AEGIS: Using Conditional Multi-View Diffusion Model to Achieve Angiographic Enhancement in Non-Contrast CT

Jiahao Xia, Xiaolei Zhang, Yuting He, Yaolei Qi, Yutao Hu, Pascal Haigron, Chunxiang Tang, Longjiang Zhang, and Guanyu Yang, *Senior Member, IEEE*

***Abstract*—Angiographic enhancement of non-contrast CT (NCCT) using AI techniques is essential for diagnosing patients unable to use contrast agents. However, AI angiography remains a challenging task because of the feature fragility, structural complexity, and spatial continuity. In this paper, we propose an angiographic framework based on a conditional multi-view diffusion model called AEGIS with three innovations: multi-view hybrid learning (MHL), conditional angiographic diffusion estimation (CADE), and multi-view map fusion (MMF). 1) MHL targets Contrast Map (CM), the difference between NCCT and CT angiography, from multiple views to perceive 3D features in 2D space, enhancing the stability of feature representation. 2) CADE is a conditional diffusion model using NCCT as spatial guidance, providing crucial information for CM generation. 3) MMF adopts a lightweight AutoEncoder for filtering and fusing multi-view CMs, maintaining coherence between adjacent slices while modifying slight bias in low-dimensional representations, thus optimizing data quality and accuracy. Experiments demonstrate our superior performance, which achieve state-of-the-art image quality (PSNR+ 6.69, SSIM+ 3.17, MSE-46.38), segmentation evaluation (CADIR$\times$10.49, HSDIR$\times$5.57) and feature distance (FID-64.27). Visualizations and positive evaluation scores from clinicians further reveals that AEGIS has significant potential in clinical applications.**

***Index Terms*—Conditional diffusion model, AI-generated content, contrast agent free, 3D angiographic enhancement.**

Received 7 July 2025; accepted 3 December 2025. Date of publication 10 December 2025; date of current version 7 May 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 82441021, Grant 82441019, and Grant 82171933; in part by the Key Projects of the National Natural Science Foundation of China under Grant 82230068; in part by the Outstanding Youth Fund of Jiangsu Province under Grant BK20230048; in part by the Jiangsu Province Science Foundation for Youths under Grant BK20251279; in part by the Start-Up Research Fund of Southeast University under Grant RF1028624156; and in part by the Big Data Computing Center of Southeast University. This article was recommended by Associate Editor L. Li. *(Jiahao Xia and Xiaolei Zhang are co-first authors.) (Corresponding authors: Guanyu Yang; Longjiang Zhang.)*

Jiahao Xia, Yuting He, Yaolei Qi, Yutao Hu, and Guanyu Yang are with the Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications, Southeast University, Ministry of Education, Nanjing, Jiangsu 210096, China (e-mail: xia_jiahao@seu.edu.cn; yang.list@seu.edu.cn).

Xiaolei Zhang, Chunxiang Tang, and Longjiang Zhang are with the Department of Radiology, Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing 210002, China (e-mail: kevinzhlj@163.com).

Pascal Haigron is with Inserm, LTSI-UMR1099, University of Rennes, 35000 Rennes, France.

Data is available on-line at https://github.com/jiahaoxia-list/AEGIS

This article has supplementary downloadable material available at https://doi.org/10.1109/TCSVT.2025.3642730, provided by the authors.

Digital Object Identifier 10.1109/TCSVT.2025.3642730

I. Introduction

AI-BASED contrast enhancement to simulate contrast agents (CA) is critically important and has become an increasingly popular research topic [6], [36], [38], because the usage of chemical CA is strictly limited due to various objective reasons, although CA are widely used in clinical practice to enhance the visualization of CT and MR images [36], [40]. The contrast-enhanced CT (CECT) is significant to improve diagnostic accuracy compared with non-contrast CT (NCCT). Angiography, a technique where CA is usually injected intravenously, utilizes blood flow to enhance vascular regions, resulting in what is known as CT angiography (CTA). CTA is a kind of CECT and crucial for diagnosing cardiovascular, cerebrovascular, and various other conditions [26]. However, as illustrated in Fig. 1 (a) and (b), the use of contrast agents poses challenges for certain patient populations. Specific patients suffer allergic reactions to contrast agents or have metabolic or physiological conditions that preclude the use of chemical contrast agents [33], [34]. These limitations make contrast agents unsuitable for the specific patients, complicating the acquisition of CTA images and potentially hindering subsequent diagnostic and therapeutic efforts.

Recently, several studies have focused on using AI algorithms to generate contrast-enhanced images from NCCT scans for patients [24], [32]. This approach has shown considerable promise as it provides physicians with valuable enhanced references without the need for chemical contrast agents [20]. AI contrast enhancement, including AI angiography, eliminates the restrictions and risks associated with contrast agents, and it also avoids complications like accidental contrast agent leakage during the procedure. Patients are able to undergo a simpler and safer NCCT scan to obtain a corresponding CT AI-generated Angiography (CTAA), offering critical insights to aid in diagnosis [53]. Successful implementation of AI angiography could revolutionize clinical practice by providing significant benefits to both patients and healthcare providers, enhancing diagnostic capabilities while minimizing risks and costs. Although existing efforts have made some preliminary advances, current methods fall short in addressing our specific task - *AI angiography*. Previous methods can be categorized into two types: **Generative adversarial networks (GANs)** based [6], [13], [20], [22], [24], [32], [34],

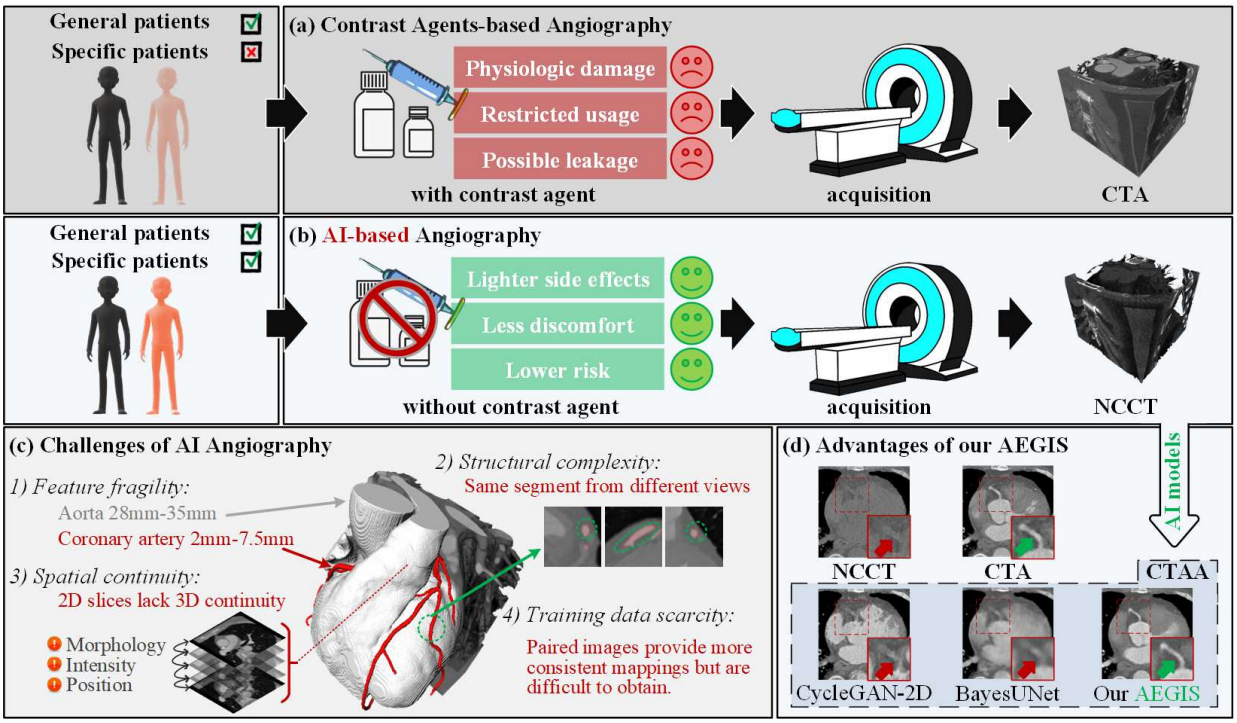

Fig. 1. Contrast agents-based versus AI-based angiography (e.g., coronary CT). (a) Contrast agents-based angiography enhances CT images by injecting a chemical contrast agent into the patient. (b) AI-based angiography relies solely on CT imaging, with subsequent AI techniques applied to achieve contrast enhancement. (c) Challenges in AI angiography include feature fragility, structural complexity, spatial continuity and training data scarcity. (d) Our AEGIS framework currently achieves state-of-the-art (SOTA) performance.

[38], [53] and **U-Net** based [10], [11], [61]. However, these methods are ineffective in accurately enhancing multi-scale structures especially with thin vessels (Fig. 1 (d)) due to the following challenges of AI angiography shown in Fig. 1 (c):

A. Feature Fragility

CT images contain rich intensity and spatial information, but the feature scales of different organs and tissues vary significantly, which makes it particularly challenging to extract fragile features in some small structures. For example, blood vessels occupy only a small portion of the scanned area, which are easily obscured by surrounding larger tissues, thereby reducing the representation capacity for small targets [14], [51]. AI algorithms struggle to accurately extract these fragile features, making the enhanced region unclear or incorrect, compromising diagnostic accuracy.

B. Structural Complexity

The structures of organs and tissues presents different features in different views (e.g., a vessel appearing circular in a transverse view may look elongated in a sagittal view) [23], [45]. Relying on a single two-dimensional view results in missing crucial details of these complex structures, leading to incomplete or inaccurate enhancement that does not fully represent the true morphology.

C. Spatial Continuity

CT images are continuous three-dimensional volume with strong correlations and spatial relationships between adjacent slices [52]. AI angiography from a single 2D plane results in discontinuities in 3D space.

D. Training Data Scarcity

High-quality, contemporaneous paired 3D NCCT and CTA data from the same patient provide algorithms with stronger mapping spatiotemporal consistency [54]. However, acquiring such datasets is challenging, and the time lag between NCCT and CTA acquisitions causes slight deformations [12]. Addressing training with such rare data remain a significant obstacle for AI angiography.

AI angiography involves an algorithm to adjust the intensity of specific regions within an NCCT scan, simulating the augmentation effect of chemical contrast agents. This process is similar to tasks such as colorization, brightness adjustment, and contrast adjustment in general image processing [31], [37], [60], [64], [67], and previous works have shown that employing generative models to perform these natural tasks is explainable and effective. In recent years, the introduction of diffusion models has injected new vitality into the field of general computer vision, demonstrating great potential in image generation, style transfer, and other applications [5], [27], [62]. However, applying diffusion models directly to AI angiography remains challenge. Medical data typically possess rigorous and complex 3D topological structures, but diffusion models operating in two-dimensional space struggle to learn high-dimensional feature information [44]. Moreover, the high computational complexity prevents the model from directly processing three-dimensional data.

To address the challenges mentioned above, in this paper, we propose an AI angiography framework based on a multi-view conditional diffusion model, called AEGIS. As far as we know, we are the first to enhance NCCT specifically targeting the complex and multi-scale structures, which incorporates strong distribution fitting ability of diffusion model and verify its

powerful capacity in medical imaging generation. The detailed contributions of include the following:

- We are the first to develop a method for angiographic enhancement of NCCT images. We also review relevant techniques of AI angiography, providing a valuable reference for future research in this field.
- We present a multi-view hybrid learning (MHL) strategy that emphasizes the features at the slice level by using the difference between paired NCCT and CTA data, i.e., contrast map (CM). MHL learns feature representation from multiple views in a hybrid manner to enhance the model's understanding of 3D CM and address the challenges of structural complexity.
- We introduce a conditional angiographic diffusion estimation (CADE) module, which encodes NCCT data as latent spatial prompts to provide continuous priori knowledge of NCCT and guide the decoding process along with fragile features. Furthermore, we derive the minimum timesteps in the CADE inverse sampling process, increasing inference efficiency by a factor of 5.
- We propose a lightweight multi-view map fusion (MMF) module to fuse CMs predicted by CADE from different views, improving the inter-slice continuity and the accuracy of representation in 3D space.
- We focus on AI angiography of coronary arteries and chambers in cardiac CT and conducted a comprehensive series of experiments on clinical dataset from five perspectives: image quality, coronary enhancement effect, cardiac substructure enhancement effect, rationality of model design, and clinical usability. Quantitative metrics and visualization results demonstrate that our AEGIS achieves SOTA performance in AI angiography, and the enhanced CTAA images are able to provide valuable assistance in clinical practice.

II. Related Work

A. Image-to-Image Task

The translation between NCCT and CECT data, which involve a certain change in pixel or voxel values preserving spatial relationships, can be regarded as a special form of Image-to-Image task. The applications of AI-generated content (AIGC) has extended across multiple domains and shown rich potential in various tasks [42]. Pix2Pix [8] and CycleGAN [15] realize style transfer of paired and unpaired images respectively, and achieved very good results in asonal change, image enhancement, and image denoising tasks.

Since Ho et al. [16] revitalized the concept of diffusion models and demonstrated their superiority over various GANs in natural image generative tasks, diffusion models have become a focal point for tackling complex generative problems. Diffusion models incrementally add noise to and subsequently denoise images via a Hidden Markov Chain to fit a particular distribution. Latent Diffusion Model [25] laid the foundation of Stable Diffusion, which greatly reduced the complexity of training. Palette [5] applied diffusion models to Image-to-Image tasks and achieved SOTA on colorization, inpainting, and other tasks. More and more text-to-image and image-to-image models, along with fine-tuning approaches based on diffusion models, have been proposed [27], [28], [35], [39], [43] and excelled in metrics related to style transfer tasks. Müller-Franzes et al. [30] observed that diffusion models generate more realistic and indistinguishable 2D fundus images, pathology slides, and X-rays than those produced by GANs. However, the work by Fang et al. [59], [66] indicated that coronary artery structures are highly complex, making them difficult to extract and analyze even in 2D X-ray images.

Meanwhile, with the introduction of the concept of diffusion models, numerous researchers [56], [57], [58], [63], [65] have adopted it for medical image analysis such as modality conversion, region enhancement, and brightness adjustment, all of which have shown effective visualization outcomes. However, despite the strong ability of diffusion models to fit target distributions stably and realistically in Image-to-Image tasks, applying them directly to 3D medical data still remain challenging due to their complex and redundant sampling process [50]. Thus, effectively adapting 2D diffusion models for 3D AI angiography is important but difficult.

B. AI Contrast Enhancement in Medical Imaging

Contrast agents and contrast enhancement play a crucial role in improving the readability of medical images, but traditional methods involving contrast agents are not viable for patients with allergies or metabolic disorders. Consequently, AI-based contrast enhancement has become a hot topic of research in medical image processing in recent years. Santini et al. [10] utilized a simple fully convolutional network, which showed some effects on large cardiac chambers but produced blurry results and neglected smaller vascular regions like coronary arteries. Pasumarthi et al. [18] applied a 2D U-Net on low-dose brain MRI scans and evaluated the efficacy by comparing Dice coefficients between real full-dose MRI and AI-enhanced MRI using a tumor segmentation model. Kleesiek et al. [11] directly enhanced 3D brain images using a 3D BayesUNet and concluded that predicting gadolinium contrast using AI algorithms is feasible, showing promise for AI-based enhancement techniques. Swin UNETR proposed by Hatamizadeh et al. [61]. demonstrates that replacing traditional convolutions with transformers achieves promising results in medical image segmentation tasks, while also exhibiting certain effectiveness in modality conversion. Müller-Franzes et al. [6] experimented on breast MRI using Pix2Pix and found that GANs could enhance images with low contrast but faced challenges with non-contrast images, indicating limitations in certain clinical applications. Song et al. [12] applied CycleGAN for the first time to achieve style transfer between CECT and NCCT in liver imaging and facilitated cross-modality segmentation. Xu et al. [13] cascaded GANs to synthesize late gadolinium enhancement images from non-contrast cine MR images, successfully achieving the segmentation of multiple tissues. Haubold et al. [20] employed Pix2Pix to systematically assess the permissible reduction of contrast agents during AI-enhanced contrast procedures, but the clinically acceptable reduction needs further validation. Besides, many other scholars [22], [24], [34], [36] have also conducted experiments and validations using data from various regions, such as the chest, abdomen, and brain, across different imaging modalities, yielding similar findings and

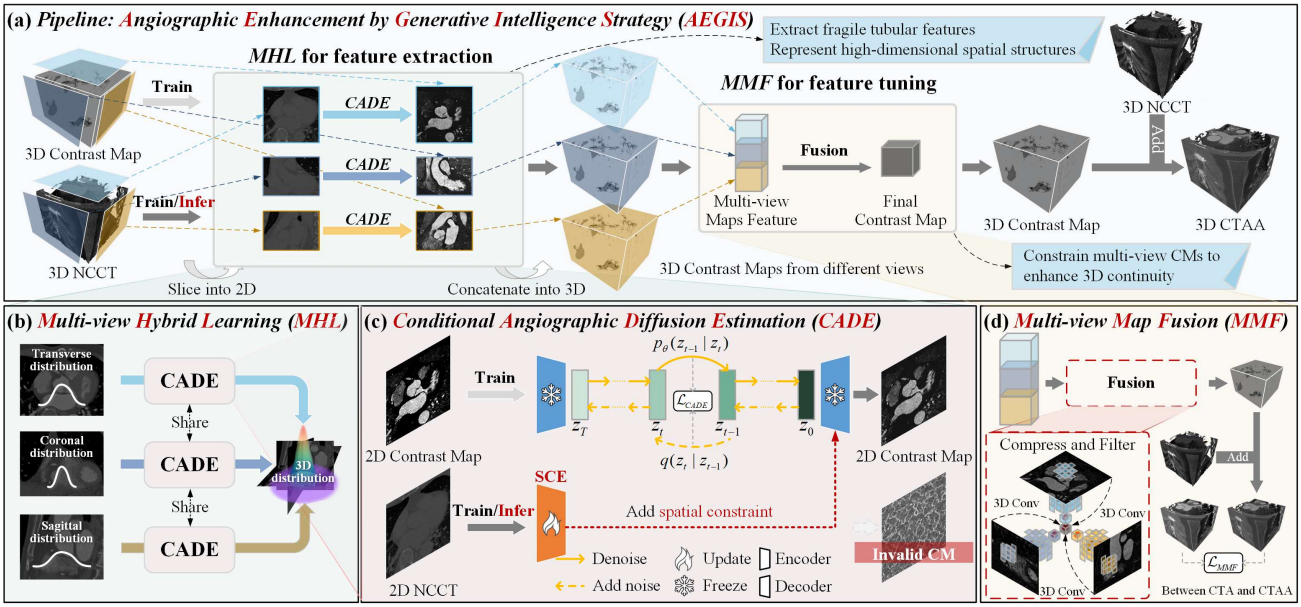

Fig. 2. The overall framework of AEGIS. (a) During the training process, AEGIS utilizes the CM as the target and NCCT as the spatial condition, whereas during inference, only NCCT is required. (b) MHL equips CADE with the ability to fit 3D spatial distributions by learning the 2D distributions of slices from multiple views. (c) CADE leverages NCCT images for spatial guidance through the diffusion process, and updates the parameters of Spatial-conditional Constraint Encoder (SCE). (d) MMF integrates multi-view CMs using an AutoEncoder to enable inter-layer feature interaction.

results. Nevertheless, these studies predominantly concentrate on larger structures (e.g., liver, aorta, cardiac chambers) rather than on smaller structures (e.g., coronary arteries).

Despite the progress made in these studies, several challenges remain. On the one hand, the effectiveness of GANs heavily depends on the discriminator's ability to differentiate between real and generated images, which is prone to make the convergence process arduous [2]. Without carefully selected hyperparameters and regularization strategies, GANs may experience collapse [17], [30]. On the other hand, methods based on U-Net architecture are proficient at feature extraction and fusion, leveraging skip connections to maintain underlying structural information. Nonetheless, direct sampling and reconstruction using U-Net often result in the loss of small structural details in deeper layers. Consequently, the final output tends to be blurred, making these methods less effective for our specific task where high quality and detail are critical.

III. Methodology

A. Overview

As shown in Fig. 2, AEGIS has three innovative designs. 1) MHL employs a weight-sharing strategy and uses the difference between the CTA and NCCT, named as Contrast Map (CM), as the training target for CADE. MHL enables the perception of 3D spatial information (Section III-C). 2) CADE module comprises a Stable Diffusion with frozen parameters and a Spatial-conditional Constraint Encoder (SCE) to be fine-tuned. SCE encodes the spatial information from NCCT images into prompts, which constrains the spatial structure of CMs (Section III-D). 3) MMF module consolidates the generated multi-view CMs by using a lightweight network to merge information from three views, enhancing the continuity and coherence of the CM in 3D space (Section III-E).

B. Training and Inference

Due to training data scarcity, we decouple the task into two stages and transform the insufficient 3D volumes into a sufficient number of 2D slices for effective training. In the first phase, we slice all paired 3D NCCT and CTA volumes along the transverse, coronal, and sagittal planes. These 2D slices are used to train CADE to accurately fit the distribution of the contrast-enhanced regions. After that, the 2D CMs are stacked into 3D volumes. The 3D CMs from three different views of the same data are concatenated along the channel dimension, providing the multi-view features. In the second phase, MMF fuses these multi-view features to produce a coherent and accurate 3D CM, which is then superimposed with the NCCT to output CTAA. This approach not only reduces training costs but also achieves excellent performance.

During inference, AEGIS requires only the 3D NCCT images as input, which serve as spatial guidance prompts for CADE. CADE autonomously predicts CMs from multiple views and then MMF optimizes these multi-view CMs to produce the CTAA without the need for CTA involvement.

C. Multi-View Hybrid Learning

MHL aims to allow CADE to perceive 3D features in 2D space, thus alleviating its inability to train directly on 3D medical data while utilizing the powerful feature extraction and distribution fitting capabilities of diffusion models to address structural complexity of AI angiography.

1) Definition of MHL: For the i^{th} 3D NCCT image $\boldsymbol{x}^i \in \mathbb{R}^{(D\times H\times W)}$ and its paired CTA image $\boldsymbol{y}^i \in \mathbb{R}^{(D\times H\times W)}$, we compute their CM by the formula $\boldsymbol{z}^i = \boldsymbol{y}^i - \boldsymbol{x}^i$, which then forms the triplet $\boldsymbol{D}^i = \langle \boldsymbol{x}^i, \boldsymbol{y}^i, \boldsymbol{z}^i \rangle$ in 3D space. Each triplet is sliced along three orthogonal planes, producing

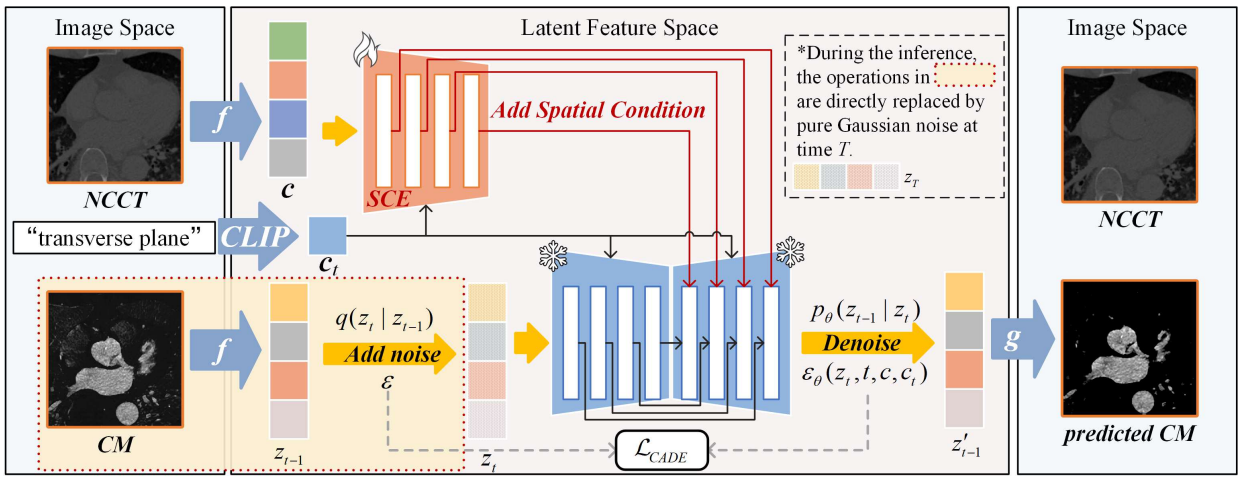

Fig. 3. Detailed Architecture of CADE. CADE translates NCCT and CM from the image space into the latent feature space and encodes NCCT images and text prompt as spatial condition using SCE. During training, CADE fine-tunes the parameters of SCE by a loss function derived from the frozen pre-trained Stable Diffusion. CADE then inversely maps the predicted features from the latent space back to the image space, generating predicted CM.

a slice-level 2D triplet $\boldsymbol{d}_v^{i,j} = \left\langle \boldsymbol{x}_v^{i,j}, \boldsymbol{y}_v^{i,j}, \boldsymbol{z}_v^{i,j} \right\rangle$, where j and v denote the j^{th} slice from the view $v \in \{transverse, sagittal, coronal\}$. CADE takes these $\boldsymbol{d}_v^{i,j}$ as training data, where $\boldsymbol{x}$ and $\boldsymbol{z}$ serve as spatial guidance and training target, respectively. To prevent overfitting to distributions from any specific view during the training process, all $\boldsymbol{d}_v^{i,j}$ are randomly sampled and mixed. Guided by $\boldsymbol{x}^i$, CADE outputs $\hat{\boldsymbol{z}}_v^i$ and each of these outputs captures the contrast features more accurately in its respective plane but may show inconsistencies in intensity between layers observed from other views. To address this, MHL concatenates the three CMs along the channel dimension to form a new vector, which describes the CM with 3D distribution by integrating the distinct 2D distributions from the three planes, i.e., multi-view features $\boldsymbol{Z}^i = \boldsymbol{z}^i_{transverse} \oplus \boldsymbol{z}^i_{sagittal} \oplus \boldsymbol{z}^i_{coronal}$, where $\oplus$ denotes $concat(\cdot)$.

2) Advantages of MHL: i) **Simplification of training target.** Most of the previous studies [32] have focused on establishing a direct mapping between NCCT and CTA, which is less effective when the target only occupies a minimal portion of space. In contrast, MHL redefines the target from CTA to the CM, which represents the intensity difference between NCCT and CTA. CM minimizes extraneous information, enhancing the representation of contrast-enhanced regions. ii) **Cross-dimensional feature perception.** Unlike methods that extract features solely from the transverse plane [24], [34], MHL processes data by slicing them along three orthogonal planes: transverse, sagittal, and coronal. This comprehensive approach compels CADE to efficiently fit the distribution of slices from all three planes, which enables CADE to achieve perception of 3D spatial distributions while learning from the 2D slices, thus bridging the gap between 2D and 3D.

D. Conditional Angiographic Diffusion Estimation

CADE is a crucial module in the MHL strategy, which aims to simulate the procedure of contrast agent by extracting the fragile spatial and intensity features based on a given NCCT image and predicting the regions to be enhanced as well as the possible amount of enhancement in these regions.

1) Using NCCT Image as Spatial Condition: As illustrated in Fig. 3, CADE is composed of three key sub-modules, including a Stable Diffusion pre-trained on natural images with frozen parameters [25], a Spatial-conditional Constraint Encoder (SCE) that shares the encoder structure and initial parameters of the Stable Diffusion, and a CLIP responsible for encoding image view information. During training, Stable Diffusion maps real CM to the latent space, adds noise iteratively, and then restores the CM from pure Gaussian noise. During inference, however, Stable Diffusion predicts the CM directly from pure Gaussian noise without incorporating any prior information from the NCCT or CM, making the results misalign with the actual spatial relationships present in the NCCT. To address this, SCE replicates the weight parameters from Stable Diffusion at the beginning and updates during subsequent iterations so that it can eventually encode the information of NCCT images into a spatially guided prompt. Consequently, the transformation $\langle \boldsymbol{\epsilon}, \boldsymbol{x} \rangle \rightarrow \boldsymbol{z}$ is necessary to achieve enhancement while ensuring more precise structural information of NCCT images compared with $\boldsymbol{\epsilon} \rightarrow \boldsymbol{z}$, where $\boldsymbol{\epsilon}$,$\boldsymbol{x}$, and $\boldsymbol{z}$ denote the pure Gaussian noise in the inference phase, NCCT images, and the predicted CM, respectively.

In each training iteration, the input 2D triplet $\boldsymbol{d}_v^{i,j}$ is encoded and decoded as detailed in (1). Here, $Encoder_{\mathrm{SD}}$, $Decoder_{\mathrm{SD}}$, and SCE refer to the Encoder and Decoder of Stable Diffusion, and SCE, respectively. $CLIP$ is an open-source, pre-trained text encoder. We use text descriptions $\boldsymbol{\tau} \in$ ["$transverse plane$", "$coronal plane$", "$sagittal plane$"] to represent the view of the input NCCT slice and encode them into textual features $\boldsymbol{c}_t$. q and $p_{\boldsymbol{\theta}}$ denote the processes of adding noise and denoising within latent space by a noise prediction model with parameters $\boldsymbol{\theta}$. The term $\boldsymbol{z}_t$ represents the CM feature in latent space after adding noise t times starting from the initial state $\boldsymbol{z}_0$, where $t \in [1, T]$. The variable $\boldsymbol{c}$ denotes the spatial condition derived from NCCT. $\mathcal{N}$ means Gaussian distribution, and $\boldsymbol{\mu}_{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ denote the mean and variance predicted by the noise prediction model based on the current information. The hyperparameter β changes with the iteration

step t, set to $\beta_1 = 10^{-4}$ and $\beta_T = 0.02$, following the settings proposed by [16]. $f(\cdot)$ and $g(\cdot)$ are the mapping functions that perform the transformation between the image space and the latent feature space, implemented by a pre-trained AutoEncoderKL [29]. The inference process involves only the last three steps, where $\boldsymbol{z}_T = \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$ is initialized as random noise from a standard normal distribution. $\hat{\boldsymbol{z}}_{t-1}$, $\hat{\boldsymbol{z}}_0$ and $\hat{\boldsymbol{z}}_v^{i,j}$ denote predictions.

$$\begin{cases} \boldsymbol{z}_0 = Encoder_{\mathrm{SD}}\left(f\left(\boldsymbol{z}_v^{i,j}\right)\right) \\ \boldsymbol{z}_t = q\left(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}\right) \sim \mathcal{N}\left(\boldsymbol{z}_t;\ \sqrt{1-\beta_t}\boldsymbol{z}_{t-1}, \beta_t \boldsymbol{I}\right) \\ \boldsymbol{c} = SCE\left(f\left(\boldsymbol{x}_v^{i,j}\right)\right) \\ \boldsymbol{c}_t = CLIP\left(\boldsymbol{\tau}_v^{i,j}\right) \\ \hat{\boldsymbol{z}}_{t-1} = p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t\right) \sim \mathcal{N}\left(\boldsymbol{z}_{t-1}; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t, t, \boldsymbol{c}, \boldsymbol{c}_t\right), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t, t, \boldsymbol{c}, \boldsymbol{c}_t\right)\right) \\ \hat{\boldsymbol{z}}_v^{i,j} = g\left(Decoder_{\mathrm{SD}}\left(\hat{\boldsymbol{z}}_0, \boldsymbol{c}, \boldsymbol{c}_t\right)\right) \end{cases}. \tag{1}$$

2) Training Loss of CADE: During the training process, the parameter $\boldsymbol{\theta}_{\mathrm{SCE}}$ of SCE is optimized according to the scheme shown as (2). The loss for CADE is shown as (3), where $\boldsymbol{\epsilon}$ is a random noise following a standard normal distribution and $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$ is the noise predicted by the denoising network with parameters $\boldsymbol{\theta}$ based on the current input information.

$$\begin{cases} \boldsymbol{\theta}_{\mathrm{SCE}} = \boldsymbol{\theta}_{Encoder_{\mathrm{SD}}} & \text{, initialize} \\ \boldsymbol{\theta}_{\mathrm{SCE}} \leftarrow \boldsymbol{\theta}_{\mathrm{SCE}} - \dfrac{\partial \mathcal{L}_{\mathrm{CADE}}\left(\boldsymbol{z}, \boldsymbol{c}, \boldsymbol{c}_t\right)}{\partial \boldsymbol{\theta}_{\mathrm{SCE}}} & \text{, update} \end{cases}. \tag{2}$$

$$\mathcal{L}_{\mathrm{CADE}} = \mathbb{E}_{\boldsymbol{z}_0, \boldsymbol{c}, \boldsymbol{c}_t, t, \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})}\left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}\left(\boldsymbol{z}_t, t, \boldsymbol{c}, \boldsymbol{c}_t\right)\|^2\right]. \tag{3}$$

3) Advantages of CADE: i) **More stable generation process.** Compared with GANs-based methods [38], diffusion models gradually learn from noise and generate targets, and such a step-by-step generation process is smoother and more stable than GANs. ii) **Richer prompt information.** In our task, the spatial relationship and intensity of tissues and organs in NCCT images fail to be described by textual prompts, while the images themselves often convey more information [37]. Inspired by ControlNet [39], we use NCCT as the image prompt for Stable Diffusion to provide spatial guidance for CM prediction, improving the precision of results.

E. Multi-View Map Fusion

MMF takes the multi-view features from CADE as input, enabling the CMs to impose mutual constraints on each other. This process reconstructs a 3D CM with greater integrity and coherence. MMF ensures the unification of continuity between neighboring slices at the spatial level and the accuracy of structural representation at the channel (view) level.

1) Lightweight 3D Refining Network With smoothL1Loss: The fusion and filtering of multi-view features $\boldsymbol{Z}$ to generate a more coherent and accurate CM is a feature reconstruction task. We achieve this using a lightweight AutoEncoder with only two stages, which is motivated by two primary reasons. i) By compressing and then decompressing the input features, it is able to filters out biased information present in the original multi-view CMs effectively, thus ensuring a tight integration and synthesis of the multi-view CMs into a unified, coherent output. ii) The use of a only two-stage structure not only reduces the loss of details due to oversampling in deeper AutoEncoder, but also reduces the necessity for data cropping or sliding windows, which will cause incomplete global information during training and inference. Due to the inevitable time gap between the acquisition of NCCT and CTA, compounded by movements like breathing and cardiac motion, there remains spatial discrepancy between the NCCT and CTA images, even after registration. These small differences in certain regions are currently unavoidable and must be tolerated to maintain model's generalization. To address these discrepancies and enhance the robustness of our architecture against outliers, we employ SmoothL1Loss [46] as the loss function for MMF, as defined in (4).

$$\mathcal{L}_{\mathrm{MMF}} = \begin{cases} \dfrac{1}{2}(\boldsymbol{y} - \hat{\boldsymbol{y}})^2, & \text{if } |\boldsymbol{y} - \hat{\boldsymbol{y}}| < 1 \\ |\boldsymbol{y} - \hat{\boldsymbol{y}}|, & \text{otherwise} \end{cases}, \tag{4}$$

where $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ denote the real CTA and the CTAA reconstructed through MMF, respectively. Notably, $\hat{\boldsymbol{y}}$ is obtained by a residual approach, i.e., $\hat{\boldsymbol{y}} = \boldsymbol{x} + \hat{\boldsymbol{z}}$. The AutoEncoder's output represent the CM of the enhanced region, but when calculating the loss function and evaluation metrics, MMF uses the CTAA obtained by superimposing the NCCT on the CM, accounting for the overall difference between CTAA and CTA.

2) Advantages of MMF: i) **Stronger reconstruction generalization capability**: MMF employs an AutoEncoder with SmoothL1Loss for reconstruction, which mitigates sensitivity to outliers during the feature filtering process, endowing the model with better robustness and reconstruction generalization. ii) **Lighter model architecture**: MMF utilizes a only two-stage AutoEncoder to filter and fuse multi-view features, reducing the loss of detail caused by oversampling. Compared to the slide-window, our lightweight MMF enables direct computation on 3D volume, further enhancing spatial continuity.

IV. Experiments Configurations

A. Evaluation Dataset

Due to the specificity of the task, to the best of our knowledge, there is no relevant public dataset available and we are working on open-sourcing the dataset. All methods in the experiments were trained and validated on a clinical dataset collected from Nanjing Jinling Hospital. The clinical dataset comprises non-contrast cardiac CT calcium score images and coronary CTA images from 282 patients, including 165 males and 117 females. The age distribution of the patients ranges from 18 to 90 years, with the majority concentrated between 50 and 70 years, as shown in Fig. 4. The acquisition intervals for NCCT and CTA data range from 48 seconds to 1816 seconds with a mean of 81 seconds. More detailed information about this dataset is presented in Table I. For each patient, NCCT images and CTA images were previously registered using Affine and B-Spline parameters facilitated by Elastix [47]. CT imaging data typically ranges from $[-1024\ HU, 1024 HU]$. However, artifacts such as metal implants or motion introduces extreme values. Thus, an offset of 1024 was add and then the thresholding was applied to remove outliers. For model compatibility, min-max normalization was also used. Then the 3D dataset was divided into

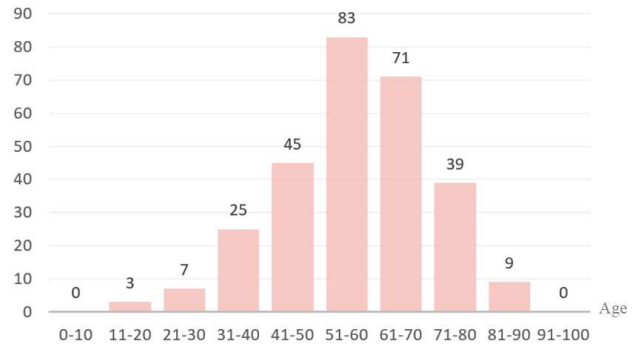

Fig. 4. Age distribution of patients in the clinical dataset.

TABLE I
DETAILED INFORMATION ABOUT THE CLINICAL DATASET

Attribution	Modality	
	NCCT	CTA
manufacturer	SIEMENS	SIEMENS
manufacturer model name	SOMATOM Definition Flash	SOMATOM Definition Flash
contrast bolus agent	/	APPLIED
slice thickness	1.0	0.75
kilovolt peak	120	100
rotation direction	CW	CW
generator power	20	64
photometric Interpretation	MONOCHROME2	MONOCHROME2

training, validation, and test sets with 7 : 1 : 2. All volumes have the size of $512 \times 512 \times D$, and D representing depth ranges from 180 to 598. The 3D volume in each subset was converted to 2D triples following the processing described in Section III-C1.

For subsequent segmentation verification, we utilized two available public datasets, ImageCAS [41] and MM-WHS [7], to pre-train the widely recognized nnU-Net [19]. We developed two specialized models: nnU-Net-Coronary and nnU-Net-Heart, optimized for the segmentation of coronary artery and whole-heart substructures, respectively.

B. Implementation

All experiments are based on the Pytorch framework and are performed on 2 NVIDIA GeForce RTX 4090 with 24 GB of memory. CADE uses AdamW as the optimizer with an initial learning rate of 10^{-5} and a total of 10 epochs of training. All slices in the sagittal and coronal planes are zero-padded or cropped during the training process to ensure a size of 512×512, while no size adjustment is performed during inference. We set the parameter *timesteps*, which affects the inverse denoising process, to 10 in the inference process, which improves the efficiency by a factor of 5 without degrading the performance compared with the default of 50. Considering the randomness of the initial noise of the generated model, we fixed the seed to 0. MMF is optimized using Adam with an initial learning rate of 2×10^{-4} and weight decay of 1×10^{-5}, and a total of 400 epochs of training are performed. MMF does not undergo any resize transformations during training and inference and does not use sliding windows. The number of training rounds for CADE and for MMF depends on the model convergence and the amount of training data (3D volume vs. 2D slice).

C. Evaluation Measures

To demonstrate the superiority of our AEGIS framework, we compared its performance with several previously utilized methods, including CycleGAN-2D [2], Pix2Pix-2D [6], CyTran [38], Pix2Pix-3D [22], CTA-GAN [53], RegGAN [55] and BayesUNet [11], and Swin UNTER [61]. Additionally, we evaluated AEGIS against advanced diffusion-based models such as ControlNet [39], UniControl [43], T2I-Adapter [35], Palette [5], CT2MRI [58], CC Net [65], Fast-DDPM [63], Self-RDB [56], and LighTDiff [57], which have achieved SOTA results in various Image-to-Image tasks. All models were trained or fine-tuned on the same dataset (2D slices for 2D methods and 3D volumes for 3D methods) under identical environmental conditions, following the configurations provided in their respective implementations. To validate the rationality of our design, we further conducted a series of ablation experiments. Consistent with prior research [9], [18], [20], we employed three image quality assessment metrics. To further assess the enhancement effects on blood vessels and cardiac substructures, we utilized the Dice Similarity Coefficient (DSC) improvement ratio for both coronary artery segmentation and cardiac substructure segmentation. To explore the relationships among NCCT, CTA, and the CTAA predicted by each method within the feature space, we also employed a distance measure to evaluate the distributional differences between these images [16]. Detailed definitions of the relevant metrics are provided below.

Mean Square Error (MSE) [49] is a metric that quantifies the voxel-wise discrepancy between two volumes by squaring the differences between the CTA and CTAA voxel values and averaging them. Larger MSE indicates greater discrepancy between the CTA and CTAA volumes.

Peak Signal-to-Noise Ratio (PSNR) [53] is a metric based on the MSE and the maximum value of the data to assess the information loss between the CTAA and CTA. Higher PSNR values signify better image quality, with a PSNR above 30 dB generally considered indicative of acceptable quality.

Structure Similarity (SSIM) [53] evaluates the contrast relationships between neighboring voxels. Higher SSIM means the CTAA is closer to the CTA in terms of three image features: luminance, contrast, and structure.

Coronary Artery Dice Increasing Ratio (CADIR) proposed by us measures the improvement in coronary artery segmentation accuracy provided by the CTAA relative to the NCCT. Using nnU-Net-Coronary, pre-trained on the ImageCAS dataset, CADIR simulates the coronary recognition of experts who have already acquired some prior knowledge when faced with these data. As shown in (5), $N(\cdot)$ denotes the segmenting process by nnU-Net-Coronary, and the DSC metric are calculated by $DSC(A,B) = 2 \times \left|A \bigcap B\right| / (|A| + |B|)$.

$$CADIR(CTA, CTAA) = \frac{DSC(N(CTA), N(CTAA))}{DSC(N(CTA), N(NCCT))}. \quad (5)$$

Heart Substructure Dice Increasing Ratio (HSDIR) proposed by us is analogous to CADIR but focuses on the segmentation of seven critical cardiac substructures - left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), myocardium (Myo), ascending aorta (Ao), and

TABLE II

THE PROPOSED AEGIS FRAMEWORK DEMONSTRATES SOTA PERFORMANCE ACROSS ALL SIX EVALUATION METRICS. THE "/" INDICATES METHODS THAT FAILED TO CONVERGE ON OUR TASK. FOR THE NCCT COLUMN, VALUES ARE COMPUTED BY SUBSTITUTING CTAA IN THE RESPECTIVE METRIC FORMULAS, HENCE NO RECORDS FOR CADIR AND HSDIR. UNDERLINED VALUES REPRESENT THE SECOND-BEST SCORES

Methods		Evaluation Metrics						
		PSNR ↑	SSIM(%) ↑	MSE ↓	CADIR↑	HSDIR ↑	FID ↓	Time(s)↓
NCCT		30.39 ± 0.97	94.05 ± 0.78	60.91 ± 14.04	-	-	80.95 ± 30.14	-
GAN-based	CycleGAN-2D [2]	35.42 ± 1.64	96.00 ± 0.84	20.11 ± 8.15	6.38 ± 5.78	4.20 ± 1.34	18.11 ± 10.47	23
	CyTran [38]	34.65 ± 1.66	95.81 ± 0.83	24.00 ± 9.51	0.36 ± 0.37	1.81 ± 0.92	55.70 ± 23.94	28
	Pix2Pix-2D [6]	/	/	/	/	/	55.87 ± 23.80	19
	Pix2Pix-3D [22]	34.93 ± 1.50	95.81 ± 0.75	22.23 ± 8.37	3.30 ± 3.59	3.73 ± 1.19	31.88 ± 17.49	5
	CTA-GAN [53]	35.68 ± 1.74	96.17 ± 0.79	19.18 ± 8.65	7.06 ± 6.42	5.09 ± 1.63	17.82 ± 10.72	21
	RegGAN [55]	35.23 ± 1.69	95.07 ± 1.22	21.15 ± 9.05	3.32 ± 3.77	4.13 ± 1.48	30.79 ± 18.08	35
U-Net-based	BayesUNet [11]	35.65 ± 2.01	96.73 ± 0.73	15.81 ± 8.27	5.13 ± 4.94	4.64 ± 1.51	25.84 ± 17.86	**2**
	Swin UNETR [61]	33.86 ± 2.12	94.97 ± 0.89	29.41 ± 10.13	2.45 ± 2.06	1.44 ± 0.32	42.79 ± 19.28	6
DM-based	ControlNet [39]	36.27 ± 1.62	96.79 ± 0.89	19.40 ± 8.09	7.70 ± 4.38	5.51 ± 1.82	17.49 ± 12.08	1014
	Palette [5]	35.93 ± 1.91	96.30 ± 1.03	18.36 ± 8.72	5.41 ± 5.67	5.07 ± 1.69	23.95 ± 13.54	1289
	T2I-Adapter [35]	36.07 ± 1.99	96.52 ± 0.80	17.90 ± 8.86	2.18 ± 2.06	3.79 ± 1.45	33.11 ± 22.13	1161
	UniControl [43]	35.61 ± 1.36	96.33 ± 1.30	20.70 ± 7.98	3.56 ± 3.89	5.48 ± 1.83	33.61 ± 21.77	1382
	CT2MRI [58]	35.47 ± 1.76	95.25 ± 1.45	22.70 ± 9.56	2.87 ± 2.41	1.03 ± 0.41	38.76 ± 17.92	872
	CC Net [65]	36.33 ± 1.92	96.61 ± 0.84	16.43 ± 8.49	7.92 ± 6.48	4.92 ± 1.43	23.47 ± 15.99	1854
	Fast-DDPM [63]	27.02 ± 4.93	94.13 ± 2.32	48.62 ± 13.58	0.87 ± 0.53	0.62 ± 0.26	65.15 ± 26.35	660
	SelfRDB [56]	36.18 ± 2.08	96.77 ± 0.73	14.88 ± 8.96	8.61 ± 6.44	5.43 ± 1.73	20.79 ± 14.22	1127
	LighTDiff [57]	36.40 ± 1.24	96.81 ± 0.80	15.89 ± 7.82	8.76 ± 7.92	5.53 ± 1.82	19.48 ± 12.70	254
	AEGIS (ours)	$\mathbf{37.08 \pm 2.20}$	$\mathbf{97.22 \pm 0.86}$	$\mathbf{14.53 \pm 8.00}$	$\mathbf{10.49 \pm 9.85}$	$\mathbf{5.57 \pm 1.88}$	$\mathbf{16.68 \pm 11.46}$	345

pulmonary artery (PA). Here, $N(\cdot)$ denotes the segmenting process by nnU-Net-Heart, and HSDIR quantifies the improvement ratio of the mean DSC across these substructures.

Fréchet Inception Distance (FID) [28] assesses the similarity between distributions in the feature space, based on the spatial relationship between two data distributions. Smaller FID indicates the closer distributions. As shown in (6), $N_{Encoder}(\cdot)$ refers to feature extraction using nnU-Net, which generates the vectors $\boldsymbol{V}_{\mathrm{A}}$ or $\boldsymbol{V}_{\mathrm{AA}}$ in the feature space. $Tr(\cdot)$ denotes the trace of the matrix, and the FID is calculated by the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of these vectors.

$$\begin{cases} \boldsymbol{V}_{\mathrm{A}} = N_{Encoder}(CTA) \\ \boldsymbol{V}_{\mathrm{AA}} = N_{Encoder}(CTAA) \\ FID(CTA, CTAA) = \left\|\boldsymbol{\mu}_{\boldsymbol{V}_{\mathrm{A}}} - \boldsymbol{\mu}_{\boldsymbol{V}_{\mathrm{AA}}}\right\|^2 + Tr\left(\boldsymbol{\Sigma}_{\boldsymbol{V}_{\mathrm{A}}} + \boldsymbol{\Sigma}_{\boldsymbol{V}_{\mathrm{AA}}}\right) - 2 \cdot \sqrt{\boldsymbol{\Sigma}_{\boldsymbol{V}_{\mathrm{A}}} \cdot \boldsymbol{\Sigma}_{\boldsymbol{V}_{\mathrm{AA}}}} \end{cases} . \quad (6)$$

V. RESULTS AND ANALYSIS

AEGIS ultimately achieves excellent AI angiography by capturing the complex and multi-scale structures in NCCT images. AEGIS predicts the location and intensity of enhancement regions with powerful fitting capabilities, while enhancing the spatial continuity of the data with a lightweight fusion network. In this section, we provide a comprehensive evaluation and analysis of AEGIS from four aspects. 1) First, we analyze the image quality of CTAA and the contrast enhancement effect by quantitative metrics (Section V-A) and qualitative visualization (Section V-B). 2) Secondly, we compare the feature distributions of CTAA obtained by different methods against NCCT and CTA in the feature space (Section V-C). 3) Then, we conduct ablation experiments on the multi-view strategy of MHL, the training target of CADE, the stage structure of MMF, and the number of timesteps in sampling process, to validate the rationality of our design and hyper-parameter choices (Section V-D). 4) Finally, we assess the practical application scenarios of AEGIS in terms of its clinical potential and usability on related tasks (Section V-E).

A. Quantitative Evaluation Shows Excellent Performance

As shown in Table II, AEGIS achieves SOTA performance on six metrics. The optimal results in PSNR, SSIM and MSE all indicate relatively higher image quality of CTAA generated by AEGIS. Compared to NCCT, the voxel value difference between AEGIS generated CTAA and real CTA is smaller, and the MSE is reduced by 46.38. For the comparison methods, CC Net, LighTDiff, and ControlNet also perform well on the metrics, but AEGIS outperforms them with a PSNR improvement of 0.68 and an SSIM improvement of 0.41. While Pix2Pix-2D exhibits strong enhancement capabilities in tasks involving larger structures, it fails to converge for our specific task. Moreover, Pix2Pix-3D converges but shows slightly less efficacy compared to diffusion model-based algorithms.

The CADIR metrics and HSDIR metrics obtained from the coronary segmentation task as well as on the whole heart substructure segmentation task respectively, simulate the variance in recognition difficulty for physicians when assessing NCCT, CTA, and CTAA. The higher CADIR and HSDIR suggest that the data are closer to the real CTA in terms of the discernibility of coronary and cardiac substructures. Compared with segmenting coronary arteries directly on the NCCT, CTA-GAN and ControlNet were able to improve the DSC metrics by a factor of 7.06× and 7.70×, and the cardiac substructures by a factor of 5.09× and 5.51×. AEGIS notably excels, achieving a boost of 10.49× in coronary and 5.57× in whole heart substructure segmentation. Further analysis of CADIR and HSDIR will be discussed in Section V-B.

In terms of feature space, using the features extracted from each type of data by nnU-Net-coronary, NCCT, and CTA are the furthest away from each other, with the FID as high as 80.95. In contrast, AEGIS, through the appropriate contrast enhancement, reduces the feature distance significantly,

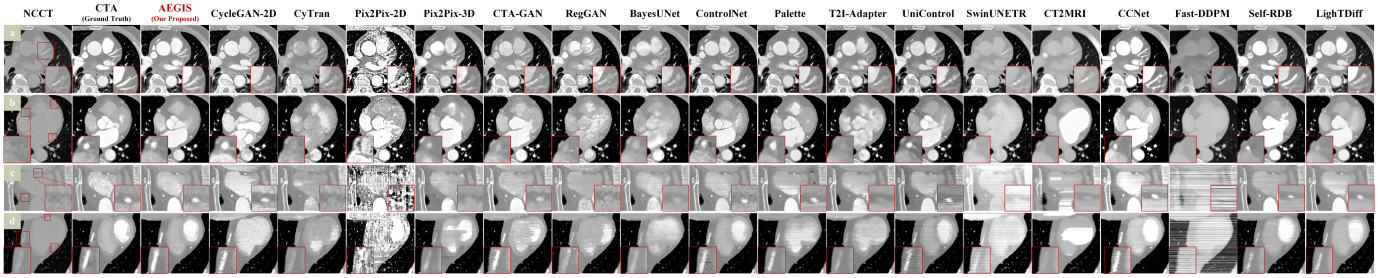

Fig. 5. Comparison of CTAA enhancement effects across different methods. (a) and (b) show the effects in the transverse plane, focusing on the LAD and LCX and the RCA. (c) and (d) illustrate the effects in the sagittal and coronal planes, respectively.

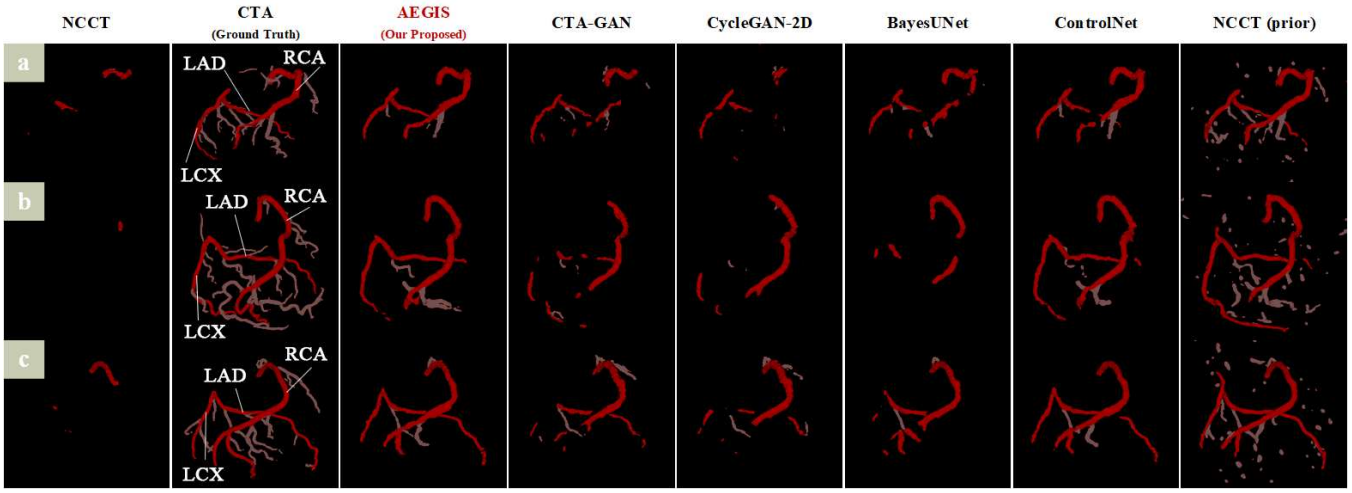

Fig. 6. Visualization of coronary segmentation. Using nnU-Net pre-trained on ImageCAS to segment coronary arteries of NCCT, CTA, and CTAA obtained by AI methods (AEGIS, CTA-GAN, CycleGAN-2D, BayesUNet, and ControlNet as examples) to compare their effects on LAD, LCX, RCA.

achieving a decrease in FID by 64.27. Further feature space analysis will be presented in Section V-C.

B. Qualitative Evaluation Shows Strong Clinical Potential

As illustrated in Fig. 5, our proposed AEGIS is able to provide CTAA that more closely matches CTA scans in both realism and accuracy. Clinically, the left anterior descending artery (LAD), the left circumflex artery (LCX) and the right coronary artery (RCA) are crucial for patient evaluation due to their significant roles in coronary circulation [48]. In Case (a) and Case (b), which depict the left coronary artery (LAD and LCX) and the right coronary artery (RCA) in the transverse plane, respectively, AEGIS excels in enhancing vascular structures of varying morphologies and orientations. Generally, the diffusion model-based approaches outperform others, showing significant enhancement in vascular regions to a certain extent and having similar visualization of cardiac chambers as CTA. AEGIS, in particular, demonstrates superior capability in accurately capturing the fine structures of coronary arteries, resulting in enhanced vascular regions with closest intensities compared to CTA. Conversely, GAN-based methods, such as CycleGAN-2D, while effective in enhancing the RCA, often lead to substantial deviations in other cardiac substructures. Also, the coronary artery region enhanced by BayesUNet is unsatisfactory and the images are blurred.

Case (c) and Case (d) further showcase the results from the sagittal and coronal planes, respectively, underscoring AEGIS's excellent performance in maintaining 3D continuity and spatial structure. Enhancement methods that operate solely on 2D transverse slices exhibit noticeable discontinuities and inconsistencies in intensity across adjacent slices when viewed from other planes. Furthermore, the complexity of the vascular structure is demonstrated by the elongated sections shown in Case (a) and Case (d) as well as the elliptical sections shown in Case (b) and Case (c). AEGIS learns slices from multiple perspectives and thus ensures the accuracy of such structures. Undeniably, although DM-based methods generally perform better, Fast-DDPM and CT2MRI demonstrate unsatisfactory performance. CT2MRI is originally designed for modality conversion between CT and MR, which exhibit significantly different feature distributions. Consequently, it underperforms on our task where most regions share similar characteristics with only subtle structural differences. As for Fast-DDPM, while it accelerates model inference by adjusting the number of sampling steps, it sacrifices stability and accuracy, leading to its failure in the AI angiography task.

As shown in Fig. 6, the performance of nnU-Net, pre-trained on CTA data, shows inferior segmentation results when applied directly to NCCT images. The red annotations represent the vascular segments of greater interest to the physicians, while the branches represented in gray have relatively lower research

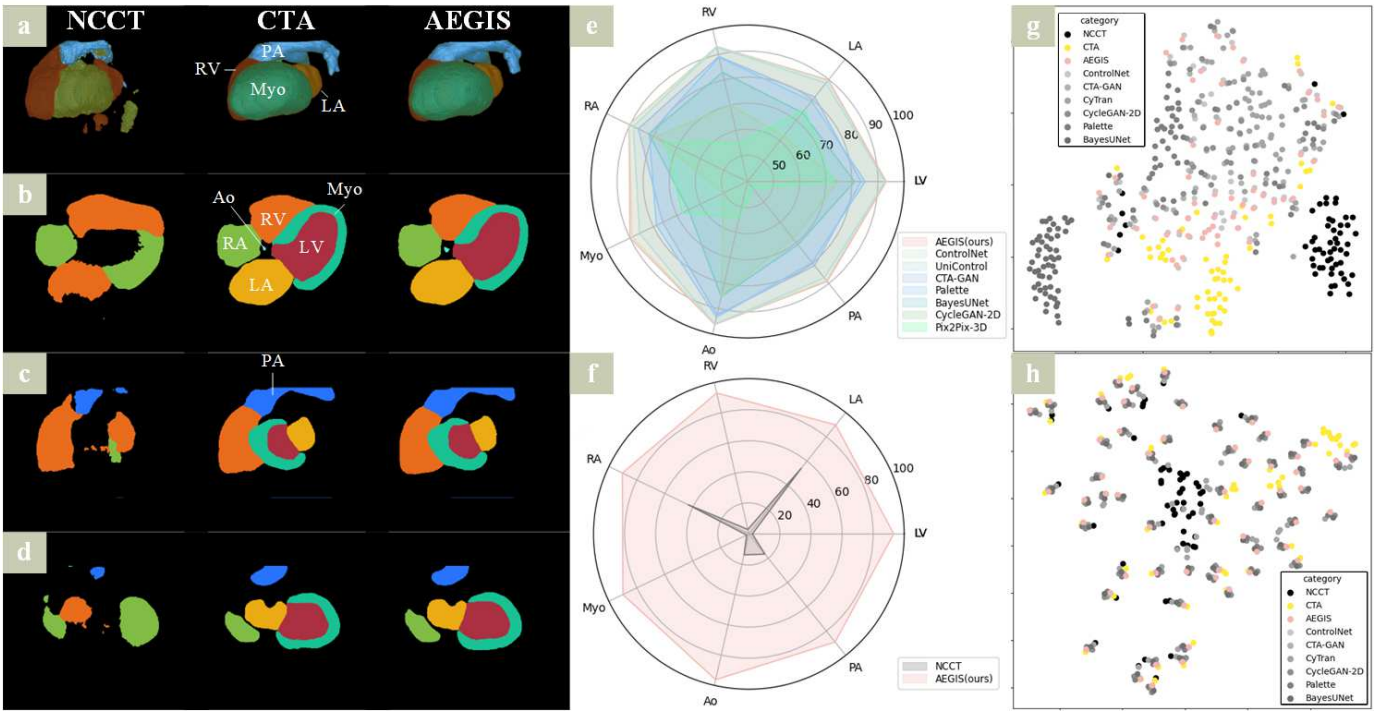

Fig. 7. Visualization of whole heart substructure segmentation results and scatter plots of feature spatial distribution. (a)-(d) present the 3D, transverse, sagittal, and coronal views of the heart substructure segmentation for a specific case, respectively. (e) displays a comparative radar chart of the DSC for segmentation results obtained by AEGIS and other methods across different targets. (f) illustrates the enhancement effects of AEGIS on seven substructures in comparison to NCCT. (g) and (h) provide scatter plots of features extracted from the intermediate hidden layers of nnU-Net-Coronary and nnU-Net-Heart, respectively, visualized in 2D space using t-SNE [4] for dimensionality reduction.

value. The figure underlines the challenges in segmenting NCCT due to its limited contrast between different tissues. In contrast, AEGIS greatly enhances the visualization and segmentation accuracy of the three main coronary branches. AEGIS not only enhances these primary arteries with high precision but also plays a certain enhancement effect on some finer branches. Comparatively, CTA-GAN, CycleGAN-2D and BayesUNet show some capability in enhancing vascular structures, however, their results exhibit a notable degree of discontinuity, leading to fragmented vessel segmentation results. ControlNet, while performing reasonably well, is limited by training on 2D slices only, leading to apparent disconnections. To further verify the superiority of AI angiography compared with the original NCCT, we also segmented the NCCT images using nnU-Net pre-trained on NCCT for coronary artery segmentation. As shown in "NCCT (prior)" of Fig. 6, due to the insufficient differentiation in Hounsfield Unit (HU) values between tissues, the nnU-Net model missegments regions outside the coronary arteries, whereas such problems are virtually absent in our AEGIS-generated CTAA.

AEGIS excels in extracting and enhancing vessel features from complex and poorly differentiated NCCT images. Through advanced algorithmic enhancement, AEGIS substantially improves the visualization of coronary vessels and their distinction from surrounding tissues, thereby offering a more precise and clinically relevant representation. As shown in Fig. 7, AEGIS shows superior performance in predicting the enhancement of the seven heart substructures alongside coronary angiography, achieving both optimal visualization and the highest HSDIR metrics. Targets for whole heart substructure segmentation include the left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), left ventricular myocardium (Myo), aorta (Ao), the pulmonary artery (PA). Fig. 7 (a)-(d) presents the 3D, transverse, sagittal, and coronal visualizations of the specific data segmentation outcome. It is evident that segmentation results on NCCT are substantially inferior, exhibiting numerous mis-segmentations and omissions, attributed to the low contrast and lack of clear boundaries between tissues in NCCT images. The segmentation results on AEGIS are greatly improved compared with NCCT and are almost the same as the segmentation results on CTA, which ensures accurate positional relationships among the substructures. (e) and (f) of Fig. 7 visualize the comparison of DSC metrics of CTAA obtained by different methods on the seven substructures in the form of radar charts, our AEGIS achieves the highest scores on each substructure, and all of them have considerable improvement compared with NCCT.

C. Feature Space Evaluation Shows Closer Distribution

To further investigate the relationships between NCCT, CTA, and CTAA generated by various methods in the feature space, we employed the t-SNE [4] to visualize the hidden features extracted by nnU-Net during both coronary segmentation and whole heart substructure segmentation. As shown in (g) and (h) of Fig. 7, the scatter plots demonstrate the distribution of features. The features of NCCT exhibit clustering due to low contrast and poor differentiation between tissues. In contrast, CTA, which enhances the visibility of vessels and

TABLE III

PERFORMANCE IMPROVEMENT OF AEGIS WITH MHL AND MMF. T, C, AND S DENOTE TRANSVERSE, CORONAL, AND SAGITTAL PLANES, RESPECTIVELY

Methods	Evaluation Metrics			
	PSNR↑	CADIR↑	HSDIR↑	FID↓
CADE (T-only)	36.37 ± 1.97	8.84 ± 4.43	5.50 ± 1.84	17.08 ± 11.68
CADE (T-only)+MMF	36.90 ± 2.12	9.51 ± 9.11	5.54 ± 1.85	16.98 ± 11.90
CADE (C-only)	35.72 ± 2.04	7.91 ± 4.27	5.48 ± 1.90	17.59 ± 11.82
CADE (C-only) + MMF	36.25 ± 1.95	8.62 ± 7.43	5.51 ± 1.87	17.04 ± 11.73
CADE (S-only)	35.66 ± 1.99	7.86 ± 4.19	5.48 ± 1.86	17.78 ± 11.91
CADE (S-only) + MMF	36.18 ± 2.01	8.59 ± 7.4	5.50 ± 1.86	17.10 ± 11.70
CADE (T+C)+MMF	37.04 ± 2.17	10.08 ± 9.41	5.56 ± 1.87	16.94 ± 11.56
CADE (T+S)+MMF	37.00 ± 2.16	10.14 ± 9.53	5.56 ± 1.86	16.95 ± 11.58
CADE (T+C+S)+MMF	$\mathbf{37.08 \pm 2.20}$	$\mathbf{10.49 \pm 9.85}$	$\mathbf{5.57 \pm 1.88}$	$\mathbf{16.68 \pm 11.46}$

other structures, shows a distinct separation between different tissue types, leading to a more dispersed distribution in the feature space. The enhancement of NCCT by various methods demonstrates a transformation from the clustered feature distribution towards a more dispersed one, aligning closer to the CTA's feature space. The results in (g) of Fig. 7 show that although several methods enhance the coronary arteries to achieve a more dispersed distribution, their distributions do not align closely with that of CTA. AEGIS, however, generates CTAA features that most closely resemble the distribution of CTA. (h) of Fig. 7, on the other hand, indicates that many methods achieve a feature distribution for cardiac substructures that is similar to CTA, which suggests that fitting larger structures like chambers is relatively easier compared with finer vascular structures. The visualization of the feature space generally matches the segmentation results and quantitative evaluation metrics, both of which confirm that AEGIS provides effective contrast enhancement on NCCT and achieves SOTA on fine structures such as coronary arteries.

D. Ablation Study and Model Analysis

1) Module Design Improves the Framework's Performance: Our proposed MHL strategy and MMF module have shown exceptional performance in the previously described experiments. To substantiate their effectiveness, we conducted validations as summarized in Table III. When CADE learns from transverse slices only, without the MMF to filter and integrate features, the performance metrics are relatively low. This is primarily because relying on a single view makes it challenging for CADE to capture the fragile features of various tissues in 3D space and to maintain spatial continuity. Upon introducing the MMF module, even with only the transverse perspective, the MMF enhances spatial continuity, leading to notable improvements in metrics. It holds consistently for both coronal and sagittal views, with quantitative analysis showing that the transverse plane contributes more substantially to CADIR compared to the other two orthogonal views. When coronal or sagittal slices are incorporated into the training process, the data benefit from spatial cross-referencing, enabling the MMF to perform structural corrections from multiple perspectives. Expanding the referenced views to include all three planes optimizes all evaluation metrics. Notably, the CADIR metric increases by 1.65 and the FID decreases by 0.4 compared with the initial CADE (T-only).

The proper utilization of NCCT, CTA, and CM is very important in the CADE training process. We investigated

TABLE IV

RATIONALITY OF CM AS A TRAINING TARGET AND NCCT AS A PROMPT

Methods	Evaluation Metrics			
	PSNR↑	CADIR↑	HSDIR↑	FID↓
CM w/o NCCT prompt	/	/	/	/
CTA as target	36.89 ± 2.12	9.40 ± 9.00	5.54 ± 1.88	17.06 ± 11.55
CM with NCCT prompt	$\mathbf{37.08 \pm 2.20}$	$\mathbf{10.49 \pm 9.85}$	$\mathbf{5.57 \pm 1.88}$	$\mathbf{16.68 \pm 11.46}$

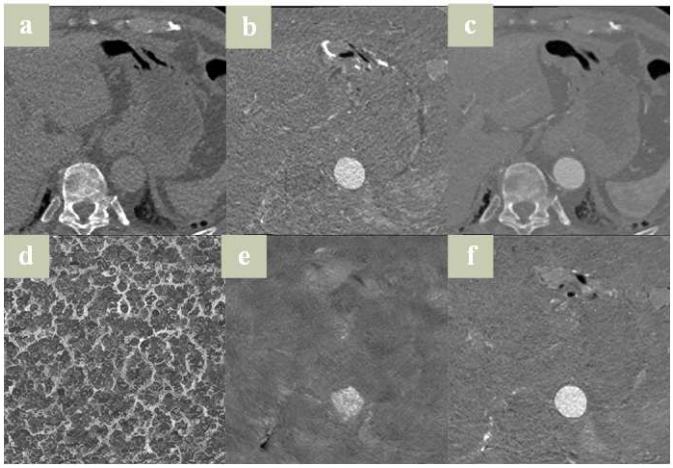

Fig. 8. Effect of guidance scale provided by NCCT on CM generation. (a)-(c) are the NCCT, CM, and CTA, respectively. (d)-(f) are the CMs obtained by the NCCT with guidance scale= 0, 0.5, and 1, respectively.

the differences between using CTA and CM as a generative target first. As shown in Table IV, although they produce similar results in terms of HSDIR metrics, using CM as the target yields an improvement of 1.09 CADIR. This suggests that directly using CTA as a target, while achieving similar enhancement for larger-volume substructures in the heart, struggles with capturing the very small percentage of coronary arteries due to the complex information within non-enhanced regions. CM, on the other hand, allows the model to concentrate more on fragile feature extraction. This is accomplished by removing most of the information from non-enhanced regions, increasing the relative proportion of small structures and reducing the learning complexity for the model.

However, the drawback of using CM as the generation target is that it also removes the rich spatial information from the non-enhanced regions in NCCT, leading to a lack of spatial constraints in the generated CM. To address this, we further incorporate NCCT as a spatial prompt. The results, shown in (d) of Fig. 8, reveal that this approach fails to reliably generate CM or CTA results from a given NCCT when we removed the NCCT prompt and instead directly denoised pure noise generated by corrupting the NCCT 1000 times. Combining the results of Table IV and Fig. 8, we conclude that the design of using CM as the training target, supplemented by NCCT as the spatial prompt, is both reasonable and effective.

The correlation between MMF structure and AEGIS performance is detailed in Table V. As the number of MMF stages increases, the hidden layer channels for feature extraction are adjusted to accommodate the requirements of downsampling and upsampling. Although more stages allow the model to extract features, with an excessive number of stages, the loss of information due to downsampling within the MMF also

TABLE V
IMPACT OF MMF STRUCTURAL VARIATIONS ON AEGIS PERFORMANCE

Hidden channels	Evaluation Metrics				
	PSNR↑	CADIR↑	HSDIR↑	FID↓	Params
1: [8]	36.67 ± 2.03	7.99 ± 7.42	5.52 ± 1.85	17.03 ± 11.92	$0.01M$
2: [8, 16]	37.08 ± 2.20	$\mathbf{10.49 \pm 9.85}$	5.57 ± 1.88	$\mathbf{16.68 \pm 11.46}$	$0.02M$
3: [8, 16, 32]	37.12 ± 2.21	10.43 ± 9.80	5.58 ± 1.90	16.70 ± 11.51	$0.06M$
4: [8, 16, 32, 64]	$\mathbf{37.18 \pm 2.24}$	10.33 ± 9.76	$\mathbf{5.58 \pm 1.89}$	16.81 ± 11.52	$0.23M$

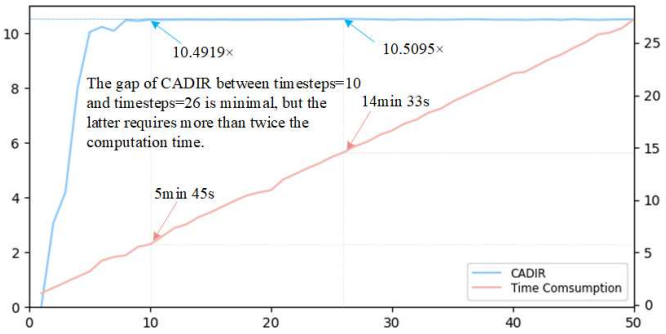

Fig. 9. Impact of timesteps on CADIR and time spent on inference for a single volume ($512 \times 512 \times 362$).

increases. Therefore, the 4-stage MMF it is 0.16 CADIR lower than the 2-stage. Meanwhile, adding the number of stages raises the model's parameter count and computational complexity, leading to higher memory and time requirements. Further, with more stages, the 4D multi-view features are also difficult to be fed directly into the model, which negatively affects the 3D spatial continuity. Consequently, we did not explore MMF with more than four stages. Balancing the metric improvements with efficiency, we concludes that the 2-stage MMF is the optimal configuration for our AEGIS.

2) Appropriate Reduction of Timesteps Boosts the Framework's Efficiency: As shown in Table II, we conducted a comparative evaluation of the inference runtime across all methods. All approaches were tested under identical conditions for AI angiography on volumes sized $512 \times 512 \times 362$. While GAN-based and U-Net-based methods generally exhibited faster inference speeds, they demonstrated inferior performance in terms of image quality and visualization outcomes. Among DM-based approaches, both Fast-DDPM and LighTDiff attempted algorithmic acceleration, but Fast-DDPM proved incompatible with our specific task. Although our application isn't designed for intraoperative use and thus has no stringent real-time requirements, we nevertheless optimized the inference speed to the greatest possible extent.

timesteps is a crucial hyperparameter of diffusion models for controlling the number of sampling steps, directly impacting both the quality and efficiency of generation. Setting *timesteps* too low results in insufficient denoising, whereas excessively high *timesteps* offers minimal quality improvement while boosting the computational time. As shown in Fig. 9, we investigated the changes in CADIR and time consumption as *timesteps* were raised from 1 to 50 (the default parameter value). The computational time is measured as the minutes required by AEGIS to perform contrast enhancement on a single NCCT volume with a size of $512 \times 512 \times 362$. As *timesteps* increase, CADIR boosts fast

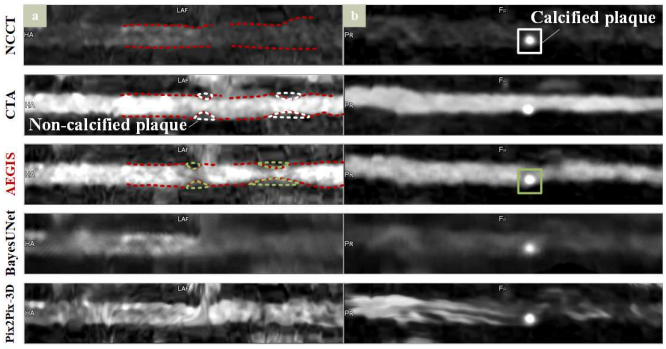

Fig. 10. Angiography in regions with calcified and non-calcified plaques. (a) shows the visualization of coronary stenosis caused by non-calcified plaques, while (b) demonstrates the retention of calcified plaques during enhancing NCCT by different methods.

and reaches a first peak value of 10.4919 at *timesteps* = 10. Subsequently, although CADIR fluctuates slightly and reaches an overall maximum value of 10.5095 at *timesteps* = 26, the corresponding time is more than doubled. Consequently, we select *timesteps* = 10 as the optimal hyperparameter, which balances the need for high-quality image enhancement with the necessity of reducing computational burden.

E. Clinical Validation Shows Practical Potential

Furthermore, we applied curved planar reformation (CPR) to all coronary data using manually annotated labels and extracted centerlines to observe stenotic lesions and assess the clinical applicability of AI angiography. As shown in Fig. 10, non-calcified plaques and calcified plaques represent two distinct but highly detrimental lesion types in coronary artery disease diagnosis. Calcified plaques often exhibit abnormally high intensity and are easily observable on NCCT, whereas non-calcified plaques can be located primarily through CPR on CTA. An effective AI angiography should retain the original calcification points and enhance the visibility of the non-calcified plaques. Due to the high continuity requirements of CPR processing, we selected Pix2Pix-3D and BayesUNet, both of which enhance NCCT in 3D space, for comparison. It should be noted that all enhancement experiments were conducted on NCCT, with CTA registered as closely as possible, leading to some spatial discrepancies in the CPR images.

Fig. 10 (a) demonstrates AEGIS's potential in distinguishing non-calcified plaques, showing increased intensity contrast between plaque and coronary regions compared with NCCT. In contrast, Pix2Pix-3D and BayesUNet not only tend to mis-enhance or miss-enhance but also produce more ambiguous results, making observation more difficult. In Fig. 10 (b), the calcified plaques are highly prominent in NCCT due to their higher intensity. Both AEGIS and Pix2Pix-3D retained the calcification points, whereas BayesUNet disrupted the original plaque characteristics during the sampling process. In conclusion, the CTAA obtained using AEGIS is able to meet the clinical requirements for lesion visualization enhancement, providing a feasible reference for patients unable to undergo CTA due to objective reasons to some extent.

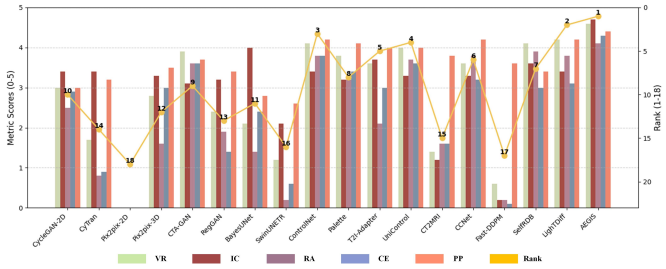

Fig. 11. Clinical evaluation results were obtained from five aspects along with the overall ranking: Visual Realism (VR), Inter-layer Continuity (IC), Region Accuracy (RA), Coronary Enhancement (CE), and Plaque Preservation (PP).

Moreover, we also invited two clinical experts to evaluate and score the CTAA from different methods. As shown in Fig. 11, where we assessed five metrics on a 5-point Likert scale (1=worst, 5=best): Visual Realism (VR), Inter-layer Continuity (IC), Region Accuracy (RA), Coronary Enhancement (CE), and Plaque Preservation (PP). AEGIS achieved the highest scores across all metrics, with an average score exceeding 4. Through random sampling and comprehensive quality ranking of the synthesized data from each method, AEGIS obtained the best rank among all approaches, which means that our method currently provides optimal clinical utility.

VI. Discussion and Conclusion

In this paper, we propose and verify AEGIS, a novel multi-view conditional diffusion model-based framework for AI angiography of NCCT. AEGIS introduces three innovations to address the challenges of feature fragility, structural complexity, and spatial continuity, as well as the difficulty of directly training diffusion models on 3D medical data. MHL leverages the difference between CTA and NCCT as the training target, and learns 2D slices from three different planes to enhance the extraction of 3D features in 2D space. CADE employs a conditional diffusion model using NCCT as a spatial prompt, imposing constraints during CM generation to fit more closely with the real spatial distribution. MMF finally integrates and filters the multi-view features with a lightweight AutoEncoder to optimize continuity between adjacent slices.

Experimental results confirm that the CTAA predicted by AEGIS more accurately reflects the clinical representation of real CTA than NCCT, demonstrating significant contrast enhancement in multi-scale structures such as coronary arteries and cardiac substructures. Ablation experiments further validate the rationale of MHL, CADE, and MMF in detail, demonstrating that multi-view hybrid learning helps 2D generative models fit 3D spatial feature distributions and that fusing multi-view features optimizes their continuity and detail accuracy. Furthermore, as illustrated in Fig. 10, the experiments also validated that our AEGIS is capable of preserving the calcification points in NCCT during the enhancement process and improving the visualization of non-calcified plaque, which shows considerable clinical value.

Future Work

Our AEGIS greatly improves the angiographic results and also shows the powerful capabilities of diffusion models, providing a promising solution for future research in this field. However, it is undeniable that the training and inference efficiency of diffusion models is still relatively low compared with GAN-based and U-Net-based algorithms, despite a series of adjustments and efforts to reduce the computational complexity (as shown in Table V and Fig. 9). Fortunately, the angiographic task of NCCT prioritizes image quality and accuracy over real-time performance, making it acceptable to invest more processing time to achieve superior CTAA. Meanwhile, ongoing research [1], [3] aims to develop faster and lighter diffusion models, indicating an important direction for future advancements. With improved hardware, this limitation may also be mitigated in the future.

Moreover, AI angiography is still in its infancy. As noted in previous work [11], [36], although it is able to enhance NCCT and provide some level of assistance, it is not yet capable of fully replacing chemical contrast agents. Further research and validation are needed to explore this potential in the future.

References

[1] Y. Li et al., "SnapFusion: Text-to-image diffusion model on mobile devices within two seconds," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 20662–20678.

[2] 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, Feb. 2019.

[3] Z. Wang et al., "Patch diffusion: Faster and more data-efficient training of diffusion models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 72137–72154.

[4] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.

[5] C. Saharia et al., "Palette: Image-to-image diffusion models," in *Proc. ACM SIGGRAPH Conf.*, Aug. 2022, pp. 1–10.

[6] G. Müller-Franzes et al., "Using machine learning to reduce the need for contrast agents in breast MRI through synthetic images," *Radiology*, vol. 307, no. 3, May 2023, Art. no. 222211.

[7] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, "A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI," *IEEE Trans. Med. Imag.*, vol. 29, no. 9, pp. 1612–1625, Sep. 2010.

[8] 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. (CVPR)*, Jul. 2017, pp. 5967–5976.

[9] H. Chen et al., "Low-dose CT with a residual encoder–decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[10] G. Santini et al., "Synthetic contrast enhancement in cardiac CT with deep learning," 2018, *arXiv:1807.01779*.

[11] J. Kleesiek et al., "Can virtual contrast enhancement in brain MRI replace gadolinium?: A feasibility study," *Investigative Radiol.*, vol. 54, no. 10, pp. 653–660, 2019.

[12] C. Song, B. He, H. Chen, S. Jia, X. Chen, and F. Jia, "Non-contrast CT liver segmentation using CycleGAN data augmentation from contrast enhanced CT," in *Proc. Int. Workshop Interpretability Mach. Intell. Med. Image Comput.*, 2020, pp. 122–129.

[13] C. Xu, L. Xu, P. Ohorodnyk, M. Roth, B. Chen, and S. Li, "Contrast agent-free synthesis and segmentation of ischemic heart disease images using progressive sequential causal GANs," *Med. Image Anal.*, vol. 62, May 2020, Art. no. 101668.

[14] Y. He et al., "Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101722.

[15] 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. (ICCV)*, Oct. 2017, pp. 2242–2251.

[16] J. Ho, A. N. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2024, pp. 6840–6851.

[17] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 8780–8794.

[18] S. Pasumarthi, J. I. Tamir, S. Christensen, G. Zaharchuk, T. Zhang, and E. Gong, "A generic deep learning model for reduced gadolinium dose in contrast-enhanced brain MRI," *Magn. Reson. Med.*, vol. 86, no. 3, pp. 1687–1700, Sep. 2021.

[19] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[20] J. Haubold et al., "Contrast agent dose reduction in computed tomography with deep learning using a conditional generative adversarial network," *Eur. Radiol.*, vol. 31, no. 8, pp. 6087–6095, 2021.

[21] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.

[22] J. W. Choi et al., "Generating synthetic contrast enhancement from non-contrast chest computed tomography using a generative adversarial network," *Sci. Rep.*, vol. 11, no. 1, p. 20403, Oct. 2021.

[23] Y. He et al., "Learning better registration to learn better few-shot medical image segmentation: Authenticity, diversity, and robustness," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2588–2601, Feb. 2024.

[24] T. Hu et al., "Aorta-aware GAN for non-contrast to artery contrasted CT translation and its application to abdominal aortic aneurysm detection," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, no. 1, pp. 97–105, Jan. 2022.

[25] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10674–10685.

[26] J. J. Cavallo and J. K. Pahade, "Practice management strategies for imaging facilities facing an acute iodinated contrast media shortage," *Amer. J. Roentgenology*, vol. 219, no. 4, pp. 666–670, Oct. 2022.

[27] Z. Li, Y. Zhou, H. Wei, C. Ge, and J. Jiang, "Toward extreme image compression with latent feature guidance and diffusion prior," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 1, pp. 888–899, Jan. 2025.

[28] M. Jian, R. Wang, X. Yu, F. Xu, H. Yu, and K.-M. Lam, "UniFRD: A unified method for facial image restoration based on diffusion probabilistic model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 12, pp. 13494–13506, Dec. 2024.

[29] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[30] G. Müller-Franzes et al., "Diffusion probabilistic models beat GANs on medical images," 2022, *arXiv:2212.07501*.

[31] W. Zhang et al., "Underwater image enhancement via weighted wavelet visual perception fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2469–2483, Apr. 2024.

[32] Z. Zhou et al., "Artificial intelligence–based full aortic CT angiography imaging with ultra-low-dose contrast medium: A preliminary study," *Eur. Radiol.*, vol. 33, no. 1, pp. 678–689, Jul. 2022.

[33] K. Higashigaito et al., "CT angiography of the aorta using photon-counting detector CT with reduced contrast media volume," *Radiol., Cardiothoracic Imag.*, vol. 5, no. 1, Feb. 2023, Art. no. 220140.

[34] A. Chandrashekar et al., "A deep learning approach to visualize aortic aneurysm morphology without the use of intravenous contrast agents," *Ann. Surg.*, vol. 277, no. 2, pp. e449–e459, 2023.

[35] C. Mou et al., "T2I-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 4296–4304.

[36] C. A. Mallio et al., "Artificial intelligence to reduce or eliminate the need for gadolinium-based contrast agents in brain and cardiac MRI: A literature review," *Investigative Radiol.*, vol. 58, no. 10, pp. 746–753, 2023.

[37] X. Wang, W. Wang, Y. Cao, C. Shen, and T. Huang, "Images speak in images: A generalist painter for in-context visual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 6830–6839.

[38] N.-C. Ristea et al., "CyTran: A cycle-consistent transformer with multi-level consistency for non-contrast to contrast CT translation," *Neurocomputing*, vol. 538, Jun. 2023, Art. no. 126211.

[39] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 3813–3824.

[40] G. Azarfar, S.-B. Ko, S. J. Adams, and P. S. Babyn, "Applications of deep learning to reduce the need for iodinated contrast media for CT imaging: A systematic review," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 18, no. 10, pp. 1903–1914, Mar. 2023.

[41] A. Zeng et al., "ImageCAS: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images," *Computerized Med. Imag. Graph.*, vol. 109, Oct. 2023, Art. no. 102287.

[42] S. Li et al., "Introduction to the special issue on AI-generated content for multimedia," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 6809–6813, Aug. 2024.

[43] C. Qin et al., "UniControl: A unified diffusion model for controllable visual generation in the wild," 2023, *arXiv:2305.11147*.

[44] C. Cao et al., "High-frequency space diffusion model for accelerated MRI," *IEEE Trans. Med. Imag.*, vol. 43, no. 5, pp. 1853–1865, May 2024.

[45] Y. Qi, Y. He, X. Qi, Y. Zhang, and G. Yang, "Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 6047–6056.

[46] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[47] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[48] S. Masrochah, J. Ardiyanto, R. Rasyid, and S. Nurbaiti, "Correlation of calcium scoring with image anatomy of cardiac angiography CT examination in coronary heart disease," *J. Phys., Conf. Ser.*, vol. 2498, no. 1, May 2023, Art. no. 012009.

[49] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Mach. Learn.*, vol. 36, nos. 1–2, pp. 105–139, Jul. 1999.

[50] A. Kazerouni et al., "Diffusion models in medical imaging: A comprehensive survey," *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102846.

[51] Y. Qi et al., "Examinee-examiner network: Weakly supervised accurate coronary lumen segmentation using centerline constraint," *IEEE Trans. Image Process.*, vol. 30, pp. 9429–9441, 2021.

[52] Y. He et al., "Meta grayscale adaptive network for 3D integrated renal structures segmentation," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102055.

[53] J. Lyu et al., "Generative adversarial network–based noncontrast CT angiography for aorta and carotid arteries," *Radiology*, vol. 309, no. 2, Nov. 2023, Art. no. 230681.

[54] Y. He et al., "Geometric visual similarity learning in 3D medical image self-supervised pre-training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 9538–9547.

[55] L. Kong, C. Lian, D. Huang, Z. Li, Y. Hu, and Q. Zhou, "Breaking the dilemma of medical image-to-image translation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 1964–1978.

[56] F. Arslan, B. Kabas, O. Dalmaz, M. Ozbey, and T. Çukur, "Self-consistent recursive diffusion bridge for medical image translation," 2024, *arXiv:2405.06789*.

[57] T. Chen et al., "LighTDiff: Surgical endoscopic image low-light enhancement with T-diffusion," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2024, pp. 369–379.

[58] K. Choo, Y. Jun, M. Yun, and S. J. Hwang, "Slice-consistent 3D volumetric brain CT-to-MRI translation with 2D Brownian bridge diffusion model," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Jun. 2024, pp. 657–667.

[59] H. Fang et al., "Greedy soft matching for vascular tracking of coronary angiographic image sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 5, pp. 1466–1480, May 2020.

[60] G. Gao, S. Yang, X. Hu, Z. Xia, and Y.-Q. Shi, "Reversible data hiding-based local contrast enhancement with nonuniform superpixel blocks for medical images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 2, pp. 1745–1757, Feb. 2025.

[61] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," in *Proc. Int. MICCAI Brainlesion Workshop*, 2022, pp. 272–284.

[62] Y. Hu, L. Zhang, X. Luo, and X. Cao, "Diffusion self-distillation for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5626315, doi: 10.1109/TGRS.2025.3569616.

[63] H. Jiang et al., "Fast-DDPM: Fast denoising diffusion probabilistic models for medical image-to-image generation," *IEEE J. Biomed. Health Informat.*, vol. 29, no. 10, pp. 7326–7335, Oct. 2025.

[64] X. Lu, Y. Yuan, X. Liu, L. Wang, X. Zhou, and Y. Yang, "Low-light salient object detection by learning to highlight the foreground objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 7712–7724, Aug. 2024.

[65] R. Osuala et al., "Towards learning contrast kinetics with multi-condition latent diffusion models," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2024, pp. 713–723.
[66] S. Song et al., "Spatio-temporal constrained online layer separation for vascular enhancement in X-ray angiographic image sequence," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3558–3570, Oct. 2020.
[67] G. Yue, J. Gao, R. Cong, T. Zhou, L. Li, and T. Wang, "Deep pyramid network for low-light endoscopic image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 5, pp. 3834–3845, May 2024.

Jiahao Xia received the B.S. degree in software engineering from Central South University, Hunan, China. He is currently pursuing the master's degree in computer science and technology with Southeast University, Nanjing, China. His research interests include computer vision, medical image analysis, and artificial intelligence-generated content.

Xiaolei Zhang is currently a Radiologist Investigator with the Clinical Research Center, Department of Radiology, Nanjing Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University. Her research interests include the application of multimodal imaging in the diagnosis, risk stratification, and prognosis assessment of cardiovascular diseases.

Yuting He received the B.E. degree from Hefei University of Technology, Hefei, China, in 2018, and the Ph.D. degree from Southeast University, Nanjing, China, in 2023. He visited Western University, London, Canada, in 2022. His research interests include developing novel methodologies to learn more efficient extraction of knowledge in medical information for computer-aided diagnosis, surgery, and medical imaging.

Yaolei Qi received the B.E. degree from Southeast University, Nanjing, China, in 2019, where he is currently pursuing the Ph.D. degree with the Laboratory of Image Science and Technology. His research interests include the intersection of deep learning and artificial intelligence in medical image processing, with a special focus on dealing with the tough task in tubular structure analysis that is knowledge-driven and data-efficient. His current research interests include innovative network design, weakly-supervised learning, and computer-assisted diagnosis.

Yutao Hu received the B.S. and Ph.D. degrees in electronics and information engineering from Beihang University, Beijing, China, in 2017 and 2022, respectively. From 2022 to 2024, he was a Post-Doctoral Research Fellow at HKU-MMLab, The University of Hong Kong. He is currently an Associate Professor with the School of Computer Science and Engineering, Southeast University. His research interests include machine learning and computer vision.

Pascal Haigron received the Ph.D. degree in signal processing and telecommunications from the University of Rennes 1, Rennes, France, in 1993. He is currently a Professor with the University of Rennes 1. He is also the Head of the Images and Models for Planning and Assistance to Therapy and Surgery (IMPACT) Team, LTSI-Inserm U1099, University of Rennes. His experience is related to 3-D shape reconstruction, virtual endoscopy, medical image registration, and image-guided therapy.

Chunxiang Tang is currently a Principal Investigator and the Executive Director of the Clinical Research Center, Department of Radiology, Nanjing Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University. Her research interests include the precise diagnosis and treatment of cardiovascular diseases from a multi-level perspective and the prevention and control of chronic diseases.

Longjiang Zhang is currently the Chair of the Department of Radiology, Nanjing Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University. His research interests include the clinical and translational research of molecular imaging and cardiovascular imaging.

Guanyu Yang (Senior Member, IEEE) received the B.S. and M.S. degrees in biomedical engineering from Southeast University in 2002 and 2005, respectively, and the Ph.D. degree in signal and image processing from Leiden University Medical Center (LUMC), Leiden University, The Netherlands, in 2008. He held a post-doctoral fellowship at Leiden University from 2009 to 2011. After returning to China in 2011, he joined Southeast University as a Faculty Member, where he is currently the Vice Dean of the School of Computer Science and Engineering, a Professor, the Doctoral Supervisor, and the Head of the Department of Imaging Science and Technology. In 2018, he was a member of the Medical Imaging Professional Committee of Chinese Society of Graphics.