

Data-driven Metal Artifact Correction in Computed Tomography using conditional Generative Adversarial Networks

Nele Blum¹, Thorsten M. Buzug^{1,2} and Maik Stille^{1,2},
Institute of Medical Engineering, University of Lübeck, 23556 Germany¹,
Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering IMTE, 23556
Germany²

Abstract

Metal objects in the field of view cause artifacts in the image, which manifest as dark and bright streaks and degrade the diagnostic value of the image. Standard approaches for metal artifact reduction are often unable to correct these artifacts sufficiently or introduce new artifacts. We propose a new data-based method to reduce metal artifacts in CT images applying conditional Generative Adversarial Networks to the corrupted data. A generator network is applied directly to the corrupted projections by the metal objects to learn the corrected sinogram data. Further, two discriminator networks are used to evaluate the image quality of the enhanced data from the generator. The method was initially developed based on a supervised approach. However, there is usually no ground truth for actual clinical data without artifacts, which is needed to train the networks. Therefore, the method was further improved to train an unsupervised network, i.e., without the ground truth. In addition the input data, the neighboring slices and the stochastic components of the image are included using the latent space representation of the data. The results show that the trained generator network can reasonably replace the missing projection data and reduce the artifacts in the reconstructed image.

I. INTRODUCTION

ARTIFACTS are particularly apparent in computed tomography (CT) when high-density objects like metal implants or surgical instruments are present in the field of view. Various metal artifact reduction (MAR) methods have been proposed since the first publications on MAR [1]. Projection completion methods are often used due to their simplicity and fast application. They treat the data affected by metal objects as missing image information and replace them with synthetic data, usually obtained by interpolation [2]. A significant drawback of this method, especially in inhomogeneous image regions, is the loss of information in the metal trace. An alternative to this approach are iterative methods [3], yet they have the disadvantage of high computation times, especially for large 3D images. Recently, neuronal networks have been used to correct corrupted data [4], [5]. To train the networks effectively, the definition of a suitable loss function is crucial. Still, loss functions are often designed to optimize specific, quantifiable image parameters, even if a selection of image properties to describe good image quality is usually hard to define. However, instead of specifying parameters for a good image quality, it is possible to use an additional network. Besides the so-called generator network, which generates the improved projection data, a second network can be trained, the discriminator, to distinguish the actual projections from the synthetic data.

The first results using this method have already been published and showed the potential of the new method on the reduction of artifacts in the image compared to conventional approaches [6]. One major drawback of the published method, however, is the use of ground truth data to train the networks.

Prajot et al. propose a model for image inpainting, which allows learning the distribution of reconstructed images in a completely unsupervised setting by integrating the stochastic component that introduces an explicit dependency between this component and the generated output in the learning process [7]. We further developed this approach and incorporated it into our previous work. Different from Prajot's approach, it can be assumed that additional information is available in the neighborhood of the corrupted metal projections. This information can be used in the training process to get more realistic predictions from the generator network.

N.Blum is with the Institute of Medical Engineering, University of Lübeck, 23556 Germany, e-mail: (see blum@imt.uni-luebeck.de).

T.M.Buzug and M.Stille are with Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering IMTE, Lübeck, 23556 Germany and the Institute of Medical Engineering, University of Lübeck, 23556 Germany

II. METHODS

To train the networks, a combination of different loss functions is used. In the objective function

$$G^* = \arg \min_{G,E} \max_{D1,D2} L^{\text{GAN}}(G, D1) + L^{\text{GAN}}(G, D2) + \lambda_z L^z(G, E) + \lambda_y L^y(G, D1, E) \quad (1)$$

several networks interact with each other. For the training process, the loss function is minimized by the generator G

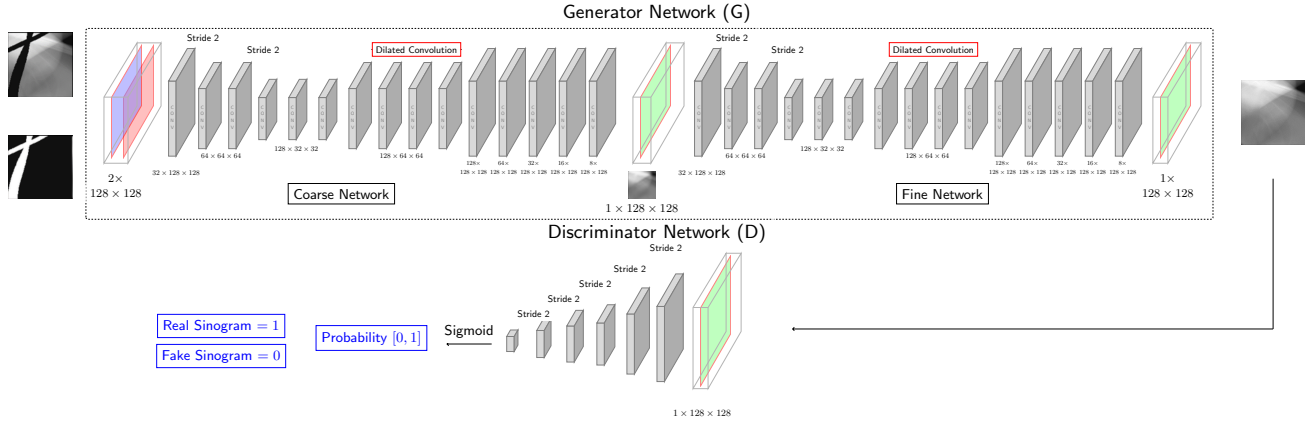


Fig. 1: Schematic representation of the used network architectures.

and the encoder E and maximized by the discriminators $D1$ and $D2$. No access to the metal-free projection data is usually available for the networks training. This is especially the case if the networks are to be trained on actual clinical data. Hence, in the first part of the loss function, the output data of the generator network are not used directly. Instead, a binary mask of the metal trace is first applied to the generated data using a function F , allowing the output data to be compared with the original input data. This is realized by applying an adversarial loss function

$$L_{\text{GAN}}(G, D1) = \mathbb{E}_y[\log(D1(y))] + \mathbb{E}_{y,m,z}[\log(1 - D1(F[G(y, z) + y, m]))] \quad (2)$$

in which the generator and the first discriminator $D1$ compete against each other. Here, both networks are optimized based on a Min-Max game over the expectation values \mathbb{E} of the input data y and the masked output data of the generator. In addition to the incomplete or corrupted sinogram sections, there are also sections without the influence of a metal object. By searching for slices in the image in front of and behind the metal trace, these sections can be used as a reference for another discriminator network $D2$. In the second part of the loss function, again an adversarial loss is used in which the generator and the second discriminator compete against each other. The loss

$$L_{\text{GAN}}(G, D2) = \mathbb{E}_{x^*}[\log(D2(x^*))] + \mathbb{E}_{y,m,z,x^*}[\log(1 - D2(x^*, G(y, z)))] \quad (3)$$

is calculated now over the expectation values \mathbb{E} of the input data y and the reference data x^* from the neighbor slices to integrate the information from the uncorrupted projection data into the learning process. In both parts of the objective function, the generator receives the incomplete sinogram data and the corresponding metal trace as an input. Using a two-stage architecture consisting of a coarse and a fine network [8], the generator generates an output image that is used as an input to the discriminator networks. Each discriminator network, a binary classification network, assigns a probability between $[0, 1]$, indicating whether the data are real or artificially generated compared to the original input data or the data from the neighbor slices, respectively. The structures of the generator and discriminator network architectures are represented in Figure 1. To stabilize the training process, the loss is extended by two complementary losses, which help to enforce the dependency on the stochastic component weighted by the parameter λ . First, an encoding z -loss is

used to force the generator to use information from the stochastic component z from the latent representation of the given data and the masked output data of the generator network by adding

$$L^z(G, E) = \mathbb{E}_{y,m,z} \|z - E(F[G(y, z) + y, m])\|^2 \quad (4)$$

to the objective function. However, as shown in [9] and [10], this loss exhibits a “steganography behavior” and is not sufficient alone to stabilize the network training.

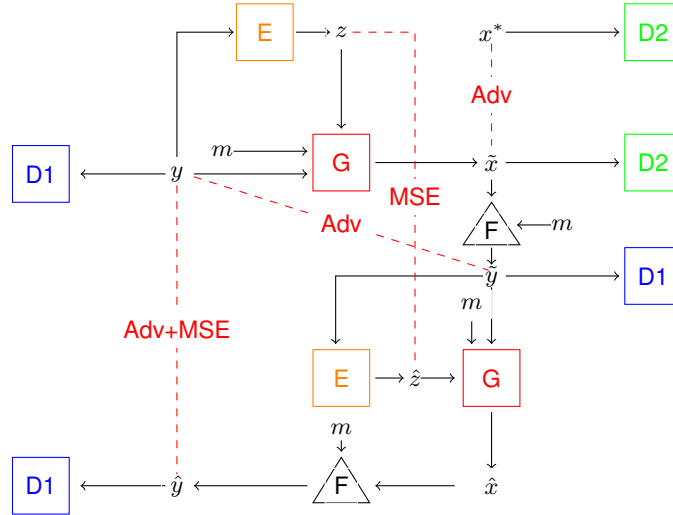


Fig. 2: Schematic representation of the used network interactions. The input data y are given together with a binary mask m and the latent representation into the generator G . The output data \hat{x} are then used as an input to the discriminator $D1$, by applying as masked function F , and the discriminator $D2$. Additionally, the masked output data \tilde{y} are given again into the generator together with the latent representation \tilde{z} and the binary mask. The output \hat{y} of the second pass is given into $D1$, and an adversarial loss is calculated together with the MSE-error to the original input data. Further, both latent vectors z and \tilde{z} are used to calculate an MSE-loss to update the generator and encoder network.

Therefore, an additional loss function, the encoding y -loss, is used. Here, the output data of the generator \tilde{y} are not used directly but are given after the application of the function F , again into the generator. The masked output data of the second pass are then compared again with the original input data y by calculating

$$\begin{aligned} L^y(G, D1, E) = & \mathbb{E}_y [\log(D1(y))] \\ & + \mathbb{E}_{\tilde{y}, \tilde{z}} [\log(1 - D(F[G(\tilde{y}, \tilde{z}) + \tilde{y}, m]))] \\ & + \mathbb{E}_{y, \tilde{z}} \|y - F[G(\tilde{y}, \tilde{z}) + \tilde{y}, m]\|^2 \end{aligned} \quad (5)$$

using an MSE-loss to constrain $G(\tilde{y}, \tilde{z})$ to be close to y and let their distribution be similar via an adversarial loss. Figure 2 shows the interaction between the four networks and the before described realization of the calculated loss functions for the network training.

III. EXPERIMENTS

Simulated data from a software phantom [11] with varying parameter settings were used to train and test the networks. Objects with different shapes were inserted in 120 different data sets at random positions. Using simulations for 3D cone-beam CT, the metal affected projection data were generated. In the image domain, the metal objects were segmented and projected forward to obtain the metal trace, which is used to remove the metal corrupted data. The generated data were divided into training (90), validation (15), and test (15) data. To generate more training data and data with a higher variation, image sections of 128×128 were used instead of the full 3D sinograms. Training was performed using the training parameters listed in table I.

TABLE I: training parameters

Parameter	Value
number training data	183 000
number validation data	31 000
batch size	32
number epochs	250
learning rate	$1 \cdot 10^{-5}$

IV. RESULTS

After successfully training the networks, the generator network can be applied to the test data. As shown in Figure 3, the network is able to reconstruct the missing data from the input data consisting of the data with the missing metal trace and the binary mask. Compared to the ground truth, hardly any differences can be seen.

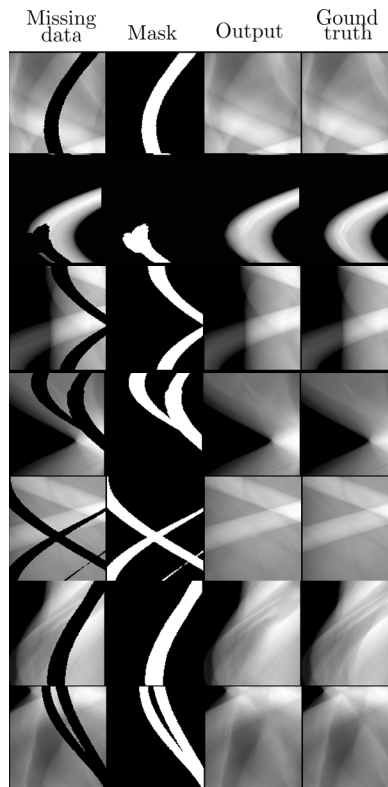
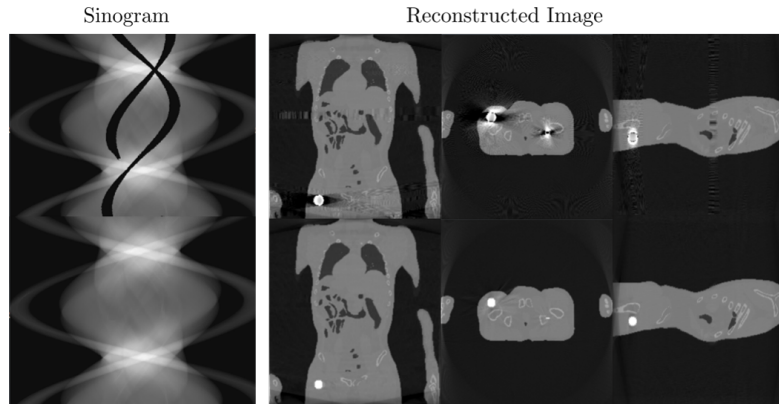


Fig. 3: Example data from the test data set completed by the generator network.

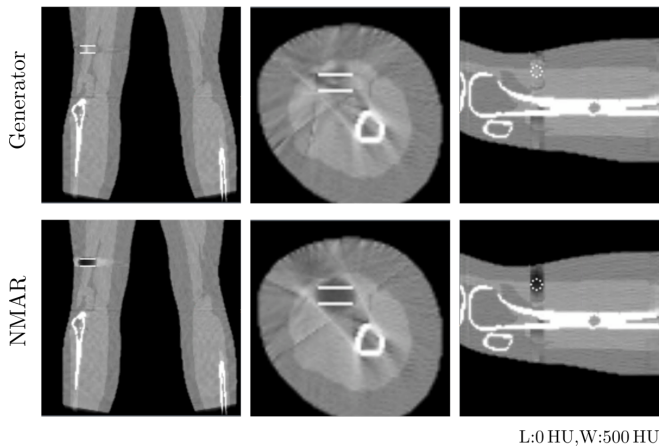
In the left part of Figure 4, an example slice from a completed 3D sinogram from the trained generator network is shown compared to the original input data above. Thereby, the artificially generated data by the generator network can barely be distinguished from the real surrounding data, making it difficult to detect the previous existing metal trace. The right side of the figure shows the resulting reconstruction obtained by an FDK reconstruction compared to the reconstruction from the original metal inserted data. Compared to the original data, it can be shown that the presence of artifacts could be reduced significantly by applying the generator network.

In Figure 5a, the results of the generator network are shown compared to a linear interpolation for cylindrical hollow objects, which the network had not seen yet in the training process. It can be shown that the network, in contrast to the interpolation, is able to reconstruct the inner structures of these kinds of objects. The mean squared error (MSE) between the reconstructed images and the ground truth was calculated on the test data set for validation. As a result, the

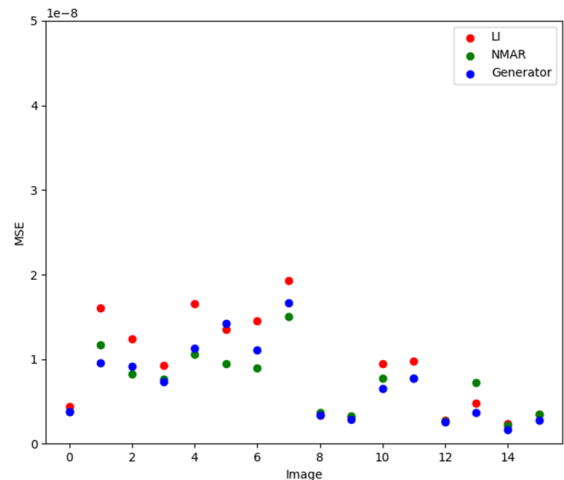


L:0 HU; W:1500 HU

Fig. 4: Data before (top) and after application of the network (below).



(a) Generator results (top) compared to NMAR (below) for a cylindrical hollow object.



(b) Calculated MSE values.

Fig. 5: Reconstruction results of the generator and the ground truth compared to NMAR and linear interpolation.

network was able to learn most of the missing structures, but some streak artifacts remain visible in the reconstructed image (Figure 5b).

Compared to linear interpolation, the average MSE value overall test images is recognizably lower with $7.7 \cdot 10^{-9}$ versus $1.0 \cdot 10^{-8}$. As well compared to the normalized metal artifact reduction (NMAR), the average MSE value is slightly smaller with $7.7 \cdot 10^{-9}$ versus $7.9 \cdot 10^{-9}$.

V. DISCUSSION

The developed method shows the first promising results on the simulated test data set. However, a detailed study of the influence of different training parameters is still missing. In particular, the weighting of the individual components of the loss function could further improve the training results.

One of the next steps is the integration of actual clinical data into the training process of the networks. Here, both the influence of different acquisition geometries on the network training, as well as the used resolution should be investigated. Increased or decreased resolution could lead to the necessity of adjusting the patch size for training the networks. Also, a validation method must be developed that is applicable to actual clinical data where, unlike for the simulated data, the ground truth is not available.

Further, the corrupted data should be integrated into the training process, as there might be more information about the

image structures near the metal object available. One way to realize this is to replace the binary mask as an input of the network with the original data of the metal trace. Here, for example, a weighting depending on the position would also be conceivable, which could be integrated into the network convolution. Besides the integration of the original data, it is also possible to take advantage of the fact that 3D data are available so that the input data used as well as the networks can be adapted to 3D.

VI. CONCLUSION

The results demonstrate that the generator is able to replace the missing image information in the sinogram and reduce a vast number of artifacts in the reconstructed image. Furthermore, by developing a method to train the networks without using any ground truth, the networks can be applied to real clinical data in the next step without the need for major modifications. In the future, the corrupted projection data should be used as an input of the generator network instead of using the binary mask of the metal trace. The network should be able to use this information from the metal trace to avoid introducing false image structures.

ACKNOWLEDGMENT

This research was partially supported by TANDEM the competence center for medical technology supported by the European Union and the State of Schleswig-Holstein (grant no. 122-09-024) and TOMEDEx (Federal Ministry of Education and Research under grant number BMBF 13GW0371C).

REFERENCES

- [1] G. H. Glover and N. J. Pelc, "An algorithm for the reduction of metal clip artifacts in CT reconstructions," *Medical Physics*, vol. 8, no. 6, pp. 799–807, 1981. [Online]. Available: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.595032>
- [2] W. J. H. Veldkamp, R. M. S. Joemai, A. J. van der Molen, and J. Geleijns, "Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT," *Medical Physics*, vol. 37, no. 2, pp. 620 – 628, 2010. [Online]. Available: <https://doi.org/10.1118/1.3276777>
- [3] M. Stille and T. M. Buzug, "Augmented likelihood image reconstruction with non-local prior image regularization," in *Proc. 4th Intl. Mtg. on image formation in X-ray CT*, 2016, pp. 145–8.
- [4] M. U. Ghani and W. Karl, "Fast Enhanced CT Metal Artifact Reduction using Data Domain Deep Learning," *IEEE Transactions on Computational Imaging*, vol. PP, pp. 1–1, 2019.
- [5] C. Peng, B. Li, M. Li, H. Wang, Z. Zhao, B. Qiu, and D. Z. Chen, "An irregular metal trace inpainting network for x-ray CT metal artifact reduction," *Medical Physics*, vol. 47, no. 9, pp. 4087–4100, 2020. [Online]. Available: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.14295>
- [6] N. Blum, T. Buzug, and M. Stille, "Projection Domain Metal Artifact Reduction in Computed Tomography using Conditional Generative Adversarial Networks," *MIDL*, 2021.
- [7] A. Pajot, E. de Bézenac, and P. Gallinari, "Unsupervised adversarial image inpainting," *CoRR*, vol. abs/1912.12164, 2019. [Online]. Available: <http://arxiv.org/abs/1912.12164>
- [8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative Image Inpainting with Contextual Attention," *arXiv preprint arXiv:1801.07892*, 2018.
- [9] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. C. Courville, "Augmented cyclegan: Learning many-to-many mappings from unpaired data," *CoRR*, vol. abs/1802.10151, 2018. [Online]. Available: <http://arxiv.org/abs/1802.10151>
- [10] C. Chu, A. Zhmoginov, and M. Sandler, "Cyclegan, a master of steganography," 2017. [Online]. Available: <https://arxiv.org/abs/1712.02950>
- [11] W. Segars, M. Mahesh, T. Beck, E. Frey, and B. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Medical Physics*, vol. 35, pp. 3800–3808, 2008. [Online]. Available: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2955743>