

Multiple Adversarial Learning Based Angiography Reconstruction for Ultra-Low-Dose Contrast Medium CT

Weiwei Zhang , Zhen Zhou, Zhifan Gao , Guang Yang , Senior Member, IEEE, Lei Xu, Weiwen Wu , and Heye Zhang , Senior Member, IEEE

Abstract—Iodinated contrast medium (ICM) dose reduction is beneficial for decreasing potential health risk to renal-insufficiency patients in CT scanning. Due to the low-intensity vessel in ultra-low-dose-ICM CT angiography, it cannot provide clinical diagnosis of vascular diseases. Angiography reconstruction for ultra-low-dose-ICM CT can enhance vascular intensity for directly vascular diseases diagnosis. However, the angiography reconstruction is challenging since patient individual differences and vascular disease diversity. In this paper, we propose a Multiple Adversarial Learning based Angiography Reconstruction (i.e., MALAR) framework to enhance vascular intensity. Specifically, a bilateral learning mechanism is developed for mapping a relationship between source and target domains rather than the image-to-image mapping. Then, a dual correlation constraint is introduced to characterize both distribution uniformity from across-domain features and sample inconsistency within domain simultaneously. Finally, an adaptive fusion module by combining multi-scale information and long-range interactive dependency is explored to alleviate the interference of high-noise metal. Experiments are

Manuscript received 16 April 2022; revised 28 August 2022; accepted 2 October 2022. Date of publication 11 October 2022; date of current version 5 January 2023. This work was supported in part by the National Natural Science Foundation of China under Grants U1801265, U1908211, 82271986, 62201628, 62101606, and 62276282, in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2019B151502031, in part by the National Youth Talent Support Program under Grant RC2020-01, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515011384, in part by the ERC IMI under Grant 101005122, in part by the H2020 under Grant 952172, in part by the MRC under Grant MC/PC/21013, in part by the Royal Society under Grant IEC/NSFC/211235, and in part by the UKRI Future Leaders Fellowship under Grant MR/V023799/1. (Corresponding authors: Weiwen Wu; Heye Zhang.)

Weiwei Zhang, Zhifan Gao, Weiwen Wu, and Heye Zhang are with the School of Biomedical Engineering, Sun Yat-sen University, Shenzhen 518107, China (e-mail: mr.ww.zhang@gmail.com; gaozhifan@gmail.com; wuweiw7@mail.sysu.edu.cn; zhangheye@mail.sysu.edu.cn).

Zhen Zhou and Lei Xu are with the Department of Radiology, Beijing Anzhen Hospital, Capital Medical University, Beijing 100029, China (e-mail: alexzhou@ccmu.edu.cn; leixu2001@hotmail.com).

Guang Yang is with the Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, U.K., and also with the National Heart and Lung Institute, Imperial College London, London SW7 2AZ, U.K. (e-mail: g.yang@imperial.ac.uk).

The code and model are available at <https://github.com/HIC-SYSU/MALAR>.

This article has supplementary downloadable material available at <https://doi.org/10.1109/JBHI.2022.3213595>, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2022.3213595

performed on CT sequences with different ICM doses. Quantitative results based on multiple metrics demonstrate the effectiveness of our MALAR on angiography reconstruction. Qualitative assessments by radiographers confirm the potential of our MALAR for the clinical diagnosis of vascular diseases.

Index Terms—Angiography CT reconstruction, ultra-low-dose, iodinated contrast medium, multiple adversarial learning, adaptive fusion, customized windowing.

I. INTRODUCTION

IODINATED contrast medium (ICM) plays an important role in the diagnosis of clinical aortic diseases by promoting aortic imaging in computed tomography angiography (CTA) [1]. Compared with aortography, CTA has better visualization performance in terms of aortic arch, brachiocephalic arterial branches and adjacent aneurysms [2]. Meanwhile, the intravenous administration of iodinated contrast medium enables the visualization of true and false luminal flow, and slow perigraft flow around stent grafts [3]. The clear anatomical structure of vascular system in CTA can help physicians to improve the assessment of aortic abnormalities.

ICM in aorta CTA imaging can result in increasing health risk to patients [4]. The application of ICM not only significantly increases the amount of peripheral lymphocyte DNA radiation damage, but also easily causes contrast-induced acute kidney injury [5], [6]. In addition, intravenous administration of ICM potentially worsens the renal function of people with chronic kidney disease [7]. In particular, some ICM-related hypersensitivity reactions are life-threatening in some cases [8]. Thus, ultra-low-dose-ICM aorta CT (UDCT) can be significant for reducing the potential health risk [9].

UDCT is difficult to be directly used for the clinical diagnosis of aortic diseases due to the low intensity of aorta. Again, ultra-low-dose ICM is difficult to ensure the aorta to be fully enhanced [10]. Lower contrast between aorta and surrounding organs in UDCT prevents physicians from identifying the delineation of aorta, which further results in the difficulty of discriminating aortic aneurysms and pseudoaneurysms [2]. It is also difficult to clearly observe the anatomical morphology of aorta even if performing multiplanar reconstruction for UDCT. Consequently, physicians cannot identify whether there are morphological abnormalities in the aorta, including slow perigraft

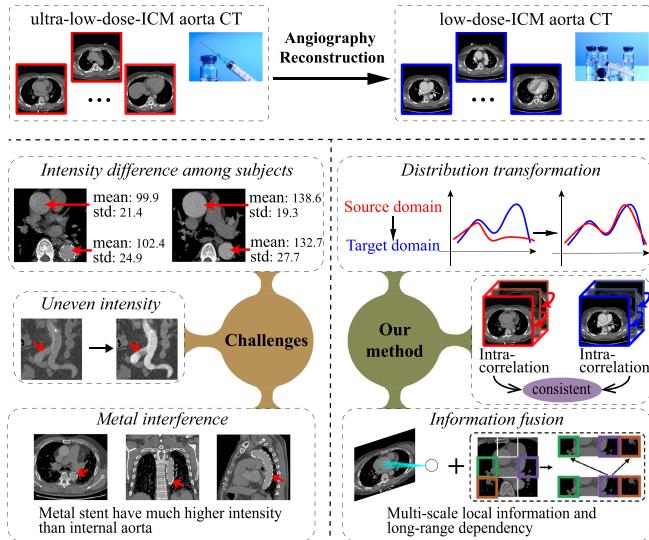


Fig. 1. Our proposed framework overcomes the physical and physiological characteristics of UDCT aortic reconstruction. Aortic intensity differences among subjects and among slices will prevent aorta reconstruction. And high-intensity metal also influences the identification and reconstruction of internal aorta. Our framework employs three strategies to address corresponding challenges.

flow and aortic dissection. The aorta reconstruction for UDCT, i.e., enhancing aortic intensity to the routine CTA level, makes it possible to directly diagnose aortic disease from UDCT.

It is challenging to reconstruct pseudo-normal aorta with UDCT limited by imaging environment variation and physiological tissues diversity, as shown in Fig. 1. First, aortic intensity differences among different UDCT sequences can prevent aorta reconstruction. In clinical, there are subject individual differences and aortic disease diversity. To achieve the rationalization of clinical CT scanning scheme, radiologists adaptively adjust the tube voltage and tube current according to the body type, disease, and examination part of patients. Hence aortic intensity varies by clinical subjects. Second, the uncertainty of ICM concentration in aorta increases the difficulty of aorta reconstruction. Specifically, the ICM will cause the aortic intensity fluctuations in UDCT when aorta gets diseased. Besides, when ICM travels with blood in human vessels after intravenous administration, it will be gathered passing through true lumen or aneurysm. Conversely, the ICM concentration will be decreased passing through distal blood vessels, such as inferior clavicular artery and abdominal aorta. Again, ICM shows the intensity distribution differences in aorta. Thirdly, aortic diseases further interfere aortic reconstruction. For example, for aortic dissection UDCT, the torn intimal of blurry aortic wall and an uncertain shape cause the true and false lumens mixed to be difficultly discriminated [11]. For postoperative cases of vascular stenosis, the implanted stent has a high Hounsfield unit (HU) in UDCT, which strongly interferes with aorta reconstruction (Fig. 1).

In this paper, we propose a Multiple Adversarial Learning based Angiography Reconstruction framework named MALAR to realize the aortic reconstruction in UDCT, which promotes the diagnosis of aortic diseases from UDCT directly. Our MALAR is verified by isolated UDCT and low-dose-ICM aorta CT (i.e.,

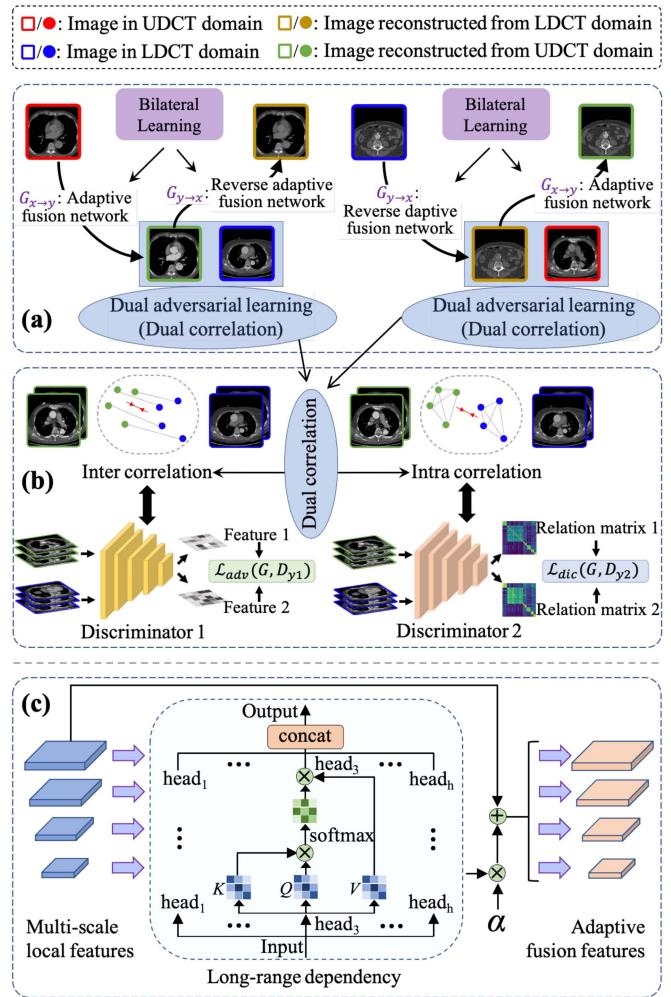


Fig. 2. (a) Flow chart of our proposed MALAR framework for aorta reconstruction in UDCT. The bilateral learning applies an adaptive fusion ($G_{x \rightarrow y}$) network and a reverse adaptive fusion network ($G_{y \rightarrow x}$) to achieve mutual transformation of source and target domains. The two dual adversarial learning adopt the same learning process and implementation, but for different images. (b) Learning process and implementation of the dual correlation constraint. It includes an inter-correlation between domains and an intra-correlation in domains, and is realized by two discriminators. (c) Fusion process of multi-scale local features from our multi-scale residual network and long-range dependency from multi-head self-attention. It constitutes the building block of $G_{x \rightarrow y}$ and $G_{y \rightarrow x}$ in (a). α is a learned parameter indicated by (6).

LDCT) from a clinical centre. Detailed information of our used datasets can be found in Section IV-A. Because UDCT and LDCT correspond to lower and higher aortic intensity, they are treated as source and target sequences to develop our MALAR network. To address the aortic intensity difference in UDCT and LDCT, here we introduce a bilateral learning mechanism to develop a mapping function between source and target domains inspired by cycle-consistent adversarial networks (CycleGAN) [12], as shown in Fig. 2(a). The bilateral learning mechanism mutually transfers aortic intensity between given two domains through forward and inverse learning processes. Specifically, the bilateral learning applies an adaptive fusion network $G_{x \rightarrow y}$ and a reverse mapping network $G_{y \rightarrow x}$ to characterize forward mapping and backward mapping for UDCT and LDCT domains. $G_{x \rightarrow y}$ translates source CTAs to the target

domain to perform pixel-level enhancement, while $G_{y \rightarrow x}$ translates target CTAs to the source domain to implement pixel-level weakening. To address the inconsistency of translated image and target image domains, we design a dual correlation constraint, as shown in Fig. 2(b) [13]. It enforces the feature space consistency and performs distribution alignment by characterizing both distribution uniformity from across-domain features and sample inconsistency within domain simultaneously. The dual correlation constraint aims to preserve aortic smoothness in a longitudinal direction for enhancing its intensity. To overcome high-intensity inference of metal stent surrounding the aorta, hence we propose an adaptive fusion module by combining the multi-scale local information and long-range dependency of aorta, as shown in Fig. 2(c) [14], [15]. The integrated module is not only beneficial to the identification of aorta, but also improves the structure visualization and completeness. Additionally, the contrast between aorta and tissue is too small to be difficultly reconstructed, here our framework uses a customized windowing technique to restrict the intensities of UDCT and LDCT at MALAR input. With such a windowing technique, the contrast between aorta and surrounding tissues are larger so that our MALAR can be easier to identify aortic intensity range. In general, our MALAR can be considered as an augmented CycleGAN framework, which promotes CycleGAN by integrating an adaptive fusion network, a reverse mapping network and a dual correlation constraint [12].

The main contributions of this work are summarized as follows:

- We propose a bilateral learning mechanism to learn a function between two domains with different CT intensity, which is beneficial to eliminating the intra-domain aortic intensity difference.
- We introduce a dual adversarial constraint to improve the inter- and intra-domain correlations, which is good for enhancing pixel-level aorta intensity and eliminating the difference among enhanced slices.
- We establish an adaptive fusion module to learn both multi-scale local information and long-range dependency information, which is helpful to relieve the interference of vessel stent.

II. RELATED WORK

Contrast medium dose reduction: Haubold et al. have proposed a conditional generative adversarial network to reduce the dose of intravenous ICM based on virtual contrast-enhanced CT images [7]. The model can realize 50% and 80% contrast medium dose reduction on virtual paired data. Xie et al. have proposed a CycleGAN framework to generate contrast-enhanced CT based on non-contrast simulation CT [16]. The framework redesigns the cycle consistency loss, which consists of mean absolute error loss, gradient difference loss, and mutual information loss. However, the generated CT images are not very suitable for clinical diagnosis. Specifically, the generated images are too smooth, and the vessel edges have low sharpness. The areas without ICM are enhanced to varying intensity levels, and artifacts exist along organ edges.

Adversarial learning for domain transformation: Since the paired training data is difficult to be acquired in practice, Zhu

et al. have proposed a CycleGAN for domain transformation to address this problem [12]. CycleGAN aims to learn a mapping $G : X \rightarrow Y$ from a source domain X to a target domain Y and couples it with an inverse mapping $F : Y \rightarrow X$. It introduces a cycle consistency loss to impose a constraint $F(G(X)) \approx X$ (and vice versa). CycleGAN-based approaches have also been used in medical image translation, including low-dose CT denoising and cross-domain medical image synthesis [17], [18].

Self-attention: Self-attention extracts queries, keys, and values from the same context instead of across multiple contexts, which can model long-range dependency based on the context [19]. A theme of augmenting convolution models with self-attention has yielded gains in vision tasks of classification, detection, and segmentation [20], [21], [22]. Because self-attention is able to capture the dependency of each input token, and learn long-range dependencies between the tokens, it is emerging as a prominent building block of models. Full self-attention across the entirety of networks has gradually attracted attention on various vision tasks [23], [24]. However, self-attention focuses on the current position, and inhibits interactive information with other positions. Multi-head attention is able to improve the learning ability of self-attention in terms of interactive dependency relationships from different positions [15]. Based on its powerful representation capabilities, multi-head attention has become an important unit of Transformer [25].

III. METHODOLOGY

Given domains of UDCT and LDCT denoted by X and Y where $X = \{x_i\}_{i=1}^N$ and $Y = \{y_j\}_{j=1}^M$, we denote the data distribution as $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$. Our goal is to learn a mapping function $G_{x \rightarrow y}$ for making distributions of $G_{x \rightarrow y}(x)$ and y are consistent. The overview of our proposed reconstruction framework (i.e., MALAR) is shown in Fig. 2. It is learned by multiple adversarial learning process, where the name comes from. Additionally, the customized windowing technique is also introduced in Section III-D.

A. Bilateral Learning Mechanism

Inspired by CycleGAN [12], we introduce a bilateral learning mechanism to establish the mapping between UDCT and LDCT domains, as shown in Fig. 2(a). It consists of two forward adaptive fusion networks $G_{x \rightarrow y}$ from domain X to Y and two reverse adaptive fusion networks $G_{y \rightarrow x}$ from domain Y to X . They are learned by dual correlation constraint adversarial learning, including inter- and intra-correlation (as shown in Fig. 2(b)).

To match distributions between source and target domains in domain shift process, $G_{x \rightarrow y}$ tries its best to transform the content and style of X to Y , which can confuse discriminators to generate unsatisfied reconstruction results. Because we only focus on enhancing the aorta in UDCT, the rest image content of UDCT should be retained. Thus, reverse mapping mechanism is developed to improve enhancement performance by incorporating the reverse process of main task [26]. Here we explore a reverse mapping of $G_{y \rightarrow x}$, i.e., the mapping from LDCT to UDCT to regularize the mapping $G_{x \rightarrow y}$ for preserving source content, e.g., cardiac structure and CT intensity of adipose tissue, etc [27]. The reverse mapping inputs images from

LDCT domain and regenerates UDCT images, which would be consistent with the input UDCT in image structure and intensity. In other words, the UDCT in source domain can be transformed to LDCT within target domain, and the generated LDCT can be also transformed back to source UDCT. Additionally, based on the knowledge of cycle consistency loss and reconstruction loss [12], [28], [29], minimizing the divergence between source and regenerated UDCT can promote the model's learning ability by adding additional supervised information. The reconstruction errors are introduced, which are denoted by \mathcal{L}_{rec_x} and \mathcal{L}_{rec_y} . The reverse mapping $G_{y \rightarrow x}$ combined with reconstruction errors drives $G_{x \rightarrow y}$ to generate anatomically invariant LDCT images, which facilitates the learning of $G_{x \rightarrow y}$.

B. Dual Correlation Based on Dual Adversarial Learning

Our MALAR explores the relations of source and target domains in forward mapping from two aspects: inter- and intra-correlation. The first adversarial learning learns the feature-level inter-correlation from different domains to align their distributions. The second adversarial learning can extract the intra-correlation from two domains and further keep their consistency. Then, reverse mappings are also developed by adversarial learning for the two correlations.

1) Adversarial Learning for Inter-Correlation Across Domain:

To transfer domain X to Y , previous adversarial translation approaches can be employed to learn a generator $G_{x \rightarrow y}$, such as CycleGAN, DualGAN and DiscoGAN [12], [28], [29], [30]. The generator tries to generate fake samples of high-fidelity aorta images to fool the discriminator D_{y1} , which focuses on distinguishing generated images and true LDCT. To further explore the feature-level difference between domain X and Y , the adversarial translation formulation $\min_{G_{x \rightarrow y}} \max_{D_{y1}} \mathcal{L}_{adv}(G_{x \rightarrow y}, D_{y1})$ is improved by a least-squares loss, which is modified as [31]:

$$\begin{aligned} \min_{G_{x \rightarrow y}} \mathcal{L}_{G_{x \rightarrow y}} &= \mathbb{E}_{x \sim P_{data(x)}} [(D_{y1}(G_{x \rightarrow y}(x)) - 1)^2] \\ \min_{D_{y1}} \mathcal{L}_{D_{y1}} &= \mathbb{E}_{y \sim P_{data(y)}} [(D_{y1}(y) - 1)^2] \\ &\quad + \mathbb{E}_{x \sim P_{data(x)}} [D_{y1}(G_{x \rightarrow y}(x))^2] \end{aligned} \quad (1)$$

Similarly to Equation (1), the training objective $\min_{G_{y \rightarrow x}} \max_{D_{x1}} \mathcal{L}_{adv}(G_{y \rightarrow x}, D_{x1})$ of reverse adaptive fusion network $G_{y \rightarrow x}$ is given by:

$$\begin{aligned} \min_{G_{y \rightarrow x}} \mathcal{L}_{G_{y \rightarrow x}} &= \mathbb{E}_{y \sim P_{data(y)}} [(D_{x1}(G_{y \rightarrow x}(y)) - 1)^2] \\ \min_{D_{x1}} \mathcal{L}_{D_{x1}} &= \mathbb{E}_{x \sim P_{data(x)}} [(D_{x1}(x) - 1)^2] \\ &\quad + \mathbb{E}_{y \sim P_{data(y)}} [D_{x1}(G_{y \rightarrow x}(y))^2] \end{aligned} \quad (2)$$

2) Adversarial Learning for Intra-Correlation in Domain: Inspired by analogous entities property within graph network [32], it is feasible to learn extra correlation information for CTA images in same domain [13]. Thus, the intrinsic correlation is explored in our MALAR by developing a consistency

paradigm to characterize domain intra-correlation consistency. This consistency regularizes the generator $G_{x \rightarrow y}$ to maintain intra-correlation consistency across domains. It encourages our MALAR to retain the integrity and continuity of reconstructed aorta for 3D space.

We model the intra-correlation across-domain with a slice-level Gram Matrix [33]. Here a mini-batch with N images is given as input, and the feature responses of layer l in network D_{y2} is denoted by $F^l \in \mathbb{R}^{N \times H \times W \times C}$. The H , W , and C represent height, width, and channel number of such features respectively. Thus, the intra-correlation can be denoted by a Gram Matrix $M^l \in \mathbb{R}^{N \times N}$:

$$M^l = \sum_{i,j}^N F_i^l \cdot F_j^l \quad (3)$$

where M^l represents the results of matrix multiplication between feature responses F_i^l and F_j^l . It is noted that F_j^l should be reshaped and transposed first. It can be explained as a characterization of correlation or similarity from the feature of i^{th} slice and j^{th} slice in mini-batch input.

To generate images for matching target domain, the deviation between M_g^l from generated images and M_t^l from target images should be calculated. Therefore, the domain intra-correlation loss (Dic-loss) is defined as:

$$\mathcal{L}_{dic} = \frac{1}{N} \| M_g^l - M_t^l \|_2^2 \quad (4)$$

By minimizing \mathcal{L}_{dic} during optimization process, the generator $G_{x \rightarrow y}$ is constrained to capture additional intra-correlation representation of different neighbour slices in the same domain. This representation helps to extract the aortic integrity information, which is very important for the continuity and uniformity of aortic structure and intensity within 3D space.

Additionally, inspired by the sample relation consistency (SRC) in [13], features from the last layer of network D_{y2} are used to calculate DIC-loss. With such a strategy, low-level information, e.g., color, texture, and other characteristics can be extracted with shallow layers. Much high-level information can be fully explored by deeper layers. To emphasize these structural differences in generated images compared with corresponding LDCT, the deeper layers can efficiently extract desired intra-correlation information. Therefore, the output of D_{y2} is used as F^l in this study.

3) Overall Objective: The overall training objective of our MALAR is as follows:

$$\begin{aligned} \mathcal{L}_{MALAR}(G_{x \rightarrow y}, G_{y \rightarrow x}, D_{y1}, D_{y2}, D_{x1}, D_{x2}) &= \mathcal{L}_{adv}(G_{x \rightarrow y}, D_{y1}) + \mathcal{L}_{adv}(G_{y \rightarrow x}, D_{x1}) \\ &\quad + \lambda \mathcal{L}_{dic_y}(G_{x \rightarrow y}, D_{y2}) + \lambda \mathcal{L}_{dic_x}(G_{y \rightarrow x}, D_{x2}) \\ &\quad + \beta \mathcal{L}_{rec_x} + \beta \mathcal{L}_{rec_y} \end{aligned} \quad (5)$$

where $\mathcal{L}_{dic_y}(G_{x \rightarrow y}, D_{y2})$ denotes the Dic-loss between generated and target LDCT, $\mathcal{L}_{dic_x}(G_{y \rightarrow x}, D_{x2})$ denotes the Dic-loss between the generated and source UDCT. λ and β are hyperparameters used to balance loss terms, and are set to 1 and 10 based on [13] and [12].

C. Adaptive Fusion Module

To better alleviate the disturbance of high-intensity metal stent and artifact to aorta reconstruction, we design an adaptive fusion module as the main component of $G_{x \rightarrow y}$ and $G_{y \rightarrow x}$, as shown in Fig. 2(c). It adaptively integrates multi-scale local information and long-range interactive dependency. Multi-scale local information is of great importance to vision tasks, such as object detection, edge detection, etc., since pixel-level tasks generally require convolutional neural networks (CNNs) to extract coarse-to-fine multi-scale features for locating holistic objects and their region details [34]. Based on the understanding of multi-scale representations, residual learning is introduced to construct a multi-scale residual network [14]. This network is able to extract rich aorta information for improving the identification of holistic aorta and its border in UDCT. Long-range dependency enables global content-dependent interactions among different image regions to further learn detailed context information [35]. Multi-head self-attention is an effective way to extract long-range dependency, and is widely validated in vision tasks [15]. In this paper, long-range dependency learned by multi-head self-attention is integrated with our multi-scale residual network, to enrich aortic features, distill key aortic information, and exclude noise interference.

Given a local feature $F_s \in \mathbb{R}^{N \times H_s \times W_s \times C_s}$ of a certain scale, it is transformed into three feature spaces Q, K, V based on linear transformation W_Q, W_K, W_V . The adaptive fusion process can be formulated as follows:

$$\begin{aligned} Q &= F_s \cdot W_Q, \quad K = F_s \cdot W_K, \quad V = F_s \cdot W_V \\ \text{head}_h &= \text{softmax} \left(Q \cdot K / \sqrt{d_k} \right) \cdot V \\ \text{mhsa} &= \text{concat} (\text{head}_1, \text{head}_2, \dots, \text{head}_h) \\ F_{af} &= \alpha * \text{mhsa} + F_s \end{aligned} \quad (6)$$

where \cdot and $*$ denote matrix multiplication and element-wise multiplication operation respectively. $\sqrt{d_k}$ represents scaling factor, which prevents extremely small gradients [15]. F_{af} represents the adaptive fusion feature. α is a learned scalar and is initialized as 0.

D. Customized Windowing

To better characterize the intensity differences between aorta and tissues, we propose a customized windowing to change the intensity distribution of UDCT and LDCT. In UDCT, the intensity ranges from -1024 HU to 3071 HU. The intensity of aorta is about 120 HU, and adjacent vessels are about 60 HU. It is difficult to find their intensity difference in the context of a wide range of CT intensity, where covering unnecessary attenuation features in our task [36]. In clinical, radiologists routinely adjust window level and window width to change image contrast and brightness for better revealing aortic anatomy or pathology [37]. Inspired by this procedure, here we automatically set the window level and window width of UDCT and LDCT before inputting our model. In this process, the theoretically possible maximum value of 700 in LDCT should be covered. Therefore, the window level is set to 200 HU, and the window width is empirically set to 1000 . Again, the modified CT intensity range is set from

-300 HU to 700 HU, which can be calculated by following formula [38]:

$$ui = wl + ww/2, \quad li = wl - ww/2 \quad (7)$$

where ui and li denote upper and lower intensity, wl and ww denote the window level and window width, respectively.

Then, we truncate the CT intensity outside the set narrow window width [38]. Specifically, the CT intensity less than -300 HU is reset to -300 HU, and that greater than 700 HU is reset to 700 HU. The truncated pixels are saved with their coordinates and intensities. After implementing the aortic reconstruction, the intensity of these truncated pixels are restored with a series of reverse operations. LDCT is restricted in the same way as UDCT. The truncated UDCT and LDCT are used as inputs of our reconstruction network.

IV. EXPERIMENTAL SETUP

A. Dataset

To evaluate the performance of our MALAR, 170 in total contrast-enhanced aortic CTA scans have been collected from the Beijing Anzhen Hospital, Capital Medical University. The aortic CTA scanning ranges from lung apex to lesser trochanter of the femur according to clinical routine standards strictly. All the clinical subjects have ethical approval from the institutional review board. Each subject have underwent twice CTA scans. The first scan, i.e., UDCT protocol, is performed using a 256-row multi-detector CT scanner (Revolution CT, GE Healthcare, USA) with a body mass index (BMI)-adapted tube voltage of $80\text{--}120$ kV. Automatic tube current modulation is provided by CareDose 4D adaptive with patient anatomy. $40\text{--}70$ ml mixed ICM (Iodixanol 320 (320 mgI/ml):saline=1:2) is injected intravenously. All the acquired CT slices have an in-plane resolution of $0.3\text{--}0.5$ mm and slice thickness of 0.625 mm. The second scan, i.e., LDCT protocol, is with the same dose of pure ICM ($40\text{--}70$ ml, Iodixanol 320, 320 mgI/ml, GE Healthcare, Buckinghamshire, U.K.). It is executed one hour after UDCT protocol. The ICM administrated in UDCT scan has been washed out when the LDCT scan started, and thus avoiding the potential residual enhancement as much as possible. Except for the relative ICM dose, other scanning parameters of LDCT protocol are consistent with the scanning configuration of UDCT protocol. The subjects of UDCT are randomly divided into two subsets for training (70% with 66 subjects) and testing (30% with 30 subjects). The other 74 UDCT subjects collected the second time are used for independent validation.

B. Training Details

Our MALAR has been trained and validated using TensorFlow 1.15.0 on a Tesla V100 GPU (32 GB GPU memory). It is trained using an Adam solver with $\beta_1 = 0.5$, $\beta_2 = 0.999$, and learning rate of 0.0002. The learning rate is constant at 0.0002 for the first 30 epochs, and then linearly decays to zero over the next epochs. The model stops training at 40 epochs by an early stopping strategy. The batch size is set to 8. All the generators and discriminators in MALAR are updated in every optimization step. To save memory and speed up training, a

data preprocessing of fixed-position cropping is performed to DICOM images. Specifically, the original DICOM images have a size of 512×512 . A centre patch with a size of 384×384 is cropped to be used for the input, which can cover the entire aortic structure. UDCT and LDCT sequences are both performed the above-mentioned cropping operation at once before training. The pixel values of input images are normalized to $[-1, 1]$ after the customized windowing (Section III-D) before inputting the adversarial model.

$G_{x \rightarrow y}$ and $G_{y \rightarrow x}$ adopt an encoder-decoder architecture to integrate residual learning and self-attention for merging local features and long-range dependency information [14], [15], [39]. The encoders use the ResNet-18 as backbone networks, in which the last three building blocks are improved with a multi-head self-attention [14], [15]. For the multi-head self-attention, we reduce its channel number c to $c/8$ in preliminary linear transformation for memory efficiency without significant performance decrease [40]. The attention map is added back to the input feature map by multiplying a learnable scalar parameter [22]. After that, layer normalization is used for the above-mentioned features fusion. The size of feature maps in each scale of the encoders is half of the one in previous scale, gradually changing from 384×384 to 12×12 . The decoders use continuous bilinear interpolation and convolutions to resize feature maps. All convolutional layers inside networks are followed by an instance normalization layer and a ReLU activation function, while the final convolutional layer is followed by a Tanh activation function. In $G_{x \rightarrow y}$ and $G_{y \rightarrow x}$, the first and last convolutional layers are 7×7 spatial filters with stride 2 and stride 1, respectively. The other convolutional layers are 3×3 spatial filters with stride 1 [14].

D_{y1} and D_{x1} adopt the same architecture as PatchGAN, which try to distinguish the generated or target LDCT image patches [41]. Let C_k denotes a Convolution-InstanceNorm-LeakyReLU layer with k filters. The detailed architecture of D_{y1} and D_{x1} consists of C64-C128-C256-C512. All convolution layers are 4×4 spatial filters with stride 2. All ReLUs are configured as leaky with a slope of 0.2. After the last layer C1024, a 4×4 convolution layer with stride 1 is applied to map output, where the activation function is ignored. D_{y2} and D_{x2} adopt pre-trained 16-layer VGGNet. The feature maps from layer *block5_conv3* of VGGNet are employed to calculate our DIC-loss (Section III-B2).

C. Evaluation Scheme

Qualitative evaluation: Two board-certified radiographers (with 5 and 7 years of experience in CTA imaging) independently have assessed whether image quality fulfilled clinical diagnostic requirements guided by a 5-point scored scale [42]. The subjective evaluation criterion is as follows: 5 = excellent contrast ratio, good vascular delineation, rare artifacts, easy to diagnose; 4 = good contrast ratio, normal vascular delineation, slight artifacts, adequate for diagnosis; 3 = acceptable contrast ratio, moderate artifacts, partially blurred vascular delineation, sufficient for diagnosis; 2 = poor contrast ratio, high artifacts, blurred vascular delineation, not sufficient

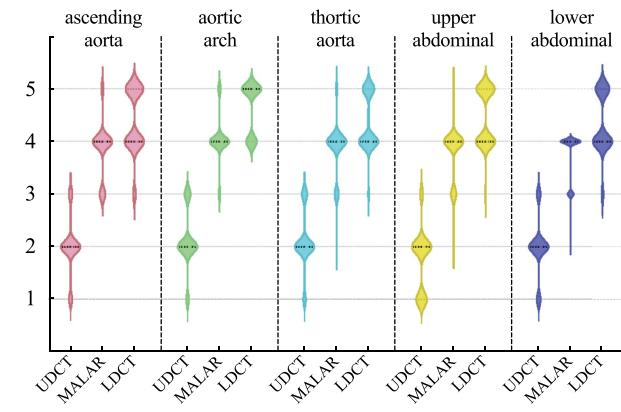


Fig. 3. Box-plot analysis for qualitative scores of UDCT, MALAR, and LDCT. Mean scores of aortic quality of MALAR exceeds those of UDCT, and approaches to those of LDCT.

TABLE I
FIVE ANATOMICAL STRUCTURES ARE SELECTED TO EVALUATE THE RECONSTRUCTION QUALITY OF AORTA (UPPER/LOWER ABDOMINAL AORTA, UPPER/LOWER SEGMENT OF ABDOMINAL AORTA)

Structure	Location
Ascending aorta	2 cm above the aortic sinus
Aortic arch	The middle level of aortic arch
Thoracic aorta	The level of 8-th thoracic vertebra
Upper abdominal aorta	The level of celiac trunk opening
Lower abdominal aorta	2 cm above the iliac bifurcation

for diagnosis; 1 = uninterpretable contrast ratio, severe artifacts, indistinguishable vascular delineation, diagnosis was impossible. All subjects are anonymized and randomly shuffled. The two radiographers are blinded to scanning and processing conditions. When their evaluation results cannot reach agreement (i.e., the same qualitative score), a senior cardiovascular radiologist (with > 15 years of experience in radiology) makes a final decision.

Quantitative evaluation: To evaluate aorta quality, five anatomical structures related to aortic disease diagnosis are chosen, as shown in Table I. Signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) are adopted to assess the noise level and smoothness of different anatomical structures in reconstructed CTA and LDCT, which are defined as follows [43]:

$$\text{SNR} = \text{mean}_1 / \text{std}_2$$

$$\text{CNR} = (\text{mean}_1 - \text{mean}_2) / [(\text{std}_1 + \text{std}_2) / 2] \quad (8)$$

where mean_1 , std_1 , mean_2 , and std_2 denote mean intensities and standard deviations of regions of interest (ROI) and background region respectively. Peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and root mean square error (RMSE) are also introduced to compare the image quality of reconstructed CTA and LDCT [44]. Because in aorta CTA imaging, radiologists adjust imaging parameters to adapt to diagnostic requirements, which results in aortic intensity variability among subjects. Aortic intensities of LDCT have no uniform standard, so we assess whether the reconstructed aorta are enhanced by judging their mean CT intensity can satisfy clinical requirement (i.e., 400 HU).

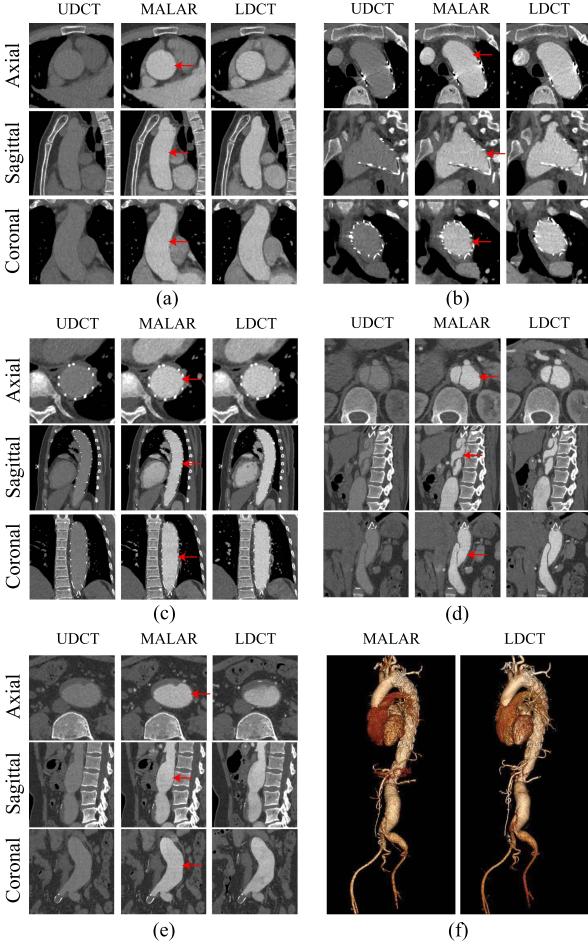


Fig. 4. The multi-view visualization of reconstructed aorta from a post-operative subject with metal stent and artifacts by comparing with UDCT and LDCT in five sites of (a) ascending aorta, (b) aortic arch, (c) thoracic aorta, (d) upper and (e) lower abdominal aorta. (f) shows the aortic volume rendering (VR) image of MALAR and LDCT. Because UDCT has a low-intensity aorta, it is difficult to obtain the aortic VR image. Red arrows indicate the enhanced aorta. Additionally, the MALAR columns are consistent with corresponding UDCT columns in morphological structures, while the LDCT columns are slightly inconsistent with the other two corresponding columns due to the organ motion of patients.

V. RESULTS AND DISCUSSIONS

In this section, we conduct comprehensive experiments to evaluate the feasibility and effectiveness of our MALAR on the angiography reconstruction for UDCT. The experiment results include the comparison with UDCT and LDCT, MALAR with single adversarial learning, MALAR without adaptive fusion module, MALAR without customized windowing, and the state-of-the-art methods.

A. Effectiveness of Our MALAR in Angiography Reconstruction

Fig. 4 shows the structural results in terms of ascending aorta, aortic arch, descending aorta, upper and lower segments of abdominal aorta from three orthogonal views. They further demonstrate the reconstruction advantages of our proposed MALAR by comparing with UDCT and LDCT results. As shown in

TABLE II
COMPARISON OF QUALITATIVE EVALUATION BETWEEN UDCT, RECONSTRUCTION RESULTS, AND LDCT BASED ON SELECTED STRUCTURES

Structures	UDCT	MALAR	LDCT	p-Value
Ascending aorta	2.0±0.5	3.9±0.5	4.4±0.6	<0.001
Aortic arch	2.1±0.5	4.0±0.4	4.6±0.5	<0.001
Thoracic aorta	2.1±0.5	3.9±0.5	4.3±0.5	<0.001
Upper abdominal aorta	1.8±0.6	3.8±0.5	4.4±0.6	<0.001
Lower abdominal aorta	2.0±0.5	3.7±0.5	4.3±0.6	<0.001

p-value denotes difference levels representing qualitative scores of MALAR and LDCT.

TABLE III
COMPARISON OF QUALITATIVE EVALUATION BETWEEN UDCT FOR INDEPENDENT VALIDATION, RECONSTRUCTION RESULTS, AND LDCT BASED ON SELECTED STRUCTURES

Structures	UDCT	MALAR	LDCT	p-Value
Ascending aorta	1.5±0.5	3.6±0.7	4.3±0.6	<0.001
Aortic arch	1.6±0.5	4.1±0.6	4.6±0.5	<0.001
Thoracic aorta	1.6±0.5	4.0±0.5	4.5±0.5	<0.001
Upper abdominal aorta	1.4±0.5	3.5±0.6	4.5±0.5	<0.001
Lower abdominal aorta	1.3±0.5	3.6±0.8	4.4±0.6	<0.001

p-value denotes difference levels representing qualitative scores of MALAR and LDCT.

Fig. 4, our MALAR can comprehensively enhance the aorta with excellent contrast, smooth vascular interior and normal vascular delineation. More importantly, our reconstructed aorta has similar CT intensity to the aorta in LDCT. Additionally, the volume rendering images in Fig. 4(f) also show the structural completeness of our MALAR results.

Table II, Fig. 3 and Table IV are qualitative and quantitative results, which demonstrate the distinguished aorta reconstruction performance of our MALAR in UDCT. The metrics in Table IV show quantitatively results in terms of CT intensity value (Attenuation), noise level (Image noise), aorta contrast (SNR, CNR), and similarity (PSNR, SSIM, RMSE). The average attenuations from different ROIs in our reconstructed aorta are about 400 HU, which are sufficient for providing accurate clinical diagnosis of aortic diseases. Although the average image noises are greater than 30 HU, they are still relatively smaller compared to average aortic intensity of 400 HU, and they cannot influence the clinical diagnosis. Indeed, the enhanced aorta structures have high SNRs and CNRs, which indicates the reconstructed images have higher smoothness and contrast. Additionally, the average PSNR, SSIM and RMSE are about 33, 0.970 and 90 HU respectively. These metrics indicate that the reconstructed CTAs have similar aorta quality with LDCT. Therefore, the reconstructed aorta by our MALAR have an adequate potential for clinical diagnosis of aortic diseases like LDCT. Table II and Fig. 3 show the qualitative scores by clinical radiographers according to 5-point scale (Section IV-C). For MALAR results in Table II, mean scores of selected structures are about 4, which indicate the reconstructed aorta are able to be used for diagnosis of aortic diseases. As shown in Fig. 3, only little results have low qualitative scores

TABLE IV
QUANTITATIVE RESULTS (MEAN \pm STD) ASSOCIATED WITH SELECTED STRUCTURES

Structures		Attenuation	Image noise	SNR	CNR	PSNR	SSIM	RMSE
Ascending aorta	MALAR	430.2 \pm 45.9	29.0 \pm 2.8	22.2 \pm 5.0	14.7 \pm 2.6	33.8 \pm 4.5	0.973 \pm 0.004	96.3 \pm 54.6
	LDCT	379.7 \pm 82.0	31.4 \pm 3.4	17.4 \pm 4.3	11.7 \pm 3.0			
Aortic arch	MALAR	417.9 \pm 44.1	32.5 \pm 4.9	21.0 \pm 4.0	13.4 \pm 2.2	34.4 \pm 4.4	0.974 \pm 0.005	89.4 \pm 51.9
	LDCT	389.1 \pm 79.0	31.2 \pm 4.1	17.1 \pm 4.1	12.0 \pm 2.8			
Thoracic aorta	MALAR	408.5 \pm 47.5	30.9 \pm 4.0	20.7 \pm 3.6	13.6 \pm 2.0	33.9 \pm 3.7	0.970 \pm 0.004	90.9 \pm 38.7
	LDCT	387.7 \pm 73.5	33.9 \pm 3.0	17.3 \pm 3.4	11.3 \pm 2.6			
Upper abdominal aorta	MALAR	370.1 \pm 57.9	34.1 \pm 4.7	18.7 \pm 4.4	11.3 \pm 2.3	33.7 \pm 3.2	0.965 \pm 0.005	90.5 \pm 34.7
	LDCT	393.6 \pm 75.7	37.6 \pm 3.7	17.9 \pm 4.1	11.1 \pm 2.5			
Lower abdominal aorta	MALAR	384.0 \pm 67.5	33.3 \pm 6.6	22.0 \pm 5.5	12.6 \pm 2.7	33.6 \pm 3.6	0.968 \pm 0.008	94.7 \pm 46.0
	LDCT	389.2 \pm 70.9	34.9 \pm 7.2	19.8 \pm 4.4	12.3 \pm 2.7			

Attenuation (HU) and image noise denote mean value and standard deviation of CT intensity in selected structures respectively.

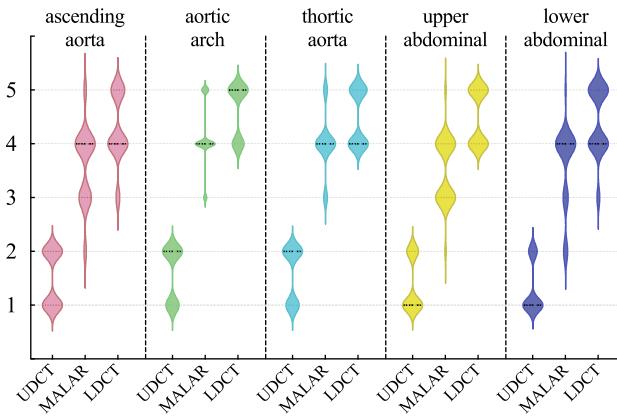


Fig. 5. Box-plot analysis for qualitative scores of UDCT for independent validation, MALAR, and LDCT. Mean scores of aortic quality of MALAR exceeds those of UDCT, and approaches to those of LDCT.

(score = 2) for their poor image quality in UDCT (too much noise, extremely low contrast). Table III and Fig. 5 describe the qualitative scores of independent validation dataset. In Table III, mean scores of selected structures by MALAR are larger than 3.5, and their standard deviations are smaller than 0.8. “Acceptable” scores (score = 3, as shown in Fig. 5) demonstrate the feasibility of our reconstruction results in clinical aorta diagnosis.

Additionally, our MALAR can be considered as a universal framework. The backbones of ResNet-18 and VGGNet can be replaced with other network structures, as shown in Table V. The adopted ResNet-50 and DenseNet should be reduced channel numbers to fit hardware memory. The last convolutional features of AlexNet and Inception-v3 are employed to calculate the DIC-loss. In general, the quantitative results by different backbones are close, and any of the listed networks can be used as the backbone network of our MALAR.

B. Effectiveness of Multiple Adversarial Learning

In this section, the effectiveness of our multiple adversarial learning is evaluated by comparing with the reconstruction

TABLE V
COMPARISON OF DIFFERENT CONFIGURATIONS IN OUR MALAR FRAMEWORK FOR ASCENDING AORTA. RESNET-50 [14] AND DENSENET [45] ARE USED TO REPLACE THE ADOPTED RESNET-18. ALEXNET [46] AND INCEPTION-V3 [47] ARE USED TO REPLACE THE ADOPTED 16-LAYER VGGNET

Configuration	SNR	CNR	PSNR	SSIM	RMSE
MALAR	22.2 \pm 5.0	14.7 \pm 2.6	33.8\pm4.5	0.973\pm0.004	96.3 \pm 54.6
ResNet-50	22.0 \pm 5.1	14.8 \pm 2.8	33.5 \pm 4.4	0.970 \pm 0.004	97.6 \pm 53.9
DenseNet	20.5 \pm 4.3	13.9 \pm 2.1	33.1 \pm 4.8	0.964 \pm 0.010	89.9\pm50.2
AlexNet	19.9 \pm 4.6	14.4 \pm 2.3	33.4 \pm 4.3	0.970 \pm 0.008	95.9 \pm 57.2
Inception-v3	22.3\pm5.0	14.8\pm2.5	33.6 \pm 4.5	0.973 \pm 0.006	96.6 \pm 55.3

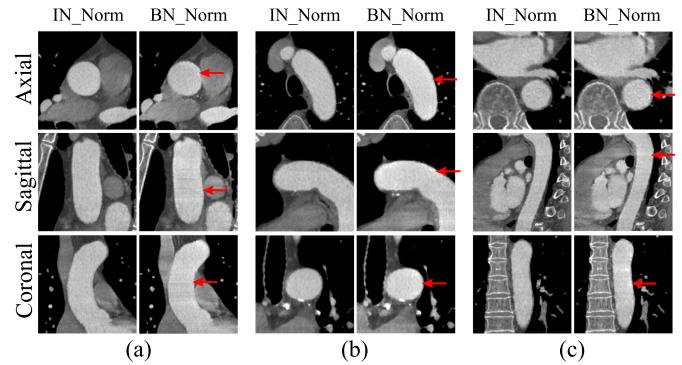


Fig. 6. The multi-view visualization of reconstructed aorta in three sites of (a) ascending aorta, (b) aortic arch and (c) thoracic aorta. IN_Norm and BN_Norm represent the reconstructions by our MALAR with instance normalization and batch normalization, respectively. Red arrows point to the enhanced aorta. Some brighter pixels exist in the reconstruction results of BN_Norm, resulting in less smoothness of reconstructed aorta than the results of IN_Norm.

method equipped with single adversarial learning. Our model introduces a dual adversarial constraint to learn the relationships of inter- and intra- domains. In this process, the constraint of inter-domain is essential to explore the relationship between UDCT and LDCT domains. Therefore, we need to evaluate the validity of intra-domain constraint. Specifically, the single

adversarial model is treated as the baseline, where only the discriminators D_{x1} and D_{y1} are used.

Fig. 7 shows aortic intensity changing results of the typical sagittal slices from two subjects, where the S-model represents the baseline with single adversarial learning. It can be observed that S-model results have significant intensity variation among slices. Ascending aorta in **Fig. 7(a)** and abdominal aorta in **Fig. 7(b)** occur segmental darker in some places. Aortic arch in **Fig. 7(b)** has relative low-intensity regions. As shown in line charts of **Fig. 7**, mean intensities (blue and red dotted lines) of slices in MALAR results are very close, while the intensity difference between slices in S-model results are much greater than the one in MALAR results. In our MALAR, it is difficult to explore the relationship between UDCT and LDCT domains by developing an image-to-image mapping. Because the image intensity of each subject in LDCT domain may be different, it will result in the instability of reconstruction direction. Fortunately, an intra-domain constraint is introduced into MALAR to learn inter-slice consistency for stabilizing reconstruction performance. In this way, our MALAR can further ensure the continuous and over-smoothness of image structure within 3D space. As shown in **Table IV**, the intensities of different aorta structures are about 400 HU. From the MALAR results in **Fig. 7**, the intensity fluctuation is greatly alleviated. The comparison of MALAR and S-model in **Fig. 7** demonstrates that the introduced intra-domain correlation constraint is effective in suppressing intensity fluctuation as well as providing high-fidelity reconstruction results.

C. Effectiveness of Adaptive Fusion Module

In this section, we evaluate the effectiveness of our adaptive fusion module by comparing our MALAR, MALAR without the adaptive fusion module and MALAR with different parameters. The proposed adaptive fusion module adaptively combines multi-scale local information and long-range dependency to improve aorta reconstruction. Therefore, we treat the model without long-range dependency as baseline for comparison, where the baseline only uses the multi-scale residual network. Additionally, to validate the effect of long-range dependency, multiple linear transformations (c) and head numbers (h) are adopted in multi-head self-attention for further comparison. These configurations are trained using the same strategy as our MALAR, and the visualization results are compared as shown in **Fig. 8** and **9**.

As depicted in **Fig. 8**, our MALAR performs better than baseline in aorta reconstruction (a-c. SNR-MALAR: 22.6, 20.2, 21.6, SNR-Baseline: 21.9, 18.3, 21.3. PSNR-MALAR: 34.8, 33.4, 34.5, PSNR-Baseline: 34.4, 32.7, 34.0), especially the sharpness of aortic borders and the intensity of regions nearby stents and artifacts. Specifically, in **Fig. 8(a)**, the aortic arch of baseline not only has noticeable noise, but also has poor smoothness in its border than the results of MALAR and LDCT. In **Fig. 8(b)**, partial region nearby metal artifact is failed to be reconstructed. In **Fig. 8(c)**, metal stent disturbs the reconstruction of its adjacent areas in baseline. To the best of our knowledge, aortic intensity is lower, and metal stent and artifacts usually have high intensity.

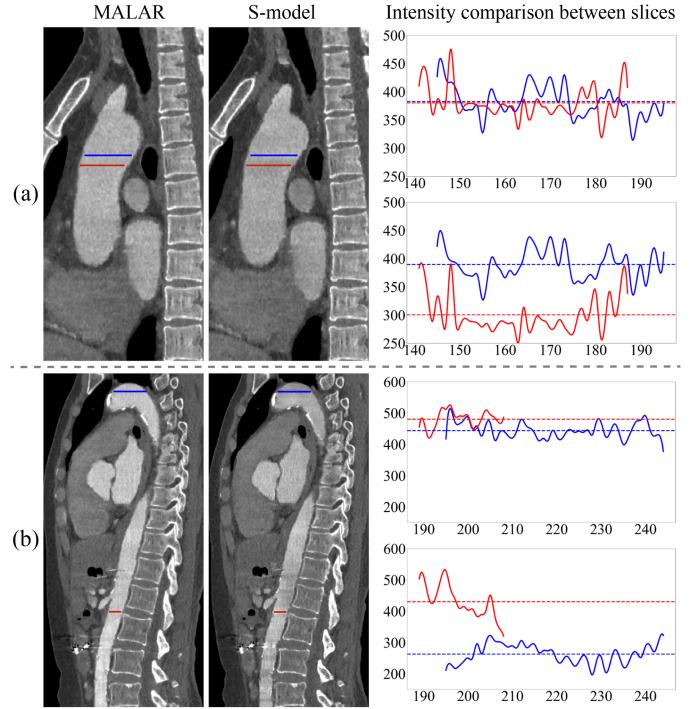


Fig. 7. The visualization comparison between our MALAR and S-model (MALAR with single adversarial learning). The figures are from different subjects, which are shown in sagittal view. (a) and (b) illustrate the results of ascending aorta, aortic arch, thoracic aorta, and abdominal aorta. The profiles on the right of each sub-figure denote intensity values along the marked lines on left aorta (upper profile: MALAR, lower profile: S-model). The horizontal red and blue dotted lines denote corresponding mean intensity values of different slices. x-axis represents the horizontal coordinates of the marks in left visualization images, and y-axis represents corresponding intensity values. For both visualization results and profiles of different axial levels, S-model results have greater intensity differences than our MALAR results.

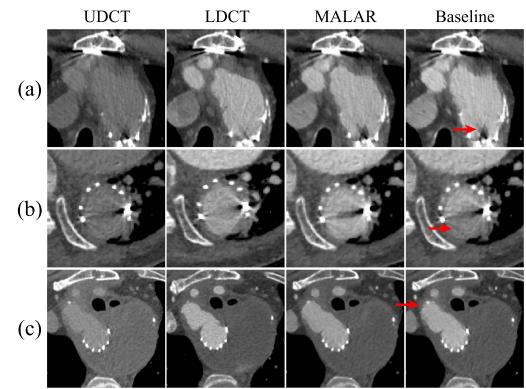


Fig. 8. The visualization comparison between our MALAR and MALAR with the only multi-scale residual network (Baseline). MALAR performs better than Baseline in the reconstruction of aortic border (as shown in a) and regions nearby metal (as shown in b and c). Red arrows point to regions that are significantly poor than MALAR.

Therefore, the intensities of aorta and metal vary toward to their mean value in computing MSE error in Equation (1). Long-range dependency is able to model interactions among image positions, and decreases the weight of non-relevant regions [15]. Consequently, the combination of local information and long-range

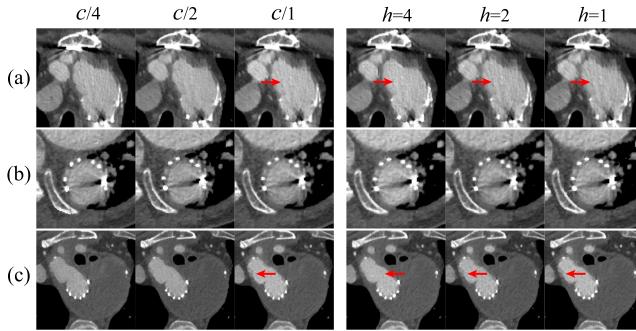


Fig. 9. The visualization comparison between our MALAR and MALAR with different parameters. c/num denotes the channel number after scaling num times in preliminary linear transformations. h denotes the head number in our multi-head self-attention. These results pointed by red arrows are slightly worse than MALAR in **Fig. 8**.

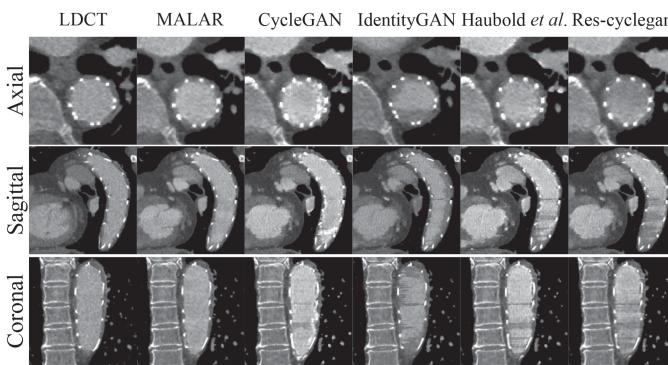


Fig. 10. The visualization comparison between our MALAR and state-of-art methods from axial, sagittal, and coronal views.

dependency can relieve the disturbance of high-intensity metal, and further promote aorta reconstruction.

Fig. 9 shows comparison of MALAR with different parameters. Our multi-head self-attention adopts linear transformation to scale channel numbers inspired by [40]. MALAR($c/4$) and MALAR($c/2$) do not have significant performance decrease compared with MALAR in **Fig. 8**, which scales channel numbers from c to $c/8$. For memory efficiency, our MALAR adopts the scale parameter of 8. MALAR($h = num$) have more streak artifacts than our MALAR in **Fig. 8**. In a multi-head self-attention layer, the keys, values, and queries come from different positions. Head number represents the complexity of information from different representation subspaces. To balance model complexity and information richness, the head number is set to 8 in our MALAR.

D. Effectiveness of Customized Windowing

In this study, we evaluate the effectiveness of our customized windowing technique by comparing MALAR without customized windowing and our MALAR. The proposed customized windowing changes intensity distribution in UDCT to highlight the intensity range of aorta, thus the proposed MALAR can easily train the reconstruction network. Therefore, we further treat the model without customized windowing as baseline for comparison, where the baseline directly normalizes UDCT and LDCT to input our MALAR network. Additionally, to demonstrate the superiority of our customized windowing,

multiple window levels and window widths are employed for further comparison.

As summarized in **Table VI**, our MALAR outperforms the baseline and other models with different window levels and window widths in terms of quantitative metrics. It numerically demonstrates the effectiveness of our customized windowing relative to other settings of window level and window width. For the results of our customized windowing in **Table VI**, aortic mean attenuations are about 400, and image noises are close to those of LDCT. However, unconstrained intensity range will result in larger attenuation values and higher noise levels of reconstructed aorta, as shown in **Table VI-w/o_wl_ww**. Additionally, $wl_350_ww_700$ has relative smaller intensity range and larger contrast, which causes that high-intensity pixels of LDCT input are more highlighted. These pixels guide the model to learn redundant features. Therefore, few highlighted pixels appear in the result images of $wl_350_ww_700$ and further lead to a higher noise level. In UDCT images, the raw intensity range is from -1024 HU to 3071 HU, while the contrast of aorta with neighboring muscle and fat is about 60 HU. It is different to distinguish between aorta and tissues under the overall intensity distribution. However, a customized windowing technique narrows the CT intensity range, which is equivalent to changing intensity distribution. Such a process makes it easier to identify and enhance the aorta in UDCT. The comparison of customized windowing and w/o_wl_ww in **Table VI** shows the customized windowing can improve our MALAR reconstruction performance.

E. Comparison With State-of-the-Art Methods

In this section, we evaluate our MALAR through the comparison with state-of-the-art methods, including CycleGAN [12], IdentityGAN [17], model of Haubold et al. [7] and Res-cyclegan [16]. The previous two methods are conducted with their original architecture, and the last two methods are realized by our own reimplementation based on their descriptions. They are trained and tested on our UDCT dataset.

Fig. 10 shows the multi-view comparison of different methods on a postoperative subject with metal stent. **Table VII** illustrates the quantitative comparison among different methods with evaluation metrics of SNR, CNR, PSNR and SSIM. For the results of CycleGAN in **Fig. 10**, unnecessarily highlighted pixels exist near the metal stents, which cause that the reconstructed aorta has a high noise level. Therefore, the quantitative metrics are inferior to our MALAR in **Table VII**. CycleGAN model does not integrate long-range dependency information, so it cannot effectively eliminate the interference of metal stents in angiography reconstruction task [12]. In the results of IdentityGAN, regions close to stents are not successfully reconstructed. The aortic region is notably not smooth, coupled with lower intensity, so the evaluation metrics showed in **Table VII** are significantly inferior to other methods. The two above-mentioned methods only consider the relationship between domains, so they lack reconstruction ability for aorta with high-intensity stents. Res-cyclegan introduces additional gradient difference loss and mutual information loss to establish domain relationship compared to CycleGAN [16]. It achieves slightly better quantitative

TABLE VI
EFFECTIVENESS OF CUSTOMIZED WIDOWING. $wl_num1_ww_num2$ INDICATES THAT THE MODEL CONDUCTS CUSTOMIZED WINDOWING WITH WINDOW LEVEL OF ‘ $num1$ ’ AND WINDOW WIDTH OF ‘ $num2$ ’ FOR INPUT

Configuration	Ascending aorta		Aortic arch		Thoracic aorta	
	Attenuation	Image noise	Attenuation	Image noise	Attenuation	Image noise
customized windowing	430.2±45.9	29.0±2.8	417.9±44.1	32.5±4.9	408.5±47.5	30.9±4.0
$wl_300_ww_800$	408.1±38.5	32.1±3.5	424.4±40.1	35.6±4.7	392.8±34.1	35.5±4.9
$wl_350_ww_700$	416.1±42.3	76.6±14.6	458.4±47.5	61.9±13.0	416.1±42.8	90.1±20.5
w/o- wl_ww	458.3±36.8	37.1±6.0	465.1±45.8	39.3±7.0	430.3±38.6	36.4±6.0
UDCT	141.8±18.9	19.5±1.8	142.9±18.8	20.8±2.6	143.1±18.3	22.7±2.5
LDCT	379.7±82.0	31.4±3.4	389.1±79.0	31.2±4.1	387.7±73.5	33.9±3.0

W/O- wl_ww represents the model without setting window level and window width for input.

TABLE VII
QUANTITATIVE EVALUATION (MEAN ± STD) ASSOCIATED WITH STATE-OF-THE-ART METHODS FROM SELECTED STRUCTURES

Structures	Metrics	UDCT	LDCT	MALAR	CycleGAN	IdentityGAN	Haubold et al.	Res-cyclegan
Ascending aorta	SNR	7.3±1.7	17.4±4.3	22.2±5.0	19.8±3.7	15.6±3.8	20.2±4.1	18.8±3.6
	CNR	4.3±1.1	11.7±3.0	14.7±2.6	11.5±2.2	9.3±2.0	11.8±2.7	10.2±2.0
	PSNR	/	/	33.8±4.5	30.8±3.9	28.9±4.2	31.8±4.3	30.6±3.6
	SSIM	/	/	0.973±0.004	0.949±0.005	0.941±0.007	0.959±0.005	0.953±0.007
Aortic arch	SNR	7.1±1.6	17.1±4.1	21.0±4.0	19.1±2.7	14.0±3.1	18.2±3.8	17.4±2.4
	CNR	4.1±1.1	12.0±2.8	13.4±2.2	12.3±1.6	6.7±2.0	12.0±1.9	10.3±1.7
	PSNR	/	/	34.4±4.4	31.6±3.8	29.2±4.7	32.6±4.0	33.0±3.6
	SSIM	/	/	0.974±0.005	0.953±0.003	0.947±0.009	0.962±0.004	0.962±0.007
Thoracic aorta	SNR	7.1±1.3	17.3±3.4	20.7±3.6	19.5±3.1	13.5±2.7	20.1±3.4	18.1±3.0
	CNR	4.1±0.9	11.3±2.6	13.6±2.0	12.0±2.2	5.4±1.4	11.7±1.7	10.8±1.7
	PSNR	/	/	33.9±3.7	31.1±3.5	28.1±3.7	32.1±3.7	30.7±2.9
	SSIM	/	/	0.970±0.004	0.948±0.004	0.942±0.008	0.954±0.005	0.950±0.004
Upper abdominal aorta	SNR	7.3±1.8	17.9±4.1	18.7±4.4	17.9±3.6	11.3±2.9	18.1±4.1	17.5±3.1
	CNR	3.8±1.2	11.1±2.5	11.3±2.3	11.4±2.3	5.6±2.3	9.9±2.2	8.7±1.7
	PSNR	/	/	33.7±3.2	30.0±3.7	26.7±5.0	30.9±3.3	30.4±3.6
	SSIM	/	/	0.965±0.005	0.944±0.015	0.933±0.015	0.951±0.012	0.947±0.010
Lower abdominal aorta	SNR	8.2±1.9	19.8±4.4	22.0±5.5	19.1±2.7	11.4±3.7	20.5±6.7	21.1±7.0
	CNR	4.4±1.2	12.3±2.7	12.6±2.7	10.1±3.6	5.1±1.8	9.7±3.3	10.1±3.0
	PSNR	/	/	33.6±3.6	30.5±4.0	27.5±5.1	29.7±4.3	30.1±3.9
	SSIM	/	/	0.968±0.008	0.948±0.035	0.934±0.017	0.953±0.026	0.950±0.019

results than CycleGAN as shown in Table VII. However, the model only trained and validated on a small dataset, which will result in a poor generalization. Haubold et al. and Res-cyclegan do not learn slices correlation, resulting in an inferior performance of larger intensity difference among slices. Additionally, both of them are implemented based on virtual data and not validated on true clinical CT data. Our MALAR combines multi-scale local features and long-range dependency, and shows superior reconstruction capabilities against high-intensity metal stent and artifacts. And it also introduces a dual correlation constraint for the integrity and smoothness of reconstructed aorta.

VI. CONCLUSION

In this paper, we proposed a multiple adversarial learning angiography reconstruction framework named MALAR to

enhance the aorta in ultra-low-dose-ICM CT angiography. The proposed MALAR adopted a customized windowing to optimize the intensity distribution of CTA to identify aorta. Then, it introduced a dual correlation constraint adversarial learning to solve the challenge of low-intensity aorta reconstruction, where integrates the multiple adversarial learning processes to effectively establish a mapping function from UDCT to unpaired LDCT. The compelling results obtained from ultra-low-dose contrast medium CTAs could demonstrate the feasibility and effectiveness of our MALAR in clinical diagnosis of aortic diseases.

The proposed method is one of the supervised learning models with unpaired data. It does not rely on accurately paired labeling from physicians, which saves significant time and workload. However, the proposed model still needs another high-intensity aorta sequence to guide the optimization direction of model. In the future, we plan to explore the possibilities of self-supervised

learning for aorta reconstruction without or reducing normal CTA sequences. In addition, our method can be also applied to other tissues reconstruction or enhancement (e.g., coronary artery, veins, and cardiac chambers) with full exploration, which is also listed on our future plans. In summary, our proposed MALAR will be significant for ultra-low-dose-ICM CT angiography.

REFERENCES

- [1] T. Yu, X. Zhu, L. Tang, D. Wang, and N. Saad, "Review of CT angiography of aorta," *Radiologic Clin. North Amer.*, vol. 45, no. 3, pp. 461–483, 2007.
- [2] G. D. Rubin, "Helical CT angiography of the thoracic aorta," *J. Thoracic Imag.*, vol. 12, no. 2, pp. 128–149, 1997.
- [3] K. Lakhal, S. Ehrmann, and V. Robert-Edan, "Iodinated contrast medium: Is there a re (n) al problem? A clinical vignette-based review," *Crit. Care*, vol. 24, no. 1, pp. 1–12, 2020.
- [4] M. Andreucci, R. Solomon, and A. Tasanarong, "Side effects of radiographic contrast media: Pathogenesis, risk factors, and prevention," *Pathophysiology*, vol. 4, no. 6, 2014, Art. no. 7.
- [5] M. Fähling, E. Seeliger, A. Patzak, and P. B. Persson, "Understanding and preventing contrast-induced acute kidney injury," *Nature Rev. Nephrol.*, vol. 13, no. 3, pp. 169–180, 2017.
- [6] E. I. Piechowiak, J.-F. W. Peter, B. Kleb, K. J. Klose, and J. T. Heverhagen, "Intravenous iodinated contrast agents amplify DNA radiation damage at CT," *Radiology*, vol. 275, no. 3, pp. 692–697, 2015.
- [7] J. Haubold et al., "Contrast agent dose reduction in computed tomography with deep learning using a conditional generative adversarial network," *Eur. Radiol.*, vol. 31, pp. 6087–6095, 2021.
- [8] M. J. Cha et al., "Hypersensitivity reactions to iodinated contrast media: A multicenter study of 196081 patients," *Radiology*, vol. 293, no. 1, pp. 117–124, 2019.
- [9] G. C. Kane, B. J. Doyle, A. Lerman, G. W. Barsness, P. J. Best, and C. S. Rihal, "Ultra-low contrast volumes reduce rates of contrast-induced nephropathy in patients with chronic kidney disease undergoing coronary angiography," *J. Amer. College Cardiol.*, vol. 51, no. 1, pp. 89–90, 2008.
- [10] G. D. Rubin, "CT angiography of the thoracic aorta," in *Multidetector-Row CT of the Thorax*, Berlin, Germany: Springer, 2016, pp. 311–340.
- [11] J. Golledge and K. A. Eagle, "Acute aortic dissection," *Lancet*, vol. 372, no. 9632, pp. 55–66, 2008.
- [12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
- [13] Q. Liu, L. Yu, L. Luo, Q. Dou, and P. A. Heng, "Semi-supervised medical image classification with relation-driven self-ensembling model," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3429–3440, Nov. 2020.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [15] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [16] H. Xie et al., "Generation of contrast-enhanced CT with residual cycle-consistent generative adversarial network (Res-CycleGAN)," *Proc. SPIE*, vol. 11595, 2021, Art. no. 1159540.
- [17] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, "Cycle-consistent adversarial denoising network for multiphase coronary CT angiography," *Med. Phys.*, vol. 46, no. 2, pp. 550–562, 2019.
- [18] C. Wang et al., "DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis," *Inf. Fusion*, vol. 67, pp. 147–160, 2021.
- [19] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 68–80.
- [20] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.
- [21] I. Bello, B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le, "Attention augmented convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 3286–3295.
- [22] J. Fu, J. Liu, J. Jiang, Y. Li, Y. Bao, and H. Lu, "Scene segmentation with dual relation-aware attention network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2547–2560, Jun. 2021.
- [23] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 108–126.
- [24] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10076–10085.
- [25] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–41, 2021.
- [26] C. Yu et al., "Multitask learning for estimating multitype cardiac indices in MRI and CT based on adversarial reverse mapping," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 493–506, Feb. 2021.
- [27] D. He et al., "Dual learning for machine translation," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 820–828, 2016.
- [28] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1857–1865.
- [29] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2849–2857.
- [30] I. J. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [31] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2794–2802.
- [32] P. W. Battaglia et al., "Relational inductive biases, deep learning, and graph networks," 2018, *arXiv:1806.01261*. [Online]. Available: <http://arxiv.org/abs/1806.01261>
- [33] L. Gatys, A. Ecker, and M. Bethge, "A neural algorithm of artistic style," *J. Vis.*, vol. 16, no. 12, pp. 326–326, 2016.
- [34] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.
- [35] J. Yang et al., "Focal attention for long-range interactions in vision transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 30008–30022.
- [36] M. H. Lev et al., "Acute stroke: Improved nonenhanced CT detection—benefits of soft-copy interpretation by using variable window width and center level settings," *Radiology*, vol. 213, no. 1, pp. 150–155, 1999.
- [37] P. Turner and G. Holdsworth, "CT stroke window settings: An unfortunate misleading misnomer?," *Brit. J. Radiol.*, vol. 84, no. 1008, pp. 1061–1066, 2011.
- [38] Z. Xue, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Window classification of brain CT images in biomedical articles," in *Proc. AMIA Annu. Symp. Proc.*, 2012, pp. 1023–1029.
- [39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.
- [40] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.
- [41] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.
- [42] Q. Zhang et al., "Toward replacing late gadolinium enhancement with artificial intelligence virtual native enhancement for gadolinium-free cardiovascular magnetic resonance tissue characterization in hypertrophic cardiomyopathy," *Circulation*, vol. 144, no. 8, pp. 589–599, 2021.
- [43] R. Usamentiaga, C. Ibarra-Castanedo, and X. Malague, "More than fifty shades of grey: Quantitative characterization of defects and interpretation using SNR and CNR," *J. Nondestruct. Eval.*, vol. 37, no. 2, pp. 1–17, 2018.
- [44] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369.
- [45] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
- [46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [47] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.