ZERO-SHOT MEDICAL IMAGE ARTIFACT REDUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Medical images may contain various types of artifacts with different patterns and mixtures, which depend on many factors such as scan setting, machine condition, patients' characteristics, surrounding environment, etc. However, existing deep learning based artifact reduction methods are restricted by their training set with specific predetermined artifact type and pattern. As such, they have limited clinical adoption. In this paper, we introduce a "Zero-Shot" medical image Artifact Reduction (ZSAR) framework, which leverages the power of deep learning but without using general pre-trained networks or any clean image reference. Specifically, we utilize the low internal visual entropy of an image and train a light-weight image-specific artifact reduction network to reduce artifacts in an image at test-time. We use Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) as vehicles to show that ZSAR can reduce artifacts better than state-of-the-art both qualitatively and quantitatively, while using shorter execution time. To the best of our knowledge, this is the first deep learning framework that reduces artifacts in medical images without using *a priori* training set.

1 INTRODUCTION

Deep learning has demonstrated its great power in artifact reduction, a fundamental task in medical image analysis to produce clean images for clinical diagnosis, decision making, and accurate quantitative image analysis. Existing deep learning based frameworks (Yang et al., 2018; Jiang et al., 2018; Yuan et al., 2019; Yi & Babyn, 2018) use training data sets that contain paired images (same images with and without artifacts) to learn the artifact features. Simulations are often needed to generate the data set for these methods, which may be different from clinical situations and lead to biased learning (Yang et al., 2018; Veraart et al., 2016). To address this issue, Kang et al. (2018) used cycle-consistent adversarial denoising network (CCADN) which no longer requires paired data.

However, all these methods still suffer from two mainstays: First, they require clean image references, which can be hard to obtain clinically. For example, motion artifacts in Magnetic Resonance Imaging (MRI) are almost always present due to the lengthy acquisition process (Zaitsev et al., 2015). In such situations, simulation is still the only way to generate the data set. Second, although the trained networks outperform non-learning based algorithms such as Block Matching 3D (BM3D) (Dabov et al., 2007), they can only be applied to scenarios where the artifacts resemble what are in the training set, lacking the versatility that non-learning based methods can offer.

To attain the performance of deep learning based methods and the versatility of non-learning based ones, we introduce a "Zero-Shot" image-specific artifact reduction network (ZSAR), which builds upon deep learning yet does not require any clean image reference or *a priori* training data. Based on the key observation that most medical images have areas that contain artifacts on a relatively uniform background and that the internal visual entropy of an image is much lower than that among images (Zontak & Irani, 2011). At test-time, ZSAR extracts an artifact pattern directly and synthesizes paired image patches from input image to iteratively train a light-weight image-specific autoencoder for artifact reduction. Experimental results on clinical CT and MRI data with a variety of artifacts show that it outperforms the state-of-the-art methods both qualitatively and quantitatively, using shorter execution time. To the best of our knowledge, ZSAR is the first deep learning based method that reduces artifacts in medical images without *a priori* training data.

2 Related Works

We limit our discussion to CT and MRI as they are the vehicles to demonstrate the efficacy of our method. For CT, artifacts can be classified into patient-based (e.g., motion artifact), physics-based (e.g., noise, metal artifact), and helical and multi-channel artifacts (e.g., cone beam effect) according to the underlying cause (Boas & Fleischmann, 2012). For MRI, the types include truncation artifacts, motion artifacts, aliasing artifacts, Gibbs ringing artifacts, and others (Krupa & Bekiesińska-Figatowska, 2015). These artifacts are influenced by a number of factors, including scan setting, machine condition, patient size and age, surrounding environment, etc. These artifacts may occur at random places in an image. In addition, an image can contain multiple artifacts and mixture simultaneously. We also limit our discussion to medical image specific deep learning based methods only, although some general-purpose denoising methods such as Deep image prior (DIP) Ulyanov et al. (2018) and non deep-learning based methods such as BM3D can also be readily applied but inferior in this problem.

For noise and artifacts reduction, Chen et al. (2017) proposed a convolution neural network (CNN) to denoise low-dose CT images and reconstruct the corresponding routine-dose CT images. Kang et al. (2017) proposed a directional wavelet transform on the CNN to suppress CT-specific noise. Wolterink et al. (2017) designed a generative adversarial network (GAN) with CNN. Yang et al. (2018) adopted Wasserstein distance and perceptual loss to ensure the similarity between input and the generated image. Manjón & Coupe (2018) proposed a simple CNN network for 2D MRI artifact reduction and Jiang et al. (2018) explored multi-channel CNN for 3D MRI. The use of deep learning to reduce other artifacts has also been explored (Gjesteby et al., 2017; Zhang & Yu, 2018). For these methods, simulations are often used to generate the paired data, which may lead to biased learning when simulated artifacts differ from the real ones (Kang et al., 2018). To address this issue, Kang et al. (2018); You et al. (2018) used cycle-consistent adversarial denoising network that learns the mapping between the low and routine dose cardiac phase without matched image pairs. Yet this method still requires clean image reference, which may be hard to obtain in some situations.

3 Methods

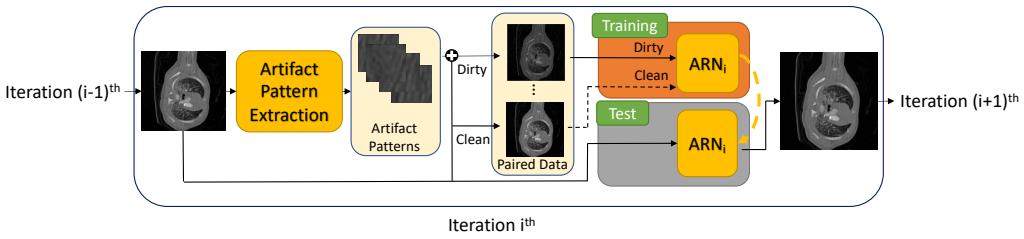


Figure 1: The overall structure of ZSAR composed of Artifact Pattern Extraction and Artifact Reduction Network (ARN). Note that we treat the original input image as the output of "0^{th}" iteration.

3.1 Overview

The main motivation of our work lies behind the fact that it is almost always possible to identify small regions of interests where significant artifacts exist over a relatively uniform background in any medical images. As such, it is possible to synthesize the paired dirty-clean patches from the exact image with artifacts to be reduced.

The overall framework of the proposed ZSAR is shown in Fig. 1, which is an iterative process. The framework works with 2D images, so 3D volumes are sliced first, similar to many existing works (Yang et al., 2018). For clarity, we call the phase where the model is trained to obtain the weights as "training", and the phase that applies the trained model to the input image to reduce artifacts as "test". Note that both phases are done on the spot for each specific input image and no pre-training is conducted.

For every iteration, ZSAR first extracts artifact patterns and synthesizes the paired dirty and clean images using the patterns (the details will be explained in Section 3.2). Note that the artifact pattern

extraction in the 1^{st} iteration is different from those in the subsequent $(i+1)^{th}$ iterations ($i \geq 1$). Later, the synthesized image is then used to train a light-weight artifact reduction network (ARN), which is used to reduce the artifact in input image (the details will be explained in Section 3.3). We terminate the iterative process when the artifact level doesn't decrease. Our experiments show that the number of iteration needed is usally not more than four (the details will be explained in Section 4.3).

For a 3D volume, we only perform the above training process for one 2D slice of it. Since the artifacts are usually similar across all slices in the same 3D volume, the remaining slices can directly use the trained ARNs to iteratively reduce the artifact (by following the test path only in Fig. 1).

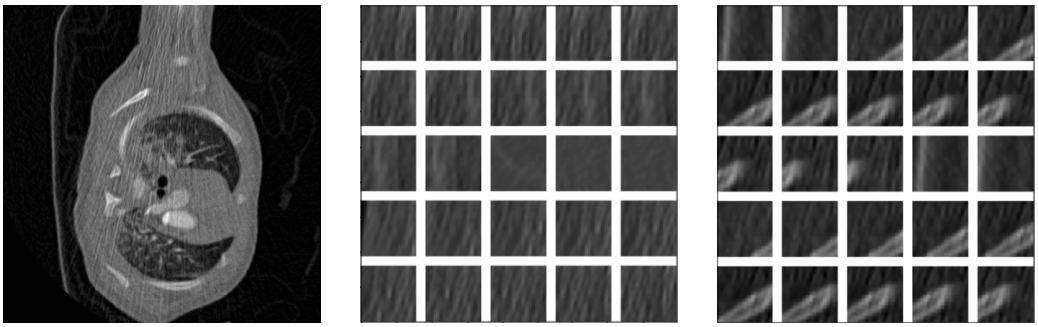
3.2 Artifact Pattern Extraction and Training Data Synthesis

For the 1^{st} iteration, since no clean reference image is provided, we extract the artifact pattern from the input image itself through an unsupervised approach. This is made possible based on the fact that for most artifacts in medical images, we can always identify areas where only artifacts exist (Boas & Fleischmann, 2012; Krupa & Bekiesińska-Figatowska, 2015). As such, we need to identify areas where the background is relatively uniform yet significant artifacts are present.

Towards this, we first crop the input image into patches with size 32×32. After that, an unsupervised clustering method, K-means (Fahim et al., 2006) is applied. The main idea is to classify the patches into two clusters, one containing patches without structure boundaries (i.e., relatively uniform background), and the other containing patches with structure boundaries. Such a classification is possible as the patches in these two clusters will exhibit significant differences in terms of standard distributions of the pixel values: when structure boundaries are present, significant mean shift and large yet localized variations in pixel values can be observed. The feature of each patch is thus extracted as follows: the overall standard deviation of all the pixel values in the patch, and the mean value of all standard deviations extracted by a 8×8 sliding window. Fig. 2 shows an example of the clustering process. It can be clearly seen that one of the clusters contains patches with only uniform backgrounds (either with or without artifacts), while the other one contains all the patches with structure boundaries. As the patches in the former cluster always contain relatively uniform background, a zero-mean artifact pattern can be extracted by subtracting the mean pixel value of each patch. Note that in the patches without artifacts, the pattern extracted will just be empty. On the other hand, as long as some of the patches contain artifacts, our framework can utilize them to further synthesize the training data, which will be discussed later.

In the subsequent $(i+1)^{th}$ iteration ($i \geq 1$), we observed that the difference between the clean image and the output image of $(i)^{th}$ iteration can be seen as the reduced artifact. The zero-mean artifact pattern is again generated by subtracting the mean pixel value of the difference.

To reflect artifacts of different intensities, we randomly scale each pattern following standard normal distribution. Those scaled artifact patterns are then superposed to random areas in the input image to form dirty images, and we use the input image as the corresponding clean image so that paired dirty-clean data set is formed. Note that this synthesis process is conducted in every iteration.



(a) Input image (b) Artifact cluster (c) Other cluster

Figure 2: An input image and examples of the two clusters after K-means is applied to the patches.

3.3 ARTIFACT REDUCTION NETWORK

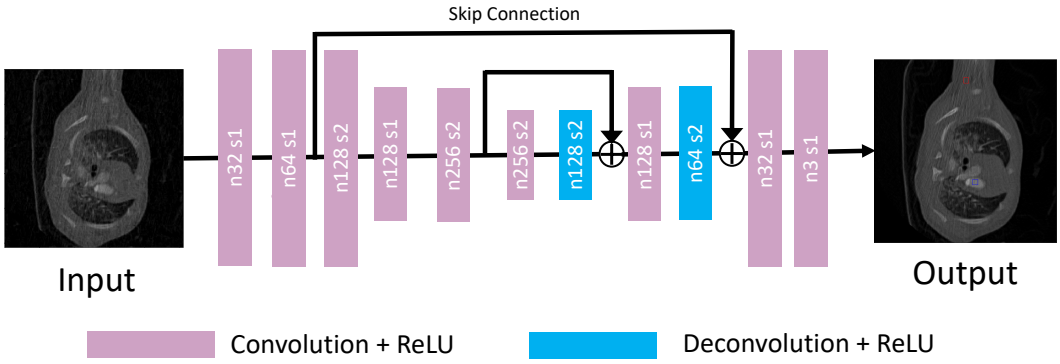


Figure 3: Artifact Reduction Network (ARN) architecture. Note that n and s of each layer stand for the number of kernels and strides, respectively.

After the paired data is synthesized, it can be used to train any existing neural networks for artifact reduction. Considering the need of test-time training, we design a compacted network as shown in Fig. 3, which is formed by a 11-layer contextual autoencoder to reduce artifacts and restore the structural information. With the skip connection, these decoder layers can capture more contextual information extracted from different encoder layers. With such a light-weight network structure, it requires only a few epochs to converge. The pixel-wise mean square error (MSE) is used as the loss function to preserve structural and substance information:

$$Loss = L_{MSE}(O, G) \tag{1}$$

where O and G are the output of the contextual autoencoder and the clean image reference, respectively.

Through experiments, we find that ARN should be initialized and retrained in every iteration, which is more effective than incremental training based on the network from previous iterations. This is because each iteration is essentially a new artifact reduction procedure and the model only needs to learn the artifact level in the input of the current iteration. Also, in each iteration, only a single pair of images are used for training due to speed consideration. Since essentially, the same image is used during training and test, there is no overfitting concern.

4 EXPERIMENTS AND RESULTS

4.1 CARDIAC DATA SET AND EVALUATION METRICS

Our CT data set is a collection of 48 3D cardiac CT volumes from 24 patients, which consists of 11,616 gray-scale 2D images (512x512) acquired through a wide detector 256-slice multiple detector computed tomography scanner. The dosages are set between 80 kVp/55 mAs (low-dose) and 100 kVp/651 mAs (routine-dose). Our MRI data set contains 286 pulse sequences as 17,844 2D MRI images from 11 patients, scanned by a 3T system. Along with long-axis planes, a stack of short-axis single-shot balanced standard steady-state in free-precession sequence images from apex to basal were collected.

All CT and MRI images are qualitatively evaluated by our radiologists on structural preservation and artifact level. For quantitative evaluation, due to the lack of ground truth, for CT, we follow existing work (Wolterink et al., 2017; Yang et al., 2018) and select the most homogeneous area in regions of interest selected by our radiologists. The standard deviation (artifact level) of the pixels in the area should be as low as possible, and the mean (substance information) discrepancy after artifact reduction should not be too large to cause information loss. For MRI, in addition to the mean of the pixel values in the most homogeneous areas, similar to Goerner & Clarke (2011); Kellman & McVeigh (2005); Rajalakshmi et al. (2018) we divide the mean by the standard deviation of the pixel values in the area and use the resulting Signal-to-Noise ratio (SNR) as the metric.

4.2 EXPERIMENTAL SETUP

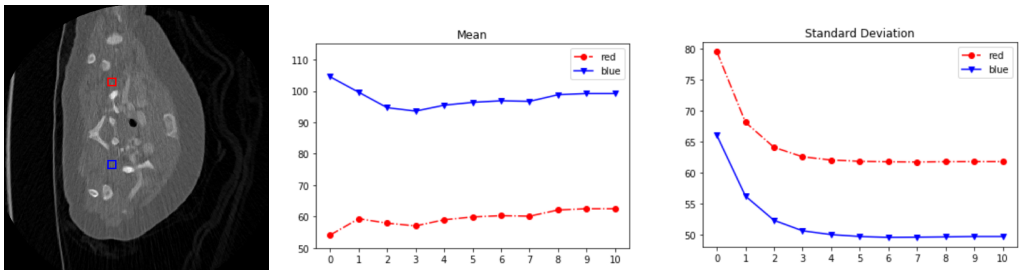
ZSAR was implemented in Python3 with TensorFlow library. NVIDIA GeForce GTX 1080 Ti GPU was used to train and test the networks. For every convolution and deconvolution layer, Xavier initialization (Glorot & Bengio, 2010) was used for the kernels and the filter size is set to 3 and 4, respectively. Adam optimization (Kingma & Ba, 2014) method was applied to train ARN by setting learning rate as 0.0005. Training phase was performed by minimizing loss function with the number of epoch set to 1,000.

We compare ZSAR with CCADN, a state-of-the-art deep learning based method for medical image artifact reduction (Kang et al., 2018), which unlike other deep learning based method does not require paired training data. We train it with exactly the same settings as used in (Kang et al., 2018) for the data set generated through training data synthesis algorithm (excluding 10% of the images used for test). We would like to emphasize again that although ZSAR is also based on deep learning, it does not require any prior training data. We also compare ZSAR with Deep image prior (DIP) (Ulyanov et al., 2018), a state-of-the-art general-purpose denoising method, and a non-learning based algorithm BM3D. For DIP, we follow the setting recommended in the paper. For each image, we tuned the parameters in BM3D to attain the best quality.

Due to the space limit, in the paper we only present comparisons between the four methods using a limited number of test images. Additional results can be found in the appendix. Similar conclusions are drawn there.

4.3 EFFICIENCY OF THE ITERATIVE PROCESS

We use the input image as shown in Fig. 4 (a) to study the efficiency of the iterative process in ZSAR. In the image, our radiologists selected the most homogeneous areas inside the regions marked with red and blue rectangles for quantitative evaluation. The resulting mean (substance information) and standard deviation (artifact level) are shown in Fig. 4 (b) and (c), respectively. Note that the"0^{th}" iteration contains the data for the input image in Fig. 4 (a). From Fig. 4 (b) we can see that as the number of iterations increase, small fluctuations in mean can be observed for both regions. On the other hand, in Fig. 4 (c), the standard deviation of both regions stops to decrease around the 4^{th} iteration. This experiment demonstrates the efficiency of the iterative process.



(a) (b) (c)

Figure 4: (b) Mean and (c) standard deviation vs. iteration number for the red and blue regions in CT image (a).

4.4 COMPARISONS WITH THE STATE-OF-THE-ART

We start our experiments with the ideal scenario where the artifacts in both training set of CCADN and test CT images contain Poisson noise only. The qualitative results for CCADN, BM3D, DIP, and ZSAR are shown in Fig. 5 (a). In the figure, we found CCADN, BM3D, and ZSAR can all preserve structure well, while DIP is oversmoothed visually. Our radiologists then selected the largest homogeneous areas inside the regions marked with red and blue rectangles for quantitative comparison, and the results are summarized in Table 1 (a). From the table, it shows that ZSAR achieves results comparable with that of CCADN, better than BM3D. This is expected as CCADN is trained on the same type of images and artifact patterns, so it is as effective as ZSAR in this ideal

		Original	ZSAR	CCADN	BM3D	DIP
(a) Red	Mean	54.1	61.9	68.2	48.8	161.7
	SD	79.5	62.9	65.8	72.4	23.7
(a) Blue	Mean	104.4	98.2	118.5	99.3	214.5
	SD	66.0	51.3	48.9	53.4	6.8
(b) Red	Mean	27.4	36.9	78.3	21.9	213.7
	SD	95.7	75.9	81.5	85.6	30.9
(b) Blue	Mean	63.3	83.5	115.8	57.3	271.0
	SD	138.6	127.0	130.9	132.8	27.8
(c) Red	Mean	54.0	53.8	88.8	48.6	158.7
	SD	95.2	65.2	80.1	86.2	47.7
(c) Blue	Mean	1603.9	1537.5	1661.9	1596.9	1831.7
	SD	229.1	197.0	208.2	226.3	156.3

Table 1: Mean and standard deviation (SD) for the largest homogeneous areas inside the marked regions of the CT images in Fig. 5 (in Hounsfield Unit).

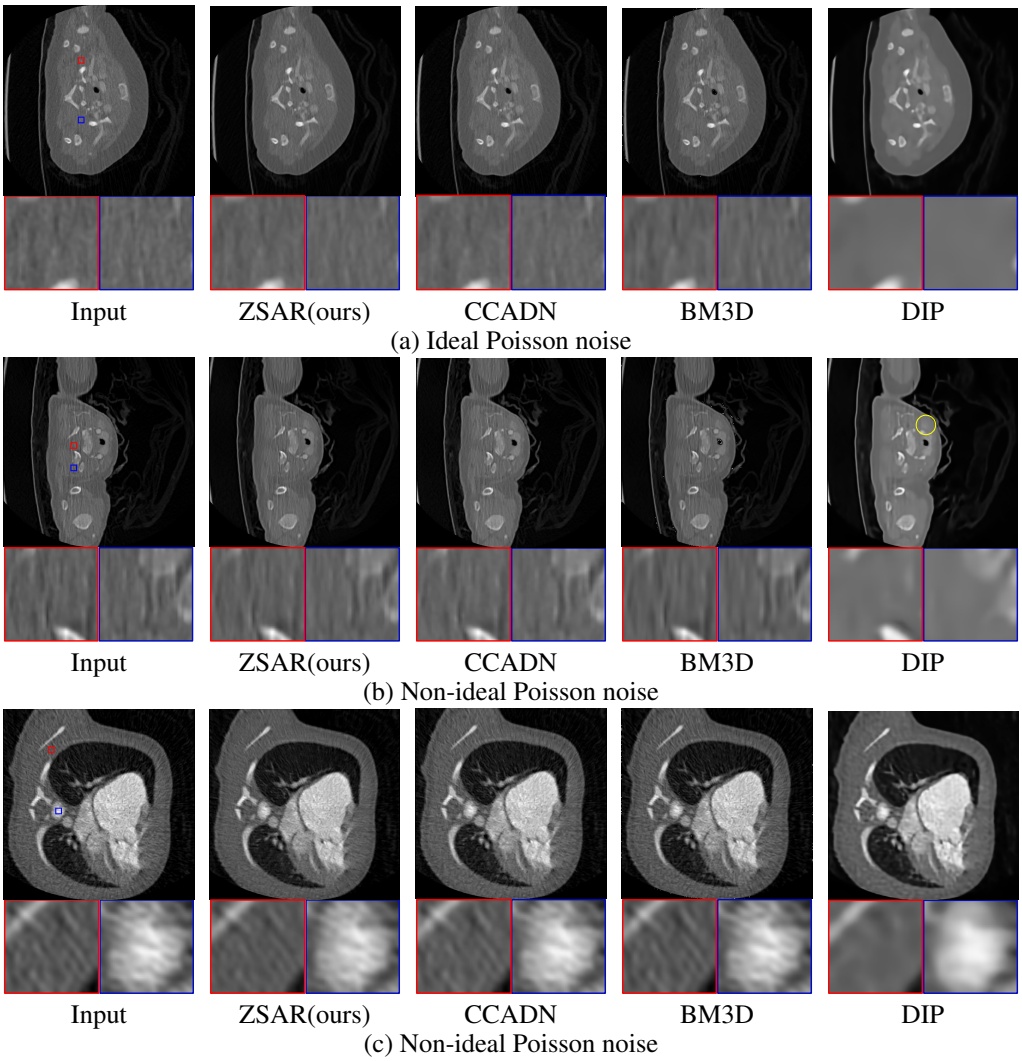


Figure 5: Artifact reduction in cardiac CT images. Quantitative comparisons for the largest homogeneous areas inside the marked regions can be found in Table 1.

scenario. Although DIP seems to reduce artifact the most effectively, it yields over 200% mean discrepancy, which is a critical problem for CT images.

We further applied the four methods to MRI motion artifact reduction in the ideal scenario that both training set of CCADN and the test images contain motion artifact pattern only. The qualitative results are shown in Fig. 6 (a). All the methods preserve structures well, and ZSAR leads to the best motion artifact reduction. The corresponding statistics for the largest homogeneous areas inside the marked regions are reported in Table 3 (a). From the table, though CCADN has the largest SNR, it suffers from large mean discrepancy, which again can be problematic. ZSAR achieves up to 50% higher SNR than BM3D. When comparing with DIP, ZSAR achieves similar SNR but less mean discrepancy.

Next, we studied the non-ideal scenario where different artifact patterns or noise level of artifacts absent from the training set of CCADN appeared in the test image. For CT denoising, the qualitative results are shown in Fig. 5 (b-c) and the corresponding mean and standard deviation numbers are presented in Table 1 (b-c). The results for MRI with different artifact patterns are shown in Fig. 6 (b-c) and Table 3 (b-c), respectively. Qualitatively, we can see that ZSAR outperforms CCADN and BM3D, while all the three methods preserve the structures well. On the other hand, DIP has oversmoothing issue in all the non-ideal cases of CT images, which causes the disappearance of tissue (region highlighted using yellow circle in case (b)). Quantitatively, for CT images, ZSAR beats CCADN and BM3D in all the four cases, achieving up to 19% and 25% lower standard deviation, respectively. For MRI, ZSAR attains up to 77% and 74% higher SNR compared with CCADN and BM3D, respectively. In addition, since CCADN was trained in different scenario, in the region marked with red in case (c), it obtains SNR even smaller than the input image. This clearly demonstrates the advantage of ZSAR, which does not rely on and thus is not limited to artifacts in a training set. Compared with DIP, ZSAR achieves lower SNR on region marked with red in case (b) and region marked with blue in case (c). However, this is in fact due to the large mean discrepancy in those two cases from DIP, which is not acceptable.

Finally, to show that test-time training is feasible, we compare the average execution time of ZSAR (which include both training and test) with CCADN (which only include test), BM3D, and DIP on the 3D CT and MRI images above. The results are shown in Table 2. From the table, ZSAR requires less time than the other three methods despite the fact that it is trained on the spot for each input image. The fast speed of ZSAR is brought by two factors: 1) Its training usually converges within four iterations, and with few training data. Each iteration only takes about 1,000 epochs to converge. 2) It is much simpler than CCADN or DIP in structure and thus takes less time to test each 2D slice of the 3D images.

	ZSAR	CCADN	BM3D	DIP
	Training+Test	Test	-	-
CT (484 slices)	305+111	3533	1825	32641
MRI (360 slices)	134+58	1742	716	22308

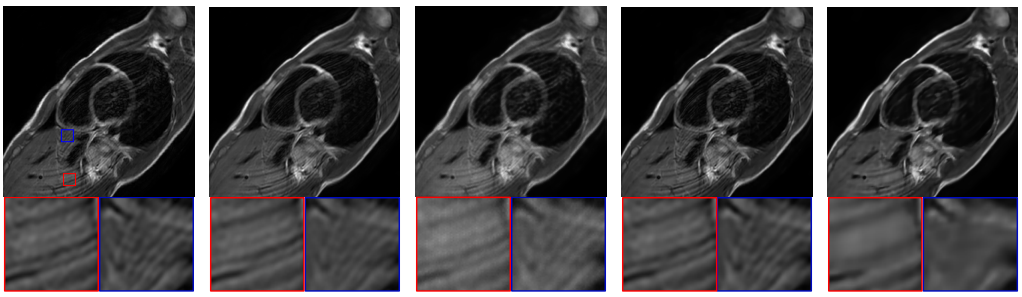
Table 2: Execution time comparison between ZSAR and the three methods, CCADN, BM3D, and DIP for 3D cardiac CT (512×512) and MRI images (320×320) (in sec.).

5 CONCLUSION

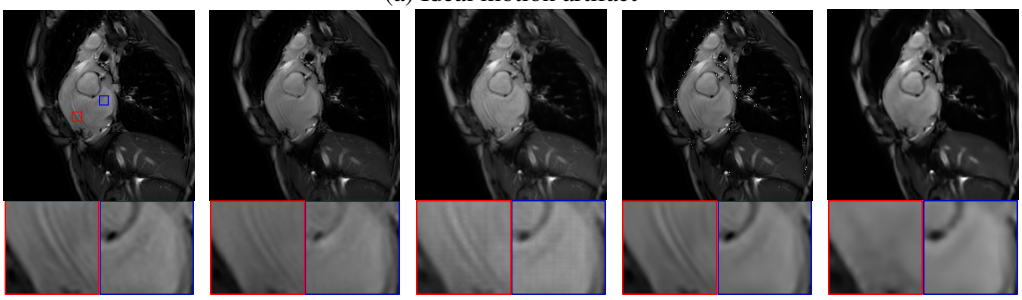
In this paper, we introduced ZSAR, a "Zero-Shot" medical image artifact reduction framework, which exploits the power of deep learning to suppress artifacts in a medical image without using general pre-trained networks. Unlike previous state-of-the-art methods which are restricted by the training data, our method can be adapted for almost any medical images that contain varying or unknown artifacts. Experimental results on cardiac CT and MRI images have shown that our framework can reduce noises and motion artifacts qualitatively and quantitatively better than the state-of-the-art, using shorter execution time.

		Original	ZSAR	CCADN	BM3D	DIP
(a) Red	Mean	477.3	494.9	627.4	476.1	509.3
	SNR	6.2	7.1	9.0	6.2	6.9
(a) Blue	Mean	381.9	396.1	495.4	380.3	404.4
	SNR	7.6	11.7	12.3	7.8	11.8
(b) Red	Mean	994.4	969.0	1280.4	989	1089.5
	SNR	9.5	15.6	12.8	9.8	18.7
(b) Blue	Mean	1277.0	1279.8	1503.2	1272.2	1433.0
	SNR	16.3	25.9	21.7	17.8	18.7
(c) Red	Mean	1265.0	1123.8	1204.3	1260.7	1333.7
	SNR	13.7	17.9	13.0	14.2	15.8
(c) Blue	Mean	1329.4	1206.9	1302.9	1325.8	1373.3
	SNR	9.8	17.6	9.9	10.1	20.5

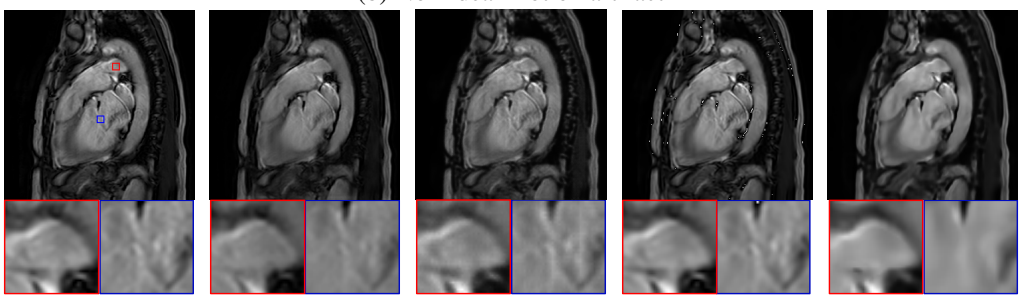
Table 3: Mean and SNR (Signal-to-Noise Ratio) for the largest homogeneous areas inside the marked regions of the MRI images in Fig. 6.



(a) Ideal motion artifact



(b) Non-ideal motion artifact



(c) Non-ideal motion artifact

Figure 6: Comparison using MRI test images under (a) ideal and both (b) and (c) non-ideal scenarios. Both contain motion artifacts but the pattern in (a) appeared in the training set and but not in (b) and (c).

REFERENCES

F Edward Boas and Dominik Fleischmann. Ct artifacts: causes and reduction techniques. *Imaging Med*, 4(2):229–240, 2012.

Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12):2524–2535, 2017.

Kostadin Dabov, Alessandro Foi, and Karen Egiazarian. Video denoising by sparse 3d transform-domain collaborative filtering. In *2007 15th European Signal Processing Conference*, pp. 145–149. IEEE, 2007.

AM Fahim, AM Salem, F Af Torkey, and MA Ramadan. An efficient enhanced k-means clustering algorithm. *Journal of Zhejiang University-Science A*, 7(10):1626–1633, 2006.

Lars Gjesteby, Qingsong Yang, Yan Xi, Bernhard Claus, Yannan Jin, Bruno De Man, and Ge Wang. Reducing metal streak artifacts in ct images via deep learning: Pilot results. In *The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pp. 611–614, 2017.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, 2010.

Frank L Goerner and Geoffrey D Clarke. Measuring signal-to-noise ratio in partially parallel imaging mri. *Medical physics*, 38(9):5049–5057, 2011.

Dongsheng Jiang, Weiqiang Dou, Luc Vosters, Xiayu Xu, Yue Sun, and Tao Tan. Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network. *Japanese journal of radiology*, 36(9):566–574, 2018.

Eunhee Kang, Junhong Min, and Jong Chul Ye. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *Medical physics*, 44(10), 2017.

Eunhee Kang, Hyun Jung Koo, Dong Hyun Yang, Joon Bum Seo, and Jong Chul Ye. Cycle consistent adversarial denoising network for multiphase coronary ct angiography. *arXiv preprint arXiv:1806.09748*, 2018.

Peter Kellman and Elliot R McVeigh. Image reconstruction in snr units: a general method for snr measurement. *Magnetic resonance in medicine*, 54(6):1439–1447, 2005.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Katarzyna Krupa and Monika Bekiesińska-Figatowska. Artifacts in magnetic resonance imaging. *Polish journal of radiology*, 80:93, 2015.

José V Manjón and Pierrick Coupe. Mri denoising using deep learning. In *International Workshop on Patch-based Techniques in Medical Imaging*, pp. 12–19. Springer, 2018.

N Rajalakshmi, K Narayanan, and P Amudhavalli. Wavelet based weighted median filter for image denoising of mri brain images. *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, 10(1):201–206, 2018.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

Jelle Veraart, Dmitry S Novikov, Daan Christiaens, Benjamin Ades-Aron, Jan Sijbers, and Els Fieremans. Denoising of diffusion mri using random matrix theory. *NeuroImage*, 142:394–406, 2016.

Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose ct. *IEEE transactions on medical imaging*, 36(12):2536–2545, 2017.

Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 2018.

Xin Yi and Paul Babyn. Sharpness-aware low-dose ct denoising using conditional generative adversarial network. *Journal of digital imaging*, 31(5):655–669, 2018.

Chenyu You, Yi Zhang, Xiaoliu Zhang, Guang Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, Punam K Saha, and Ge Wang. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *arXiv preprint arXiv:1808.04256*, 2018.

Nimu Yuan, Jian Zhou, and Jinyi Qi. Low-dose ct image denoising without high-dose reference images. In *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, volume 11072, pp. 110721C. International Society for Optics and Photonics, 2019.

Maxim Zaitsev, Julian Maclaren, and Michael Herbst. Motion artifacts in mri: a complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4):887–901, 2015.

Yanbo Zhang and Hengyong Yu. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE transactions on medical imaging*, 37(6):1370–1381, 2018.

Maria Zontak and Michal Irani. Internal statistics of a single natural image. In *CVPR 2011*, pp. 977–984. IEEE, 2011.

A Appendix

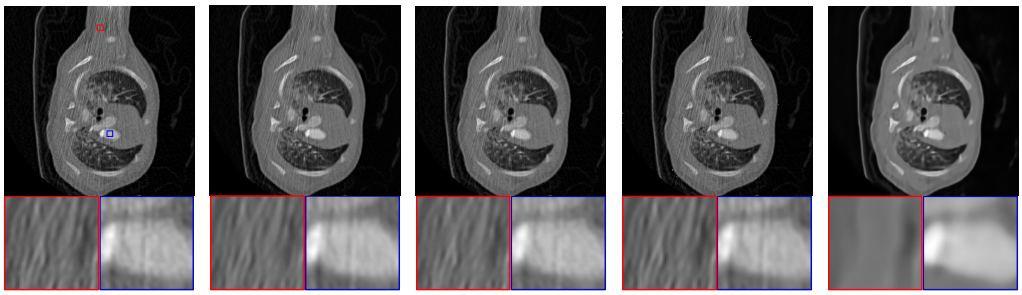
In this section, we expand our results in Section 4 of the paper, and show more comparison between ZSAR and the state-of-the-art deep learning based methods for medical image artifact reduction, CCADN, a general-purpose deep learning based denoising framework, DIP, and the non-learning based method, BM3D.

For CT denoising, the qualitative and quantitative results are shown in Fig. 7 and Table 4, respectively. Qualitatively, we can see that ZSAR outperforms CCADN and BM3D, while all the three methods preserve the structures well. On the other hand, DIP has oversmoothing issue in all those cases of CT images. Quantitatively, for CT images, ZSAR beats CCADN and BM3D in all the four cases, achieving up to 14% and 22% lower standard deviation, respectively.

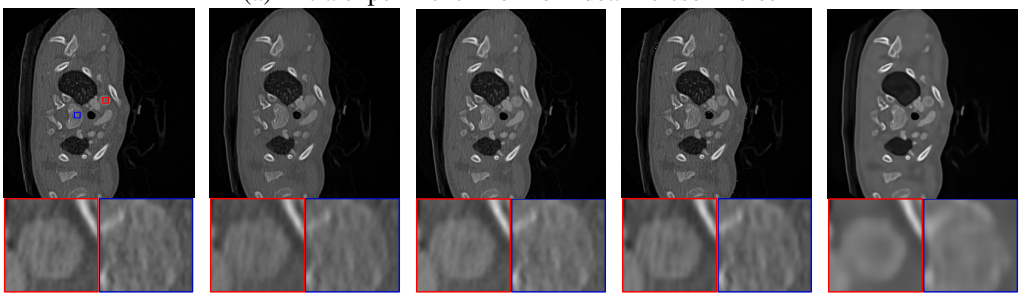
For MRI, the results are shown in Fig. 8 and Table 5, respectively. From the table, ZSAR attains up to 25% and 52% higher SNR compared with CCADN and BM3D, respectively. Compared with DIP, ZSAR achieves higher SNR in most cases, except the region marked with red in case (a). However, in this case, DIP obtains a larger mean discrepancy than ZSAR, which is not acceptable.

		Original	ZSAR	CCADN	BM3D	DIP
(a) Red	Mean	82.9	95.8	96.8	78.5	169.3
	SD	146.3	112.4	129.3	143.1	19.3
(a) Blue	Mean	790.4	827.9	823.9	786.0	970.3
	SD	89.4	73.9	77.7	84.0	67.3
(b) Red	Mean	371.9	312.0	431.0	365.8	429.1
	SD	105.7	79.0	89.1	96.7	82.2
(b) Blue	Mean	333.0	300.9	390.8	324.5	458.2
	SD	103.6	83.1	88.4	96.0	53.8

Table 4: Mean and standard deviation (SD) for the largest homogeneous areas inside the marked regions of the CT images in Fig. 7 (in Hounsfield Unit).



(a) Extra experiment 1 for non-ideal Poisson noise

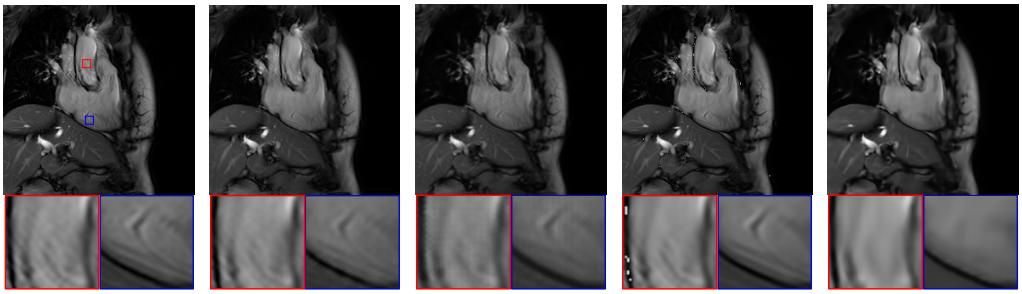


(b) Extra experiment 2 for non-ideal Poisson noise

Figure 7: Artifact reduction in cardiac CT images. Quantitative comparisons for the largest homogeneous areas inside the marked regions can be found in Table 4.

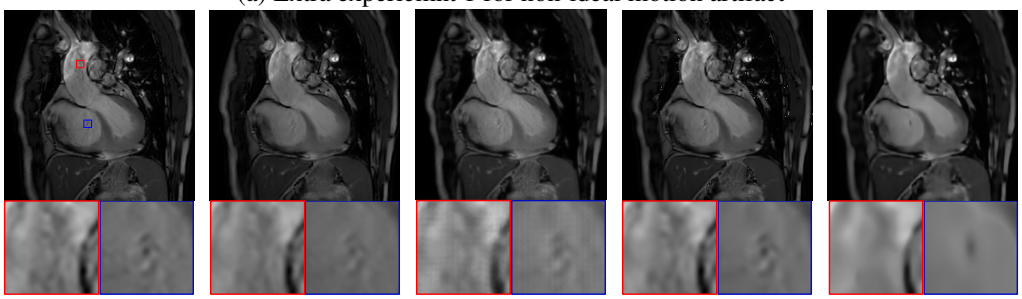
		Original	ZSAR	CCADN	BM3D	DIP
(a) Red	Mean	842.7	830.0	779.5	839.8	873.7
	SNR	17.4	21.1	19.0	19.5	23.3
(a) Blue	Mean	633.8	636.2	592.9	631.0	665.0
	SNR	9.0	13.0	10.4	9.4	12.8
(b) Red	Mean	1067.7	1036.4	1163.3	1063.6	1157.3
	SNR	14.6	18.2	16.2	15.6	17.0
(b) Blue	Mean	971.4	940.3	1093.5	966.8	1091.9
	SNR	9.1	14.2	11.9	9.3	11.7

Table 5: Mean and SNR (Signal-to-Noise Ratio) for the largest homogeneous areas inside the marked regions of the MRI images in Fig. 8.



Input ZSAR(ours) CCADN BM3D DIP

(a) Extra experiemnt 1 for non-ideal motion artifact



Input ZSAR(ours) CCADN BM3D DIP

(b) Extra experiment 2 for non-ideal motion artifact

Figure 8: Comparison using MRI test images. Quantitative comparisons for the largest homogeneous areas inside the marked regions can be found in Table 5.