# Generating Artificial Artifacts for Motion Artifact Detection in Chest CT

Guus van der Ham[1,2], Rudolfs Latisenko[2], Michail Tsiaousis[2], and Gijs van Tulder[1]

[1] Radboud University, Nijmegen, The Netherlands
[2] Thirona, Nijmegen, The Netherlands
http://www.thirona.eu

**Abstract.** Motion artifacts can have a detrimental effect on the analysis of chest CT scans, because the artifacts can mimic or obscure genuine pathological features. Localising motion artifacts in the lungs can improve diagnosis quality. The diverse appearance of artifacts requires large quantities of annotations to train a detection model, but manual annotations can be subjective, unreliable, and are labour intensive to obtain. We propose a novel method (code is available at https://github.com/guusvanderham/artificial-motion-artifacts-for-ct) for generating artificial motion artifacts in chest CT images, based on simulated CT reconstruction. We use these artificial artifacts to train fully convolutional networks that can detect real motion artifacts in chest CT scans. We evaluate our method on scans from the public LIDC, RIDER and COVID19-CT datasets and find that it is possible to train detection models with artificially generated artifacts. Generated artifacts greatly improve performance when the availability of manually annotated scans is limited.

**Keywords:** Artifact generation · Motion artifact detection · Fully convolutional network · Chest CT · Thoracic CT · ASTRA toolbox

## 1 Introduction

Motion artifacts are a common problem in computed tomography (CT) imaging. They are caused by patient movement during scanning, and can complicate both manual and automated analysis of the images [2,4,10]. For example, in chest CT, breathing and heart motion can introduce artifacts such as intensity undershoots and deformations [2,15]. These artifacts can interfere with tasks such as gross tumor volume estimation [13], automated airway segmentation [4] and automated nodule detection [10].

In this paper, we focus on localizing motion artifacts in lung CT. Knowing the location and severity of artifacts could help to reduce misinterpretations and improve diagnosis quality [12]. Previously, convolutional neural networks (CNNs) have been used to detect artifacts in cardiac CT images [6,7,12]. Recently, Beri [3] applied a U-Net-based model to detect motion artifacts in chest CT images.


C. Zhao et al. (Eds.): SASHIMI 2022, LNCS 13570, pp. 12–23, 2022.
The final publication is available at Springer via https://doi.org/10.1007/978-3-031-16980-9_2

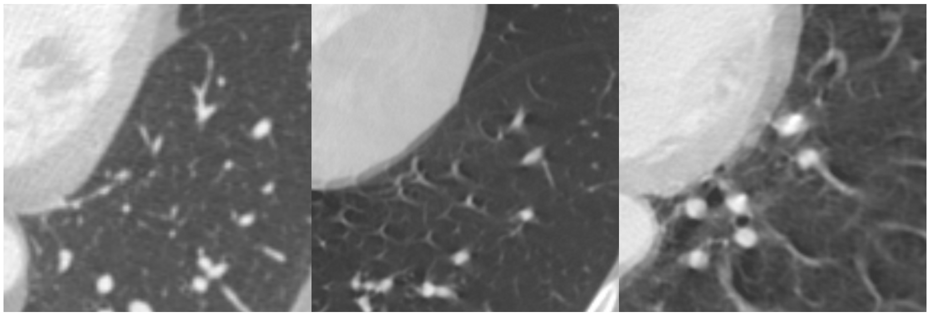



**Fig. 1.** Examples of motion artifacts in chest CT: Motion in the lungs can have multiple causes. In b) a heartbeat causes spiky artifacts near the heart. In c) the whole lung is affected with smeared-out features and intensity undershoots.

Training a detection model requires a large training set with annotated CT scans, especially because motion artifacts in the lungs have a very diverse appearance [2] and need a large number of examples to cover all variations. However, manually annotating motion artifacts is time-consuming, subjective, and error-prone. In this study, we observe a large inter-rater variability in the annotated artifacts, despite the strict annotation protocol and training of the annotators.

To address this data scarcity, we propose a novel method to generate artificial motion artifacts in chest CT. Starting from clean CT images, our method introduces motion in a simulated CT acquisition. By controlling the location, severity, and direction of the movement, we can generate a large variety of artifacts, which can then be used to train a detection model for real motion artifacts.

Our approach has several advantages. It increases the amount of training data and reduces the dependence on manual annotations. Since it is based on a simulated CT acquisition, it can use standard CT images and does not require raw projection data, which is often not available. Finally, the method provides detailed information about the motion causing the artifacts, which is not available for real artifacts and could be used to train a motion estimation model.

***Related Work.*** Motion artifacts are common during CT acquisition [2] and are difficult to correct in post-processing. Most motion artifact removal methods require raw CT projection data (e.g., [10,22]), although some recent methods work directly on reconstructed images (e.g., using generative adversarial networks [9]). Methods for motion correction often have high computational requirements, produce imperfect results, and can even introduce new artifacts [7].

As an alternative to removing motion artifacts, detection models can be used to detect them, e.g., for quality assurance. Many detection methods rely on raw CT projection data, for example, by looking for discontinuities in the sinograms [14]. Other detection approaches use the final reconstructed image. For example, Sun et al. [20] developed a pre-processing method that enhances motion artifacts

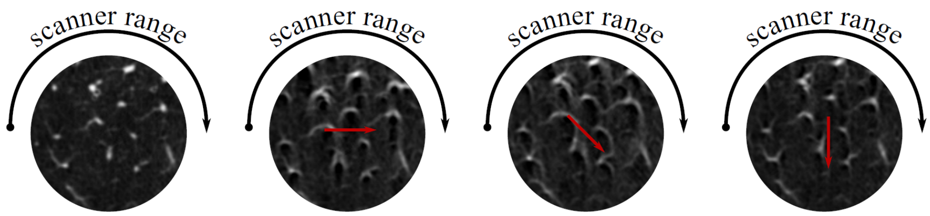



**Fig. 2.** Effects of motion direction: The motion direction relative to the scanner range changes how blood vessels are deformed. Parallel movement results in arc-shaped artifacts while perpendicular movement produces star-shaped artifacts with intensity undershoots on three sides. Figure inspired by Fig. 6 from Lossau et al. [12]

features for easier detection and trained CNNs that classify images containing artifacts. Beri [3] applied a U-Net to detect artifacts in lung CT images.

Detection models can be trained using artificial artifacts. One way to generate these artifacts is by using a generative adversarial network (GAN) [19,24], but these approaches provide little to no control over the resulting artifacts. Furthermore, GANs require a large amount of training data and can lead to unreliable results (e.g., [5]).

In this paper, we use a more principled approach to generate artifacts, by introducing motion in a simulated CT acquisition with a predefined motion model. This is somewhat similar to the work by Lossau et al. [6,7,12], who generate artifacts in blood vessels in the heart and use those to train an artifact detection CNN for coronary CT angiography. However, their method relies on perturbing projection data that is often not available in chest CT. Our approach uses simulated CT acquisition, which allows us to use reconstructed images directly and avoid this limitation.

## 2 Methods

We generate images with artificial artifacts by introducing motion during simulated CT acquisition. Figure 3 shows an overview of our method.

***Motion Parameters.*** The appearance of motion artifacts depends on two key factors: the direction, and the severity of the motion. The direction of the motion, relative to the rotation of the detector, determines the shape of the artifact. For example, small blood vessels that appear as white dots can turn out as spiky arc or star-shaped artifacts (Fig. 2), depending on the direction of movement. The severity of the motion determines the size of the artifacts. This motion is often not equal throughout the lung: for example, heartbeats cause large movements and strong artifacts near the heart, but have weaker effects elsewhere (Fig. 1b). Further examples of generated artifacts are shown in Fig. 6.

***Artifact Generation.*** The ASTRA Toolbox [21] for Python simulates CT acquisition based on previously acquired scans. It rotates a virtual detector around

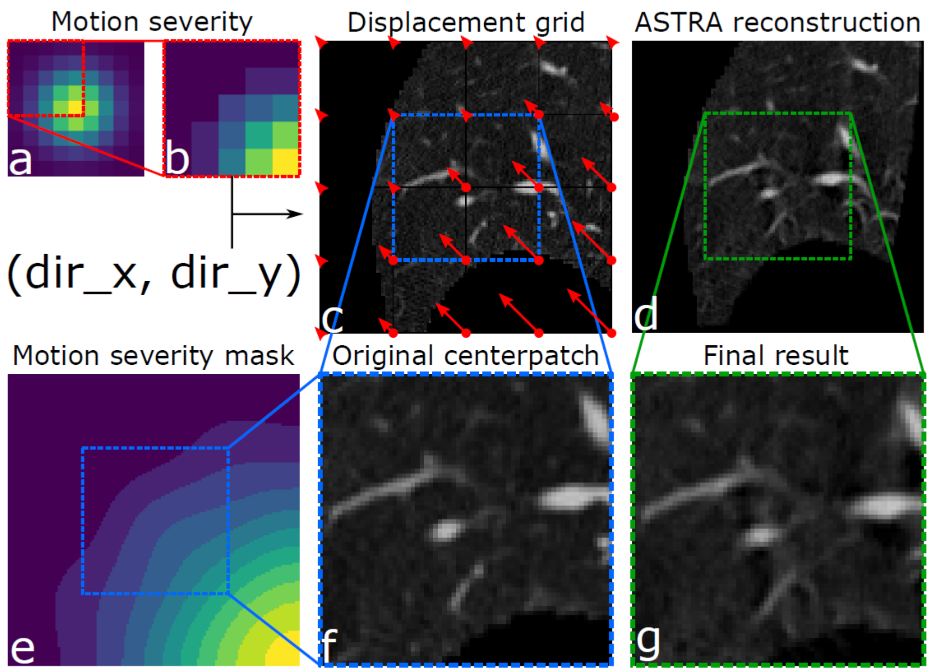



**Fig. 3.** Generating motion artifacts: a) $M \times M$ Gaussian matrix ($M = 9$, $\sigma = 0.4$) to simulate decreasing motion severity away from source, b) random $N \times N$ submatrix ($N = 5$) for varying source location, multiplying severity matrix with a vector $(-10, 10)$ results in c) $N \times N$ displacement vector grid over input image determining the movement at different locations, d) the CT simulation with motion produces the ASTRA reconstruction of the input image containing artifacts and e) an output from the elastic deformation algorithm: a mask of movement severity detailing the movement of every pixel in the input image, f) original center patch (for reference post-processed the same as final result), g) post-processed final center patch containing artifacts.

an existing CT image, makes projections of the tissue density from different angles, and then computes a reconstruction of the original image using filtered back projection [16]. We adapt this process by introducing motion during the simulated acquisition process. For several projection angles, we replace the input image with an image that has been moved slightly. This mimics a fluid movement, resulting in realistic-looking motion artifacts in the final reconstruction.

***Motion Simulation.*** To simulate realistic motion, we adapt an existing elastic deformation library [23] to create deformed images and compute a pixel-wise deformation mask showing the location, strength, and direction of movement.

The elastic deformation method defines a displacement grid $D$ with $N \times N$ control points to cover the input image. For each control point $i, j$, a displacement vector $D_{i,j}$ defines the deformation in the $x$ and $y$ directions. The displacement vectors are then interpolated for each pixel in the image to compute a smooth deformation field, which is used to map the input image to the deformed output.

For our purpose, we define the displacement grid to describe the severity and direction of motion (Fig. 3). We generate a Gaussian matrix of size $M \times M$, with $M > N$ and random standard deviation $\sigma$, and crop this to $N \times N$ to define the motion severity at each control point. For small $\sigma$, this results in a strong, local motion concentrated at one control point, whereas a large $\sigma$ produces movement in a larger area. To obtain the direction of the movement, we sample a random motion vector from a uniform distribution $U(-d, d)$. By multiplying this motion vector with the severity at each grid point, we obtain a smooth displacement grid $D$. During the simulated CT projection, we simulate a fluid movement by slowly changing the displacement grid: from zero displacement in the first projection step to the full displacement $D$ at the end, with linear interpolation in between.

***Detection Network.*** We evaluate our artificial artifacts by training detection models. We train fully convolutional networks (FCN) [11] that take a $64 \times 64$ patch as input and output a single probability. The architecture is shown in Fig. 5. The fully convolutional nature of the network allows for input images of arbitrary size. Inputs larger than the original input size produce a probability map in which each output pixel corresponds to a $64 \times 64$ input region. The boundaries of this region can be calculated based on the number of max pooling operations in the network. We upsample the prediction to match the original input resolution, averaging the probabilities in regions with overlapping receptive fields. This produces a coarse heatmap with high probabilities in the core areas containing artifacts and decreasing probabilities around the borders.

## 3 Experiments

For our evaluation, we generated motion artifacts to train the FCN detection model under varying conditions. In all settings, we used the Adam optimizer with binary cross-entropy loss, a minibatch size of 128, and a learning rate of $10^{-4}$ decaying by 2% each epoch to $10^{-5}$. After training for 128 epochs, the model with best validation performance is selected. All components were implemented in Python using the Tensorflow framework.

***Data.*** We used the public LIDC dataset [1] for most of our experiments. We created random subsets of 75, 32, 25 scans for training, validation, and testing. To measure generalizability, we used two additional test sets of 9 randomly selected scans each from the RIDER [17] and COVID19-CT [18] datasets.

***Annotations.*** Each scan was annotated by one of three analysts following an annotation protocol. For the 75 training scans, the analysts were supervised by an analyst with more experience in annotating motion artifacts. We only used annotations located within the lungs.

To measure the inter-annotator agreement, 25 scans in the LIDC test set were annotated by all three analysts. The agreement was computed using Fleiss'

$\kappa$ [8], taking every voxel inside the lung masks as a sample (motion / no motion). This resulted in a $\kappa$ of 0.35. This relatively low agreement illustrates the difficult and subjective nature of the manual annotation task.

To strengthen the reliability and consistency of the evaluation, each scan in the LIDC, RIDER and COVID19-CT test sets was also annotated by the more experienced analyst. By going through all scans, checking and combining the different annotations, the experienced analyst obtained a more consistent and complete ground truth. We used these final annotations in our evaluation.

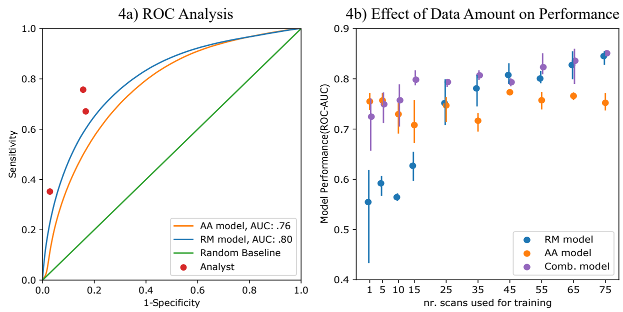
***Patch Sampling and Artifact Generation.*** Scans were normalized and masked using lung masks. We sampled 'real motion' (RM) patches of $64 \times 64$ pixels from areas with annotated motion artifacts, and 'clean' patches of the same size from slices with no annotated motion artifacts. To generate 'artificial artifact' (AA) patches, we sampled clean patches of $128 \times 128$ pixels and generated motion artifacts using our proposed method. For motion simulation we used $M = 9$, $N = 5$ and a random $\sigma$ between 0.1 and 1.2. Finally, we extracted the $64 \times 64$ center patch to avoid border artifacts. We observed that the ASTRA reconstruction algorithm sometimes produced slightly pixelated outputs. We applied mild Gaussian filtering ($\sigma = 0.5$) to remove this effect. We also applied smoothing to all other training patches and test images, to prevent the model from learning to recognize AA patches by the amount of smoothing.

***Evaluation.*** We evaluated the models by comparing their pixel-wise predictions for all CT slices in all test scans with the ground-truth annotations. Only the predictions inside the lung masks were evaluated. We computed the area under the ROC-curve as our main metric, as this shows the model's ability to separate positive from negative samples over a broad range of operating points. Based on a visual verification of the predictions, we found that an AUC over .80 represents excellent performance at localizing the core areas that are affected by motion, with most errors located near the (subjective) boundaries of the annotations.

***Experiment Set-Up.*** We performed three experiments: training AA models using AA patches and clean patches, RM models using RM patches and the same sets of clean patches, and combined models using all three types of patches.

In the first experiment, we looked at how the artifacts generated by our method compare against real motion artifacts. Both models were trained on 100 000 clean patches and 100 000 AA or RM motion patches.

In the second experiment, we explored how the models perform with limited training data. One of the advantages of the proposed method is its ability to generate a large amount of AA patches from a small set of clean patches. In this experiment, we compared the performance of AA, RM and combined models trained on subsets of our full training set, by extracting patches from a random subset of 1 to 75 training scans. The RM models have to rely on the small amount of RM patches that they can sample from a set of scans. In contrast, using our method, the AA models have access to a wide variety of AA patches generated from a much larger set of clean patches. For comparison, we also




4c) Model Comparison

| Models | LIDC | | RIDER | | COVID | |
|---|---|---|---|---|---|---|
| | AUC | std. | AUC | std. | AUC | std. |
| AA | 0.752 | 0.018 | 0.661 | 0.138 | 0.696 | 0.140 |
| RM | 0.845 | 0.015 | 0.828 | 0.007 | 0.827 | 0.015 |
| Combined | **0.851** | 0.005 | **0.856** | 0.037 | **0.829** | 0.022 |

**Fig. 4.** Experiment results: a) An AA and RM model trained with 200 000 patches each show comparable performance, signifying that artificial artifacts are relatively good representations of real artifacts. Analyst performance relative to ground truth is shown. b) Generating artificial artifacts improves performance greatly when access to scans is limited. c) The RM models and the combined model show better generalizability than the AA models.

trained 'combined' models' by sampling 50% of training patches from the pool of AA patches and the other 50% from the pool of RM patches that are available in each condition. We repeated this experiment three times for each condition, and report the mean and range of performances over all runs.

In the final experiment, we evaluated the generalizability of our models. We trained three AA, RM and combined models using the complete LIDC training set, and report their mean performance and standard deviation on the LIDC, RIDER and COVID19-CT test sets.

## 4 Results

***Experiment 1: Training with Equal Amounts of Data.*** In the experiment with an equal amount of training data (Fig. 4a), the performance of the AA model (AUC of .765) trained with artificial motion patches comes close to that of the RM model (AUC of .800) trained with real motion patches. This indicates that the proposed method generates motion artifacts that look sufficiently convincing to train a model to detect real motion artifacts in unseen scans. The slightly lower performance of the AA model suggests that the artificial artifacts are not perfect imitations of real motion artifacts. We suspect this might be caused by pixelation introduced by the ASTRA-reconstruction algorithm, by a limitation of the relatively simple motion model used in our experiments, or because the automatic segmentations do not perfectly match the manual annotations of the test set (e.g., because of a consistent over- or undersegmentation).

***Experiment 2: Advantage of Generating Data.*** When training data is scarce, generating and training with artificial artifacts improves performance (Fig. 4b). The RM models trained with real motion artifacts perform poorly when only 1 to 20 training scans are available. In contrast, the AA models trained with artificial artifacts perform much better. As expected, adding more scans improves the performance of the RM model, but this improvement levels off after about 45 scans. The AA models are less dependent on the amount of available training scans. Using our method we can train an effective model with very little training data. If enough data is available, training with real motion artifacts is slightly better. Combining both types of motion patches usually works best: it outperforms the RM model if the available RM data is limited, and does not perform worse than the RM model if the amount of RM data is large enough.

***Experiment 3: Model Comparisons and Generalizability.*** As in the previous experiments, we find that given enough data a model trained using real artifacts outperforms a model trained using only artificial motion artifacts. We find that that both the combined and RM models generalize well (Fig. 4c), while the AA model shows decreased performance on the RIDER and COVID test sets. In this experiment, where the number of real artifacts is sufficiently large, combining artificial and real artifacts does not strongly improve the performance.

## 5 Discussion

In this work we proposed a novel method to generate artificial motion artifacts by introducing motion in a simulated CT acquisition. The artificial artifacts can be used to train artifact detection networks without requiring manual annotations. We evaluated our method on three public datasets by training detection networks with real or artificial artifacts. We found that the artificial artifacts were realistic enough to train models that can detect real motion artifacts. Our method was especially effective when the availability of annotated CT scans was limited.

By using simulated CT acquisition, our method can be applied on CT images directly, without the need for projection data. By visual inspection we found that the generated artifacts look very convincing, although post-processing was needed to remove some pixelation introduced by the reconstruction algorithm. Our method allows precise control over the simulated motion. By varying the motion hyperparameters we can easily generate additional training data with a great diversity of artifacts, starting from a small set of CT images.

In addition to the reconstructed image, our method generates a 'motion mask' that specifies by how much and in which direction each pixel in the input image has moved during the simulation. This information can not be obtained from manual annotations, but would be very useful for training motion severity estimation models. These models can determine the severity of artifacts in real CT images [12]. Some lung analysis tasks are more sensitive to motion than others. For these applications, motion estimation models trained with artificial artifacts could be used to automate quality assurance decisions.

Our finding that models trained using only artificial artifacts performed slightly worse than models trained using only real motion artifacts suggests that, even though the proposed method generates realistic artifacts, the distribution of

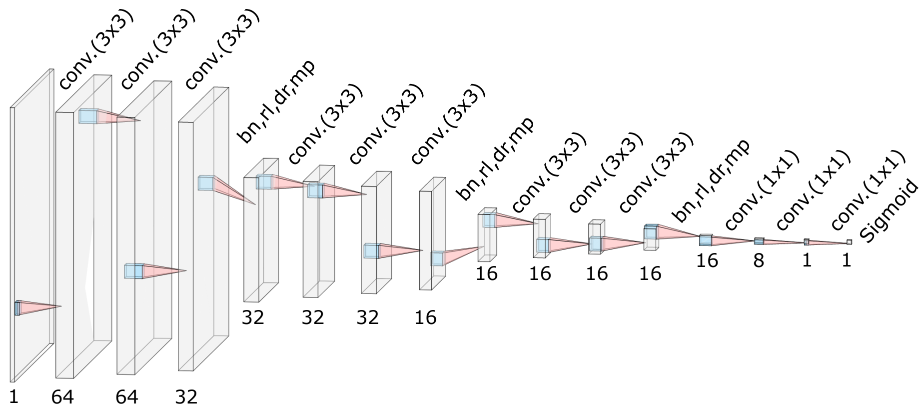



**Fig. 5.** Network architecture: The FCN takes a $64 \times 64$ patch and passes it through 3 convolutional blocks. Each block consists of 3 convolutional layers. Kernel sizes are indicated at the top and number of filters are indicated at the bottom. Each block ends with a Batch Normalization (bn), ReLU (rl), Dropout (dr) and MaxPooling (mp) layer. After the three blocks an additional three 1 by 1 convolutions and Sigmoid activation produce a single probability.

the generated artifacts does not perfectly match the distribution of real motion artifacts. This may be caused by pixelation introduced by the ASTRA-Toolbox, or by a mismatch between the artificial segmentations and the manual ground truth, but it might also be related to the diversity of the generated artifacts.

The diversity of artificial artifacts depends on the motion model. The proposed motion simulation can generate a wide range of artifacts but it is not exhaustive: the generated artifacts only represent a subset of all possibilities. Given a sufficient number of real artifacts, the method might observe a greater diversity in artifact appearances, which could explain why detection models trained with a large number of real motion artifacts outperform the models trained with artificial artifacts. This could be resolved with a more advanced motion model.

Motion artifact detection is a challenging task, and determining the boundaries of areas affected by motion can be subjective. This is also reflected in the low inter-annotator agreement of our manual annotations. In our experiments, we evaluated voxel-level predictions, which gives a good idea of the performance of the model but is sensitive to over- and undersegmentation. By visual inspection, we found that the models were often able to locate the affected areas, but had a tendency to overpredict. This might be solved by improving the detection model, e.g., using false-positive mining [3] or other forms of regularization.

In summary, the proposed method for motion artifact generation can be used to obtain additional training data with a great diversity of artifacts. This contributes to better artifact detection and motion severity estimation models. In turn, these models could improve scan and diagnosis quality for both automated analysis pipelines and human analysts.

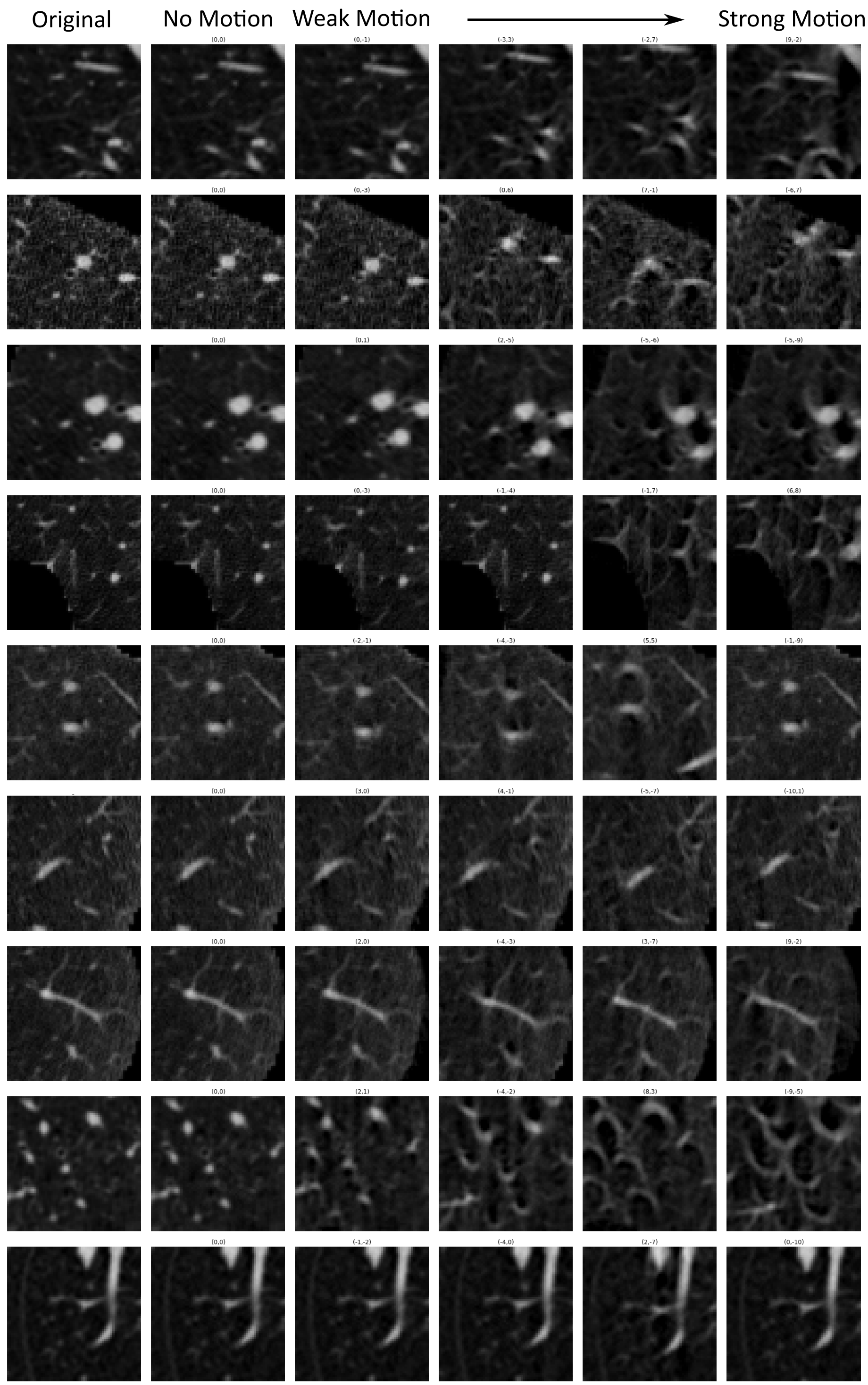



**Fig. 6.** Further examples of artificial motion artifacts generated with the proposed method: The first column shows the center patch of the original input image. The second column shows the ASTRA reconstruction without any motion. Columns 3 to 6 show ASTRA reconstructions with increasing motion severity. The motion vector is indicated above the patch. All patches received the same post-processing as described in the paper.

***Acknowledgments.*** G. van Tulder was financially supported by EFRO/OP-Oost (PROJ-00887).

## References

1. Armato III, S.G., McLennan, G., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical physics **38**(2), 915–931 (2011)
2. Barrett, J.F., Keat, N.: Artifacts in CT: recognition and avoidance. Radiographics **24**(6), 1679–1691 (2004)
3. Beri, P.: Detection of motion artifacts in thoracic CT scans (2020)
4. Bian, Z., Charbonnier, J.P., Liu, J., Zhao, D., Lynch, D.A., van Ginneken, B.: Small airway segmentation in thoracic computed tomography scans: A machine learning approach. Physics in Medicine & Biology **63**(15), 155024 (2018)
5. Cohen, J.P., Luck, M., Honari, S.: Distribution matching losses can hallucinate features in medical image translation. In: International conference on medical image computing and computer-assisted intervention. pp. 529–536. Springer (2018)
6. Elss, T., Nickisch, H., Wissel, T., Bippus, R., Morlock, M., Grass, M.: Motion estimation in coronary CT angiography images using convolutional neural networks. MIDL (2018)
7. Elss, T., Nickisch, H., Wissel, T., Schmitt, H., Vembar, M., Morlock, M., Grass, M.: Deep-learning-based CT motion artifact recognition in coronary arteries. In: Medical Imaging 2018: Image Processing. vol. 10574, p. 1057416. (2018)
8. Fleiss, J.L., Levin, B., Paik, M.C.: Statistical methods for rates and proportions. John Wiley & Sons (2013)
9. Jiang, C., Zhang, Q., Ge, Y., Liang, D., Yang, Y., Liu, X., Zheng, H., Hu, Z.: Wasserstein generative adversarial networks for motion artifact removal in dental CT imaging. In: Medical Imaging 2019: Physics of Medical Imaging. vol. 10948, p. 1094836. International Society for Optics and Photonics (2019)
10. Kim, D., Choi, J., Lee, D., Kim, H., Jung, J., Cho, M., Lee, K.Y.: Motion correction for routine X-ray lung CT imaging. Scientific reports **11**(1) (2021)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4), 640651 (2014)
12. Lossau, T., Nickisch, H., Wissel, T., Bippus, R., Schmitt, H., Morlock, M., Grass, M.: Motion artifact recognition and quantification in coronary CT angiography using convolutional neural networks. Medical image analysis **52**, 68–79 (2019)
13. Persson, G.F., Nygaard, D.E., af Rosenschöld, P.M., Vogelius, I.R., Josipovic, M., Specht, L., Korreman, S.S.: Artifacts in conventional computed tomography (CT) and free breathing four-dimensional CT induce uncertainty in gross tumor volume determination. International Journal of Radiation Oncology-Biology-Physics **80**(5), 1573–1580 (2011)
14. Rad, Z.I., Peyvandi, R.G., Heshmati, R.: Motion detection in CT images with a novel fast technique. Instruments and Experimental Techniques **56**(3), 276–282 (2013)
15. Rohkohl, C., Bruder, H., Stierstorfer, K., Flohr, T.: Improving best-phase image quality in cardiac CT by motion correction with mam optimization. Medical physics **40**(3), 031901 (2013)

16. Schofield, R., King, L., Tayal, U., Castellano, I., Stirrup, J., Pontana, F., Earls, J., Nicol, E.: Image reconstruction: Part 1–understanding filtered back projection, noise and image acquisition. Journal of cardiovascular computed tomography **14**(3), 219–225 (2020)
17. Zhao, B., Schwartz, L.H., Kris, M.G.: Data From RIDER Lung CT. The Cancer Imaging Archive. (2015) https://doi.org/10.7937/K9/TCIA.2015.U1X8A5NR
18. Shakouri, S., Bakhshali, M.A., Layegh, P., et al.: Covid19-CT-dataset: an open-access chest CT image repository of 1000+ patients with confirmed covid-19 diagnosis. BMC Research Notes **14**(1), (2021)
19. Su, K., Zhou, E., Sun, X., Wang, C., Yu, D., Luo, X.: Pre-trained StyleGan based data augmentation for small sample brain CT motion artifacts detection. In: International Conference on Advanced Data Mining and Applications. (2020)
20. Sun, X., Huang, F., Lai, G., Yu, D., Zhang, B., Guo, B., Ma, Z.: Motion artifacts detection from computed tomography images. In: International Conference on Advanced Data Mining and Applications. pp. 347–359. Springer (2020)
21. Van Aarle, W., Palenstijn, W.J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K.J., Sijbers, J.: The ASTRA toolbox: A platform for advanced algorithm development in electron tomography. Ultramicroscopy **157**, 35–47 (2015)
22. Van Stevendaal, U., Von Berg, J., Lorenz, C., Grass, M.: A motion-compensated scheme for helical cone-beam reconstruction in cardiac CT angiography. Medical physics **35**(7Part1), 3239–3251 (2008)
23. Van Tulder, G.: elasticdeform: Elastic deformations for n-dimensional images. https://doi.org/10.5281/zenodo.4569691 (2021), accessed: 2022-03-01
24. Wang, C., Sun, X., Zhang, B., Lai, G., Yu, D., Su, K.: Brain CT image with motion artifact augmentation based on PGGAN and FBP for artifact detection. In: International Conference on Advanced Data Mining and Applications. pp. 370–378. Springer (2020)