

SCMA: Structure-Conditioned and Metal-Aware Flow Matching for CT Metal Artifact Reduction

Heran Wang, Jianing Sun, Xu Jiang, Genwei Ma, Xing Zhao, and Jigang Duan

Abstract—In X-ray computed tomography (CT), strong attenuation by metallic objects causes beam hardening, photon starvation, and scattering, making measured projections deviate from the ideal imaging model and producing streaks, dark bands, and structural distortions that compromise clinical diagnosis and quantitative analysis. Existing metal artifact reduction (MAR) methods remain limited: optimization-based methods may leave residual artifacts or blur structures, regression networks may generalize poorly across scenarios, and generative models without sample-specific structural guidance and physical constraints may produce anatomically inconsistent structures. Flow Matching learns a continuous-time velocity field that deterministically transports a source distribution to a target distribution, providing a flexible prior for MAR. However, standard unconditional Flow Matching does not exploit sample-specific structure, the spatially nonuniform nature of metal-induced degradation, or original projection measurements. To address these limitations, we propose SCMA, a structure-conditioned and metal-aware Flow Matching framework. First, a linear-interpolation-corrected image is jointly fed into the velocity network with the intermediate state as a sample-specific structural condition, guiding inference toward artifact-free CT images while preserving anatomy. Second, time-varying spatial weights derived from the metal mask and its distance transform are incorporated into the Flow Matching loss to emphasize severe degradation within and around metal regions. Finally, conditional Flow Matching updates alternate with projection-consistency correction during inference, allowing reliable measurements outside metal traces to constrain predictions. Experiments on simulated and real CT data demonstrate that SCMA more effectively suppresses metal artifacts, preserves local anatomical structures, and reduces hallucination-like structures inconsistent with projection measurements than representative MAR methods.

Index Terms—Computed tomography, metal artifact reduction, conditional Flow Matching, metal-mask-guided weighting.

I. INTRODUCTION

COMPUTED tomography (CT) has been widely used in medical diagnosis, radiotherapy planning, industrial nondestructive testing, and security inspection because of its

noninvasiveness, high spatial resolution, and strong structural depiction capability. When metallic objects are present in the scanned object, X-rays passing through the metal undergo severe attenuation, leading to beam hardening, photon starvation, and scattering. These effects cause projection measurements along metal-intersecting paths to deviate from the ideal imaging model, producing metal artifacts such as streaks, dark bands, and local structural distortions in reconstructed images. Metal artifacts not only reduce image readability and quantitative accuracy but also affect downstream tasks such as segmentation and registration. Therefore, effectively suppressing metal artifacts while preserving genuine structures remains an important problem in CT imaging [1], [2].

A. Existing Methods and Limitations

Traditional metal artifact reduction (MAR) methods mainly rely on projection completion, physical modeling, and regularized reconstruction. Projection completion methods, represented by linear interpolation (LI) [3] and normalized metal artifact reduction (NMAR) [4], typically identify metal traces in the projection domain from image-domain metal regions and then restore the corrupted projections through interpolation [5], normalization-based completion [6], or frequency decomposition [7], [8]. Physical modeling methods reduce model mismatch by explicitly accounting for polychromatic X-ray attenuation, beam hardening, and related effects [9], [10]. Regularized reconstruction methods improve reconstruction stability using iterative correction [11], total variation minimization [12], superiorized iteration [13], region-adaptive regularization [14], or sparse priors [15]. Although traditional MAR methods offer good physical interpretability, their performance is limited by the accuracy of metal-trace estimation, physical modeling, and manually designed priors. When the metal objects are large, numerous, or surrounded by complex structures, these methods may still produce residual artifacts, secondary artifacts, and structural blurring.

Recent deep learning methods learn nonlinear relationships between metal-corrupted data and artifact-free CT images in a data-driven manner. According to the processing domain, existing methods can be categorized as image-domain, projection-domain, and dual- or multi-domain approaches. Image-domain methods typically take artifact-affected images or conventional MAR results as input and directly predict corrected images [16]–[19]. Projection-domain methods mainly complete or correct corrupted data within the metal traces [20]–[22]. Dual- and multi-domain methods further combine image-domain structural information with projection-

This work was supported by the National Natural Science Foundation of China under Grant 12426308 and the Beijing High Innovation Plan, Capital High-End Leading Talents Aggregation and Cultivation Program, under Grant 202504841094. (Corresponding authors: Xing Zhao (e-mail: zhaoxing_1999@126.com) and Jigang Duan (e-mail: hiduanjigang@163.com).)

Heran Wang, Xu Jiang, Xing Zhao, and Jigang Duan are with the School of Mathematical Sciences, Capital Normal University, Beijing 100048, China (e-mail: 2250501041@cnu.edu.cn; chianghsu97@gmail.com; zhaoxing_1999@126.com; hiduanjigang@163.com).

Jianing Sun is with the College of Mathematics Science, Inner Mongolia Normal University, Hohhot 010022, China (e-mail: 20260057@imnu.edu.cn).

Genwei Ma is with the National Center for Applied Mathematics Beijing, Capital Normal University, and the Academy for Multidisciplinary Studies, Capital Normal University, Beijing 100048, China (e-mail: magenwei@126.com).

domain measurements to improve artifact suppression and structure preservation [23]–[29].

From the perspective of modeling paradigms, existing learning-based MAR methods mainly employ regression and generative models. Regression-based methods typically learn a deterministic mapping from corrupted data to artifact-free images. However, many of these methods rely on synthetically paired data, and the discrepancy between synthetic degradation and real measurements may limit their generalizability across scanners, acquisition protocols, and real-world scenarios [30], [31]. Generative methods provide stronger image priors by modeling the distribution of high-quality CT images or conditional restoration distributions. Generative adversarial networks [32], diffusion models [33]–[37], score-based generative models [38], [39], and other continuous generative models [40]–[42] have increasingly been applied to MAR and related CT image restoration tasks. However, when sample-specific conditions are insufficient or measurement constraints are absent from the generative process, these models may produce visually plausible structures that are inconsistent with the anatomy of the current sample or the original projection measurements. Therefore, generative MAR should exploit not only image-distribution priors but also sample-specific structural information and reliable projection consistency.

Flow Matching (FM) is a continuous normalizing-flow-based generative modeling method that learns a time-dependent velocity field to continuously transport samples from a source distribution to a target distribution through a deterministic ordinary differential equation [43]. Rather than learning a one-step mapping from degraded images to target images, FM learns the local transport direction of each intermediate state along a continuous probability path, providing a direct training objective, an explicit generative trajectory, and a controllable inference process. However, standard unconditional FM provides only a population-level prior over the target image distribution. The transport from a Gaussian distribution to artifact-free CT images neither describes metal-induced degradation nor incorporates the structure, metal location, or original projection measurements of the current sample. Therefore, a continuous generative path alone does not make FM a restoration model tailored to MAR. Applying FM to CT metal artifact reduction requires sample-specific structural conditions, metal-region priors, and CT projection measurement constraints.

B. Motivation

The objective of CT metal artifact reduction is not merely to generate visually plausible artifact-free images, but to recover the genuine structure of the current sample in severely degraded regions while maintaining consistency with reliable projection measurements. Based on this objective, we adapt Flow Matching to MAR from three perspectives: sample-specific structural conditioning, metal-region learning, and projection measurement constraints.

First, as shown in Fig. 1(A), standard FM starts from a Gaussian sample $z \sim \mathcal{N}(0, \mathbf{I})$, where $\mathbf{I}$ denotes the identity matrix, and transports it toward the artifact-free image distribution along the learned velocity field. Without a condition

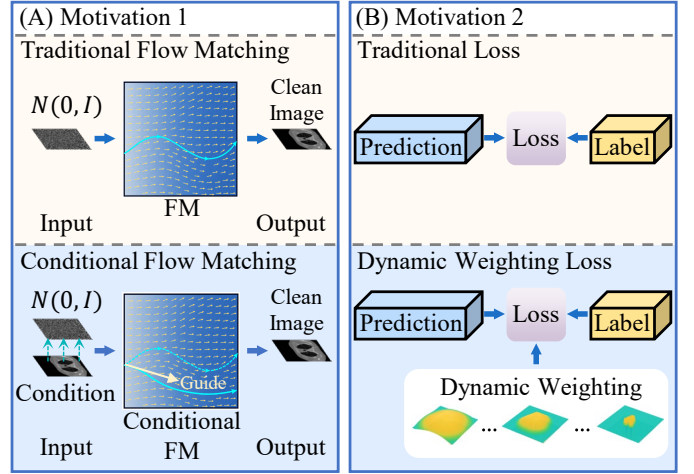


Fig. 1. Two Flow Matching designs tailored to MAR in SCMA. (A) LI-guided conditional Flow Matching uses an LI image to provide a coarse structural condition for the current sample. (B) Metal-region dynamic weighting constructs a time-varying two-dimensional pixel-wise weight map using the metal mask and distance information.

derived from the current sample, the model must infer the complete structure solely from the population-level image prior, introducing uncertainty into the restoration. Although the LI image still contains interpolation errors, residual artifacts, and structural blurring and therefore cannot serve as the final corrected result, it retains the coarse structure of the current sample. We thus jointly feed the intermediate state x_t and the LI image x_{Linear} into the velocity network at each time point, as indicated by the orange arrow. In this way, the FM trajectory is transformed from unconditional population-level generation into a conditional restoration process constrained by the structure of the current sample.

Second, metal artifacts are spatially nonuniform. Severe degradation is typically concentrated within and around the metal regions, whereas areas farther from the metal are relatively reliable. A spatially uniform FM loss can be dominated by large areas with mild degradation, weakening the model’s ability to learn the severe artifacts near the metal. As shown in Fig. 1(B), we construct a two-dimensional pixel-wise weight map W_t at each time point using the metal mask M and its distance-transform map D , and apply it to the velocity prediction error. During the early stage near the noise endpoint, the weighting covers a broader neighborhood around the metal to enhance the learning of large-scale streaks and dark bands. As the state approaches the artifact-free image endpoint, the weighted region gradually contracts, allowing the model to focus on the metal regions and their immediate neighborhoods while reducing disturbances to reliable structures farther away.

Finally, relying solely on an image-domain generative prior may still produce hallucination-like structures inconsistent with the CT measurements. We therefore introduce projection-consistency correction (PCC) at each inference stage. PCC uses the original metal-affected projection y_{MA} , the metal trace T , and the forward-projection operator $\mathcal{F}(\cdot)$ to constrain the current prediction with reliable observations outside the metal traces, and feeds the correction residual back into sub-

sequent Flow Matching updates. By combining LI-based structural conditioning, metal-region dynamic spatial weighting, and projection-consistency correction, SCMA imposes task-specific constraints on sample structure, degraded regions, and physical measurements, rather than directly applying generic Flow Matching to CT image generation.

C. Our Contribution

Based on the above designs, the main contributions of this work are summarized as follows:

- We propose SCMA, a structure-conditioned and metal-aware Flow Matching framework for CT metal artifact reduction. SCMA introduces an LI-corrected image into velocity-field learning as a sample-specific structural condition, enabling the generative trajectory to restore artifact-free CT images under the coarse structural constraint of the current sample and thereby reducing the structural uncertainty associated with unconditional generation.
- We propose a dynamically spatially weighted Flow Matching loss guided by the metal mask. The proposed loss constructs time-varying spatial weights from the metal mask and its distance transform, enabling the model to adaptively focus on the metal regions and their neighborhoods at different generative stages and improving the restoration of locally severe degradation.
- We develop an alternating inference framework that integrates the Flow Matching generative prior with projection-consistency correction. The same velocity network is shared across all time steps, while reliable projections outside the metal traces constrain the current prediction. This design requires no separate network for each inference time step, reduces hallucination-like structures inconsistent with the original measurements, and improves the physical reliability of the MAR results.

II. RELATED WORK

A. Flow Matching Generative Models

Flow Matching (FM) is a class of generative modeling methods based on continuous normalizing flows. It learns a time-dependent velocity field that continuously transports samples from a source distribution to a target data distribution through a deterministic ordinary differential equation [43]. Let $\mathbf{x}_1$ denote a sample drawn from the target data distribution and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ denote Gaussian noise, where $\mathbf{I}$ is the identity matrix. FM constructs a linear probability path between them as

$$\mathbf{x}_t = (1 - t)\mathbf{z} + t\mathbf{x}_1, \quad t \in [0, 1], \quad (1)$$

where $t = 0$ corresponds to the noise endpoint and $t = 1$ to the data endpoint. The target velocity associated with this path is

$$\mathbf{u}_t = \frac{d\mathbf{x}_t}{dt} = \mathbf{x}_1 - \mathbf{z}. \quad (2)$$

The velocity network $\mathbf{v}_\theta(\mathbf{x}_t, t)$ is trained by regressing the target velocity, with the basic objective defined as

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{\mathbf{x}_1, \mathbf{z}, t} \left[\|\mathbf{v}_\theta(\mathbf{x}_t, t) - \mathbf{u}_t\|_2^2 \right]. \quad (3)$$

Once trained, samples from the target distribution can be obtained by starting from $\mathbf{x}_0 = \mathbf{z}$ and solving

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{v}_\theta(\mathbf{x}_t, t). \quad (4)$$

FM directly learns the local transport directions along a probability path and therefore provides an explicit training objective, a continuous inference trajectory, and flexible numerical solvers.

However, standard unconditional FM learns only the population-level distribution of target images and does not establish a correspondence between a random initial sample and a specific metal-affected image. Its velocity field incorporates neither the structural condition and metal-region information of the current sample nor constraints from the original CT projections. Consequently, unconditional FM can provide only a distributional prior over artifact-free CT images and cannot directly serve as a sample-specific MAR model. Building on this formulation, we further introduce LI-based structural conditioning, dynamic spatial weighting over metal regions, and projection-consistency correction, such that the FM inference trajectory is jointly constrained by the structure of the current sample, the spatial distribution of artifacts, and reliable projection measurements.

B. Metal Segmentation in MAR

Metal-region segmentation is a preprocessing step in many MAR methods. Its results are commonly used to locate metal regions in the image domain and determine the corresponding metal traces through forward projection. Traditional methods mainly employ CT-number thresholding and morphological processing to identify metal regions, but they are prone to undersegmentation or oversegmentation in the presence of severe streak artifacts, high-density bone structures, or ambiguous boundaries. To improve localization robustness, Hegazy et al. used U-Net to directly predict metal traces from projection data [44]. U-Net employs an encoder-decoder architecture with skip connections to integrate multiscale semantic information and spatial details [45]. nnU-Net further uses a self-configuring mechanism to automatically determine the preprocessing procedure, network architecture, training strategy, and postprocessing pipeline, demonstrating strong adaptability across various medical image segmentation tasks [46].

Metal segmentation serves different purposes in different MAR modules. In projection completion methods such as LI and NMAR, the segmentation result directly determines the extent of the metal traces to be replaced. Undersegmentation may retain corrupted measurements, whereas oversegmentation may discard reliable projections. These methods are therefore generally sensitive to metal-boundary localization. In this work, we employ an existing nnU-Net to obtain the image-domain metal mask $\mathbf{M}$ rather than developing a new segmentation network. During training, $\mathbf{M}$ and its distance information are used to adjust the loss weights at different spatial locations. During preprocessing and inference, $\mathbf{M}$ is forward-projected to obtain the metal trace $\mathbf{T}$, which is used to generate the LI-corrected conditioning image and delineate the reliable measurement regions for projection-consistency correction. In

loss weighting, the mask serves as a spatial attention prior, whereas the projection-domain operations remain affected by the accuracy of metal-trace localization. We therefore regard metal segmentation as an auxiliary preprocessing step rather than a methodological contribution of SCMA.

C. Linear Interpolation-Based MAR

Linear interpolation-based metal artifact reduction (LI-MAR) is a classical projection completion method [3], [5]. It first generates projection-domain metal traces from image-domain metal regions and regards measurements within these traces as unreliable. Linear interpolation is then performed using the nonmetal measurements on both sides of each metal trace at the same projection angle. The completed projections are subsequently reconstructed to obtain an LI-corrected image. LI-MAR is simple to implement and computationally efficient, and it can effectively reduce severe streak and dark-band artifacts. However, it essentially replaces missing or distorted measurements within the metal traces with manually interpolated values. When the metal traces are wide or the projections on their two sides differ substantially, the interpolated values cannot accurately represent the true projection variation, potentially introducing secondary artifacts, intensity bias, and structural blurring.

Normalized metal artifact reduction (NMAR) uses the forward projection of a prior image to normalize the original projections, performs interpolation in the normalized domain, and subsequently restores the projection magnitudes, thereby reducing the influence of anatomical variations on interpolation [4], [6]. Nevertheless, the performance of NMAR still depends on the quality of the prior image and the accuracy of metal-trace localization. Unlike methods that directly use the LI result as the final output, we employ $\mathbf{x}_{\text{Linear}}$ only as a coarse, sample-specific structural condition. Although it still contains residual artifacts and local structural errors, its preserved principal anatomical contours can constrain the Flow Matching inference trajectory, while subsequent restoration is jointly performed by the conditional velocity field and projection-consistency correction.

III. METHOD

We propose SCMA, a structure-conditioned and metal-aware Flow Matching framework for CT metal artifact reduction. As illustrated in Fig. 2, SCMA consists of three stages: preprocessing, conditional velocity-field training, and inference with projection-consistency correction. Standard Flow Matching learns only the transport from a Gaussian source distribution to the distribution of artifact-free CT images and therefore cannot directly establish a correspondence between a random initial state and the current metal-affected sample. SCMA uses a linear-interpolation-corrected image to provide a sample-specific structural condition, employs dynamic spatial weighting over the metal region to enhance the learning of locally severe degradation, and constrains the generated result using the original projections during inference. In this way, a generic image-distribution prior is transformed into a conditional restoration model tailored to MAR.

A. Preprocessing

As shown in Fig. 2(A), the preprocessing stage generates an image-domain metal mask, a projection-domain metal trace, and a linear-interpolation-corrected image. Following [3], we use the linear-interpolation-corrected image as the conditioning image and denote it by $\mathbf{x}_{\text{Linear}}$. Given the original metal-affected projection $\mathbf{y}_{\text{MA}}$ and its reconstructed image $\mathbf{x}_{\text{MA}}$, we first employ a pretrained segmentation network $\mathcal{S}_{\phi}(\cdot)$ to obtain the image-domain metal mask:

$$\mathbf{M} = \mathcal{S}_{\phi}(\mathbf{x}_{\text{MA}}), \quad M(\mathbf{p}) \in \{0, 1\}, \quad \mathbf{p} \in \Omega, \quad (5)$$

where Ω denotes the image domain. Specifically, $M(\mathbf{p}) = 1$ indicates that pixel $\mathbf{p}$ belongs to a metal region, whereas $M(\mathbf{p}) = 0$ otherwise. This work does not develop a new metal segmentation model; instead, the mask $\mathbf{M}$ obtained using an existing segmentation model serves as a regional prior for subsequent processing.

The mask $\mathbf{M}$ is then forward-projected into the projection domain to obtain the binary metal trace:

$$\mathbf{T} = \mathbb{I}[\mathcal{F}(\mathbf{M}) > 0], \quad (6)$$

where $\mathcal{F}(\cdot)$ denotes the CT forward-projection operator and $\mathbb{I}[\cdot]$ is the indicator function. Here, $T(\mathbf{r}) = 1$ indicates that the ray corresponding to projection-domain position $\mathbf{r}$ intersects a metal region.

Based on $\mathbf{T}$, the projection measurements within the metal trace are linearly interpolated using the nonmetal measurements on both sides of the trace. The resulting projection is then reconstructed using the reconstruction operator $\mathcal{R}(\cdot)$ to obtain the conditioning image:

$$\mathbf{x}_{\text{Linear}} = \mathcal{R}((1 - \mathbf{T}) \odot \mathbf{y}_{\text{MA}} + \mathbf{T} \odot \mathcal{I}_{\text{Linear}}(\mathbf{y}_{\text{MA}}, \mathbf{T})), \quad (7)$$

where $\odot$ denotes element-wise multiplication and $\mathcal{I}_{\text{Linear}}(\cdot)$ represents linear interpolation within the metal trace. The image $\mathbf{x}_{\text{Linear}}$ may still contain interpolation errors, residual artifacts, and structural blurring. It is therefore not treated as the final corrected result but as a conditioning image that preserves the coarse structure of the current sample. The mask $\mathbf{M}$ is used for spatial weighting during training, whereas the corresponding metal trace $\mathbf{T}$ is used both to construct the conditioning image and to perform projection-consistency correction during inference. Consequently, metal segmentation errors may still affect the restoration performance of SCMA.

B. Structure-Conditioned and Metal-Aware Flow Matching Training

As illustrated in Fig. 2(B), the training stage learns a sample-specific conditional velocity field guided by the conditioning image and employs dynamic spatial weighting over the metal region to adjust the contributions of different pixels to the training objective. The artifact-free CT image, conditioning image, and metal mask in the training data are denoted by $\mathbf{x}_1$, $\mathbf{x}_{\text{Linear}}$, and $\mathbf{M}$, respectively.

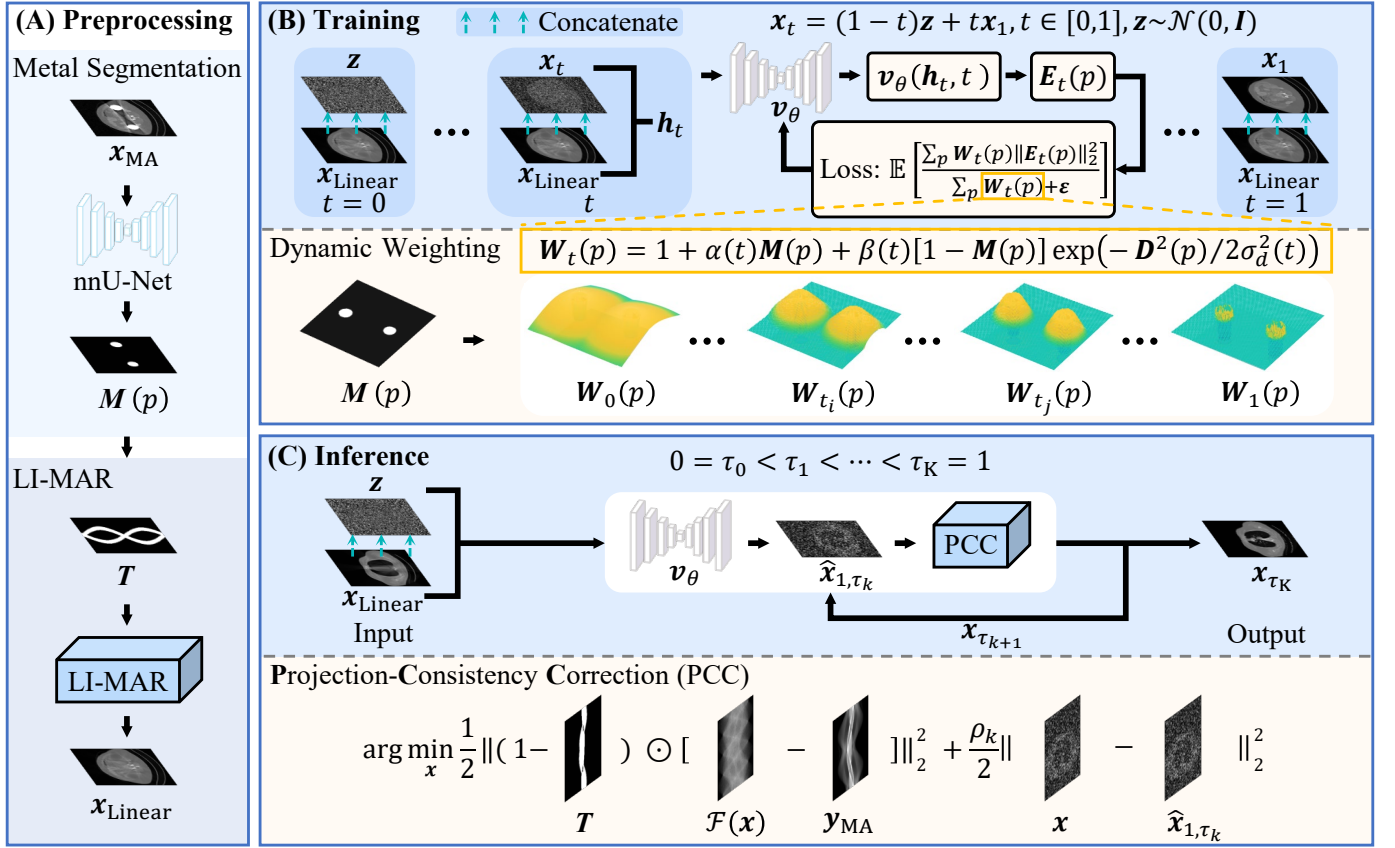


Fig. 2. Overall workflow of SCMA. (A) In the preprocessing stage, metal segmentation is performed to obtain the image-domain metal mask, and LI-MAR is applied to generate the conditioning image x_{Linear} . (B) During training, intermediate states are densely sampled along the Flow Matching path, with x_{Linear} used as the conditioning input and the metal mask used to construct a dynamically weighted loss. (C) During inference, the process starts from Gaussian noise and evolves under the fixed condition x_{Linear} , with projection-consistency correction (PCC) applied at each stage.

1) *Structure-Conditioned Flow Matching*: Given an artifact-free target image x_1 , Gaussian noise $z \sim \mathcal{N}(0, I)$, and a randomly sampled time $t \sim \mathcal{U}(0, 1)$, where I denotes the identity matrix, we construct the intermediate state x_t according to Eq. (1), with its target velocity u_t given by Eq. (2). Under this definition, $t = 0$ corresponds to the Gaussian-noise endpoint, whereas $t = 1$ corresponds to the artifact-free image endpoint.

A standard Flow Matching velocity network $v_\theta(x_t, t)$ predicts the transport direction using only the current state and time and therefore cannot determine the specific sample structure to be restored. To introduce a sample-specific condition, we concatenate x_t with its corresponding conditioning image x_{Linear} along the channel dimension:

$$h_t = \text{Concat}(x_t, x_{Linear}), \quad (8)$$

and use the time-conditioned velocity network to predict

$$\hat{u}_t = v_\theta(h_t, t). \quad (9)$$

Because x_{Linear} remains fixed across all time states, it provides the coarse anatomical structure associated with the current metal-affected sample. The velocity network consequently learns a conditional restoration trajectory corresponding to the structure of the current sample, rather than an unconditional CT image generation process.

2) *Metal Mask-Guided Dynamic Weighting Loss*: Metal artifacts exhibit pronounced spatial nonuniformity. To increase the contribution of the metal region and its neighborhood to velocity-field training, we first compute a distance-transform map from the image-domain metal mask M :

$$D(p) = \min_{q: M(q)=1} \|p - q\|_2, \quad p \in \Omega, \quad (10)$$

where $D(p)$ denotes the Euclidean distance from pixel p to the nearest metal region.

Based on M and D , we construct a two-dimensional pixel-wise weight map W_t for each time point t :

$$W_t(p) = 1 + \alpha(t)M(p) + \beta(t)[1 - M(p)] \exp\left(-\frac{D^2(p)}{2\sigma_d^2(t)}\right), \quad (11)$$

where $\alpha(t) \geq 0$ and $\beta(t) \geq 0$ control the weighting strengths within the metal region and its neighborhood, respectively, while $\sigma_d(t)$ controls the spatial extent of the neighborhood weighting. The neighborhood scale is defined as

$$\sigma_d(t) = (1-t)\sigma_{\max} + t\sigma_{\min}, \quad \sigma_{\max} > \sigma_{\min} > 0. \quad (12)$$

Near the noise endpoint, W_t therefore assigns elevated weights to a broader neighborhood around the metal. As t increases, the spatial extent of the elevated weights gradually contracts, allowing the model to focus more strongly on the

Algorithm 1 Preprocessing and Training of SCMA

-
- 1: **Input:** metal-affected image $\mathbf{x}_{\text{MA}}$, projection $\mathbf{y}_{\text{MA}}$, and artifact-free target $\mathbf{x}_1$
 - 2: Obtain the metal mask $\mathbf{M}$ using Eq. (5)
 - 3: Generate the metal trace $\mathbf{T}$ using Eq. (6)
 - 4: Compute the conditioning image $\mathbf{x}_{\text{Linear}}$ using Eq. (7)
 - 5: **for** each training iteration **do**
 - 6: Sample $(\mathbf{x}_1, \mathbf{x}_{\text{Linear}}, \mathbf{M}) \sim \mathcal{P}_{\text{train}}$
 - 7: Sample $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $t \sim \mathcal{U}(0, 1)$
 - 8: Compute $\mathbf{x}_t$ and $\mathbf{u}_t$ using Eqs. (1) and (2)
 - 9: Construct $\mathbf{h}_t$ using Eq. (8)
 - 10: Compute $\mathbf{D}$ and $\mathbf{W}_t$ using Eqs. (10)–(12)
 - 11: Predict $\hat{\mathbf{u}}_t = \mathbf{v}_\theta(\mathbf{h}_t, t)$
 - 12: Compute $\mathbf{E}_t(\mathbf{p}) = \hat{\mathbf{u}}_t(\mathbf{p}) - \mathbf{u}_t(\mathbf{p})$
 - 13: Update θ by minimizing $\mathcal{L}_{\text{wFM}}$ in Eq. (14)
 - 14: **end for**
 - 15: **return** the trained structure-conditioned velocity network $\mathbf{v}_\theta$
-

metal region and its immediate neighborhood as the state approaches the artifact-free image endpoint. The weight maps shown at different times in Fig. 2(B) represent separate two-dimensional weight maps rather than multiple feature channels at the same time point.

Using the predicted velocity obtained from Eq. (9), we define the velocity prediction error at time t and pixel position $\mathbf{p}$ as

$$\mathbf{E}_t(\mathbf{p}) = \hat{\mathbf{u}}_t(\mathbf{p}) - \mathbf{u}_t(\mathbf{p}), \quad \mathbf{p} \in \Omega. \quad (13)$$

The dynamically spatially weighted Flow Matching loss of SCMA is defined as

$$\mathcal{L}_{\text{wFM}} = \mathbb{E}_{(\mathbf{x}_1, \mathbf{x}_{\text{Linear}}, \mathbf{M}) \sim \mathcal{P}_{\text{train}}, \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), t \sim \mathcal{U}(0, 1)} \left[\frac{\sum_{\mathbf{p} \in \Omega} W_t(\mathbf{p}) \|\mathbf{E}_t(\mathbf{p})\|_2^2}{\sum_{\mathbf{p} \in \Omega} W_t(\mathbf{p}) + \varepsilon} \right], \quad (14)$$

where $\mathcal{P}_{\text{train}}$ denotes the training-sample distribution and ε is a numerical stability term. The denominator reduces the effect of variations in the overall scale of different weight maps on the loss magnitude. Compared with a spatially uniform Flow Matching loss, Eq. (14) increases the relative contribution of velocity prediction errors within and around the metal region, enabling the network to learn more effectively from locally severe degradation.

Algorithm 1 summarizes the preprocessing and training procedures of SCMA.

C. Projection-Consistency Correction Inference

As shown in Fig. 2(C), the inference stage solves the learned Flow Matching ordinary differential equation under the guidance of the fixed conditioning image $\mathbf{x}_{\text{Linear}}$ and performs projection-consistency correction (PCC) at each time step using the original projection measurements outside the metal trace. Let the discrete inference time sequence be

$$0 = \tau_0 < \tau_1 < \dots < \tau_K = 1, \quad (15)$$

where K denotes the number of inference steps. The initial state is

$$\mathbf{x}_{\tau_0} = \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (16)$$

Algorithm 2 SCMA Inference with Projection-Consistency Correction

-
- 1: **Input:** trained network $\mathbf{v}_\theta$, conditioning image $\mathbf{x}_{\text{Linear}}$, projection $\mathbf{y}_{\text{MA}}$, metal trace $\mathbf{T}$, and time sequence $\{\tau_k\}_{k=0}^K$
 - 2: Sample $\mathbf{x}_{\tau_0} = \mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
 - 3: **for** $k = 0, \dots, K - 1$ **do**
 - 4: Construct $\mathbf{h}_{\tau_k} = \text{Concat}(\mathbf{x}_{\tau_k}, \mathbf{x}_{\text{Linear}})$
 - 5: Predict $\hat{\mathbf{u}}_{\tau_k} = \mathbf{v}_\theta(\mathbf{h}_{\tau_k}, \tau_k)$
 - 6: Estimate $\hat{\mathbf{x}}_{1, \tau_k}$ using Eq. (19)
 - 7: Compute $\mathbf{x}_{\text{PCC}, \tau_k}$ using Eq. (21)
 - 8: Update $\mathbf{x}_{\tau_{k+1}}$ using Eq. (23)
 - 9: **end for**
 - 10: **return** $\mathbf{x}_{\tau_K}$
-

At the k th inference stage, $\mathbf{x}_{\tau_k}$ is the current state updated throughout the inference process, whereas the conditioning image $\mathbf{x}_{\text{Linear}}$ remains fixed across all time steps. They are concatenated along the channel dimension:

$$\mathbf{h}_{\tau_k} = \text{Concat}(\mathbf{x}_{\tau_k}, \mathbf{x}_{\text{Linear}}), \quad (17)$$

and the trained conditional velocity network predicts

$$\hat{\mathbf{u}}_{\tau_k} = \mathbf{v}_\theta(\mathbf{h}_{\tau_k}, \tau_k). \quad (18)$$

According to the linear probability path, the artifact-free endpoint corresponding to the current state is estimated as

$$\hat{\mathbf{x}}_{1, \tau_k} = \mathbf{x}_{\tau_k} + (1 - \tau_k) \hat{\mathbf{u}}_{\tau_k}. \quad (19)$$

Because the original measurements within the metal trace are severely corrupted, only the relatively reliable projections outside the metal trace are used to constrain the endpoint estimate. The reliable-region mask in the projection domain is defined as

$$\mathbf{W}_p = \mathbf{1} - \mathbf{T}. \quad (20)$$

The PCC problem at the k th inference stage is formulated as

$$\mathbf{x}_{\text{PCC}, \tau_k} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{W}_p \odot [\mathcal{F}(\mathbf{x}) - \mathbf{y}_{\text{MA}}]\|_2^2 + \frac{\rho_k}{2} \|\mathbf{x} - \hat{\mathbf{x}}_{1, \tau_k}\|_2^2, \quad (21)$$

where $\rho_k > 0$ controls the trade-off between projection consistency and the Flow Matching image prior. The first term enforces consistency between the corrected result and the original measurements outside the metal trace, while the second prevents the corrected result from deviating excessively from the current endpoint estimate. For notational simplicity, the solution of Eq. (21) is expressed as

$$\mathbf{x}_{\text{PCC}, \tau_k} = \mathcal{C}_{\text{PCC}}(\hat{\mathbf{x}}_{1, \tau_k}; \mathbf{y}_{\text{MA}}, \mathbf{T}, \rho_k). \quad (22)$$

The correction produced by PCC for the endpoint estimate is subsequently fed back into the current Flow Matching state, and the state at the next time point is obtained using a first-order Euler update:

$$\mathbf{x}_{\tau_{k+1}} = \mathbf{x}_{\tau_k} + (\tau_{k+1} - \tau_k) \hat{\mathbf{u}}_{\tau_k} + \omega_k (\mathbf{x}_{\text{PCC}, \tau_k} - \hat{\mathbf{x}}_{1, \tau_k}), \quad (23)$$

where $0 \leq \omega_k \leq 1$ denotes the correction feedback strength. The updated state $\mathbf{x}_{\tau_{k+1}}$ is fed back into the velocity network as the variable state for the next inference stage, while $\mathbf{x}_{\text{Linear}}$

TABLE I

QUANTITATIVE COMPARISON OF DIFFERENT MAR METHODS FOR SMALL, MEDIUM, AND LARGE METAL IMPLANTS. THE AVERAGE PSNR (DB) $\uparrow$ AND SSIM (%) $\uparrow$ ARE REPORTED. THE BEST AND SECOND-BEST RESULTS ARE HIGHLIGHTED IN BOLD AND UNDERLINED, RESPECTIVELY.

Method	Small metal		Medium metal		Large metal		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FBP	18.1664	60.3932	17.6839	77.4033	5.1595	33.9951	13.6699	57.2639
NMAR	39.6144	86.7410	39.1023	93.3539	37.3240	85.0119	38.6802	88.3689
DICDNet	<u>50.0476</u>	<u>99.5179</u>	<u>43.5349</u>	<u>99.2870</u>	<u>41.6628</u>	<u>97.7062</u>	<u>45.0818</u>	<u>98.8370</u>
InDuDoNet+	47.3177	99.0293	42.3334	99.2426	34.7378	96.1927	41.4630	98.1549
CALIMAR	35.4715	87.9920	38.6385	93.3364	32.5506	83.9158	35.5535	88.4148
ADN	35.8166	95.4741	36.9620	96.8683	27.8158	88.0662	33.5315	93.4696
DuDoDp-MAR	36.1534	85.6365	39.3705	92.9689	34.7557	84.0931	36.7599	87.5662
Proposed	51.0738	99.6571	48.8946	99.4946	42.7066	97.9768	47.5583	99.0429

remains the fixed condition. This process jointly constrains the inference trajectory using the sample-specific structural condition, the artifact-free image-distribution prior, and reliable projection measurements. After the iteration reaches $\tau_K = 1$, $\mathbf{x}_{\tau_K}$ is taken as the final MAR result.

Algorithm 2 summarizes the SCMA inference procedure with PCC.

IV. EXPERIMENTS

A. Experimental Setup

1) *Datasets*: Following the data simulation protocols adopted by existing MAR methods [17], [18], [24], we randomly selected 1,200 metal-free CT images from the DeepLesion dataset [47] and collected 100 metal masks with different sizes and shapes to synthesize metal-affected images. Among them, 90 metal masks and 1,000 metal-free CT images were used to synthesize the training samples. The remaining 10 metal masks and 200 metal-free CT images were used to construct the test set, yielding 2,000 paired metal-affected and metal-free test images. The sizes of the 10 test metal implants were [32, 53, 112, 115, 115, 242, 448, 878, 879, 2054] pixels. According to the metal size, the test samples were divided into small-, medium-, and large-metal categories to evaluate MAR performance under different metal sizes. Specifically, small metal implants were defined as those containing [0, 100] pixels, medium metal implants as those containing [101, 500] pixels, and large metal implants as those containing more than 500 pixels.

All CT images were resized to 416×416 pixels with a pixel spacing of 0.369 mm. Projection data were simulated using a two-dimensional fan-beam CT geometry. The source-to-object distance (SOD) and source-to-detector distance (SDD) were 396.92 mm and 793.85 mm, respectively. The one-dimensional detector contained 641 detector elements with a total width of 434.45 mm, corresponding to a detector-element width of approximately 0.678 mm. For each sample, 640 projection views were uniformly distributed over the range of 0–360°. Following the same simulation procedure, we generated a metal-affected image, an LI-corrected image, a metal mask, a metal trace, and the corresponding metal-free reference image for each sample.

To further evaluate the generalization capability of the proposed method in real metal-artifact scenarios, we selected three clinical CT volumes containing real metal implants from the COLONOG subset of the CTSpine1K dataset [48]. The case identifiers were 0064, 0131, and 0313. This subset was originally derived from the publicly available CT COLONOGRAPHY data in The Cancer Imaging Archive (TCIA). The original in-plane image size of all three cases was 512×512 pixels, with 507, 663, and 554 slices, respectively. Their interslice spacing was 0.8 mm, while their in-plane pixel spacing was 0.703 mm, 0.781 mm, and 0.859 mm, respectively. We extracted 72, 153, and 175 axial slices containing metal implants and associated artifacts from these three volumes, yielding a total of 400 real metal-affected CT images. It should be emphasized that these images were obtained from three three-dimensional cases rather than 400 independent patients.

To match the network input size, all slices were resized to 416×416 pixels while preserving the complete field of view. A pretrained nnU-Net was employed to generate the corresponding image-domain metal masks. Across all 400 slices, the segmented metal-region sizes ranged from 29 to 2,099 pixels. Using the same metal-size criteria as those applied to the synthetic test set, the 400 valid metal-containing slices comprised 10 small-metal samples, 143 medium-metal samples, and 247 large-metal samples.

2) *Training Details*: The proposed SCMA method was implemented in PyTorch. The conditional Flow Matching framework was trained on an NVIDIA GeForce RTX 5080 GPU. Its input was constructed by concatenating the current intermediate state $\mathbf{x}_t$ and the linear-interpolation-corrected image $\mathbf{x}_{\text{Linear}}$ along the channel dimension, i.e., $(\mathbf{x}_t, \mathbf{x}_{\text{Linear}})$ was used as the conditional input.

The network was optimized using AdamW with the default momentum parameters in PyTorch. The initial learning rate was set to 5×10^{-5} , the weight decay was set to 0, and the gradient-clipping threshold was set to 1.0. The batch size was set to 1, and the number of Flow Matching training time steps was set to 1,000. A conditional U-Net with 64 base channels was employed as the backbone, with FFT Transformer modules incorporated into its low-resolution feature levels. At each training iteration, one metal-free CT image and one synthetic metal mask were randomly selected from the pool of 1,000 training images and the pool of 90 training masks, respectively,

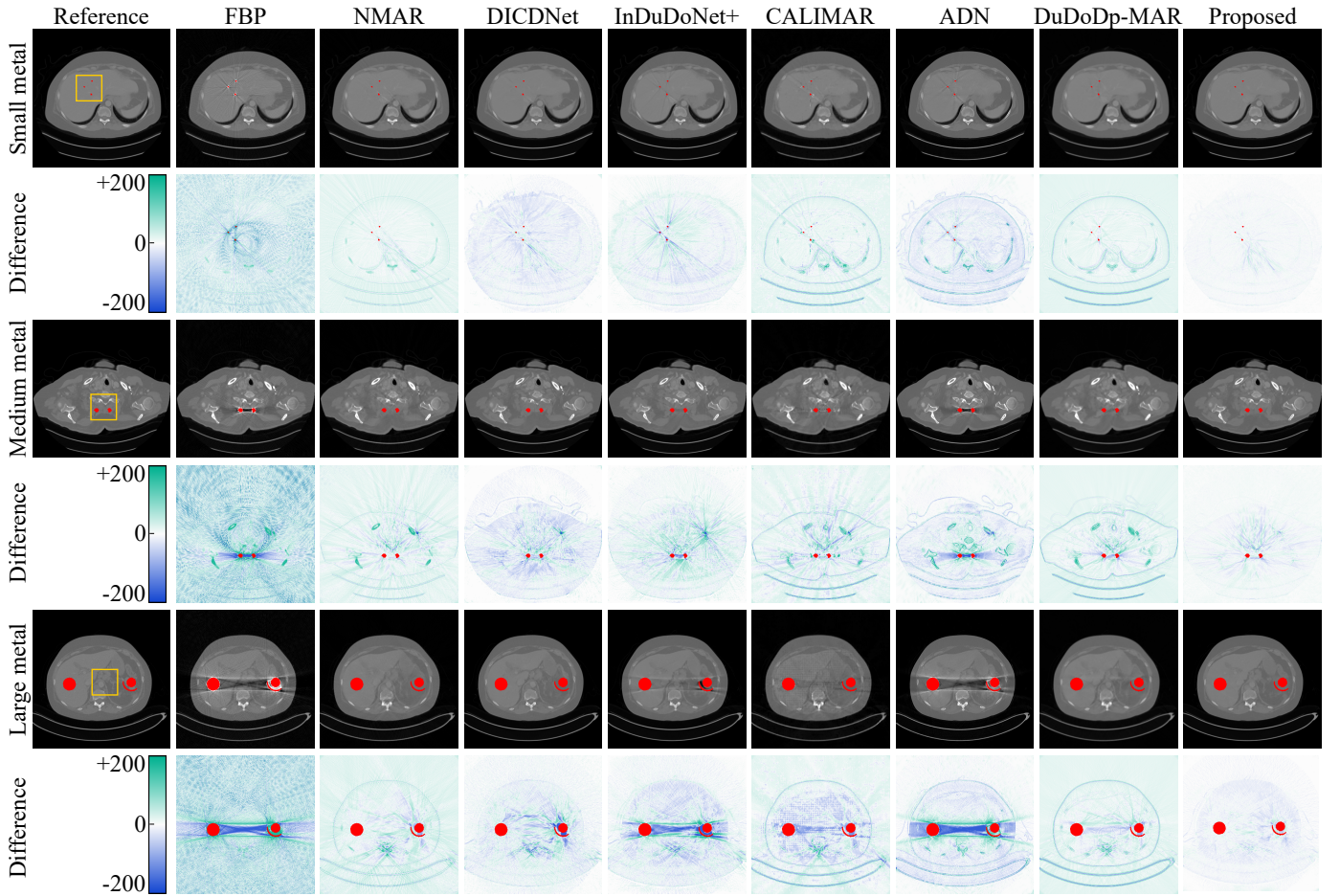


Fig. 3. Visual comparison of different MAR methods for small, medium, and large metal implants. The columns from left to right show Reference, FBP, NMAR, DICDNet, InDuDoNet+, CALIMAR, ADN, DuDoDp-MAR, and the proposed SCMA. For each metal-size category, the upper row shows the corrected CT images displayed within $[-1000, 800]$ HU, and the lower row shows the residual maps relative to the reference images displayed within $[-200, 200]$ HU. The metal regions are marked in red.

to synthesize a metal-affected CT image.

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

We compared SCMA with representative methods from several major categories of MAR techniques. In addition to FBP as the analytical reconstruction baseline, we selected NMAR [4] as a model-based projection-domain completion method; DICDNet [18] and InDuDoNet+ [24] as supervised deep MAR methods; CALIMAR [49] and ADN [17] as unpaired or unsupervised MAR methods; and DuDoDp-MAR [34] as a generative-prior-based MAR method. These methods cover the principal technical paradigms and representative advanced approaches for CT metal artifact reduction. All methods were evaluated on the same test set using PSNR and SSIM as the quantitative metrics.

Table I presents the quantitative results for different metal sizes. Because FBP does not correct the projection inconsistencies caused by metal, its overall performance is substantially limited. NMAR considerably reduces streak artifacts but remains affected by interpolation errors and the quality of its prior image. The deep learning-based methods generally outperform the conventional methods, with DICDNet achieving the most competitive performance among the compared

methods. In contrast, SCMA achieves the best performance for small, medium, and large metal implants, with average PSNR and SSIM values of 47.5583 dB and 99.0429%, respectively. Compared with the second-best method, DICDNet, SCMA improves the average PSNR by 2.4765 dB, with a particularly pronounced advantage for medium-sized metal implants. These results demonstrate that LI-based conditioning, dynamically weighted training over the metal region, and projection-consistency correction effectively improve restoration robustness across different levels of artifact severity.

Figure 3 presents the visual comparison of the different methods. FBP exhibits pronounced radial streaks and dark-band artifacts, whose severity increases with the metal size. Although NMAR removes some severe artifacts, residual errors and interpolation-induced secondary artifacts remain visible around the metal regions. DICDNet and InDuDoNet+ substantially improve the overall image quality but retain structured residuals near the metal regions and high-contrast structural boundaries. CALIMAR, ADN, and DuDoDp-MAR further reduce some artifacts, but local intensity shifts, abnormal textures, or structural oversmoothing can still be observed. In comparison, SCMA produces weaker and more localized residual responses and exhibits greater stability around the

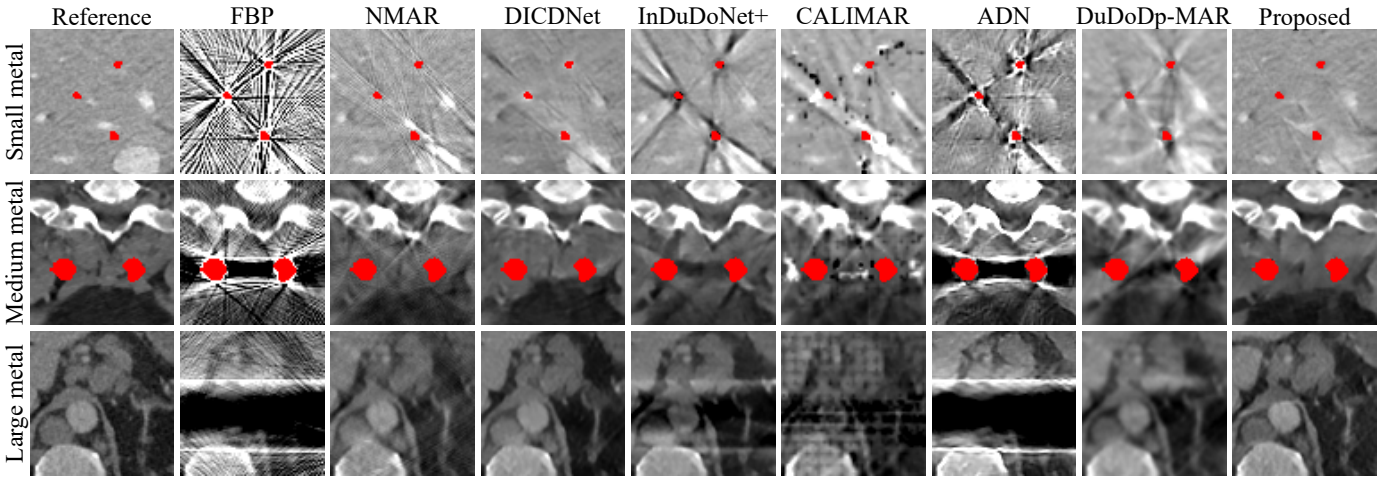


Fig. 4. Local ROI comparison of different MAR methods for small, medium, and large metal implants. The ROIs in the first row are displayed within $[-250, 50]$ HU, whereas those in the second and third rows are displayed within $[-350, 100]$ HU. The metal regions are marked in red. The proposed SCMA more effectively suppresses residual streak and dark-band artifacts around the metal regions while preserving the soft-tissue background and bone boundaries.

metal regions, in soft-tissue backgrounds, and along bone boundaries. The local ROI comparison in Fig. 4 further shows that SCMA preserves the continuity and clarity of local anatomical structures while suppressing severe artifacts, consistent with the quantitative results.

C. Experiments on Real-World Data

To further evaluate the generalization capability of SCMA under real acquisition conditions, we conducted experiments on real-world CT data containing small, medium, and large metal implants and compared the results with those obtained using FBP, NMAR, DICDNet, InDuDoNet+, CALIMAR, ADN, and DuDoDp-MAR. Because strictly registered metal-free reference images were unavailable for the real-world data, the different methods were evaluated qualitatively using both the complete CT images and the metal-adjacent ROIs. The same display windows were used for all methods to ensure a fair visual comparison.

As shown in Fig. 5, the radial streaks and bright-dark distortions in the FBP images become substantially more severe as the metal size increases, strongly interfering with tissue structures around the metal regions. NMAR, DICDNet, CALIMAR, and DuDoDp-MAR reduce the global streak artifacts to some extent, but residual streaks, shading artifacts, or structural blurring remain visible in the magnified regions. InDuDoNet+ produces noticeable intensity nonuniformity and structural distortion in some regions, particularly in the medium- and large-metal cases. Although ADN restores some tissue information, pronounced intensity variations and secondary artifacts remain around the metal regions.

In comparison, SCMA provides more stable artifact correction across all three metal sizes. For the small metal implant, SCMA effectively suppresses the surrounding radial streaks while preserving the boundaries of adjacent tissues. For the medium-sized implant, it produces a more uniform intensity distribution and clearer local tissue structures. In the more challenging large-metal case, SCMA still substantially reduces bright-dark streaks and shading artifacts around the metal

region while preserving the continuity of bone structures and soft-tissue boundaries. These results demonstrate that SCMA adapts well to different metal sizes and achieves a favorable balance between artifact suppression and local structure preservation on real-world data.

D. Generalization Across Metal Sizes

To further evaluate the generalization capability of different methods across varying metal sizes, we selected 10 metal masks containing 32, 53, 112, 115, 115, 242, 448, 878, 879, and 2,054 pixels, respectively. For each metal mask, 10 test slices were randomly selected to synthesize metal-affected images. The average PSNR and SSIM between the corrected and reference images were then calculated for each method. The results are presented in Fig. 6.

In general, as the metal size increases, more projection information within the metal traces becomes missing or unreliable, thereby increasing the difficulty of MAR. Consequently, both PSNR and SSIM exhibit an overall decreasing trend. In comparison, SCMA consistently maintains high reconstruction quality across different metal sizes and generally outperforms the competing methods in terms of both PSNR and SSIM, demonstrating improved restoration robustness across different metal scales. DICDNet and InDuDoNet+ also perform well for some metal sizes but exhibit performance degradation in cases involving larger or more complex metal implants. The curves of CALIMAR, ADN, DuDoDp-MAR, and NMAR fluctuate more substantially, indicating greater sensitivity to metal size, location, and local anatomical complexity.

It should be noted that MAR difficulty is not determined monotonically by metal-mask area alone but is also affected by metal shape, location, and the complexity of the occluded anatomy. For example, the average PSNR and SSIM obtained for the 878-pixel metal mask are higher than those obtained for the 448-pixel mask. Although the former has a larger area, it is primarily located within a relatively homogeneous tissue region, making it easier to recover structures close to the reference after metal removal. In contrast, the smaller

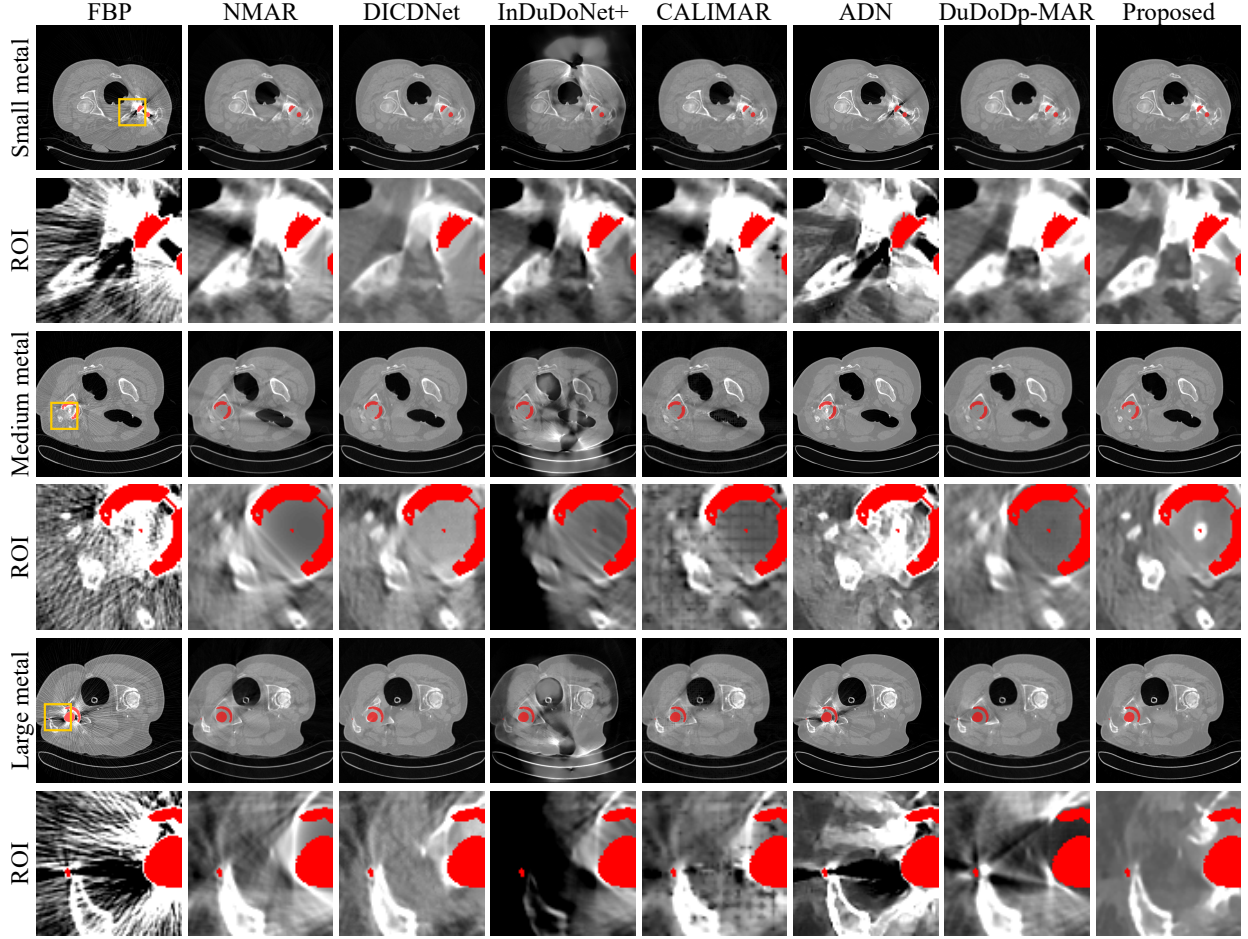


Fig. 5. Visual comparison of different MAR methods on real-world CT data containing small, medium, and large metal implants. The columns from left to right show FBP, NMAR, DICDNet, InDuDoNet+, CALIMAR, ADN, DuDoDp-MAR, and the proposed SCMA. For each metal-size category, the upper row shows the corrected CT images displayed within $[-1000, 800]$ HU, and the lower row shows the corresponding metal-adjacent ROIs displayed within $[-350, 250]$ HU. The metal regions are marked in red.

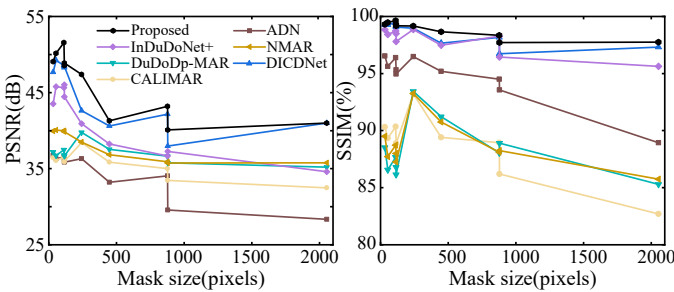


Fig. 6. Comparison of the generalization performance of different MAR methods across metal sizes. The horizontal axis indicates the number of pixels in each metal mask, while the vertical axes indicate PSNR (dB) $\uparrow$ and SSIM (%) $\uparrow$, respectively. For each metal mask, 10 test slices were randomly selected to synthesize metal-affected images, and the average PSNR and SSIM between the corrected and reference images were calculated for each method.

448-pixel mask occludes a more complex anatomical region and therefore results in greater restoration difficulty and lower average metric values. Similarly, although the 878- and 879-pixel metal masks have nearly identical areas, their performance differs substantially because the 878-pixel metal has a relatively simple structure, whereas the 879-pixel metal has a

more complex shape, leading to markedly different artifact distributions and restoration difficulties. A similar phenomenon is observed for the two 115-pixel metal masks, indicating that identical or similar metal areas do not necessarily correspond to the same level of MAR difficulty. These results demonstrate that SCMA generalizes not only across metal sizes but also across variations in metal shape and local anatomy.

E. Local Structural Fidelity Analysis

TABLE II
QUANTITATIVE RESULTS FOR LOCAL STRUCTURAL FIDELITY WITHIN TWO ROIS OBTAINED USING DIFFERENT GENERATIVE AND UNPAIRED MAR METHODS. SSIM $\uparrow$ MEASURES STRUCTURAL SIMILARITY, WHEREAS HFEN $\downarrow$ MEASURES HIGH-FREQUENCY STRUCTURAL ERRORS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Method	Metal-adjacent ROI		Spine ROI	
	SSIM	HFEN	SSIM	HFEN
CALIMAR	86.5036	0.9291	91.1514	0.3583
ADN	77.7021	1.1265	91.9477	0.3344
DuDoDp-MAR	93.6403	0.3939	93.3840	0.2658
Proposed	97.0522	0.2367	96.9391	0.1429

TABLE III

QUANTITATIVE ABLATION RESULTS FOR DIFFERENT COMPONENTS OF SCMA UNDER SMALL-, MEDIUM-, AND LARGE-METAL CONDITIONS. PSNR (dB) $\uparrow$ AND SSIM (%) $\uparrow$ ARE REPORTED. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Variant	LI	PCC	DWL	Small metal		Medium metal		Large metal		Average	
				PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
W/o LI	×	✓	✓	17.4651	41.1641	16.6351	25.9686	16.3352	29.2517	16.8118	32.1281
W/o PCC	✓	×	✓	45.0270	89.8063	42.8753	86.3775	39.4970	85.6275	42.4664	87.2704
W/o DWL	✓	✓	×	47.2420	95.7863	43.5153	93.1681	37.6677	89.6246	42.8083	92.8597
Proposed	✓	✓	✓	51.0738	99.6571	48.8946	99.4946	42.7066	97.9768	47.5583	99.0429

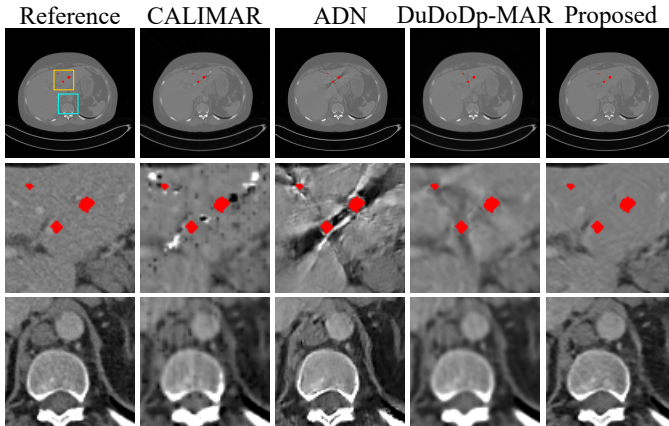


Fig. 7. Comparison of hallucination-like structures produced by generative and unpaired MAR methods. The columns from left to right show Reference, CALIMAR, ADN, DuDoDp-MAR, and the proposed SCMA. The first row shows the complete CT images displayed within $[-1000, 800]$ HU. The second and third rows show the metal-adjacent ROI indicated by the yellow box and the spine ROI indicated by the cyan box, respectively, both displayed within $[-350, 100]$ HU. The metal regions are marked in red.

Generative priors can improve the ability of MAR models to represent the distribution of artifact-free CT images. However, in regions with severe information loss, such as those adjacent to metal implants, they may also generate hallucination-like structures that are inconsistent with the true anatomy. We therefore evaluated the local structural reliability of different generative or unpaired MAR methods using both visual results and quantitative ROI metrics. Specifically, SSIM was used to measure local structural similarity, whereas HFEN was used to quantify errors in high-frequency structures such as edges and textures [50].

As shown in Fig. 7, although CALIMAR and ADN reduce some global artifacts, abnormal bright-dark textures and streak-like residuals remain around the metal regions. DuDoDp-MAR improves the overall visual quality but still exhibits local oversmoothing and detail deviations. In comparison, SCMA more effectively suppresses spurious textures around the metal region in the metal-adjacent ROI while better preserving bone boundaries and the surrounding soft-tissue background in the spine ROI. The quantitative results in Table II further support these observations. SCMA achieves the highest SSIM and lowest HFEN in both ROIs, indicating that its restored local structures are more consistent with the reference and contain smaller high-frequency structural errors.

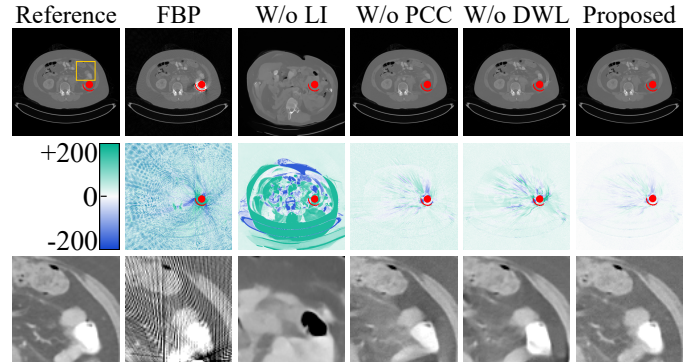


Fig. 8. Visual ablation comparison of the different components of the proposed SCMA. The columns from left to right show Reference, FBP, W/o LI, W/o PCC, W/o DWL, and the complete SCMA. The first row shows the corrected CT images displayed within $[-1000, 800]$ HU; the second row shows the residual maps relative to the reference image displayed within $[-200, 200]$ HU; and the third row shows the local ROIs displayed within $[-450, 100]$ HU. The metal regions are marked in red.

F. Ablation Studies

To evaluate the effectiveness of the individual components of SCMA, we constructed three ablation variants: W/o LI, which removes the linear-interpolation-based structural condition; W/o PCC, which removes projection-consistency correction; and W/o DWL, which removes the dynamic weighting loss. Except for the component being removed, all variants used identical training and testing settings. The quantitative and visual results are presented in Table III and Fig. 8, respectively.

As shown in Table III, removing any component results in performance degradation. W/o LI exhibits the most substantial deterioration, indicating that the coarse structural condition provided by the LI-corrected image is essential for stabilizing the Flow Matching inference trajectory. After PCC is removed, the model can still recover the principal structures using the image-domain generative prior, but its quantitative performance decreases considerably. This result indicates that relying solely on an image-domain prior cannot ensure consistency with the original projection measurements. Removing DWL weakens the model's restoration capability within and around the metal regions, demonstrating that a spatially uniform training objective is insufficient for adequately learning from regions affected by locally severe artifacts. In comparison, the complete SCMA achieves the best performance across all metal sizes, confirming the complementary roles of LI, PCC,

and DWL.

Figure 8 further illustrates the local differences among the variants. W/o LI exhibits pronounced global structural distortion and widespread residual errors, indicating that the generative process may deviate from the anatomy of the current sample in the absence of structural conditioning. W/o PCC and W/o DWL preserve the overall structure more effectively but retain more pronounced residual artifacts and local intensity deviations around the metal regions. The complete SCMA produces the weakest residual response, and its structures around the metal regions in the ROI are more consistent with the reference image. These results demonstrate that LI-based conditioning, the dynamic weighting loss, and projection-consistency correction improve MAR performance through complementary structural, regional-learning, and physical-consistency constraints.

V. DISCUSSION AND CONCLUSION

A central challenge in generative MAR is to exploit the distributional prior of artifact-free CT images while preserving the patient-specific anatomy and local fidelity around metal implants. SCMA uses the linear-interpolation-corrected image as a structural condition rather than treating it as the final correction result, thereby providing a sample-specific structural anchor for the Flow Matching trajectory. This design constrains the otherwise uncertain mapping from a Gaussian initial state to an artifact-free CT image and guides the restoration process toward the anatomy of the current input.

The metal-aware dynamic weighting strategy further addresses the spatially nonuniform distribution of metal artifacts. Because severe degradation is concentrated within and around the metal regions, a spatially uniform loss may underemphasize these locally challenging areas. SCMA constructs time-dependent spatial weights from the metal mask and its distance transform, enabling velocity-field training to focus more strongly on metal-related regions while limiting unnecessary changes to reliable anatomical structures farther from the metal. During inference, projection-consistency correction constrains the generated result using the relatively reliable measurements outside the metal trace, thereby further reducing the risk of physically inconsistent hallucination-like structures.

The experimental results validate the effectiveness of these designs. SCMA consistently achieves effective artifact suppression and anatomical structure preservation across different metal sizes. It also provides higher local structural fidelity within both the metal-adjacent and spine ROIs. The ablation results show that the linear-interpolation-based structural condition is particularly important for stabilizing the conditional generation trajectory, whereas dynamic weighting and projection-consistency correction provide further improvements through localized artifact-aware learning and measurement-based physical constraints, respectively.

SCMA nevertheless has several limitations. First, the method relies on a metal mask to construct the spatial weight map and the corresponding projection-domain metal trace. Segmentation errors may therefore affect both model training and inference stability. Future work could investigate metal-aware representations that are more robust to inaccurate or

uncertain masks. Second, projection-consistency correction improves physical reliability but introduces additional computational cost. More efficient approximate solvers or learned consistency operators could be explored to accelerate inference. In addition, the present method is primarily developed for two-dimensional slices and two-dimensional projection geometry. Extending SCMA to three-dimensional cone-beam or helical CT and validating it using larger, multicenter clinical datasets constitute important directions for future research.

In conclusion, we proposed SCMA, a structure-conditioned and metal-aware Flow Matching framework for CT metal artifact reduction. By integrating sample-specific structural conditioning, metal-region-aware training, and projection-consistency correction into a unified generative restoration process, SCMA effectively suppresses metal artifacts, preserves local anatomical structures, and reduces the risk of hallucination-like structures in generative MAR.

REFERENCES

- [1] M. Selles, J. A. C. van Osch, M. Maas, M. F. Boomsma, and R. H. H. Wellenberg, "Advances in metal artifact reduction in ct images: A review of traditional and novel metal artifact reduction techniques," *European Journal of Radiology*, vol. 171, p. 111276, 2024.
- [2] P. J. Withers, C. Bouman, S. Carmignato, V. Cnudde, D. Grimaldi, C. K. Hagen, E. Maire, M. Manley, A. Du Plessis, and S. R. Stock, "X-ray computed tomography," *Nature Reviews Methods Primers*, vol. 1, p. 18, 2021.
- [3] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of ct artifacts caused by metallic implants," *Radiology*, vol. 164, no. 2, pp. 576–577, 1987.
- [4] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Normalized metal artifact reduction (NMAR) in computed tomography," *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, 2010.
- [5] D. Prell, Y. Kyriakou, M. Beister, and W. A. Kalender, "A novel forward projection-based metal artifact reduction method for flat-detector computed tomography," *Physics in Medicine and Biology*, vol. 54, no. 21, pp. 6575–6591, 2009.
- [6] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Frequency split metal artifact reduction (FSMAR) in computed tomography," *Medical Physics*, vol. 39, no. 4, pp. 1904–1916, 2012.
- [7] J. A. Anhaus, P. Killermann, A. H. Mahnen, and C. Hofmann, "Nonlinearly scaled prior image-controlled frequency split for ct metal artifact reduction," *Medical Physics*, 2022.
- [8] —, "A nonlinear scaling-based normalized metal artifact reduction to reduce low-frequency artifacts in energy-integrating and photon-counting ct," *Medical Physics*, 2023.
- [9] J. M. Verburg and J. Seco, "Ct metal artifact reduction method correcting for beam hardening and missing projections," *Physics in Medicine and Biology*, vol. 57, no. 9, pp. 2803–2818, 2012.
- [10] H. S. Park, D. Hwang, and J. K. Seo, "Metal artifact reduction for polychromatic x-ray ct based on a beam-hardening corrector," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 480–487, 2016.
- [11] G. Wang, D. L. Snyder, J. A. O'Sullivan, and M. W. Vannier, "Iterative deblurring for ct metal artifact reduction," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 657–664, 1996.
- [12] X. Zhang, J. Wang, and L. Xing, "Metal artifact reduction in x-ray computed tomography (CT) by constrained optimization," *Medical Physics*, vol. 38, no. 2, pp. 701–711, 2011.
- [13] T. Humphries and B. Wang, "Superiorized method for metal artifact reduction," *Medical Physics*, vol. 47, no. 9, pp. 3984–3995, 2020.
- [14] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recogn.*, vol. 92, pp. 64–81, 2019.
- [15] A. Mehranian, M. R. Ay, A. Rahmim, and H. Zaidi, "X-ray ct metal artifact reduction using wavelet domain L_0 sparse regularization," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1707–1722, 2013.
- [16] Y. Zhang and H. Yu, "Convolutional neural network based metal artifact reduction in x-ray computed tomography," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1370–1381, 2018.

- [17] H. Liao, W.-A. Lin, S. K. Zhou, and J. Luo, "ADN: Artifact disentanglement network for unsupervised metal artifact reduction," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 634–643, 2020.
- [18] H. Wang, Y. Li, N. He, K. Ma, D. Meng, and Y. Zheng, "DICDNet: Deep interpretable convolutional dictionary network for metal artifact reduction in ct images," *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 869–880, 2022.
- [19] H. Wang, Q. Xie, D. Zeng, J. Ma, D. Meng, and Y. Zheng, "OSCNet: Orientation-shared convolutional network for ct metal artifact learning," *IEEE Transactions on Medical Imaging*, vol. 43, no. 1, pp. 180–192, 2024.
- [20] M. U. Ghani and W. C. Karl, "Fast enhanced ct metal artifact reduction using data domain deep learning," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 181–193, 2019.
- [21] L. Yu, Z. Zhang, X. Li, and L. Xing, "Deep sinogram completion with image prior for metal artifact reduction in ct images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 228–238, 2021.
- [22] L. Yu, Z. Zhang, X. Li, H. Ren, W. Zhao, and L. Xing, "Metal artifact reduction in 2d ct images with self-supervised cross-domain learning," *Physics in Medicine and Biology*, vol. 66, no. 17, p. 175003, 2021.
- [23] B. Zhou, X. Chen, S. K. Zhou, J. S. Duncan, and C.-C. Liu, "DuDoDR-Net: Dual-domain data consistent recurrent network for simultaneous sparse view and metal artifact reduction in computed tomography," *Medical Image Analysis*, vol. 75, p. 102289, 2022.
- [24] H. Wang, Y. Li, H. Zhang, D. Meng, and Y. Zheng, "InDuDoNet+: A deep unfolding dual domain network for metal artifact reduction in ct images," *Medical Image Analysis*, vol. 85, p. 102729, 2023.
- [25] M. Du, K. Liang, L. Zhang, H. Gao, Y. Liu, and Y. Xing, "Deep-learning-based metal artefact reduction with unsupervised domain adaptation regularization for practical ct images," *IEEE Transactions on Medical Imaging*, vol. 42, no. 8, pp. 2133–2145, 2023.
- [26] Z. Li, Q. Gao, Y. Wu, C. Niu, J. Zhang, M. Wang, G. Wang, and H. Shan, "Quad-Net: Quad-domain network for ct metal artifact reduction," *IEEE Transactions on Medical Imaging*, vol. 43, no. 5, pp. 1866–1879, 2024.
- [27] B. Shi, S. Zhang, K. Jiang, and Q. Lian, "Coupling model- and data-driven networks for ct metal artifact reduction," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 415–428, 2024.
- [28] J. Su, C. Wang, Y. Li, D. Liang, and K. Shang, "F2IFlow for ct metal artifact reduction," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 1–14, 2024.
- [29] W. Wang, X.-G. Xia, and C. He, "A dual-domain deep network for high pitch CT reconstruction," *Pattern Recogn.*, vol. 161, p. 111233, 2025.
- [30] D. Li, J. Sheng, Y. Ge, Z. Duan, J. Zhu, Y. Wang, Z. Bian, J. Ma, X. Huang, and D. Zeng, "Robust image reconstruction with real-world noise modeling for low-dose photon-counting detector CT," *Pattern Recogn.*, vol. 174, p. 112942, 2026.
- [31] C. Ma, Z. Li, J. He, J. Zhang, Y. Zhang, and H. Shan, "Universal pre-training for generalizable incomplete-view CT reconstruction," *Pattern Recogn.*, vol. 178, p. 113513, 2026.
- [32] J. Lee, J. Gu, and J. C. Ye, "Unsupervised ct metal artifact learning using attention-guided β -cyclegan," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3932–3944, 2021.
- [33] G. M. Karageorgos, J. Zhang, N. Peters, W. Xia, C. Niu, H. Paganetti, G. Wang, and B. De Man, "A denoising diffusion probabilistic model for metal artifact reduction in ct," *IEEE Transactions on Medical Imaging*, vol. 43, no. 10, pp. 3521–3532, 2024.
- [34] X. Liu, Y. Xie, S. Diao, S. Tan, and X. Liang, "Unsupervised ct metal artifact reduction by plugging diffusion priors in dual domains," *IEEE Transactions on Medical Imaging*, vol. 43, no. 10, pp. 3533–3545, 2024.
- [35] T. Cai, X. Li, C. Zhong, W. Tang, and J. Guo, "DiffMAR: A generalized diffusion model for metal artifact reduction in ct images," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 11, pp. 6712–6724, 2024.
- [36] M. Luo, N. Zhou, T. Wang, L. He, W. Wang, H. Chen, P. Liao, and Y. Zhang, "Bi-constraints diffusion: A conditional diffusion model with degradation guidance for metal artifact reduction," *IEEE Transactions on Medical Imaging*, vol. 44, no. 9, pp. 3552–3562, 2025.
- [37] L. Yang, J. Huang, G. Yang, and D. Zhang, "CT-SDM: A sampling diffusion model for sparse-view ct reconstruction across various sampling rates," *IEEE Transactions on Medical Imaging*, vol. 44, no. 6, pp. 2581–2593, 2025.
- [38] W. Wu, Y. Wang, Q. Liu, G. Wang, and J. Zhang, "Wavelet-improved score-based generative model for medical imaging," *IEEE Transactions on Medical Imaging*, vol. 43, no. 3, pp. 966–979, 2024.
- [39] J. Zhang, H. Mao, X. Wang, Y. Guo, and W. Wu, "Wavelet-inspired multi-channel score-based model for limited-angle ct reconstruction," *IEEE Transactions on Medical Imaging*, 2024.
- [40] K. Xu, S. Lu, B. Huang, W. Wu, and Q. Liu, "Stage-by-stage wavelet optimization refinement diffusion model for sparse-view ct reconstruction," *IEEE Transactions on Medical Imaging*, vol. 43, no. 10, pp. 3412–3424, 2024.
- [41] W. Wu, J. Pan, Y. Wang, S. Wang, and J. Zhang, "Multi-channel optimization generative model for stable ultra-sparse-view ct reconstruction," *IEEE Transactions on Medical Imaging*, vol. 43, no. 10, pp. 3461–3475, 2024.
- [42] J. Wu, S. Pan, N. Li, B. Chen, B. An, Z. Wang, Y. Wang, and S.-T. Xia, "Universal image restoration via task-adaptive diffusion degradation oriented model," *Pattern Recogn.*, vol. 176, p. 113193, 2026.
- [43] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow matching for generative modeling," in *International Conference on Learning Representations*, 2023. [Online]. Available: <https://openreview.net/forum?id=PqvMRDCJT9t>
- [44] M. A. A. Hegazy, M. H. Cho, M. H. Cho, and S. Y. Lee, "U-net based metal segmentation on projection domain for metal artifact reduction in dental ct," *Biomedical Engineering Letters*, vol. 9, no. 3, pp. 375–385, 2019.
- [45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer, 2015, pp. 234–241.
- [46] 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, 2021.
- [47] K. Yan, X. Wang, L. Lu, L. Zhang, A. P. Harrison, M. Bagheri, and R. M. Summers, "Deep lesion graphs in the wild: Relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9261–9270.
- [48] Y. Deng, C. Wang, Y. Hui, Q. Li, J. Li, S. Luo, M. Sun, Q. Quan, S. Yang, Y. Hao, P. Liu, H. Xiao, C. Zhao, X. Wu, and S. K. Zhou, "CTSpine1K: A large-scale dataset for spinal vertebrae segmentation in computed tomography," *Machine Learning for Biomedical Imaging*, vol. 3, pp. 824–832, 2025.
- [49] R. M. Scardigno, A. Brunetti, P. M. Marvulli, R. Carli, M. Dotoli, V. Bevilacqua, and D. Buongiorno, "CALIMAR-GAN: An unpaired mask-guided attention network for metal artifact reduction in ct scans," *Computerized Medical Imaging and Graphics*, vol. 123, p. 102565, 2025.
- [50] S. Ravishanker and Y. Bresler, "MR image reconstruction from highly undersampled k -space data by dictionary learning," *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.