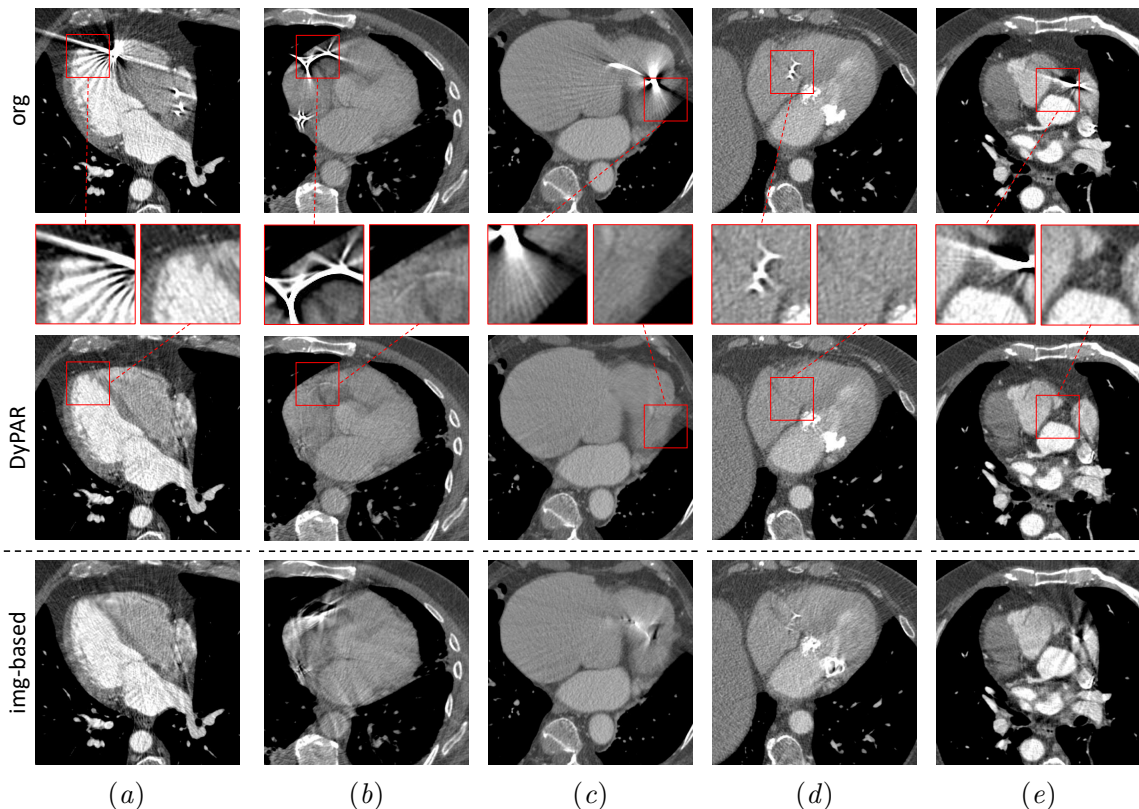
Figure 7: Axial slices of clinical test cases before and after DyPAR. For comparison, the corresponding results of the image-based MAR approach are considered.

The superiority of our method over the image-based is observable in the projection and the image domain. Figure 7 shows some results of the DyPAR with significant reduction of the streak-shaped metal artifacts. Especially in the zoomed regions, evaluation of the neighboring anatomy is facilitated. In comparison, the image-based approach is robust regarding cardiac motion and leads to increased blurring in the neighboring anatomy (see Figure 7a), incomplete metal removal (see Figure 7c,d) and introduction of new severe artifacts (see Figure 7b,e). In fact, our method also introduces new artifacts whenever the inpainting causes inconsistencies in the projection data (see Figure 7a, right). Nevertheless, the DyPAR approach demonstrates great generalization capabilities as it is transferable to different scanner types (also tested: Brilliance 64, Philips Healthcare, Cleveland, Oh , USA) and ungated CT data sets (see Figure 7b-d). In these cases, contrast enhancement and acquisition settings like the pitch and the gantry rotation speed vary from the learning data.

5. Discussion and Conclusion

Establishment or maintenance of consistent projection data is essential for artifact removal or avoidance. As already mentioned in Subsection 3.3, simple 2D linear interpolation should not be the method of choice as edge preservation is not ensured. Since neural networks also showed great results in image inpainting [12], a deep-learning-based approach should be

considered. Furthermore, metal reinsertion with respect to the cardiac motion is still an issue. So far, the metal is completely removed from the reconstructed CT image volumes.

The proposed data generation process is in principle extendable to different scanner types, acquisition modes and high-density objects. It enables on-site training for protocol-specific MAR. Therefore, Generalization to step-and-shoot data and transferability to other metal implants like artificial valves or metal clips will be part of future research.

To conclude, dynamic metal artifact removal based on metal shadow segmentation in the projection domain is feasible. CNNs trained on clinical data with synthetically introduced pacemaker leads show great generalization capabilities in the segmentation of electrode and defibrillator shadows. Quantitative validation studies are required to assess the transferability of these promising initial results to pacemaker CT artifact reduction in clinical practice.

References

[1] Olivier Ecabert, Jochen Peters, Hauke Schramm, Cristian Lorenz, Jens von Berg, Matthew J Walker, Mani Vembar, Mark E Olszewski, Krishna Subramanyan, Guy Lavi, et al. Automatic model-based segmentation of the heart in CT images. *IEEE Transactions on Medical Imaging*, 27(9):1189–1201, 2008.

[2] Lars Gjesteby, Qingsong Yang, Yan Xi, Ye Zhou, Junping Zhang, and Ge Wang. Deep learning methods to guide CT image reconstruction and reduce metal artifacts. In *Physics of Medical Imaging.* International Society for Optics and Photonics, 2017.

[3] Willi A Kalender, Robert Hebel, and Johannes Ebersberger. Reduction of CT artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987.

[4] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[5] Peter Koken and Michael Grass. Aperture weighted cardiac reconstruction for cone-beam CT. *Physics in Medicine and Biology*, 51(14):3433, 2006.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105, 2012.

[7] Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *The IEEE International Conference on Computer Vision (ICCV)*, 2017.

[8] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelrieß. Normalized metal artifact reduction (NMAR) in computed tomography. *Medical Physics*, 37(10):5482–5493, 2010.

[9] Andre Mouton, Najla Megherbi, Katrien Van Slambrouck, Johan Nuyts, and Toby P Breckon. An experimental survey of metal artefact reduction in computed tomography. *Journal of X-ray Science and Technology*, 21(2):193–226, 2013.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

[11] Shiyu Xu and Hao Dang. Deep residual learning enabled metal artifact reduction in CT. In *Medical Imaging 2018: Physics of Medical Imaging*, volume 10573, page 105733O. International Society for Optics and Photonics, 2018.

[12] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 3, 2017.

[13] Yanbo Zhang and Hengyong Yu. Convolutional neural network based metal artifact reduction in x-ray computed tomography. volume 37, pages 1370–1381. IEEE, 2018.