Cross-Domain Metal Segmentation for CBCT Metal Artifact Reduction

Maximilian Rohleder[a,b], Tristan M. Gottschalk[a,b], Andreas Maier[a], and Bjoern W. Kreher[b]

[a]Friedrich-Alexander-University, Martenstrasse 3, Erlangen, Germany
[b]Siemens Healthineers, Siemensstrasse 3, Forchheim, Germany

ABSTRACT

Metallic objects in the volume of a CBCT system can cause various artifacts after image reconstruction such as bright and dark streaks, local distortions of CT values and shadowing. In the intraoperative setting, this drastically reduces clinical value and can harden decision making. Most existing approaches to reduce such artifacts rely on a threshold-based metal segmentation in the reconstruction domain, which is prone to failure; especially in cases with extreme artifacts. Faulty metal masks impair these inpainting-based MAR methods and at times even worsen image quality by introducing secondary artifacts. In this work, a novel neural network topology is proposed to segment metallic objects in CBCT reconstruction domain by leveraging information of the given raw projection images. A reconstruction operator is embedded into this architecture, which enables the model to yield projection and reconstruction domain information during end-to-end training. This cross-domain approach is compared to the self-configuring segmentation method "nnUNet", which predicts the three dimensional metal masks directly from the artifact corrupted reconstruction. To provide a baseline, a segmentation using a global dice-optimal threshold is determined. Segmentation results on simulated data confirmed by 5-fold cross validation show that the cross-domain network yields a mean dice coefficient of 0.87 ± 0.05 at a prediction time of 4s per volume. The reference method achieves 0.86 ± 0.03 in 43s, whereas the optimal similarity using a threshold averages to 0.45 ± 0.22.

Keywords: Known Operator Integration, Metal Artifacts, Cone-Beam CT, Segmentation

1. INTRODUCTION

1.1 Motivation

Mobile C-Arm devices are an integral part of modern surgical procedures. In addition to 2D X-Ray projection images used for guidance, modern systems allow to accurately verify the placement of tools and implants via 3D imaging. One major limitation of this modality are metal artifacts. Originating from simplifications in the reconstruction model and inconsistencies in the measured data, these image artifacts appear as bright and dark streaks, local distortions of CT values, and shadowing. As they emerge especially around metal objects, they obstruct the implant placement verification and thus drastically decrease the diagnostic value during the surgical intervention.

1.2 Existing MAR Approaches

Most modern Metal Artifact Reduction (MAR) methods reduce metal artifacts by inpainting the metal traces in projection domain. These metal traces are obtained by forward projection of the volumetric segmentation of metal objects. Inpainting in this context refers to the process of substituting pixel values to remove their contribution to the reconstruction image. Over the last decades, multiple approaches for inpainting have been proposed starting with simple linear or polynomial interpolation,[1,2] frequency domain interpolation,[3] using wavelets[4] or with the help of machine learning.[5] Regardless of how elaborated the inpainting approach is designed, a faulty estimation of the metal mask can sabotage the effectiveness of said approaches, reduce the level of detail around the corrected metal, remove relevant anatomy, and even introduce new artifacts.[6]

Further author information: (Send correspondence to Maximilian Rohleder)
Maximilian Rohleder: E-mail: Maxi.Rohleder@fau.de, Telephone: +49 (0)9131 85 25246

7th International Conference on Image Formation in X-Ray Computed Tomography,
edited by Joseph Webster Stayman, Proc. of SPIE Vol. 12304, 123040K
 · 0277-786X · doi: 10.1117/12.2646382

1.3 Related work

To enhance the mask quality, a shape-model based estimation of the metallic objects has been proposed.[7] A known object's outline is registered to a coarse volumetric metal segmentation to refine its shape. However, this approach is not generally applicable, as prior knowledge about the shape of the metal is usually not available. The segmentation of metal in image domain is prone to error mainly because of the metal artifacts. The alterations of CT values around the depicted metal prohibit a purely value-based approach such as the traditional global thresholding. By including structural information, Convolutional Neural Networks (CNN) have proven to be successful in many medical segmentation scenarios. Recent advances in known operator integration have facilitated the end-to-end training of cross-domain architectures. A derivable backprojection operator can be embedded into the model and used with gradient backpropagation for supervised learning.[8] The observation that metal artifacts originate during the domain transformation suggests that neural network architectures can benefit from projection domain information. Multi-domain approaches have successfully been applied in Metal Artifact Avoidance (MAA) to estimate metallic objects from very few given projections.[9] However, the objective is fundamentally different compared to MAR. For MAA, a rough distribution of metal is sufficient to adapt the trajectory, whereas high detail masks are desirable for MAR. Furthermore, the system matrix based reconstruction from[9] cannot be applied to MAR segmentation due to the larger number of input projections. A dual domain network topology has also been demonstrated for direct, learned MAR.[10] The authors report significant improvements over other single domain MAR approaches.

1.4 Planned Contributions

In this work, a novel cross-domain segmentation network which yields both raw rotational and reconstruction domain information is presented. It is compared to a CNN approach applied to artifact-corrupted reconstruction images and a simple threshold-based method.

2. METHODS

2.1 Data

The cross-domain architecture proposed in this work requires training data, where the input consists of X-Ray projection images and the target metal mask is a volume-shaped binary array. It is crucial that the simulated projection images exhibit all physical effects which are relevant for the formation of metal artifacts. To model said effects, a simulation framework was derived from the DeepDRR method described in.[11] However, the provided pre-trained weights for material segmentation and scatter estimation could not be used as they did not generalize well on the raw data used in this project.

A total of 11 cone-beam CT volumes were acquired from 5 different specimens from a human spine cadaver study of the lumbar and thoracic region using a Siemens Cios Spin System.* To realistically model the shapes of metal objects, a library of 3D models of surgical screws, plates, k-wires, and towers are available from Nuvasive, San Diego, California. A set of metal objects is realistically positioned relative to the skeletal structures in the base volume using the 3D modelling suite Blender.† The position and orientation of each object is stored and considered during the simulation.

2.1.1 Simulation Process

First, each base volume is segmented into three materials $M \in \{air, tissue, bone\}$ using an empirically defined threshold. Then, projection images of the three material volumes and all metal objects are generated. The resulting material path-length projections are weighted with their spectrum-dependant attenuation coefficients according to the polychromatic Beer-Lambert law. To approximate the influence of scatter on the artifact formation, a constant background signal is added to the images. This simulation pipeline produces stacks of 200

*The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects, or cadavers of both kind. All data acquisitions were done in consultation with the Institutional Review Board of the University Hospital of Erlangen, Germany.

†Open Source 3D Creation Suite, https://www.blender.org/

X-Ray projection images of shape (488×488) over the angular range of 200 degrees which, after reconstruction, exhibit the desired metal artifacts. To generate the training labels, a cube-shaped binary array with side-length 256 is created from the metal objects depicted in one sample. With an isotropic voxel-size of 0.626, the binary volume resembles the standard volume size of the Siemens Cios Spin C-Arm System. Because of GPU memory limitations, the training data for this project has half the resolution and number of projection images compared to measured raw data.

2.1.2 Data Augmentation

In order to increase the dataset size, four random rotations around the z-axis are applied to each sample as an offline data augmentation step. This rotation is virtually applied prior to projection by simply appending to the projection matrices. Using this method, a total of 44 samples are generated from the 11 independent tool configurations.

2.2 Optimal Threshold-Based Method

As current product-grade MAR methods utilize a threshold-based segmentation, it serves as a baseline here. To obtain a best-case estimate of such a global thresholding method, the optimal threshold is calculated for each sample. As an optimality criterion, the dice similarity coefficient (DSC) is used. Equation 1 shows the dice similarity defined on the binarized volume X and the binary label Y.

$$DSC(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

A threshold is defined as optimal when the binarization it generates maximises the DSC metric. For each simulated CT volume, an optimal threshold is determined by testing 1000 values over the range of its histogram values.

2.3 Image Domain CNN Method

To obtain a standardized segmentation benchmark, the self-configuring framework *nnUNet* is used.[12] nnUNet automatically configures all relevant parameters of the popular architecture *U-Net* and adapts it to the so-called dataset fingerprint. The user can choose between a 2D, a 3D high-resolution, and a 3D low-resolution U-Net layout. For this project, the 3D low-resolution (*3d-lowres*) U-Net model was chosen. The training of the other models was omitted, as the cross-domain method produces lower resolution masks and the results are compared on this lower resolution anyway. This model contains 6 mio. trainable parameters and is applied patch-wise on cube-shaped sub-volumes of side-length 128. The model is trained on CT volumes reconstructed from the simulated projection data. To predict volumes of side-length 256, the model is evaluated 27 times as the patches are strided by half a patch-size.

The performance of the *3d-lowres* architecture was evaluated using 5-fold cross validation. The simulated X-Ray images were reconstructed using filtered backprojection to serve as training input to this method. The data splitting, pre-processing and evaluation using the DSC metric was handled by the framework.

2.4 Cross-Domain Architecture

Inspired by the success of the 3D U-Net architecture, an encoder-decoder layout with skip-connections is used.[13] To enable end-to-end training with input and labels from different domains, a derivable backprojector is integrated into the network. This operator is implemented as the *PyroNN* filtered backprojection layer with a non-trainable Ram-Lak filter.[8]

As seen in figure 1, this domain transform happens during the skip-connection step, such that the encoder is applied to projection domain and the decoder to reconstruction domain. The shapes of the tensors are indicated next to each stage of the architecture, whilst the number of computed feature-maps is shown above each block symbolizing a layer's output. At each stage of the network, the stack of projections is downsampled both in the number of projections and their resolution.

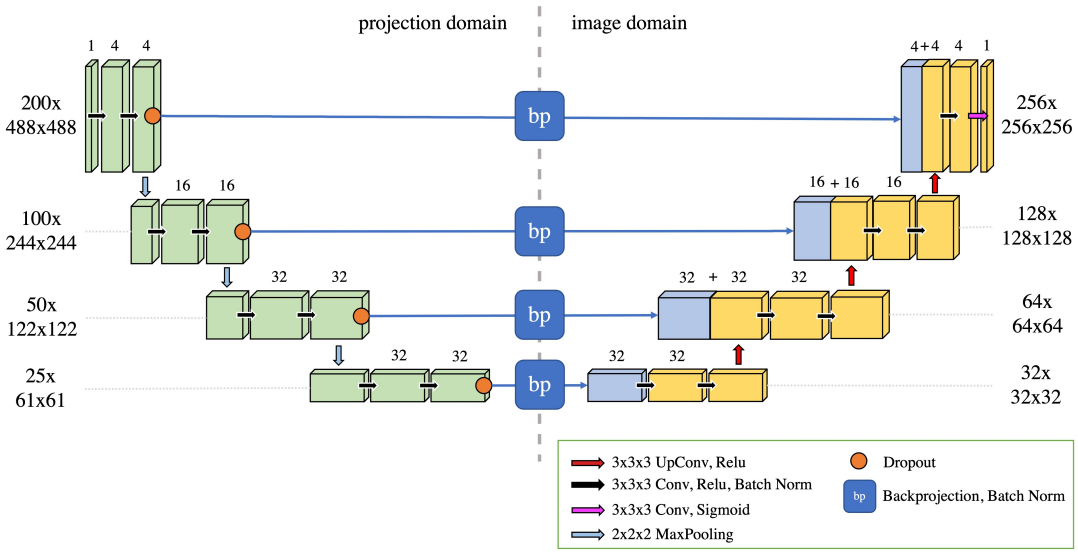

Figure 1. **Cross-domain network architecture** with an encoder in projection domain and a decoder in image/reconstruction domain. The backprojection operators embedded on the skip-connections reconstruct volumes from the input projection domain feature maps and stack them along the channel axis (blue).

To facilitate the reconstruction of differently sized volumes, the projection matrices and the resolution configurations are adapted accordingly. The convolutional layers with kernels of side length three use zero-padding to generate similarly sized feature maps. At the skip-connections, these feature maps are reconstructed individually and the resulting volume-shaped activations are stacked along the channel axis.

2.5 Cross-Domain Training Procedure

2.5.1 Preprocessing

The simulated intensity images are first converted to line integrals. A cross-validation split is defined with nine samples in the training set and the remaining two samples in the test set of each of the five splits. After splitting, the previously described offline data augmentation strategy is applied to boost the set sizes to 36/8 (train/test).

During training, different noise levels are augmented. An additive, intensity dependent noise is drawn from a Poisson distribution. Furthermore, a convolutional noise model is used to imitate the detector characteristics and low-dose effects.[14]

2.5.2 Training

The training samples are fed to the network in a batch-size of one due to GPU memory constraints. To compensate the resulting stochastic gradients, the Adam optimizer is applied with the standard parametrizations of the first and second order moments and an initial learning rate of 10^{-3}. Furthermore, the learning rate is reduced by a factor of 10 if the training loss plateaus for longer than 10 epochs. The training is terminated after the learning rate is reduced for the third time. As a loss function, the dice similarity defined in equation 1 is used.

3. RESULTS

3.1 Quantitative

The cross-validation results and the evaluation of the optimal threshold-based method are shown in Table 1. Our method achieves a dice similarity of 0.87 ± 0.05 averaged over the different data splits. The reference method applied in reconstruction domain evaluates to a DSC of 0.86 ± 0.03 across all cross-validation runs. Note, that the data splits are not identical between our method and the reference image domain CNN. The best possible segmentation using a per-sample threshold was evaluated for the entire simulated dataset. Overall, this method achieves a DSC metric of 0.45 ± 0.21.

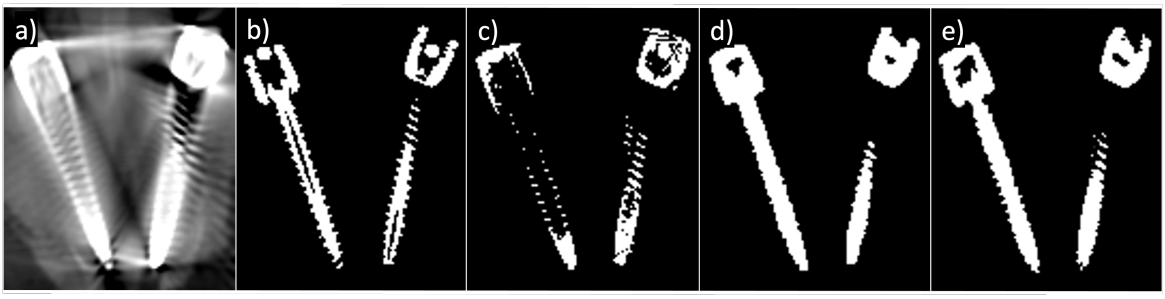

Figure 2. **Cropped slice from a volumetric metal segmentation.** *a*) shows the corresponding artifact-corrupted reconstruction, *b*) is the ground truth, *c*) depicts the best-possible threshold-based segmentation, *d*) shows the result of the reference image domain CNN, and *e*) shows our methods prediction.

Apart from the segmentation performance, other attributes of the compared methods are displayed in Table 2. The inference time of the reference image domain CNN method is about eleven times longer than our method. The auto-configuring framework *nnUNet* trains the models for about 97h which is more than six times longer than our methods, which are trained 15h on average. The cross-domain network requires about four times more GPU memory during training and inference.

3.2 Qualitative

To intuitively compare the methods performance, an unseen sample is segmented using the three compared methods (Figure 2). From the resulting volumetric metal masks, a slice is selected and cropped to a region of interest showing two screws with tulips fixated to connecting rods. For reference, the corresponding reconstruction is added. As the screws' main axes align with the acquisition trajectory, heavy artefacts are visible. The thresholding method *c*) greatly underestimates the ground truth mask *b*). The image domain CNN and the cross-domain segmentation network produce similar looking masks, which, compared to the label, yield a dice similarity of 0.88 and 0.90 respectively.

4. CONCLUSION

Summarizing the quantitative and qualitative results, it becomes evident that both learned methods largely improve the segmentation compared to the purely value-based thresholding baseline. Visually, there is no distinct difference between the cross-domain and single-domain network's predicted masks. However, the newly presented cross-domain method achieves this similar performance using 10% of trainable parameters and 15% of training time.

Table 1. Results of Model Evaluation (Dice Similarity)

	Ours	CNN	Threshold
split 0	0,8131	0,8089	-
split 1	0,8145	**0,8875**	-
split 2	0,9138	0,8574	-
split 3	0,8921	0,8699	-
split 4	**0,9140**	0,8613	-
mean±std	0,87±0,05	0,86±0,03	0,45±0,21

Table 2. Additional Method Attributes

	Ours	CNN	Threshold
Inference Time	4s	45s	<1s
Inference GPU Memory	9.6Gb	2.1Gb	-
Training Duration	15h	97h	-
Training GPU Memory	23.2Gb	6Gb	-
#Params	200k	6mio.	1

With the application in the surgical suite in mind, the cross-domain network offers the clear advantage of a faster inference time. This is largely attributable to the patch-wise application of the image domain CNN. On the downside, the novel network architecture has increased GPU memory requirements which might not be readily available on systems in the operating room due to financial cost.

Future work should focus on evaluating the effectiveness of the proposed approaches on measured data. Should this be successful, the impact of the improved metal masks on inpainting-based MAR methods needs to be investigated. Furthermore, the memory footprint of the cross-domain method can be reduced by revising the implementation of the backprojection operator.

REFERENCES

[1] Kalender, W., Hebel, R., and Ebersberger, J., "Reduction of CT artifacts caused by metallic implants," *Radiology* **164**(2), 576–577 (1987).

[2] Mahnken, A. H., Raupach, R., Wildberger, J. E., Jung, B., Heussen, N., Flohr, T. G., Günther, R. W., and Schaller, S., "A New Algorithm for Metal Artifact Reduction in Computed Tomography," *Investigative Radiology* **38**, 769–775 (12 2003).

[3] Kratz, B. and Buzug, T. M., "Metal artifact reduction in computed tomography using nonequispaced fourier transform," *IEEE Nuclear Science Symposium Conference Record* , 2720–2723 (2009).

[4] Mehranian, A., Ay, M. R., Rahmim, A., and Zaidi, H., "X-ray CT metal artifact reduction using wavelet domain L0 sparse regularization," *IEEE Transactions on Medical Imaging* **32**(9), 1707–1722 (2013).

[5] Gottschalk, T. M., Kreher, B. W., Kunze, H., and Maier, A., "Deep Learning Based Metal Inpainting in the Projection Domain: Initial Results," *Lecture Notes in Computer Science* **11905 LNCS**, 125–136 (2019).

[6] Stille, M., Kratz, B., Müller, J., Maass, N., Schasiepen, I., Elter, M., Weyers, I., and Buzug, T. M., "Influence of metal segmentation on the quality of metal artifact reduction methods," in [*Medical Imaging 2013: Physics of Medical Imaging*], **8668**, 86683C, SPIE (3 2013).

[7] Uneri, A., Zhang, X., Yi, T., Stayman, J. W., Helm, P. A., Osgood, G. M., Theodore, N., and Siewerdsen, J. H., "Known-component metal artifact reduction (KC-MAR) for cone-beam CT," *Physics in Medicine and Biology* (2019).

[8] Syben, C., Michen, M., Stimpel, B., Seitz, S., Ploner, S., and Maier, A. K., "Technical Note: PYRONN: Python reconstruction operators in neural networks," *Medical Physics* **46**, 5110–5115 (11 2019).

[9] Wu, P., Sheth, N., Sisniega, A., Uneri, A., Han, R., Vijayan, R., Vagdargi, P., Kreher, B., Kunze, H., Kleinszig, G., Vogt, S., Lo, S.-F., Theodore, N., and Siewerdsen, J. H., "Method for metal artifact avoidance in C-Arm cone-beam CT," in [*Medical Imaging 2020: Physics of Medical Imaging*], Bosmans, H. and Chen, G.-H., eds., **11312**, 78, SPIE (3 2020).

[10] Lin, W. A., Liao, H., Peng, C., Sun, X., Zhang, J., Luo, J., Chellappa, R., and Zhou, S. K., "DuDoNet: Dual domain network for CT metal artifact reduction," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2019-June**, 10504–10513 (6 2019).

[11] Unberath, M., Zaech, J. N., Lee, S. C., Bier, B., Fotouhi, J., Armand, M., and Navab, N., "DeepDRR - A Catalyst for Machine Learning in Fluoroscopy-Guided Procedures," *Lecture Notes in Computer Science* **11073 LNCS**, 98–106 (2018).

[12] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods 2020 18:2* **18**, 203–211 (12 2020).

[13] Çiçek, , Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., "3D U-net: Learning dense volumetric segmentation from sparse annotation," in [*Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*], **9901 LNCS** (2016).

[14] Wang, A. S., Stayman, J. W., Otake, Y., Vogt, S., Kleinszig, G., Khanna, A. J., Gallia, G. L., and Siewerdsen, J. H., "Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT," *Medical Physics* **41**(7) (2014).