Agrawal, Harshit; Hietanen, Ari; Särkkä, Simo

Metal artifact reduction in cone-beam extremity images using gated convolutions

Published in:
Proceedings of the IEEE 18th International Symposium on Biomedical Imaging, ISBI 2021

DOI:
10.1109/ISBI48211.2021.9434163

Published: 25/05/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Agrawal, H., Hietanen, A., & Särkkä, S. (2021). Metal artifact reduction in cone-beam extremity images using gated convolutions. In *Proceedings of the IEEE 18th International Symposium on Biomedical Imaging, ISBI 2021* (pp. 1087-1090). Article 9434163 (International Symposium on Biomedical Imaging). IEEE. https://doi.org/10.1109/ISBI48211.2021.9434163

METAL ARTIFACT REDUCTION IN CONE-BEAM EXTREMITY IMAGES USING GATED CONVOLUTIONS

Harshit Agrawal[⋆†] *Ari Hietanen*[⋆] *Simo Särkkä*[†]

[⋆] Planmeca Oy, Asentajankatu 6, FI-00880 Helsinki, Finland
[†]Department of Electrical Engineering and Automation, Aalto University, FI-02150 Espoo, Finland

ABSTRACT

Quality of cone-beam computed tomography (CBCT) images are marred by artifacts in the presence of metallic implants. Metal artifact correction is a challenging problem in CBCT scanning especially for large metallic objects. The appearance of artifacts also change greatly with the body part being scanned. Metal artifacts are more pronounced in orthopedic imaging, when metals are in close proximity of other high density materials, such as bones. Recently introduced mask incorporating deep learning networks for metal inpainting showed improvements over classical methods in CBCT image quality. However, generalization of results for more than one body part is still not investigated. We investigate, the use of gated convolutions for mask guidance inpainting to improve the filling of the corrupt metal area in projection domain. The neural network was trained with eight clinical metal affected datasets by incorporating data augmentation techniques. In the end, we validate our method on six clinical datasets. Our method shows promising results both in projections and reconstructed images.

Index Terms— cone-beam computed tomography, deep learning, metal artifact reduction, orthopedic imaging.

1. INTRODUCTION

CBCT scanners reconstruct a 3D object from its 2D projections. If high density metals are present, the projection data becomes corrupted because of multiple phenomena, including beam hardening, photon starvation, and scattering [1]. Reconstruction algorithms are often unable to account for these corruptions and thus produce artifacts such as, bright and dark streaks. In post surgical orthopedic imaging, it is important to verify the osseointegration of implants to the bones. The "large" metallic hardware significantly degrades the image quality near bones, impeding the diagnosis [2].

Metal artifact reduction (MAR) has been an active area of research for the past four decades. The standard MAR protocol includes the following steps: 1) metal segmentation in image volume, 2) forward-projection of metals to obtain metal masks in projections, 3) inpainting of the metal masks, and 4) reconstruction of the image volume and insertion of the thresholded metals from Step 1). In linear interpolation (LI-MAR) based inpainting, the corrupt projection data in the metal area is replaced by linear interpolation of non-metal neighborhood [3]. However, LI-MAR results in blurring of the bone near the implant due to loss of edge information and streak artifacts tangent to metal objects [4]. Besides LI-MAR, many different and more complex interpolation methods have been suggested but no universally accepted method exists yet due to large variations in the size, shape, and density of metallic implants [5].

In recent years, many data-driven deep learning based approaches have been proposed for metal inpainting. Many of these approaches inpaint the projections in the CT sinogram domain [6, 7, 8]. [9] used an U-net [10] for metal inpainting in the knee projections and showed that, concatenating the metal mask with the input projection improves the inpainting results. However, improvements in the corresponding image domain were not shown, which is a crucial end objective of MAR. The benefit of using metal masks has also been shown by [11] for inpainting in projections and by [12] in images.

In this paper, motivated by the success of gated convolutions in language modelling [13] and natural image inpainting [14], we used gated convolutions for CBCT metal artifact correction. We tested the developed model on a clinical dataset against and our results show improvements in image quality for different shape, size and location of metal. The results were compared against LiMAR [3], U-net [9], Mask pyramid [11], and Partial convolution [12] based methods using the same data. Our contribution is to apply deep neural gated convolutional networks based mask guidance for metal inpainting in CBCT images.

2. MATERIALS AND METHODS

2.1. Mask guidance

Our idea is to utilize a metal mask for convolutions gating and compare the results with other state-of-art mask incorporating approaches. The output of a convolutional layer for a channel C at a pixel (x, y) can be expressed as

$$\mathbf{O}_{x,y} = \sum_{i=-k}^{k} \sum_{j=-l}^{l} \mathbf{W}_{k+i,l+j} \mathbf{I}_{x+i,y+j}, \quad (1)$$

where k and l are defined by kernel height (h), $k = \frac{h-1}{2}$ and width (w), $l = \frac{w-1}{2}$, $\mathbf{W}$ is a convolutional filter, and $\mathbf{I}$ and $\mathbf{O}$ are input and output features, respectively. For brevity, bias is ignored.

In inpainting, network needs to fill the masked pixels (invalid) using surrounding pixels (valid). However, during the convolution in a given window, the same weights are applied to both invalid and valid pixels (1). For highly varying metals in the size and shape, this may cause instability during training and discrepancies on test data.

Partial convolution has been proposed to overcome this problem. Partial convolution includes a re-normalization step in the convolution to use only valid pixels, given by

$$\mathbf{O}_{x,y} = \begin{cases} \mathbf{W} \left(\mathbf{I}_{x,y} \bigodot \mathbf{M}_{x,y} \right) \frac{\text{sum}(\mathbf{1})}{\text{sum}(\mathbf{M})}, & \text{if sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $\mathbf{W}$ is a convolutional filter weights, $\mathbf{I}_{x,y}$ is the corresponding features for the current convolution (sliding) window centered around point (x, y), $\mathbf{M}_{x,y}$ is the corresponding window on binary mask and sum is the sum of all elements. $\mathbf{1}$ has same size as $\mathbf{M}$ with all elements as 1 and $\bigodot$ is element-wise multiplication. After each partial convolution, the mask at (x, y) is updated if at least one valid pixel was present in $\mathbf{M}$, that is if $\text{sum}(\mathbf{M}) > 0$.

More general, automatic soft guidance learning scheme, gated convolutions, has been proposed in [13]. The network learns also the mask (gating) which has continuous values in the range of $[0, 1]$:

$$\mathbf{O}_{x,y} = \phi(\mathbf{F}_{x,y}) \bigodot \sigma(\mathbf{G}_{x,y}) \quad (3)$$

where $\mathbf{F}_{x,y} = \mathbf{W}_f \mathbf{I}_{x,y}$ and $\mathbf{G}_{x,y} = \mathbf{W}_g \mathbf{I}_{x,y}$ are the features and gating values, respectively. $\mathbf{W}_f$, $\mathbf{W}_g$ are convolutional filters for input and gating, respectively. ϕ is an activation function (e.g., ReLU) and σ is the sigmoid activation function to squash gating values between zeros and ones.

2.2. Network Architecture

We use an U-net architecture as backbone with some modifications. There are 32 channels, four contracting paths, and partial convolution based padding for each convolution. A clipped ReLU (0-1) activation function was placed before the output layer to clip the output values between 0 and 1. Instance normalization [15] was used after each convolution+ReLU block. We did not use bias in convolutional blocks because of the similar effect caused by instance normalization. We kept backbone U-net same for mask pyramid, partial convolutions and gated convolutions-based networks. Except, for gated convolution-based model, we reduced number of channels to 24 to have comparable number of trainable parameters and maintain same efficiency as other models.

2.3. Dataset

We used unidentifiable and pseudonymized clinical dataset acquired by Planmed Verity® to create metal mask and ground truth images. The segmentation was done by manual thresholding in image domain. We forward projected metals to obtain masks for inpainting. We obtained a total of 3000 masks from eight metal volumes. 5400 projections from four knee, six ankle, and six wrist volumes (without any metal) were used to place masks. A number of random augmentations (rotation, shear, zoom, flip) were applied on masks. Augmented binary masks (zero inside metal mask) were placed in ground truth projections. All augmentations were done on-the-fly during training.

We created two test sets, synthetic and real, from twelve clinical volumes with six volumes having metal and six volumes without metal. Synthetic test set was created by inserting masks obtained from the six metal volumes on the projections of the six non-metal volumes. In this manner, we created masked projections to inpaint and corresponding ground truth projections. For real test set, we segmented metals to obtain masks in the projections of the six metal volumes. We do not have ground truth for real test set because we do not have corresponding metal free projections.

All projections were linearised using $-\ln \frac{I}{I_0}$, where I is measured intensity at the detector (projection intensities) and I_0 is the initial intensity emitted from the X-ray source. Projections were further normalised between 0-1.

2.4. Training procedure

We implemented proposed method using TensorFlow framework and trained on a single NVIDIA GeForce RTX 2080 GPU (8GB), with a batch size of 1. We used Adam for optimization with default parameters. A combination of structural similarity index (SSIM) and $l1$ loss was used with weight one. We set initial learning rate to e^{-4} and reduced it after each epoch in logarithmic steps down to e^{-6}. Learning rate was reduced during the 25 first epochs and was fixed after that. Training was stopped when the training loss did not decrease for five subsequent epochs.

3. RESULTS

3.1. Evaluation on synthetic test set

We evaluate synthetic test set for each extremity separately for inpainting in projections and corresponding image reconstructions, shown in Table 1. We use three metrics for evaluation: mean square error (MSE), mean absolute error (MAE) and peak signal to noise ratio (PSNR) for the comparison of

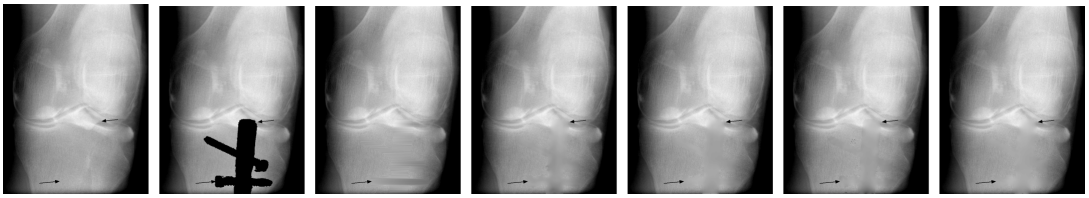

Fig. 1. Inpainted projections from synthetic knee test set. Arrows in each image show locations where major differences are visible. From left to right columns: Ground truth, Masked projection, LiMAR, U-net, Mask pyramid, Partial convolutions, Gated convolutions.

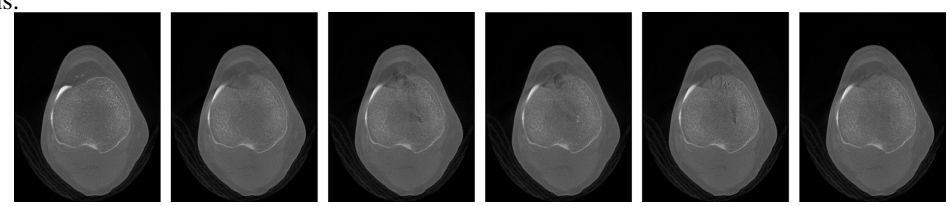

Fig. 2. Axial image slices after reconstruction of inpainted projections from synthetic test set. From left to right columns: Ground truth, LiMAR, U-net, Mask pyramid, Partial convolutions, Gated convolutions.

inpainted projections and corresponding reconstruction images. We calculate metrics for wrist, knee and ankle cases separately. Gated convolution-based model results in lowest MSE and MAE, and highest PSNR values in inpainted projections for all extremity types. In Fig. 1, a projection from knee volume is shown along with the corresponding ground truth, masked and inpainted projections.

The limited improvement in reconstruction images could be due to the inconsistencies in the inpainted area across all the projections. Nonetheless, gated convolutions are producing least artifacts in reconstructions as shown in the Fig. 2.

Table 1. Comparison for inpainted projections and corresponding image reconstructions in synthetic test set in terms of averaged metrics. All projections are linearized and normalized in the range 0-1 and reconstructions are in Hounsfield units.

Method	Extremity	Projections			Reconstructions		
		mse (10^{-3})	mae (10^{-3})	psnr	mse	mae	psnr
LiMAR	wrist	0.326	1.942	43.101	1091.82	7.55	37.42
	knee	0.149	1.516	47.955	403.30	5.33	44.24
	ankle	0.591	3.773	43.309	1457.44	12.74	37.42
U-Net	wrist	0.175	1.545	46.072	935.52	7.18	42.61
	knee	0.126	1.560	48.268	448.64	5.53	43.69
	ankle	0.388	3.446	44.736	1410.14	12.51	37.58
MaskPyramid	wrist	0.139	1.426	46.881	913.25	7.16	42.69
	knee	0.103	1.368	49.071	415.44	5.42	43.98
	ankle	0.397	3.524	44.752	1514.06	12.84	37.26
PartialConv	wrist	0.127	1.363	47.108	**819.39**	**6.90**	43.13
	knee	0.103	1.393	48.827	430.86	5.47	43.85
	ankle	0.308	3.132	45.793	1318.57	12.15	37.86
GatedConv	wrist	**0.123**	**1.328**	**47.418**	823.24	6.93	**43.14**
	knee	**0.075**	**1.112**	**50.306**	**352.91**	**5.11**	**44.76**
	ankle	**0.242**	**2.565**	**46.792**	**1168.81**	**11.57**	**38.41**

3.2. Evaluation on real test set

As we do not have ground truth for real test set, we evaluate it qualitatively. Fig. 3 shows a comparison of reconstructed sagittal and axial image slices from a wrist volume, inpainted by different methods. Gated convolution-based model removes most of multiple streaks and darkening artifacts, also the shape of the bone, near the metal, is more visible.

4. DISCUSSION AND CONCLUSION

We showed quantitative and qualitative validation of our method on clinical dataset. The method performed well on diverse shape and size of wrist, knee and ankle images. While projection correction is only one of the steps of the MAR pipeline, it is one of the crucial and difficult step to perform with classical algorithms. If the projection correction process puts wrong information in the mask area, it can cause additional artifacts in the reconstruction. Sometimes, there can be useful information available in the masked area, for example when metals are thin. This information also gets inpainted by placing merely a mask. This is a limitation of mask based inpainting methods. In future, we will acquire more CBCT clinical data from other extremities and will validate model's robustness further.

5. REFERENCES

[1] R. Schulze, U. Heil, and D. Gross, "Artefacts in cbct: a review," *Dentomaxillofacial Radiology*, vol. 45, no. 5, pp. 265–273, 2011.

[2] D. D. Robertson, P. J. Weiss, D. Magid, and P. S. Walker, "Evaluation of ct techniques for reducing artifacts in the presence of metallic orthopedic implants," *Journal of Computer Assisted Tomography*, vol. 12, no. 2, pp. 236–241, 1988.

Fig. 3. Reconstructed slices for a wrist dataset, after metal mask inpainting in projections. First row images are sagittal slices and second row images are axial slices. From left to right columns: Uncorrected, LiMAR, U-net, Mask pyramid, Partial convolutions, Gated convolutions.

[3] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of ct artifacts caused by metallic implants," *Radiolgy*, vol. 164, no. 2, pp. 576–577, 1987.

[4] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (nmar) in computed tomography," *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, 2010.

[5] L. Gjesteby, B. De Man, Y. Jin, H. Paganetti, J. Verburg, D. Giantsoudi, and G. Wang, "Metal artifact reduction in ct: Where are we after four decades," *IEEE Access*, vol. 4, pp. 5826–5849, 2016.

[6] L. Gjesteby, Q. Yang, Y. Xi, Y. Zhou, J, Zhang, and G. Wang, "Deep learning methods to guide ct image reconstruction and reduce metal artifacts," in *Medical Imaging 2017: Physics of Medical Imaging*, International Society for Optics and Photonics, 2017, vol. 10132, p. 101322W.

[7] M. U. Ghani and K. W. Clem, "Deep learning based sinogram correction for metal artifact reduction," *Electronic Imaging*, vol. 2018, no. 01, pp. 4721–4728, 2018.

[8] H. S. Park, Y. E. Chung, S. M. Lee, H. P. Kim, and J. K. Seo, "Sinogram-consistency learning in ct for metal artifact reduction," *Medical Physics*, vol. 45, no. 12, pp. 5376–55384, 2018.

[9] T. M. Gottschalk, B. Kreher, H. Kunze, and A. Maier, "Deep learning based metal inpainting in the projection domain: Initial results," in *Machine Learning for Medical Image Reconstruction*, 2019, pp. 125–136.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015, pp. 234–241.

[11] H. Liao, Wei-An Lin, Z Huo, Levon Vogelsang, W. J. Sehnert, S. K. Zhou, and J. Luo, "Generative mask pyramid network for ct/cbct metal artifact reduction with joint projection-sinogram correction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI*, 2019, pp. 77–85.

[12] D. Mangileva, A. Dokuchaev, S. Khamzin, S. Zubarev, T. Lyubimtseva, D. Lebedev, and O. Solovyova, "Removing artifacts from computed tomography images of heart using neural network with partial convolution layer," in *2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, 2020, pp. 195–198.

[13] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *34th International Conference on Machine Learning*, 2017, vol. 70, p. 933–941.

[14] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," *CoRR*, vol. abs/1806.03589, 2018.

[15] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, 2016.

6. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki.

7. ACKNOWLEDGMENTS

Harshit Agrawal and Ari Hietanen are employees at Planmeca Oy., Finland. This work was partially funded by Business Finland.