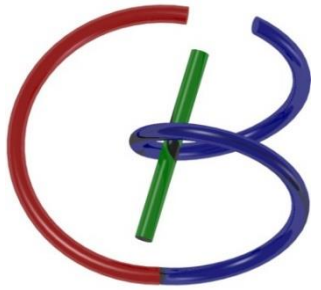


ADN: Artifact Disentanglement Network for Unsupervised Metal Artifact Reduction

Haofu Liao, *Student Member, IEEE,* Wei-An Lin, *Student Member, IEEE,*
S. Kevin Zhou, *Senior Member, IEEE,* and Jiebo Luo, *Fellow, IEEE*

Published in: IEEE Transactions on Medical Imaging (Volume: 39 , Issue: 3 , March 2020)

Chuang Niu
March 18, 2020

Problem—Metal Artifact Reduction

- Compared to body tissues, metallic materials attenuate X-rays significantly and non-uniformly over the spectrum, leading to inconsistent X-ray projections. The mismatched projections will introduce severe streaking and shading artifacts in the reconstructed CT images, which significantly degrade the image quality and compromise the medical image analysis as well as the subsequent healthcare delivery.
- Current deep neural network based approaches to computed tomography (CT) metal artifact reduction (MAR) are supervised methods which require pairs of anatomically identical CT images, one with and the other without metal artifacts , for training.
- It is clinically impractical to obtain such pairs of images, most of the supervised methods rely on synthesized images to train their models.
- The synthesized images may not fully simulate the real clinical scenarios, the performances of these supervised methods may degrade in clinical applications.

Related work

- **Conventional Metal Artifact Reduction**

Most conventional approaches address metal artifacts in X-ray projections.

One way is to directly correct the X-ray measurement of the metallic implants by modeling the underlying physical effects such as beam hardening, scatter and so on.

Another is to replace the corrupted region with estimated values, such as interpolation and prior image based method.

- **Deep metal artifact reduction**

RL-ARCNN, DesteakNet, CNNMAR, cGANMAR, SVAAE-GAN, MPN, DuDoNet.

- **Unsupervised image-to-image translation**

Image artifact reduction can be regarded as a form of image-to-image translation (CycleGAN and its variants). However, these unsupervised methods target at image synthesis and do not have suitable components for artifact reduction. Recent work that is specialized for artifact reduction is deep image prior (DIP), which, however, only works for less structured artifacts such as additive noise or compression artifacts.

Methodology—Artifact Disentanglement

Encoding and decoding between domains and spaces.

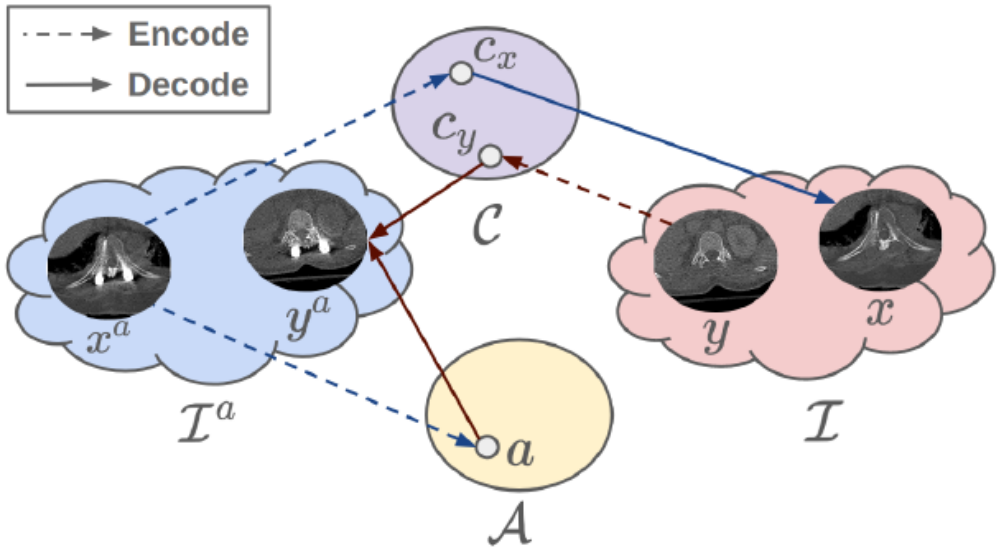


Fig. 1: Artifact disentanglement. The content and artifact components of an image x^a from the artifact-affected domain $\mathcal{I}^a$ is mapped separately to the content space $\mathcal{C}$ and the artifact space $\mathcal{A}$, i.e., artifact disentanglement. An image y from the artifact-free domain $\mathcal{I}$ contains no artifact and thus is mapped only to the content space. Decoding without artifact code removes the artifact from an artifact-affected image (blue arrows $x^a \rightarrow x$) while decoding with the artifact code adds artifacts to an artifact-free image (red arrows $y \rightarrow y^a$).

Let $\mathcal{I}^a$ be the domain of all artifact-affected CT images and $\mathcal{I}$ be the domain of all artifact-free CT images. We denote $\mathcal{P} = \{(x^a, x) \mid x^a \in \mathcal{I}^a, x \in \mathcal{I}, f(x^a) = x\}$ as a set of paired images, where $f : \mathcal{I}^a \rightarrow \mathcal{I}$ is an MAR model that removes the metal artifacts from x. In this work, we assume no such paired dataset is available and we propose to learn f with unpaired images.

As illustrated in Fig. 1, the proposed method disentangles the artifact and content components of an artifact-affected image x^a by encoding them separately into a content space $\mathcal{C}$ and an artifact space $\mathcal{A}$. If the disentanglement is well addressed, the encoded content component $c_x \in \mathcal{C}$ should contain no information about the artifact while preserving all the content information. Thus, decoding from c_x should give an artifact-free image x which is the artifact-removed counterpart of x^a. On the other hand, it is also possible to encode an artifact-free image y into the content space which gives a content code c_y. If c_y is decoded together with an artifact code $a \in \mathcal{A}$, we obtain an artifact-affected image y^a. In the following sections, we introduce an artifact disentanglement network (ADN) that learns these encodings and decodings without paired data.

Methodology—framework

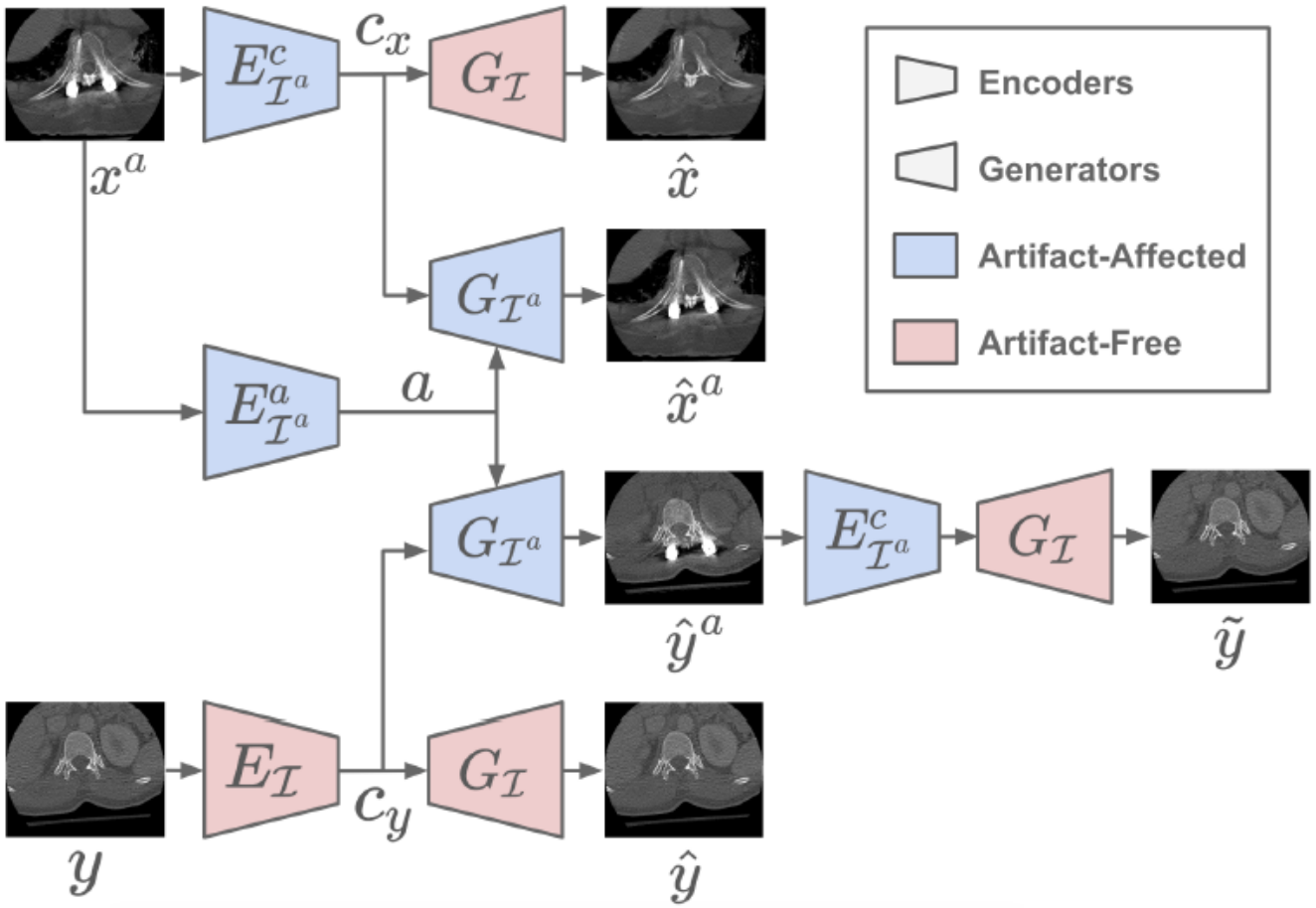


Fig. 2: Overview of the proposed artifact disentanglement network (ADN). Taking any two unpaired images, one from $\mathcal{I}^a$ and the other from $\mathcal{I}$, as the inputs, ADN supports four different forms of image translations: $\mathcal{I}^a \rightarrow \mathcal{I}$, $\mathcal{I} \rightarrow \mathcal{I}^a$, $\mathcal{I} \rightarrow \mathcal{I}$ and $\mathcal{I}^a \rightarrow \mathcal{I}^a$.

Artifact disentanglement network (ADN) aims to learn a mapping from artifact-affected image to artifact-free image without paired data.

$$c_x = E^c_{\mathcal{I}^a}(x^a), a = E^a_{\mathcal{I}^a}(x^a), c_y = E_{\mathcal{I}}(y). \quad (1)$$

$$\hat{x}^a = G_{\mathcal{I}^a}(c_x, a), \quad \hat{y}^a = G_{\mathcal{I}^a}(c_y, a) \quad (2)$$

$$\hat{x} = G_{\mathcal{I}}(c_x), \quad \hat{y} = G_{\mathcal{I}}(c_y). \quad (3)$$

$$\tilde{y} = G_{\mathcal{I}}(E^c_{\mathcal{I}^a}(\hat{y}^a)). \quad (4)$$

Methodology—Learning

Adversarial Loss

$$\mathcal{L}_{\text{adv}}^{\mathcal{I}} = \mathbb{E}_{\mathcal{I}}[\log D_{\mathcal{I}}(y)] + \mathbb{E}_{\mathcal{I}^a}[1 - \log D_{\mathcal{I}}(\hat{x})]$$
$$\mathcal{L}_{\text{adv}}^{\mathcal{I}^a} = \mathbb{E}_{\mathcal{I}^a}[\log D_{\mathcal{I}^a}(x^a)] + \mathbb{E}_{\mathcal{I},\mathcal{I}^a}[1 - \log D_{\mathcal{I}^a}(\hat{y}^a)] \quad (6)$$
$$\mathcal{L}_{\text{adv}} = \mathcal{L}_{\text{adv}}^{\mathcal{I}} + \mathcal{L}_{\text{adv}}^{\mathcal{I}^a}$$

Reconstruction Loss

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{\mathcal{I},\mathcal{I}^a}[||\hat{x}^a - x^a||_1 + ||\hat{y} - y||_1]. \quad (7)$$

Artifact Consistency Loss

$$\mathcal{L}_{\text{art}} = \mathbb{E}_{\mathcal{I},\mathcal{I}^a}[||(x^a - \hat{x}) - (\hat{y}^a - y)||_1]. \quad (8)$$

Self-Reduction Loss

$$\mathcal{L}_{\text{self}} = \mathbb{E}_{\mathcal{I},\mathcal{I}^a}[||\tilde{y} - y||_1]. \quad (9)$$

Full objective function

$$\mathcal{L} = \lambda_{\text{adv}}(\mathcal{L}_{\text{adv}}^{\mathcal{I}} + \mathcal{L}_{\text{adv}}^{\mathcal{I}^a}) + \lambda_{\text{art}}\mathcal{L}_{\text{art}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{self}}\mathcal{L}_{\text{self}}, \quad (5)$$

For ADN, learning an MAR model $f : \mathcal{I}^a \to \mathcal{I}$ means to learn the two key components $E_{\mathcal{I}^a}^c$ and $G_{\mathcal{I}}$. $E_{\mathcal{I}^a}^c$ encodes only the content of an artifact-affected image and $G_{\mathcal{I}}$ generates an artifact-free image with the encoded content code. Thus, their composition readily results in an MAR model, $f = G_{\mathcal{I}} \circ E_{\mathcal{I}^a}^c$. However, without paired data, it is challenging to directly address the learning of these two components. Therefore, we learn $E_{\mathcal{I}^a}^c$ and $G_{\mathcal{I}}$ together with other encoders and decoders in ADN. In this way, different learning signals can be leveraged to regularize the training of $E_{\mathcal{I}^a}^c$ and $G_{\mathcal{I}}$, and removes the requirement of paired data.

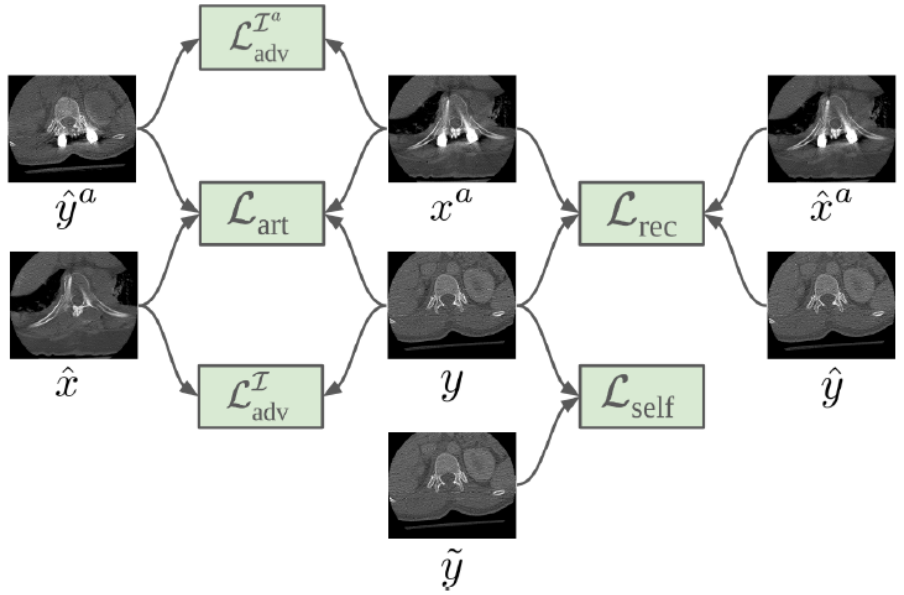


Fig. 3: An illustration of the relationships between the loss functions and ADN's inputs and outputs.

Methodology—Network Architectures

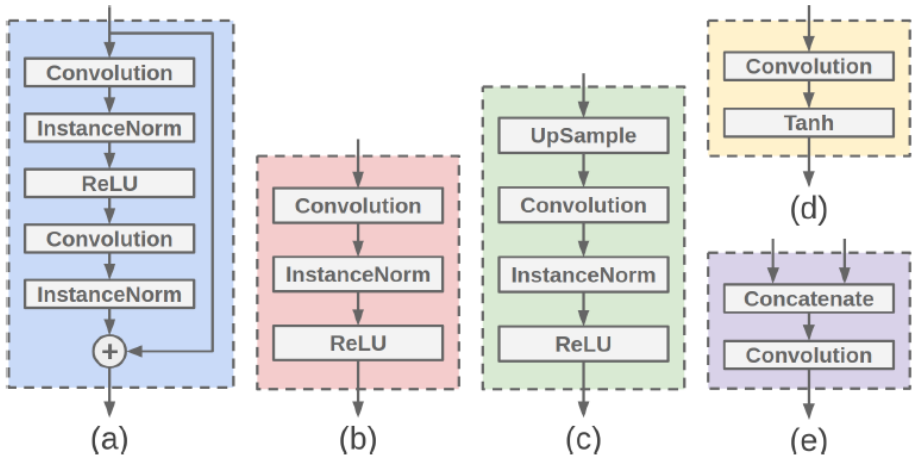


Fig. 4: Basic building blocks of the encoders and decoders: (a) residual block, (b) downsampling block, (c) upsampling block, (d) final block and (e) merging block.

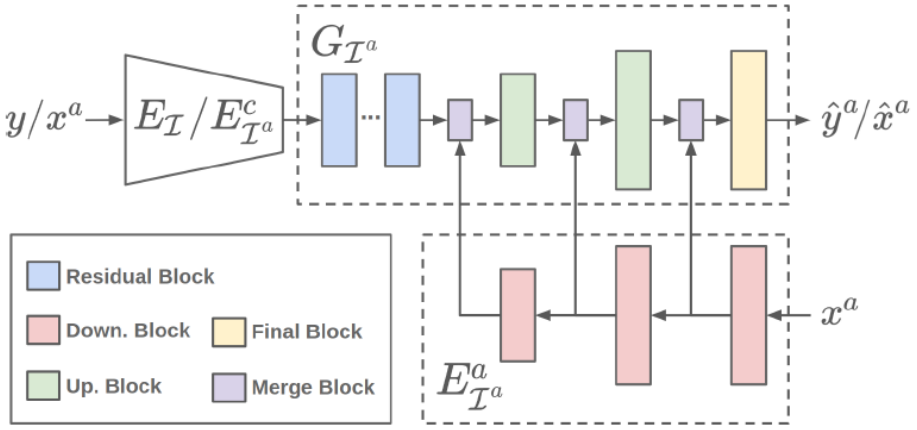


Fig. 5: Detailed architecture of the proposed artifact pyramid decoding (APD). The artifact-affected decoder $G_{\mathcal{I}^a}$ uses APD to effectively merge the artifact code from $E_{\mathcal{I}^a}$.

TABLE I: Architecture of the building components. "Channel (Ch.)", "Kernel", "Stride" and "Padding (Pad.)" denote the configurations of the convolution layers in the blocks.

Network	Block/Layer	Count	Ch.	Kernel	Stride	Pad.
$E_{\mathcal{I}}$ / $E_{\mathcal{I}^a}$	down	1	64	7	1	3
	down.	1	128	4	2	1
	down.	1	256	4	2	1
	residual	4	256	3	1	1
E_a	down.	1	64	7	1	3
	down.	1	128	4	2	1
	down.	1	256	4	2	1
$G_{\mathcal{I}}$	residual	4	256	3	1	1
	up.	1	128	5	1	2
	up.	1	64	5	1	2
	final	1	1	7	1	3
$G_{\mathcal{I}^a}$	residual	4	256	3	1	1
	merge	1	256	1	1	0
	up.	1	128	5	1	2
	merge	1	128	1	1	0
	up.	1	64	5	1	2
	merge	1	64	1	1	0
	final	1	1	7	1	3
$D_{\mathcal{I}}$ / $D_{\mathcal{I}^a}$	conv	1	64	4	2	1
	relu	1	-	-	-	-
	down.	1	128	4	2	1
	down.	1	256	4	1	1
	conv	1	1	4	1	1

Experiments—Datasets

We evaluate the proposed method on **one synthesized dataset and two clinical datasets**.

For SYN, we randomly select 4118 artifact-free CT images from DeepLesion and follow the method from CNNMAR to synthesize metal artifacts. We use 3918 of the synthesized pairs for training and validation and the rest 200 pairs for testing.

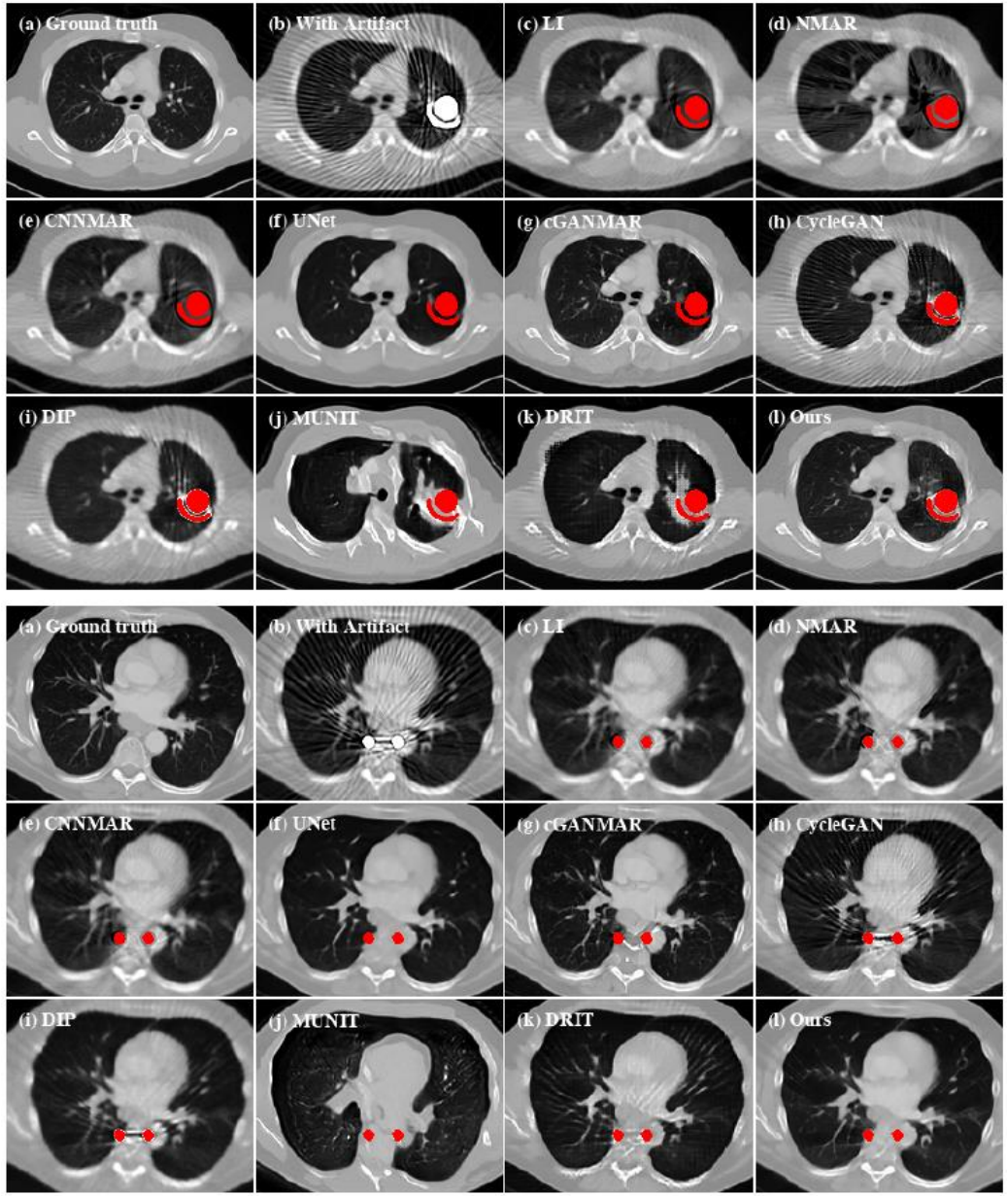
For CL1, we choose the vertebrae localization and identification dataset from Spineweb. We split the CT images from this dataset into two groups, one with artifacts and the other without artifacts. First, we identify regions with HU values greater than 2500 as the metal regions. Then, CT images whose largest-connected metal regions have more than 400 pixels are selected as artifact-affected images. CT images with the largest HU values less than 2000 are selected as artifact-free images. After this selection, the artifact-affected group contains 6270 images and the artifact-free group contains 21190 images. We withhold 200 images from the artifact-affected group for testing.

For CL2, we investigate the performance of the proposed method under a more challenging cross-modality setting. Specifically, the artifact-affected images of CL2 are from a cone-beam CT (CBCT) dataset collected during spinal interventions. Images from this dataset are very noisy and the majority of them contain metallic implants. There are in total 2560 CBCT images from this dataset, among which 200 images are withheld for testing. For the artifact-free images, we reuse the CT images collected from CL1.

Experiments- Synthesized dataset

TABLE II: Quantitative comparison with baseline methods on the SYN dataset.

	Method	Metrics	
		PSNR	SSIM
Conventional	LI [2]	32.0	91.0
	NMAR [3]	32.1	91.2
Supervised	CNNMAR [4]	32.5	91.4
	UNet [30]	**34.8**	93.1
	cGANMAR [5]	34.1	**93.4**
Unsupervised	CycleGAN [24]	30.8	72.9
	DIP [32]	26.4	75.9
	MUNIT [25]	14.9	7.5
	DRIT [33]	25.6	79.7
	Ours	<u>33.6</u>	<u>92.4</u>



Experiments—clinical datasets

Since there are no ground truths available for the clinical images, only qualitative comparisons are performed.

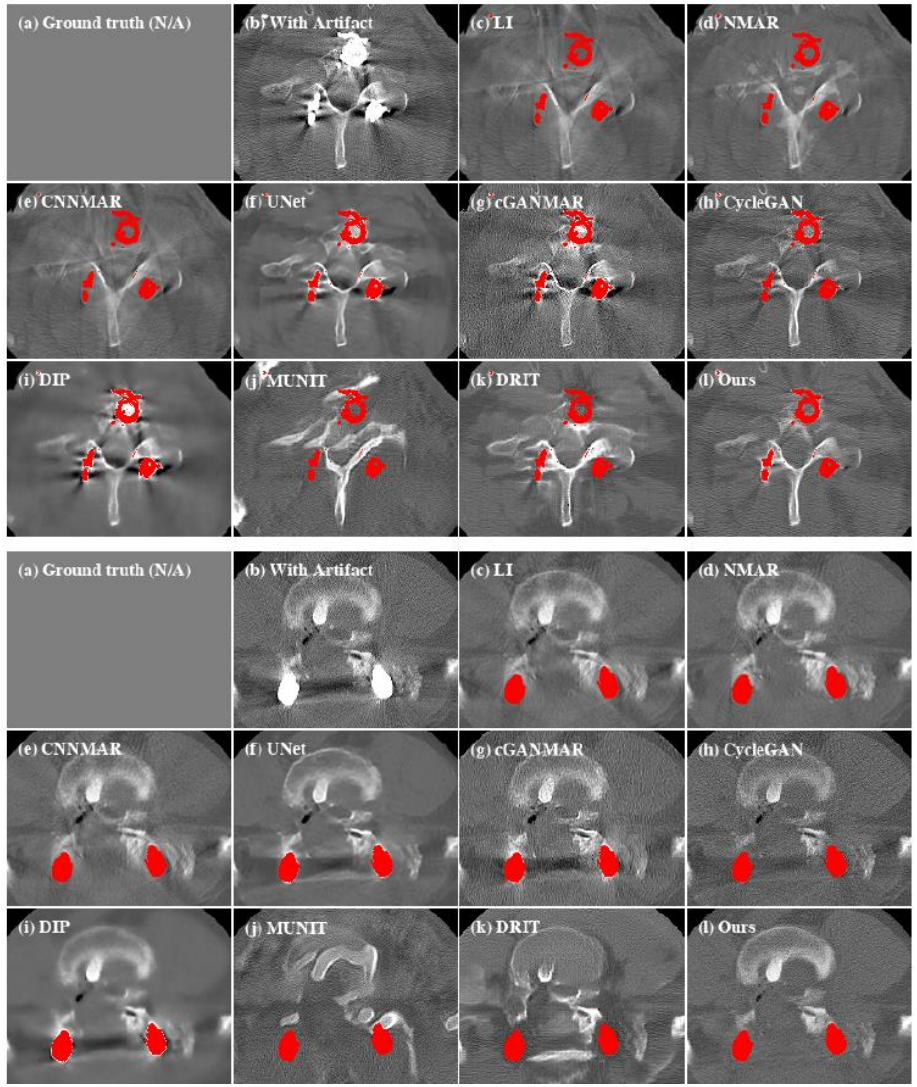


Fig. 7: Qualitative comparison with baseline methods on the CL1 dataset. For better visualization, we obtain the metal regions through thresholding and color them with red.

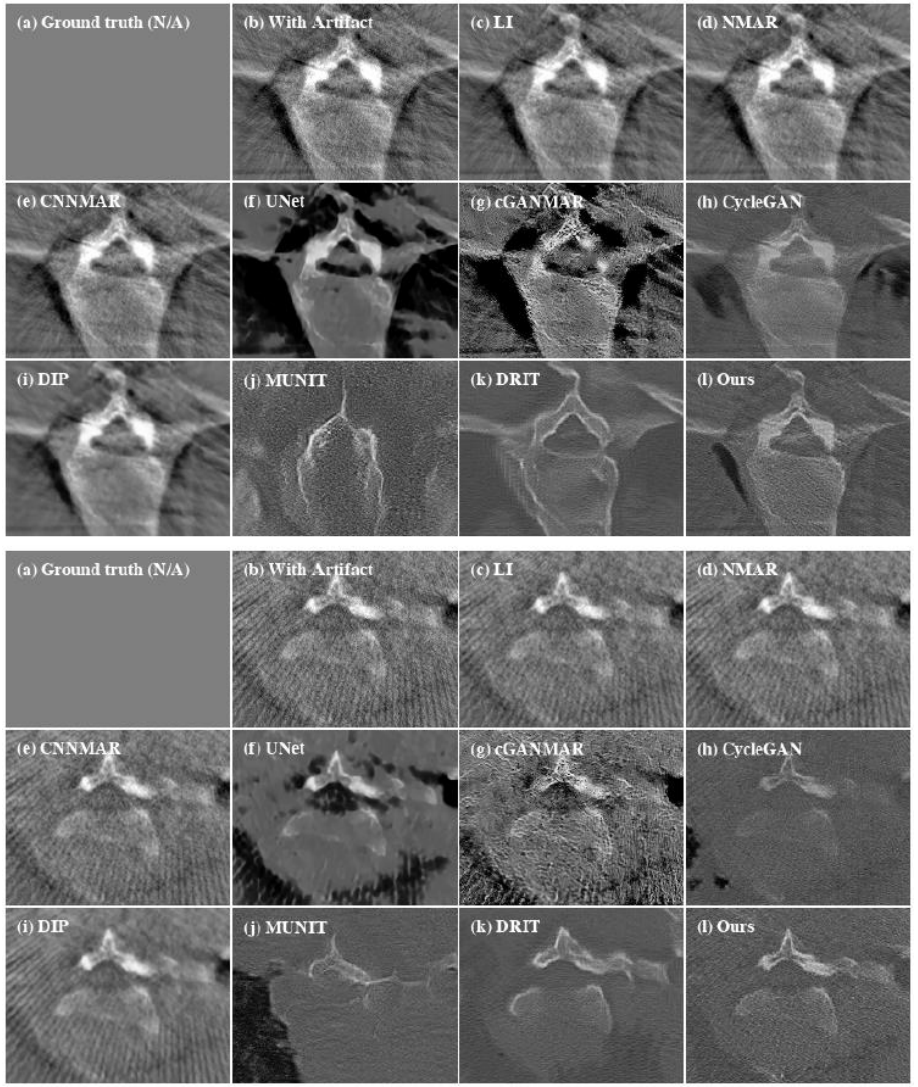
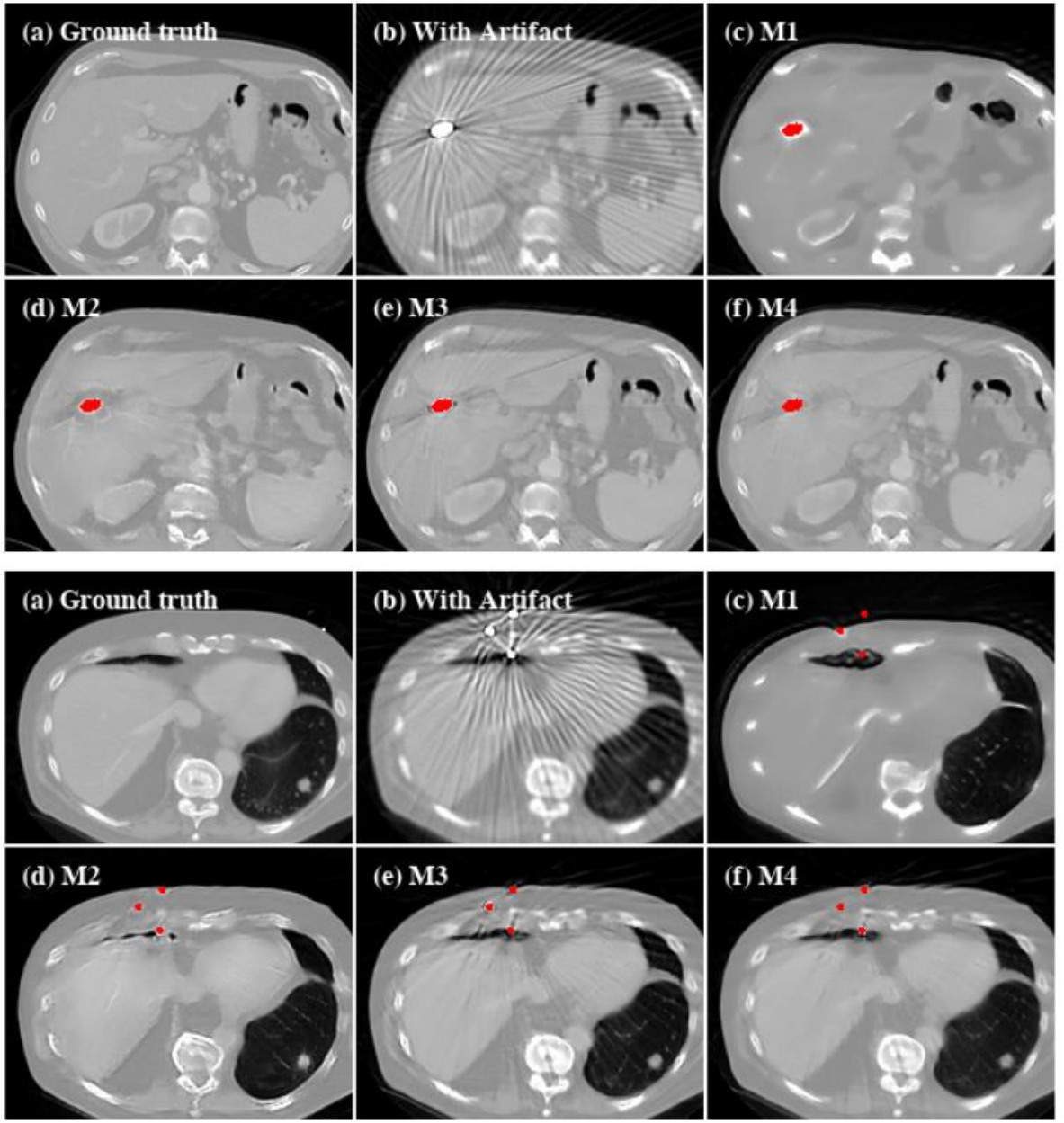


Fig. 8: Qualitative comparison with baseline methods on the CL2 dataset.

Experiments—Ablation study

TABLE III: Quantitative comparison of different variants of ADN. The compared models (M1-M4) are trained with different combinations of the loss functions discussed in Sec. III-B.

Method	Metrics	
	PSNR	SSIM
M1 ($\mathcal{L}_{adv}$ only)	21.7	61.5
M2 (M1 with $\mathcal{L}_{rec}$)	26.3	82.1
M3 (M2 with $\mathcal{L}_{art}$)	32.8	91.6
M4 (M3 with $\mathcal{L}_{self}$)	**33.6**	**92.4**



Experiments—Artifact Synthesis

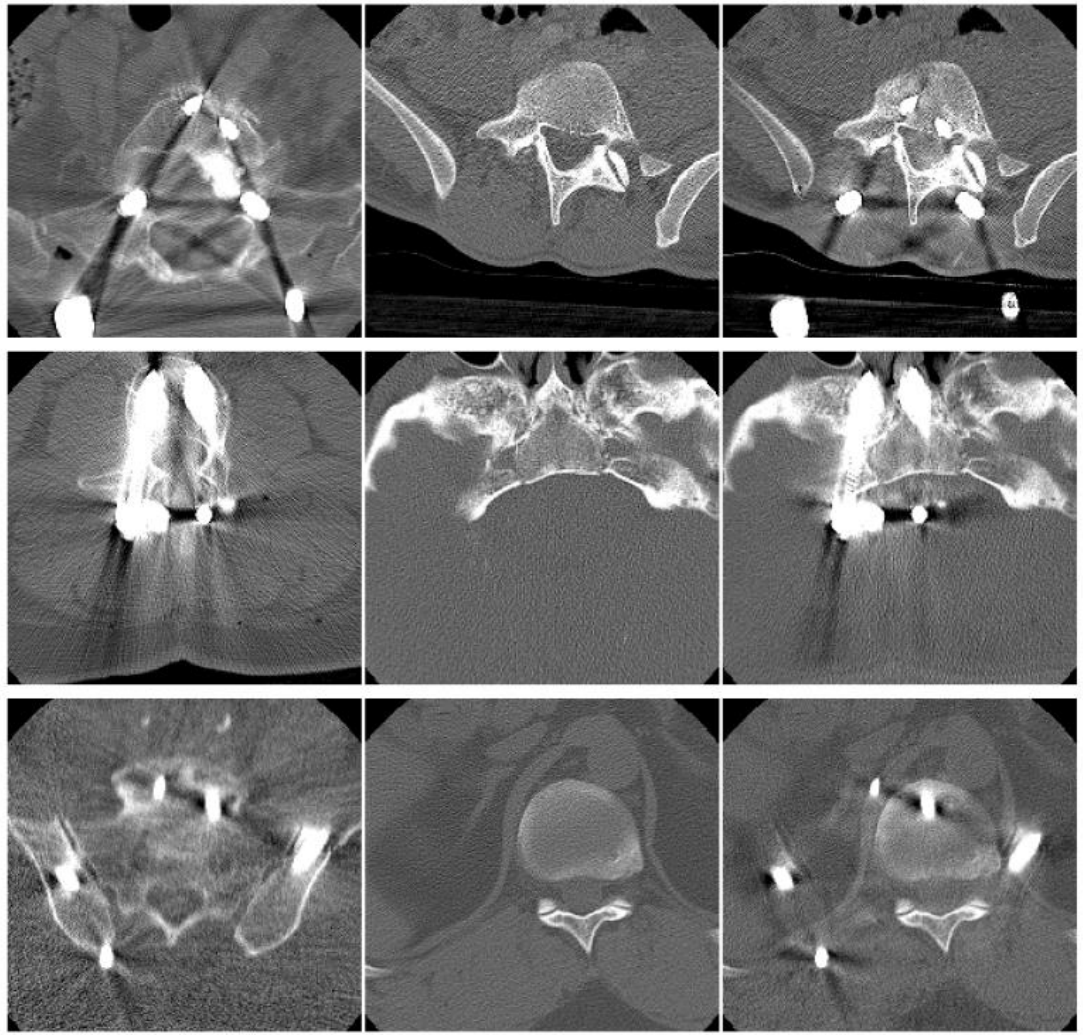


Fig. 10: Metal artifact transfer. Left: the clinical images with metal artifacts x^a. Middle: the clinical images without metal artifacts y. Right: the metal artifacts on the left column transferred to the artifact-free images in the middle $\hat{y}^a$.

As we can see, except the positioning of the metal implants, the synthesized artifacts look realistic. The metal artifacts merge naturally into the artifact-free images making it really challenging to notice that the artifacts are actually synthesized. More importantly, it is only the artifacts that are transferred and almost no content is transferred to the artifact-free images.

An application example of Artifact synthesis: A CT image segmentation model may not work well when metal artifacts are present as there are not enough metal-affected images in the dataset. By using ADN, we could significantly increase the number of metal-affected images in the dataset via the realistic metal artifact synthesis. In this way, ADN may potentially improve the performance of the CT segmentation model.

Experiments—Discussion and Conclusion

V. Discussions

Applications to Artifact Reduction. Given the flexibility of ADN, we expect many applications to artifact reduction in medicine, where obtaining paired data is usually impractical. First, as we have already demonstrated, ADN can be applied to address metal artifacts. It reduces metal artifacts directly with CT images, which is critical to the scenarios when researchers or healthcare practitioners have no access to the raw projection data as well as the associated reconstruction algorithms. For the manufacturers, ADN can be applied in a post-processing step to further improve the in-house MAR algorithm that addresses metal artifacts in the projection data during the CT reconstruction.

Second, even though our problem under investigation is MAR, ADN should work with other artifact reduction problems as well. In the problem formulation, ADN does not make any assumption about the nature of the artifacts. Therefore, if we change to other artifact reduction problems such as deblurring, destreaking, denoising, etc., ADN should also work. Actually, in the experiments, the input images from CL1 (Fig. 7b) are slightly noisy while the outputs of ADN are more smooth. Similarly, input images from CL2 (Fig. 8b) contain different types of artifacts, such as noise, streaking artifacts and so on, and ADN handles them well.

VI. Conclusions

We present an unsupervised learning approach to MAR. Through the development of an artifact disentanglement network, we have shown how to leverage artifact disentanglement to achieve different forms of image translations as well as self-reconstructions that eliminate the requirement of paired images for training. To understand the effectiveness of this approach, we have performed extensive evaluations on one synthesized and two clinical datasets. The evaluation results demonstrate the feasibility of using unsupervised learning method to achieve comparable performance to the supervised methods with synthesized dataset. More importantly, the results also show that directly learning MAR from clinical CT images under an unsupervised setting is a more feasible and robust approach than simply applying the knowledge learned from synthesized data to clinical data. We believe our findings in this work will stimulate more applicable research for medical image artifact reduction under an unsupervised setting.

Thanks for your attention!