

Fast and accurate sparse-view CBCT reconstruction using meta-learned neural attenuation field and hash-encoding regularization

Heejun Shin^a, Taehee Kim^a, Jongho Lee^b, Seyoung Chun^c, Seungryong Cho^d, Dongmyung Shin^{a,*}

^aArtificial Intelligence Engineering Division, Radisen Co. Ltd., Seoul, Republic of Korea

^bLaboratory for Imaging Science and Technology, Department of Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea

^cIntelligent Computational Imaging Laboratory, Department of Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea

^dMedical Imaging and Radiotherapy Laboratory, Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

ARTICLE INFO

Article history:

Keywords: Cone-beam CT, Image reconstruction, Implicit neural representation

ABSTRACT

Cone beam computed tomography (CBCT) is an emerging medical imaging technique to visualize the internal anatomical structures of patients. During a CBCT scan, several projection images of different angles or views are collectively utilized to reconstruct a tomographic image. However, reducing the number of projections in a CBCT scan while preserving the quality of a reconstructed image is challenging due to the nature of an ill-posed inverse problem. Recently, a neural attenuation field (NAF) method was proposed by adopting a neural radiance field algorithm as a new way for CBCT reconstruction, demonstrating fast and promising results using only 50 views. However, decreasing the number of projections is still preferable to reduce potential radiation exposure, and a faster reconstruction time is required considering a typical scan time. In this work, we propose a fast and accurate sparse-view CBCT reconstruction (FACT) method to provide better reconstruction quality and faster optimization speed in the minimal number of view acquisitions (< 50 views). In the FACT method, we meta-trained a neural network and a hash-encoder using a few scans (= 15), and a new regularization technique is utilized to reconstruct the details of an anatomical structure. In conclusion, we have shown that the FACT method produced better, and faster reconstruction results over the other conventional algorithms based on CBCT scans of different body parts (chest, head, and abdomen) and CT vendors (Siemens, Phillips, and GE).

© 2023 Elsevier B. V. All rights reserved.

1. Introduction

Cone beam computed tomography (CBCT) is an emerging medical imaging technique to visualize the internal anatomical structures of patients. Compared to a fan beam CT (FBCT), CBCT takes advantages of higher image resolution and faster scanning time (Lechuga and Weidlich, 2016). During a CBCT scan, a cone-shaped X-ray beam is diverged from a source which circularly rotates an anatomy of interest with an X-ray detector. This results in several projection images of different angles (i.e., views) that are collectively utilized to reconstruct a tomographic image. Nowadays, CBCT has been widely adopted as a practical diagnosis tool in human anatomies, including teeth (Scarfe et al., 2006), extremities (Carrino et al., 2014), and chest (Hohenforst-Schmidt et al., 2014).

However, the amount of ionizing radiation to a patient during a CBCT scan is much higher than in conventional radiography, preventing its wide applications (Lechuga and Weidlich, 2016), (Kan et al., 2008). Reducing the number of projections (i.e., sparse-view; < 100 views) in a CBCT scan can effectively circumvent the high exposure; however, preserving the quality of a reconstructed image is challenging due to the nature of an ill-posed inverse problem (Gao et al., 2014).

Conventional CBCT reconstruction methods such as the Feldkamp-Davis-Kress (FDK) algorithm (Feldkamp et al., 1984) provide best image quality in an ordinary CBCT scan with hundreds of views but suffering from streak artifacts in a sparse-view scan. Iterative optimization algorithms, such as simultaneous algebraic reconstruction technique (SART) or adaptive steepest descent-projection on convex subsets (ASD-POCS) (Andersen and Kak, 1984; Sidky and Pan, 2008), produce good image quality in a limited number of views, but they

*Corresponding author. email: shinsae11@radisentech.com;

require more computational time and produce unsatisfactory results as the number of views becomes very limited (e.g., 50 views or less).

As more recent approaches, several deep learning-based CT reconstruction techniques, primarily based on convolutional neural networks (CNNs) have successfully promoted superior image quality in a sparse-view setting. Those methods are typically categorized as one of three: sinogram-domain methods (Lee et al., 2018; Anirudh et al., 2018; Tang et al., 2019), that directly interpolate or extrapolate sinograms to fill out missing information, images-domain methods (Jin et al., 2017; Zhang et al., 2018; Georgescu et al., 2020; Liu et al., 2020) which use CNNs to recover reconstructed images with artifacts, and dual-domain methods (Hu et al., 2020; Wang et al., 2021; Lin et al., 2019; Zhou et al., 2022; Wu et al., 2021) that combine these two approaches to utilize mutual information in both domains. However, those methods, which rely on each sinogram to reconstruct each 2D CT slice, are limited to being generalized to CBCT reconstruction, which requires using multiple 2D projections simultaneously to reconstruct a 3D tomography. In addition, the methods inherently require a lot of training CT data and suffer from potential training biases.

Recently, a neural radiance field (NeRF) algorithm (Mildenhall et al., 2021) has been gaining tremendous popularity as an effective way to synthesize novel views of natural scenes or objects using a set of captured views. In NeRF, rather than a complex CNN, a simple feed-forward network is adopted to learn the relationship (i.e., implicit neural representation (Park et al., 2019)) between the spatial coordinates in a 3D volume and metrics for these coordinates, such as colors and densities, to synthesize novel views. Based on the NeRF method, Zha et al. (2022) proposed a new self-supervised method for CBCT reconstruction, which is called a neural attenuation field (NAF), demonstrating fast CBCT reconstruction with good quality using only 50 projections. Identical to the NeRF method, the NAF method does not require any training data and produces better reconstruction results than other conventional methods.

Although the NAF method reported promising results as a novel way for CBCT reconstruction, some room for improvement remains. First, a faster reconstruction time is more favorable since the convergence time of NAF is still relatively long compared to a typical CBCT scan time (Scarfe and Farman, 2008). Second, decreasing the number of projections, even less than 50 views, is more preferable, unless the image quality is acceptable for certain clinical purposes (e.g., patient positioning (Thilmann et al., 2006), emergency scan (Jacques et al., 2021), etc.), to reduce the amount of radiation exposure.

In this work, we propose a fast and accurate sparse-view CBCT reconstruction (FACT) method, which utilizes a meta-learning framework and a novel feature-encoding technique to alleviate potential disadvantages of NAF. In the FACT method, we first pre-trained a neural network using a small set of CBCT scans (= 15 scans) based on a meta-learning framework. Then, this meta-learned network is utilized to reconstruct a 3D tomography quickly using a new CBCT scan at test time. During the reconstruction, we regularized the hash-encoding process to further optimize the details of the anatomical structure even in

the minimal number of views (< 50 views).

In summary, the main contributions of this paper are the following:

- We propose a FACT method that provides faster optimization speed and better CBCT reconstruction quality than a NAF method in the minimal number of views (< 50 views) based on a meta-learning framework and a novel hash-encoding regularization process.

- Unlike previous supervised CT reconstruction methods that extensively require a large amount of CT data (> a few hundred) to train neural networks, the FACT method requires only a small number of scans (= 15 scans) to meta-train a neural network, avoiding preparation of a large amount of data and potential biases in AI training.

- Throughout the extensive experiments, we have shown that, regardless of CT vendors (Siemens, Phillips, and GE) and body parts (chest, head, and abdomen), the FACT method produced consistently better CBCT reconstruction results than the NAF method.

2. Related Works

2.1. Sparse-View CT Reconstruction

Classical CBCT reconstruction algorithms (Feldkamp et al., 1984; Andersen and Kak, 1984; Sidky and Pan, 2008) include analytical optimization methods (e.g., FDK) that estimate attenuation coefficients based on the inverse Radon transform and iterative methods (e.g., SART, ASD-POCS) that apply iterative reconstruction algorithms until convergence. However, FDK works best when hundreds of projections are available and suffers from artifacts in a sparse-view scan. In contrast, iterative methods, such as ASD-POCS, can suppress artifacts but consume more memory and computational time.

In recent years, many deep learning-based approaches have been proposed for sparse-view CT reconstruction. Those works can be categorized into three methods: sinogram-domain, image-domain, and dual domain. Sinogram-domain methods enhance projection data (i.e., sinogram) using neural networks prior to image reconstruction. Lee et al. (2018) proposed an Unet-based network to fill out the missing parts in sparsely-sampled sinograms, and Tang et al. (2019) applied a CNN-based super-resolution method to up-sample low-resolution sinograms. On the other hand, image-domain methods improve the quality of reconstructed images using neural networks. Jin et al. (2017) trained an Unet using reconstructed images from sub- and full-sampled sinograms to restore input CT images with artifacts. Zhang et al. (2018) improved the quality of reconstructed images using a DenseNet and a deconvolution-based network. Finally, dual-domain methods combine information in both domains for reconstruction. Lin et al. (2019) proposed a novel radon inversion layer that concurrently refined sinograms and corresponding images. Similarly, Zhou et al. (2022) introduced a new network layer called an image and projection data consistency layer. However, most of the works above utilized CNNs with a lot of training data, and those methods are challenging to be generalized to CBCT reconstruction.

More recently, with the success of an implicit neural representation for a novel view synthesis (Mildenhall et al., 2021), some studies (Zang et al., 2021; Zha et al., 2022) adopted a NeRF algorithm for CBCT reconstruction and proposed self-supervised methods that parameterized attenuation coefficients as continuous attenuation field, reporting promising results using only tens of projection images.

2.2. Neural Radiance Field

NeRF (Mildenhall et al., 2021) utilizes a neural network to map 3D spatial coordinates to values, such as colors and densities, using a set of 2D views. Despite the success of the original NeRF, it requires many input views and a lengthy optimization time. To solve these problems, Yu et al. (2021) and Chen et al. (2021) adopted CNNs to speed up a neural encoding process in NeRF. Tancik et al. (2021) used a meta-learning framework to pre-initialize a neural network in NeRF, and Jain et al. (2021) introduced auxiliary semantic consistency loss to supervise a neural network to synthesize a new scene with a limited number of input views. Other studies, such as Niemeyer et al. (2022); Yang et al. (2023), proposed novel regularization techniques to reduce the number of views effectively.

In NeRF, positional encoding is commonly utilized to encode low-dimensional input features for a neural network to high-dimensional features to prevent the bias of the network toward learning low frequency components in the input images (Rahaman et al., 2019; Ronen et al., 2019). This encoding process helps the neural network map high-frequency components for a novel view generation (Mildenhall et al., 2021; Tancik et al., 2020). One of the popular encoding approaches is to discretize a 3D space into multiresolution grids and interpolate trainable features in different resolutions. Takikawa et al. (2021) utilized a tree structure to save and manage multi-scale features in a 3D grid. Müller et al. (2022) adopted a hash table to efficiently control the trainable features' memory while producing small and enriched features, making the optimization process faster.

2.3. Optimization-Based Meta-Learning

Optimization-based meta-learning aims to train a pre-initialized AI algorithm that can be rapidly adopted to unseen tasks using a few examples. This method usually utilizes two models, which are an inner model (i.e., task-specific learner) and an outer model (i.e., meta-learner), to find the best initialization of the outer model that will be optimized fast within a few gradient steps. Modal-agnostic meta-learning (MAML) (Finn et al., 2017) first updates an inner model by sampling a task from a set of multiple tasks and gradually updates the outer model based on the updated weights of the inner model. For the faster adaptation of an AI model, Lee and Choi (2018) incorporated task-specific information into MAML by combining meta-learned activation layers in the inner model. Similarly, Flennerhag et al. (2019) applied a warped gradient descent optimization to the inner model. Other studies, such as (Nichol et al., 2018; Rajeswaran et al., 2019) tried to simplify the optimization process of MAML for faster adaptation. For example, a Reptile method (Nichol et al., 2018) approximated MAML using first-order gradient optimization to reduce the computational time and memory burdens.

3. Materials and Methods

3.1. CBCT Reconstruction Based on Implicit Neural Representation

The goal of CBCT reconstruction is to extract 3D tomographic information using a set of 2D X-ray projections. Compared to a FBCT scan, a CBCT scan uses a circularly rotating X-ray source with a panel detector to acquire a set of N 2D X-ray images with different views (i.e., X-ray projection set; $\mathbf{X} = \{X_1, X_2, \dots, X_N \in \mathbb{R}^{H \times W}\}$, where H and W are the height and width of images, respectively). The CBCT acquisition (φ), therefore, depends on a 3D object of interest (C), scanning angles of each view ($\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_N\}$), and other parameters (β ; e.g., distance between X-ray source and detector, detector resolution, etc.).

$$\varphi(C, \alpha, \beta) \rightarrow \mathbf{X} \quad (1)$$

Then we can formulate the CBCT reconstruction problem to find an inverse function F that accurately estimates an attenuation coefficient ($\mu \in R$) on a query point ($p \in R^3$) in a 3D object (i.e., $\mu = C(p)$) using the acquired set of 2D X-ray images (\mathbf{X}) and other parameters (α and β):

$$\mu = C(p) = F(p; \mathbf{X}, \alpha, \beta) \quad (2)$$

To achieve this goal, conventional approaches (Feldkamp et al., 1984; Andersen and Kak, 1984; Sidky and Pan, 2008), such as analytical methods (e.g., back-projection, etc.), recover attenuation coefficients in the uniform grid points of a 3D volume. In contrast, our approach, similar to Zha et al. (2022), introduces a mapping function (f_θ), which is a feed-forward neural network, to map a continuous spatial coordinate inside the 3D volume to a corresponding attenuation coefficient. In other words, the neural network was trained to implicitly learn the spatial relationship between the coordinates and attenuation coefficients (i.e., implicit neural representation (Park et al., 2019)). Before feeding the 3D coordinate into the network, we encoded it using a trainable hash-encoder (h_ϕ) (Müller et al., 2022):

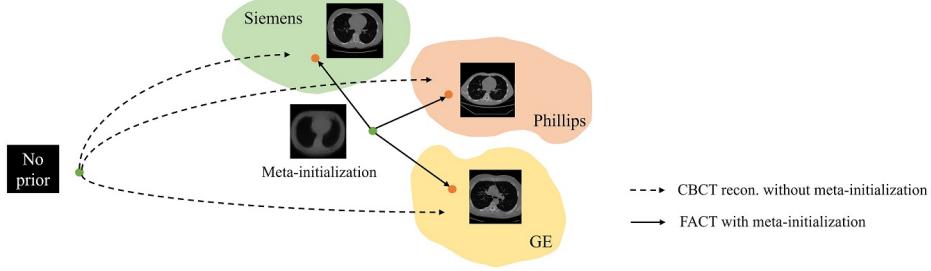
$$f_\theta(h_\phi(p)) \rightarrow \mu \quad (3)$$

To train those two functions (f_θ and h_ϕ), a virtual ray was cast from a cone beam source to a 3D space, passing through M number of coordinates (p_1, p_2, \dots, p_M) along the ray trajectory and, consequently, producing M number of attenuation coefficients ($\mu_1, \mu_2, \dots, \mu_M$) from the hash encoder (h_ϕ) and neural network (f_θ ; see Fig. 1 b). Then, we can project another ray in the same direction of the ray casting and calculate the intensity of a point after the projection (e.g., purple point on a projected X-ray image in Fig. 1 b) based on Beer's Law (Zha et al., 2022):

$$I = I_0 \exp \left(- \sum_{i=1}^M \mu_i \delta_i \right) = I_0 \exp \left(- \sum_{i=1}^M f_\theta(h_\phi(p_i)) \delta_i \right) \quad (4)$$

where I is an X-ray intensity after the projection, I_0 is an initial intensity, and δ is the distance between the two adjacent co-

(a) FACT reconstruction with meta-initialization



(b) FACT reconstruction with hash-encoding regularization

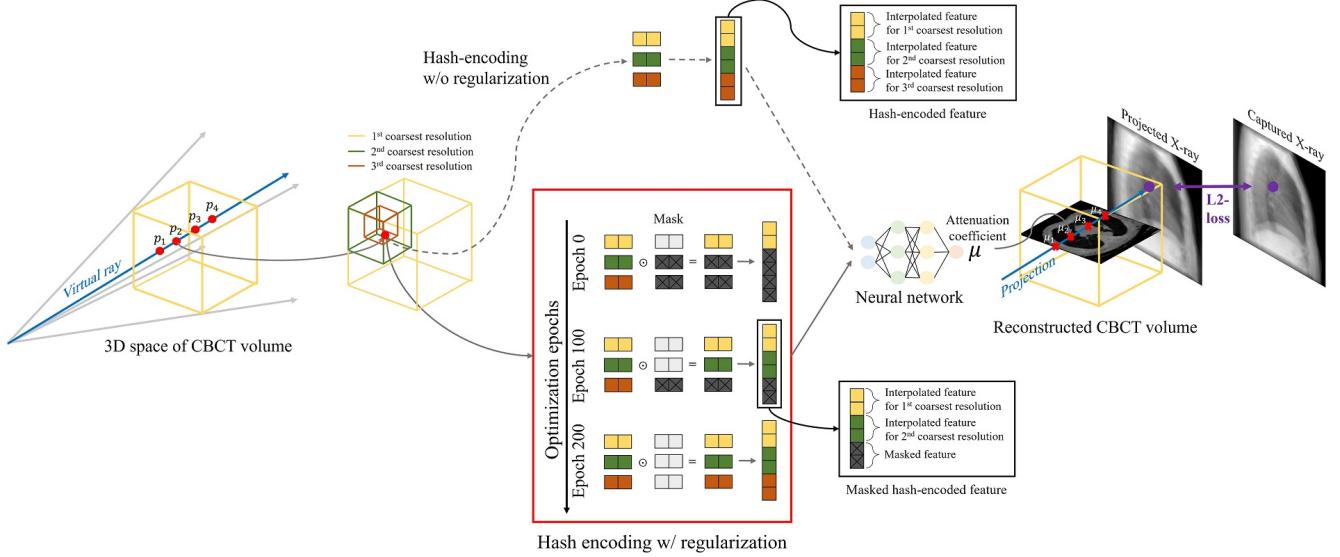


Fig. 1. Illustration of the FACT reconstruction with the meta-initialization and hash-encoding regularization. (a) By meta-learning the CBCT reconstruction functions (i.e., meta-initialization), the FACT reconstruction quickly optimizes each target scan to a specific CBCT image (dotted line vs. solid line). (b) In the FACT reconstruction, the hash-encoder, which includes feature vectors of multiresolution 3D grid points, generates feature embeddings (hash-encoded feature) of the query points (e.g., red point p_2) using tri-linear interpolation for each resolution (interpolated features). In the FACT method, a novel regularization technique (solid line and red box) for the hash-encoder is incorporated, which masks out the interpolated features depending on the optimization epochs, resulting in the masked hash-encoded feature.

ordinates. Assuming that \tilde{I} is a point intensity in an acquired 2D X-ray image and matched with the projected point, then the two mapping functions (i.e., f_θ and h_ϕ) can be optimized by minimizing the difference in the intensities between these two points using backpropagation. We can perform the optimization by iterating all the acquired X-ray images ($\mathbf{X} = \{X_1, X_2, \dots, X_N\}$), considering all the points in each image (i.e., $H \times W$ number of points), and minimizing a mean squared error as follows:

$$\text{argmin}_{\theta, \phi} L(\theta, \phi, \mathbf{X}) = \text{argmin}_{\theta, \phi} \sum_{k=1}^N L'(\theta, \phi, X_k) \quad (5)$$

$$\text{where } L'(\theta, \phi, X_k) = \frac{1}{H \times W} \sum_{j=1}^{H \times W} (\tilde{I}_{j,k} - I_{j,k})^2$$

$$= \frac{1}{H \times W} \sum_{j=1}^{H \times W} \left(\tilde{I}_{j,k} - I_0 \exp \left(- \sum_{i=1}^M f_\theta(h_\phi(p_{i,j,k})) \delta_{i,j,k} \right) \right)^2$$

Later, the optimized functions are used to fill out the attenuation coefficient values inside a 3D volume for CBCT reconstruction.

Here, the f_θ and h_ϕ are initialized with random weights, and, therefore, those functions should be optimized from scratch during a test time for a single CBCT data, requiring much computational time. To address this issue, we propose a new strategy to pre-train those two functions using a small set of CBCT scans and regularize them to improve the optimization speed and reconstruction quality.

3.2. FACT Method

3.2.1. FACT Reconstruction with Meta-Initialization

In the FACT method, a meta-learning framework was adopted to initialize learnable weights of the mapping functions, f_θ and h_ϕ (see Section 3.1), using a few CBCT scans (15 scans; see Section 3.3). For this purpose, we modified a Reptile method (Nichol et al., 2018) to update and initialize those two functions by utilizing sets of X-ray projections (i.e., no reconstructed CBCT images are required) for meta-initialization. Pseudo code for CBCT meta-initialization is shown in Algorithm 1.

Algorithm 1 CBCT Meta-Initialization

Require: Prepare P X-ray projection sets $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_P\}$

```

Randomly initialize weights  $\theta_0$  of  $f_\theta$  and  $\phi_0$  of  $h_\phi$ 
for iteration = 1, 2, ... do
    Randomly sample a projection set ( $\mathbf{X}$ )
     $\theta'_0 = \theta_0$ 
     $\phi'_0 = \phi_0$ 
    for epoch = 1, 2, ...,  $K$  do
        Update  $\theta'_0 = \theta'_0 - \gamma \nabla_{\theta'_0} L(\theta'_0, \phi'_0, (\mathbf{X}))$ 
        Update  $\phi'_0 = \phi'_0 - \gamma \nabla_{\phi'_0} L(\theta'_0, \phi'_0, (\mathbf{X}))$ 
    end for
    Update  $\theta_0 = \theta_0 - \epsilon(\theta'_0 - \theta_0)$ 
    Update  $\phi_0 = \phi_0 - \epsilon(\phi'_0 - \phi_0)$ 
end for
 $\theta_0^* \leftarrow \theta_0$ 
 $\phi_0^* \leftarrow \phi_0$ 

```

Using the P X-ray projection sets ($\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_P$), the parameters (θ_0 and ϕ_0) of the two mapping functions (f_θ and h_ϕ) are meta-initialized. To do so, based on the randomly sampled projection set (\mathbf{X}), the two optimization loops (i.e., inner loop and outer loop) are iterated sequentially, resulting in the final parameters of θ_0^* and ϕ_0^* . Here, γ is an inner loop learning rate, and ϵ is an outer loop learning rate.

Fig. 1 a. illustrates the reconstruction process of the FACT method with (solid line in Fig. 1 a) and without (dotted line in Fig. 1 a) the meta-initialization. With the meta-initialization, since the mapping functions are already pre-trained to learn CBCT prior, the FACT reconstruction quickly and accurately optimizes each target scan to a specific CBCT image. Furthermore, it requires no additional computational cost at the reconstruction time.

3.2.2. FACT Reconstruction with Hash-Encoding Regularization

In the hash-encoding process (Müller et al., 2022), 3D coordinates of points along the trajectory of a virtual ray were encoded using the hash-encoder (see dotted line in Fig. 1 b and h_ϕ in Section 3.1). The hash-encoder is a lookup table that includes

feature vectors of multi-resolution 3D grid points. In each resolution, learnable feature vectors were assigned in the corner points (e.g., corner points of the yellow or green box in Fig. 1 b). When a query point with a 3D coordinate is given (e.g., p_2 in Fig. 1 b), the feature vectors of neighboring corner points were tri-linearly interpolated to generate new feature embeddings of the query point for each resolution (e.g., interpolated features for the 1st and 2nd coarsest resolutions in Fig. 1 b). Then these interpolated features were concatenated to a single vector (i.e., hash-encoded feature) and fed into the neural network (f_θ in Section 3.1) to produce an attenuation coefficient (e.g., μ_2 in Fig. 1 b).

In our FACT method, a new way to regularize the hash-encoder is introduced by masking out the hash-encoded feature depending on the optimization epochs (see dotted line vs. solid line in Fig. 1 b). Given a hash-encoded feature vector h_ϕ ($\in R^{L \times F}$, where L is the number of resolution levels, and F is the number of feature dimensions per entry; see Table 1 in Müller et al. (2022) for notations), we multiplied a binary mask m ($\in R^{L \times F}$) to $h_\phi(p)$, so it can regularize the visible resolution through the optimization as follows:

$$\hat{h}_\phi(t, s; p) = h_\phi(p) \odot m(t, s) \quad (6)$$

$$m_i(t, s) = \begin{cases} 1 & \text{if } i \leq r \\ 0 & \text{if } i > r \end{cases} \quad \text{where } r = \left\lfloor \frac{t}{T} \right\rfloor + s$$

where \hat{h}_ϕ is the masked hash-encoded feature (see Fig. 1 b) inputted to the neural network, $m_i(t, s)$ denotes the i^{th} row of the $m(t, s)$, t is a current epoch, T is an incremental number of epochs between resolutions, and s determines the number of rows of the masked hash-encoded feature to be used in an initial epoch.

The red box in Fig. 1 b shows the case when the parameters are $L = 3$, $F = 2$, $s = 1$, and $T = 100$ as a simple example. In this case, until 100 epochs, only the interpolated feature of the 1st coarsest resolution is considered as an input to the neural network, while the other interpolated features are masked. Then, after 100 epochs, the interpolated feature of the 2nd coarsest resolution is additionally considered, and so on. This procedure regularizes the resolution levels of an input to the network to prevent high-resolution overfitting at the start of the optimization, similar to Yang et al. (2023).

3.3. Data Preparation and Implementation Details

We collected publicly available human organ CT datasets of the chest, abdomen, and head. The CT scans were from the LIDC-IDRI dataset for the chest (Armato III et al., 2011), the AbdomenCT-1K dataset for the abdomen (Ma et al., 2021), and the Head-Neck-PET-CT dataset for the head (Vallieres et al., 2017). For each dataset, we randomly sampled 15 CT scans for meta-learning and 15 CT scans for evaluation. The chest CT dataset provided detailed information about CT vendors for each scan, and, for meta-learning, we equally distributed 15 CT scans based on the three most prevalent vendors (Siemens, GE, and Phillips; 5 scans for each vendor). As a pre-processing step,

Table 1. 3D SSIM and 3D PSNR values depending on the different numbers of input views for the CBCT reconstruction methods

Optimization method	50 Views		40 Views		30 Views		20 Views		10 Views	
	PSNR	SSIM								
FDK	18.6	0.48	17.5	0.43	16.9	0.37	16.0	0.29	14.1	0.18
SART	27.2	0.72	27.5	0.72	26.7	0.68	25.5	0.62	22.9	0.51
ASD-POCS	29.1	0.79	28.7	0.77	28.0	0.74	26.8	0.69	23.7	0.56
NAF	29.9	0.81	28.9	0.77	28.0	0.73	27.0	0.69	23.6	0.59
FACT (w/o meta)	30.9	0.83	30.6	0.82	29.8	0.79	28.2	0.73	25.5	0.65
FACT (w/o reg)	31.6	0.84	31.2	0.83	30.4	0.80	29.1	0.75	26.7	0.67
FACT (Ours)	32.0	0.85	31.5	0.84	30.7	0.81	29.5	0.77	27.0	0.70

all the CT images were first resampled to have isotropic voxels (1 mm^3) and reshaped to a size of $256 \times 256 \times 256$.

To acquire X-ray projections using a cone-beam source, we generated digitally reconstructed radiographs based on a CBCT simulator (tomographic iterative GPU-based reconstruction toolbox; TIGRE) (Biguri et al., 2016). The TIGRE tool repetitively simulated different numbers of X-rays five times (50, 40, 30, 20, and 10 X-rays) for each CT scan with equally distributed views between 0° and 180° . The parameters we used for the simulation were as follows: distance source-detector = 2000, distance source origin = 1000, detector resolution = (256, 256), detector pixel size = (2.0, 2.0), CT resolution = (256, 256, 256), and CT voxel size = (1.0, 1.0, 1.0).

The neural network’s architecture to output an attenuation coefficient (see Fig. 1 b) had four layers with 32 hidden nodes. For the meta-learning, we set the number of epochs (K) = 200, an inner loop learning rate (γ) = 0.001, and an outer loop learning rate (ϵ) = 0.001 (see Algorithm 1 in Section 3.2.1). For the parameters of the hash-encoder, we set the number of levels (L) = 8, number of feature dimensions per entry (F) = 2, incremental number of epochs between resolutions (T) = 25, number of rows of the masked hash-encoded feature to be used in an initial epoch (s) = 3, coarsest resolution (N_{min}) = 8, and hash table size (T) = 2^{19} (see Section 3.2.2 in details and Table 1 in Müller et al. (2022) for notations). The other parameters include a batch size = 256 and the maximum number of epochs for optimization at test time = 1,500.

Our hardware includes GPU = NVIDIA RTX 3090, CPU = Intel(R) Xeon(R) Silver 4310, and memory = 64 GB. We used Python 3.9.16 and PyTorch 1.11.0 for implementation.

3.4. Experiments

We compared the quality of the CBCT reconstruction results using two metrics, 3D structural similarity index measure (SSIM) and 3D peak signal-to-noise ratio (PSNR), between the FDK, SART, ASD-POCS, NAF, and FACT methods for each 50, 40, 30, 20, and 10 views. The average values of those metrics were calculated over the 15 testing images for each human organ (chest, head, and abdomen). Also, to check the consistency of the reconstruction results, we reported the 3D SSIM and 3D PSNR values depending on the different CT vendors (Siemens, Phillips, and GE; five testing chest CTs for each) between the NAF and FACT methods.

To compare the optimization speed between all the methods (FDK, SART, ASD-POCS, NAF, and FACT), we measured the

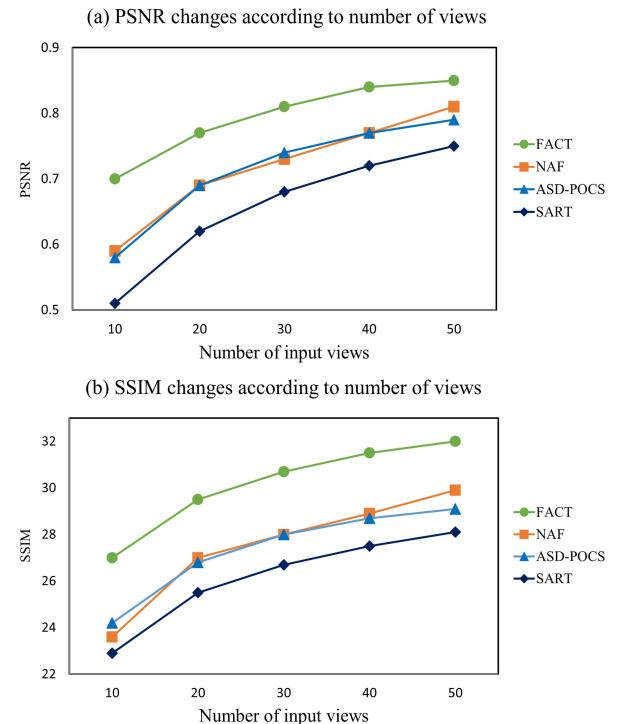


Fig. 2. 3D SSIM and 3D PSNR values according to the change in the number of input views (50, 40, 30, 20, and 10 views) for the SART, ASD-POCS, NAF, and FACT methods. Regardless of the number of views, the FACT method outperformed the others regarding SSIM and PSNR. Especially in the graph of SSIM, FACT revealed more gains as the number of views decreased.

mean running times (in minutes) of each method over the 15 testing images of the chest. To measure the running times of the conventional algorithms (FDK, SART, and ASD-POCS), we used pre-implemented functions of the FDK, SART (50 iterations), and ASD-POCS (20 iterations with 25 total variation minimization steps) in the TIGRE library. For the running times of the NAF and FACT methods, we calculated the time when the variation of 3D PSNR values was converged within 0.01 for 50 consecutive epochs (i.e., convergence time). To run the NAF algorithm, we utilized a source code from Zha et al. (2022) on the web and followed the same parameters in Section 3.3.

For the key parameters of the hash-encoding regularization, including the incremental number of epochs between resolu-

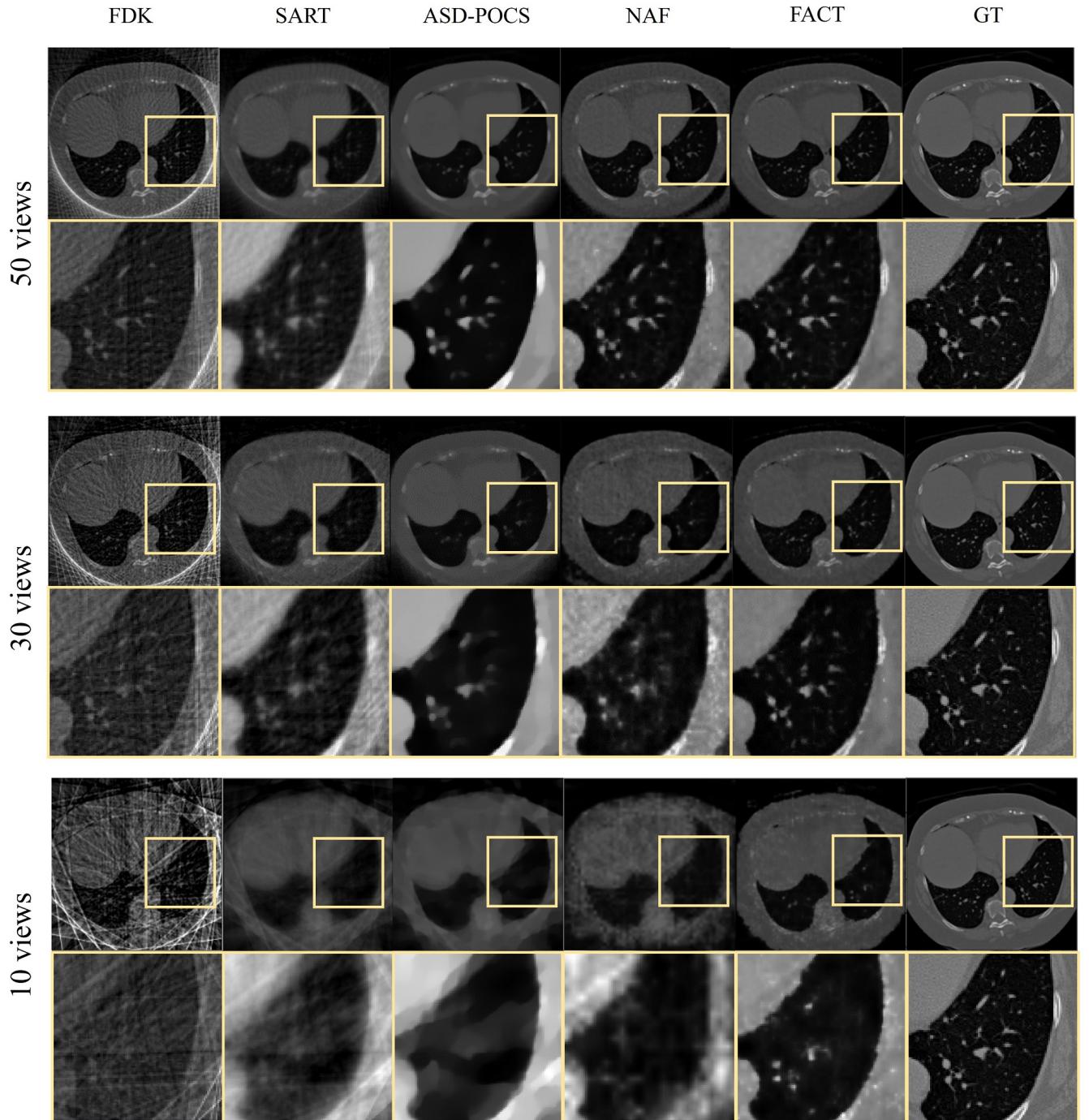


Fig. 3. Example CBCT reconstruction results between the different CBCT reconstruction methods (FDK, SART, ASD-POCS, NAF, and FACT). In the zoomed-in plots (yellow boxes), the FACT method demonstrated better image quality, successfully reconstructing anatomical structures and suppressing artifacts than the other methods (e.g., results of NAF vs. FACT in 30 views).

tions (T) and number of rows of the masked hash-encoded feature to be used in an initial epoch (s), we performed a manual search to find the optimal parameters (see Section 4.3).

4. Results

4.1. CBCT Reconstruction Results

Fig. 2 shows the graphs that report 3D SSIM and 3D PSNR values according to the change in the number of input views (50, 40, 30, 20, and 10 views) for the SART, ASD-POCS, NAF,

Table 2. 3D PSNR and 3D SSIM values for the CBCT reconstruction results of the head and abdomen

Body part	Optimization method	50 Views		40 Views		30 Views		20 Views		10 Views	
		PSNR	SSIM								
Head	FDK	20.1	0.45	20.7	0.40	19.8	0.34	18.6	0.26	15.9	0.16
	SART	29.7	0.83	29.3	0.82	28.4	0.78	26.9	0.71	24.1	0.57
	ASD-POCS	31.1	0.89	30.5	0.87	29.0	0.84	28.0	0.78	25.6	0.68
	NAF	32.3	0.90	31.6	0.89	30.4	0.86	28.7	0.79	26.2	0.72
	FACT (Ours)	33.8	0.92	32.9	0.91	31.9	0.89	30.3	0.84	27.9	0.77
Abdomen	FDK	16.5	0.49	18.5	0.48	17.9	0.43	17.0	0.36	13.8	0.21
	SART	28.1	0.81	28.3	0.80	27.6	0.77	26.4	0.72	23.4	0.61
	ASD-POCS	30.3	0.87	29.7	0.86	28.9	0.83	27.6	0.79	25.3	0.71
	NAF	31.4	0.87	30.6	0.84	29.9	0.81	28.5	0.77	26.0	0.71
	FACT (Ours)	32.9	0.91	32.2	0.89	31.3	0.86	30.0	0.82	27.2	0.76

Table 3. 3D SSIM and 3D PSNR values for the CBCT reconstruction results between the NAF and FACT methods according to the three CT vendors (Siemens, Phillips, and GE)

	Metric	Vendor	NAF	FACT
50 Views	PSNR	Siemens	28.1	30.7
		Philips	31.6	33.3
		GE	30.0	31.9
	SSIM	Siemens	0.76	0.81
		Philips	0.84	0.88
		GE	0.82	0.86
30 Views	PSNR	Siemens	26.2	29.5
		Philips	29.6	32.1
		GE	28.4	30.5
	SSIM	Siemens	0.66	0.77
		Philips	0.77	0.85
		GE	0.76	0.82
10 Views	PSNR	Siemens	21.4	29.1
		Philips	25.7	31.7
		GE	23.6	30.0
	SSIM	Siemens	0.51	0.75
		Philips	0.65	0.83
		GE	0.60	0.80

and FACT methods. Regardless of the number of views, the FACT method outperformed the others (see Table 1 for details, including FDK). Especially in the graph of SSIM (Fig. 2 a), FACT revealed more gains as the number of views decreased compared to the others. When the number of views was very limited (e.g., 30 views). Indeed, when we investigated the reconstructed images (Fig. 3), the FACT method showed better image quality, disclosing the anatomical structures and suppressing artifacts than the other methods (see zoomed-in yellow boxes in Fig. 3).

Table 1 also reports the FACT method’s 3D PSNR and 3D SSIM values, excluding the meta-initialization or hash-encoding regularization. When we ablated the hash-encoding regularization (i.e., FACT w/o reg. in Table 1), the reconstruction quality was degraded (e.g., 30.7 vs. 30.4 in 3D PSNR for 30 views). However, ablating the meta-initialization showed a more significant performance drop (e.g., 30.7 and 29.8 in 3D

PSNR for 30 views). The best performance was achieved when the meta-initialization and hash-encoding regularization were combined.

Table 2 summarizes 3D PSNR and 3D SSIM values for the CBCT reconstruction results of different body parts (head and abdomen). FACT showed the highest scores in terms of PSNR and SSIM. Fig. 4 displays the head and abdomen examples for 50 and 30 views, and the results of FACT demonstrated the best image quality, which was in line with the performance metrics in Table 2.

Table 3 summarizes 3D SSIM and 3D PSNR values of the CBCT reconstruction results between the NAF and FACT methods according to the three CT vendors (Siemens, Phillips, and GE) for the chest CBCT. Regardless of the CT vendors, the FACT method consistently outperformed the NAF method (e.g., 3D PSNR for 30 views: 26.2 vs. 29.5 for Siemens, 29.6 vs. 32.1 for Phillips, and 28.4 vs. 30.5 for GE), implying that FACT did not have a substantial bias toward a specific CT vendor after the meta-initialization. Additionally, Supplementary Fig. 1 shows example reconstruction results for 30 views.

4.2. Optimization Speed of CBCT Reconstruction

Fig. 5 compares the running times of the different reconstruction methods (SART, ASD-POCS, NAF, and FACT). The running times of the conventional algorithms (FDK, SART, and ASD-POCS; FDK took a few seconds for all cases) were relatively fast compared to those of the NAF and FACT methods. In all the configurations, NAF required the longest time for reconstruction, between 2 minutes and 12 minutes, while FACT showed comparable running times to those of the conventional methods for some cases. For example, in the case of 30 views, the running time of FACT was faster than that of ASD-POCS but slower than that of SART. However, in the case of 10 views, the running time of FACT was even slightly faster than SART.

Table 4 summarizes the number of epochs and optimization times of the NAF and FACT methods to reach the best performance of the NAF method. In the case of 30 views, the FACT method reached the best performance of NAF, spending less than 14 percent of the number of epochs and optimization times (e.g., 1456 epochs and 397 seconds for NAF vs. 197 epochs and 54 seconds for FACT in 3D SSIM). The gap was more significant in the case of 10 views where the FACT method only

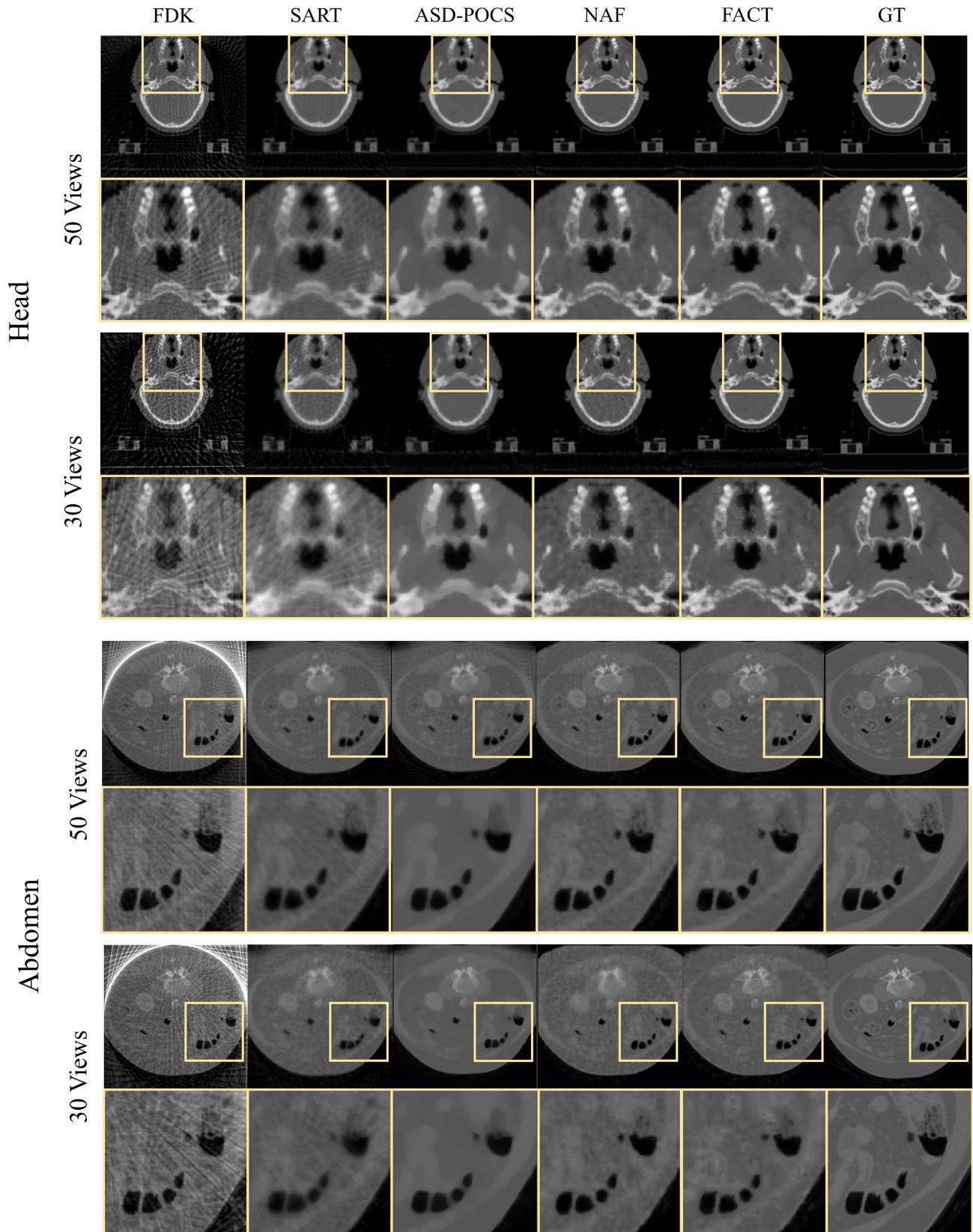


Fig. 4. Example head and abdomen CBCT reconstruction results, including FDK, SART, ASD-POCS, NAF, and FACT, for 50 and 30 views. The results of the FACT method showed the best quality compared to the others.

Table 4. Number of epochs and optimization times of the NAF and FACT methods to reach the best performance of the NAF method

Number of views	Optimization method	Number of epochs to reach best NAF performance		Optimization time to reach best NAF performance (sec.)	
		3D PSNR	3D SSIM	3D PSNR	3D SSIM
50 Views	NAF	1490	1467	677	667
	FACT	256	303	116	138
30 Views	NAF	1363	1456	371	397
	FACT	172	197	47	54
10 Views	NAF	1380	1420	125	129
	FACT	145	103	13	9

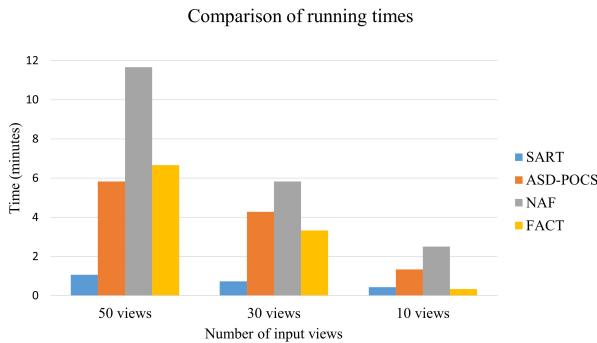


Fig. 5. Comparison of the running times of the different reconstruction methods (SART, ASD-POCS, NAF, and FACT). For some cases, FACT showed comparable running times to those of the conventional methods (SART and ASD-POCS). For example, in the case of 30 views, the running time of FACT was faster than that of ASD-POCS but slower than that of SART. However, in the case of 10 views, the running time of FACT was even slightly faster than SART.

required less than 8 percent of the number of epochs and optimization times than the NAF method to reach the same performance (e.g., 1420 epochs and 129 seconds for NAF vs. 103 epochs and 9 seconds for FACT in 3D SSIM).

Fig. 6. shows 3D SSIM and 3D PSNR values according to the change of the optimization epochs between the NAF and FACT methods for 50, 30, and 10 views (see Supplementary Fig. 2 for the graphs of the head and abdomen). In each graph, sample reconstruction results are plotted at initialization, 50 epochs, 100 epochs, and 150 epochs. The optimization graphs of the FACT method consistently are above the graphs of the NAF method, demonstrating faster optimization speed and better reconstruction quality. When we investigated the intermediate optimization results in detail, especially in the graph of 10 views (Fig. 2 c), the NAF method produced unsatisfactory results, failing to reconstruct the delicate structures of the CBCT images compared to the FACT method. At the initial optimization stage, the NAF method provided an empty image (i.e., no information), while the FACT method, which was meta-trained with a few CBCTs, still generated a CT-like image.

4.3. Key Parameter Search for Hash-Encoding Regularization

For the key parameter search of the hash-encoding regularization (see Table 5), the best performance was achieved in

Table 5. Key parameter search results for the hash-encoding regularization

Number of rows (s)	Epochs between resolutions (T)	PSNR	SSIM
2	100	30.29	0.788
2	200	30.25	0.786
3	10	30.38	0.795
3	25	30.38	0.796
3	50	30.36	0.796
3	200	30.26	0.794
4	100	30.29	0.797

terms of 3D PSNR when the incremental number of epochs between resolutions (T) was 25, and the number of rows of the masked hash-encoded feature to be used in an initial epoch (s) was three. While the other combination reported the best performance in 3D SSIM (i.e., case when $T = 100$ and $s = 4$), PSNR was largely degraded (30.38 vs. 30.29), so we decided to choose to the case when $T = 25$ and $s = 3$.

5. Discussion

The proposed FACT reconstruction method utilizes a meta-learning framework to pre-train a neural network using a small set of CBCT scans and regularizes the hash-encoding process to increase the optimization speed and the quality of the CBCT reconstruction with no additional computational cost. We proved that the 3D CBCT images of FACT had better quality than those of the other conventional algorithms. Furthermore, the FACT reconstruction speed was much faster than that of the NAF method, requiring only 14 percent (for 25 views) and 8 percent (for 10 views) of the running time to reach the same image quality. Finally, we validated that the proposed FACT method successfully reconstructed CBCTs of the multiple CT vendors (Siemens, Phillips, and GE) and body parts (chest, head, and abdomen).

As a further investigation, we also observed the intermediate CBCT reconstruction results of the different body parts at the very early stages of optimization (NAF vs. FACT in 5 and 10 epochs) to check the effect of the meta-initialization. Surprisingly, as shown in Supplementary Fig. 3, we found that the results of the FACT method produced coarse anatomical structures of each body part even after only 5 and 10 epochs, while

the results of NAF could not generate any meaningful information. This means that the meta-initialization effectively accelerates the reconstruction process even at the very early optimization stage in each body part.

In the hash-encoding, we found that the finest resolution level significantly affected the quality of the CBCT reconstruction, which motivated us to regularize the hash-encoding process. Supplementary Fig. 4 shows an example where, in the NAF method, the reconstruction results with the finest resolution of 1,024 (i.e., when the number of levels (L) = 8 and coarsest resolution (N_{min}) = 8; see Section 3.3) reveals notable noises compared to the results with the finest resolution of 64 (i.e., L = 4 and N_{min} = 8). This observation is in line with the previous work of (Yang et al., 2023), which explained the failure of the NeRF algorithm as the sudden overfitting of a neural network to high-frequency components such as noises.

This study has some limitations. First, we performed all the experiments using the simulated CBCT data since acquiring a set of raw projection images in the real world was difficult. We believe that the effectiveness of the proposed method can be further validated using real clinical data in the future. Second, although we showed that the proposed algorithm successfully reconstructed the CBCT images for some body parts, there are other body parts for CBCT imaging, such as teeth and jaws. These body parts may be tested when those CT data are publicly available. Third, some advanced encoding strategies have shown impressive results for the NeRF reconstruction, such as Dou et al. (2023). These new techniques may be integrated into the proposed FACT algorithm to enhance the CBCT reconstruction results further.

6. Conclusion

In conclusion, with the meta-initialization and the new regularization process, the FACT method demonstrated faster and more accurate sparse-view CBCT reconstruction than the other conventional (FDK, SAR, and ASD-POCS) and NAF methods. The proposed method is expected to provide benefits in some clinical situations, such as patient positioning and emergency scans (Thilmann et al., 2006; Jacques et al., 2021), assuring the consistent reconstruction quality for the multiple CT machines (Siemens, Phillips, and GE) and anatomies (chest, head, and abdomen).

Acknowledgments

Special thanks to the co-authors for their invaluable guidance and advice throughout this project. We are grateful for the contributions to the improvement of this article's readability and clarity.

Declaration of Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: H.Shin, T.Kim, and D.Shin report a relationship with Radisen Co. Ltd. that includes: employment. J.Lee and S.Cho

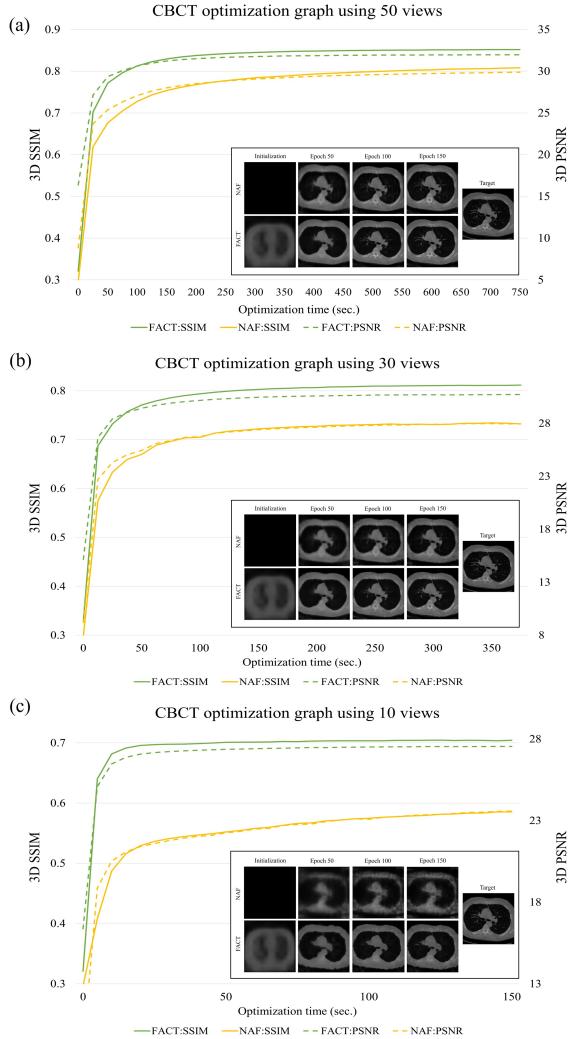


Fig. 6. CBCT optimization graphs (3D SSIM or 3D PSNR vs. optimization epochs) between the NAF and FACT methods for 50, 30, and 10 views. The graphs of the FACT method consistently are above the graphs of the NAF method, demonstrating faster optimization speed and better reconstruction quality. The intermediate reconstruction results of the FACT method are shown below in each graph. At the initial optimization stage, the NAF method produced an empty image (i.e., no information), while the FACT method generated a CT-like image.

reports a relationship with Radisen Co. Ltd. that includes: consulting or advisory. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Material

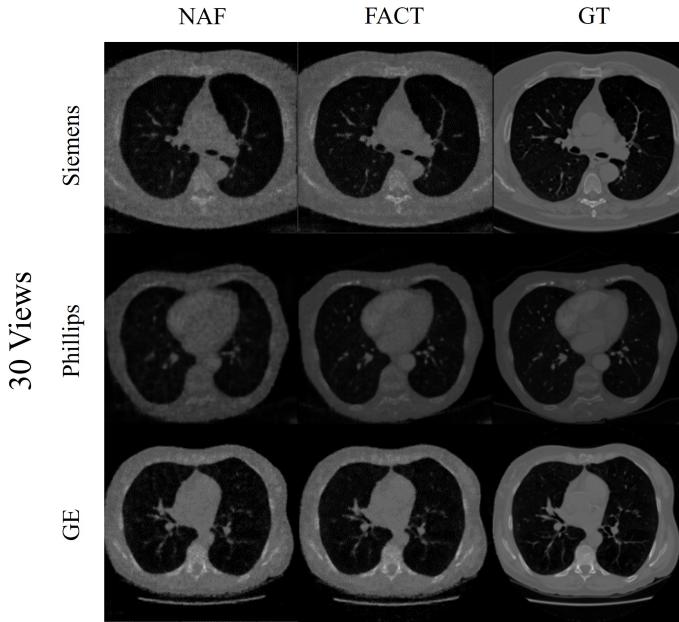
Supplementary material associated with this article can be found, in the online version.

References

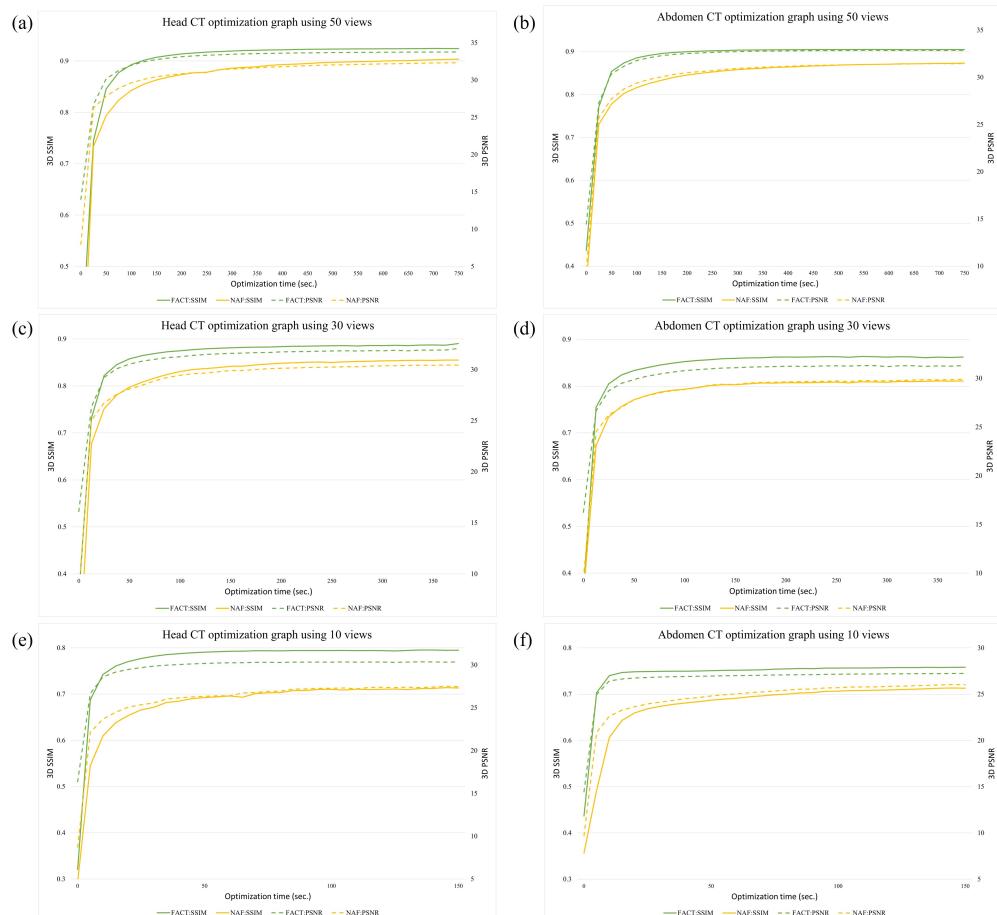
- Andersen, A.H., Kak, A.C., 1984. Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. *Ultrasonic imaging* 6, 81–94.

- Anirudh, R., Kim, H., Thiagarajan, J.J., Mohan, K.A., Champley, K., Bremer, T., 2018. Lose the views: Limited angle ct reconstruction via implicit sinogram completion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6343–6352.
- Armatto III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al., 2011. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics* 38, 915–931.
- Biguri, A., Dosanjh, M., Hancock, S., Soleimani, M., 2016. Tigre: a matlab-gpu toolbox for cbct image reconstruction. *Biomedical Physics & Engineering Express* 2, 055010.
- Carrino, J.A., Al Muhit, A., Zbijewski, W., Thawait, G.K., Stayman, J.W., Packard, N., Senn, R., Yang, D., Foos, D.H., Yorkston, J., et al., 2014. Dedicated cone-beam ct system for extremity imaging. *Radiology* 270, 816–824.
- Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H., 2021. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 14124–14133.
- Dou, Y., Zheng, Z., Jin, Q., Ni, B., 2023. Multiplicative fourier level of detail, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1808–1817.
- Feldkamp, L.A., Davis, L.C., Kress, J.W., 1984. Practical cone-beam algorithm. *Josa a* 1, 612–619.
- Finn, C., Abbeel, P., Levine, S., 2017. Model-agnostic meta-learning for fast adaptation of deep networks, in: International conference on machine learning, PMLR, pp. 1126–1135.
- Flennerhag, S., Rusu, A.A., Pascanu, R., Visin, F., Yin, H., Hadsell, R., 2019. Meta-learning with warped gradient descent. arXiv preprint arXiv:1909.00025 .
- Gao, Y., Bian, Z., Huang, J., Zhang, Y., Niu, S., Feng, Q., Chen, W., Liang, Z., Ma, J., 2014. Low-dose x-ray computed tomography image reconstruction with a combined low-mas and sparse-view protocol. *Optics express* 22, 15190–15210.
- Georgescu, M.I., Ionescu, R.T., Verga, N., 2020. Convolutional neural networks with intermediate loss for 3d super-resolution of ct and mri scans. *IEEE Access* 8, 49112–49124.
- Hohenforst-Schmidt, W., Zarogoulidis, P., Vogl, T., Turner, J.F., Browning, R., Linsmeier, B., Huang, H., Li, Q., Darwiche, K., Freitag, L., et al., 2014. Cone beam computertomography (cbct) in interventional chest medicine—high feasibility for endobronchial realtime navigation. *Journal of Cancer* 5, 231.
- Hu, D., Liu, J., Lv, T., Zhao, Q., Zhang, Y., Quan, G., Feng, J., Chen, Y., Luo, L., 2020. Hybrid-domain neural network processing for sparse-view ct reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* 5, 88–98.
- Jacques, T., Morel, V., Dartus, J., Badr, S., Demondion, X., Cotten, A., 2021. Impact of introducing extremity cone-beam ct in an emergency radiology department: a population-based study. *Orthopaedics & Traumatology: Surgery & Research* 107, 102834.
- Jain, A., Tancik, M., Abbeel, P., 2021. Putting nerf on a diet: Semantically consistent few-shot view synthesis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5885–5894.
- Jin, K.H., McCann, M.T., Froustey, E., Unser, M., 2017. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing* 26, 4509–4522.
- Kan, M.W., Leung, L.H., Wong, W., Lam, N., 2008. Radiation dose from cone beam computed tomography for image-guided radiation therapy. *International Journal of Radiation Oncology* Biology* Physics* 70, 272–279.
- Lechuga, L., Weidlich, G.A., 2016. Cone beam ct vs. fan beam ct: a comparison of image quality and dose delivered between two differing ct imaging modalities. *Cureus* 8.
- Lee, H., Lee, J., Kim, H., Cho, B., Cho, S., 2018. Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* 3, 109–119.
- Lee, Y., Choi, S., 2018. Gradient-based meta-learning with learned layerwise metric and subspace, in: International Conference on Machine Learning, PMLR, pp. 2927–2936.
- Lin, W.A., Liao, H., Peng, C., Sun, X., Zhang, J., Luo, J., Chellappa, R., Zhou, S.K., 2019. Dudonet: Dual domain network for ct metal artifact reduction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10512–10521.
- Liu, Z., Bicer, T., Kettimuthu, R., Gursoy, D., De Carlo, F., Foster, I., 2020. Tomogan: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion. *JOSA A* 37, 422–434.
- Ma, J., Zhang, Y., Gu, S., Zhu, C., Ge, C., Zhang, Y., An, X., Wang, C., Wang, Q., Liu, X., et al., 2021. Abdomenct-1k: Is abdominal organ segmentation a solved problem? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 6695–6714.
- Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* 65, 99–106.
- Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)* 41, 1–15.
- Nichol, A., Achiam, J., Schulman, J., 2018. On first-order meta-learning algorithms. arXiv preprint arXiv:1803.02999 .
- Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S., Geiger, A., Radwan, N., 2022. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5480–5490.
- Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S., 2019. Deepsdf: Learning continuous signed distance functions for shape representation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 165–174.
- Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A., 2019. On the spectral bias of neural networks, in: International Conference on Machine Learning, PMLR, pp. 5301–5310.
- Rajeswaran, A., Finn, C., Kakade, S.M., Levine, S., 2019. Meta-learning with implicit gradients. *Advances in neural information processing systems* 32.
- Ronen, B., Jacobs, D., Kasten, Y., Kritchman, S., 2019. The convergence rate of neural networks for learned functions of different frequencies. *Advances in Neural Information Processing Systems* 32.
- Scarf, W.C., Farman, A.G., 2008. What is cone-beam ct and how does it work? *Dental Clinics of North America* 52, 707–730.
- Scarf, W.C., Farman, A.G., Sukovic, P., et al., 2006. Clinical applications of cone-beam computed tomography in dental practice. *Journal-Canadian Dental Association* 72, 75.
- Sidky, E.Y., Pan, X., 2008. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology* 53, 4777.
- Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., Jacobson, A., McGuire, M., Fidler, S., 2021. Neural geometric level of detail: Real-time rendering with implicit 3d shapes, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11358–11367.
- Tancik, M., Mildenhall, B., Wang, T., Schmidt, D., Srinivasan, P.P., Barron, J.T., Ng, R., 2021. Learned initializations for optimizing coordinate-based neural representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2846–2855.
- Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R., 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems* 33, 7537–7547.
- Tang, C., Zhang, W., Li, Z., Cai, A., Wang, L., Li, L., Liang, N., Yan, B., 2019. Projection super-resolution based on convolutional neural network for computed tomography, in: 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, SPIE, pp. 537–541.
- Thilmann, C., Nill, S., Tücking, T., Höss, A., Hesse, B., Dietrich, L., Bendl, R., Rhein, B., Häring, P., Thieke, C., et al., 2006. Correction of patient positioning errors based on in-line cone beam cts: clinical implementation and first experiences. *Radiation Oncology* 1, 1–9.
- Vallieres, M., Kay-Rivest, E., Perrin, L.J., Liem, X., Furstoss, C., Aerts, H.J., Khauouam, N., Nguyen-Tan, P.F., Wang, C.S., Sultanem, K., et al., 2017. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Scientific reports* 7, 10117.
- Wang, C., Zhang, H., Li, Q., Shang, K., Lyu, Y., Dong, B., Zhou, S.K., 2021. Improving generalizability in limited-angle ct reconstruction with sinogram extrapolation, in: Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24, Springer, pp. 86–96.
- Wu, W., Hu, D., Niu, C., Yu, H., Vardhanabuti, V., Wang, G., 2021. Drone:

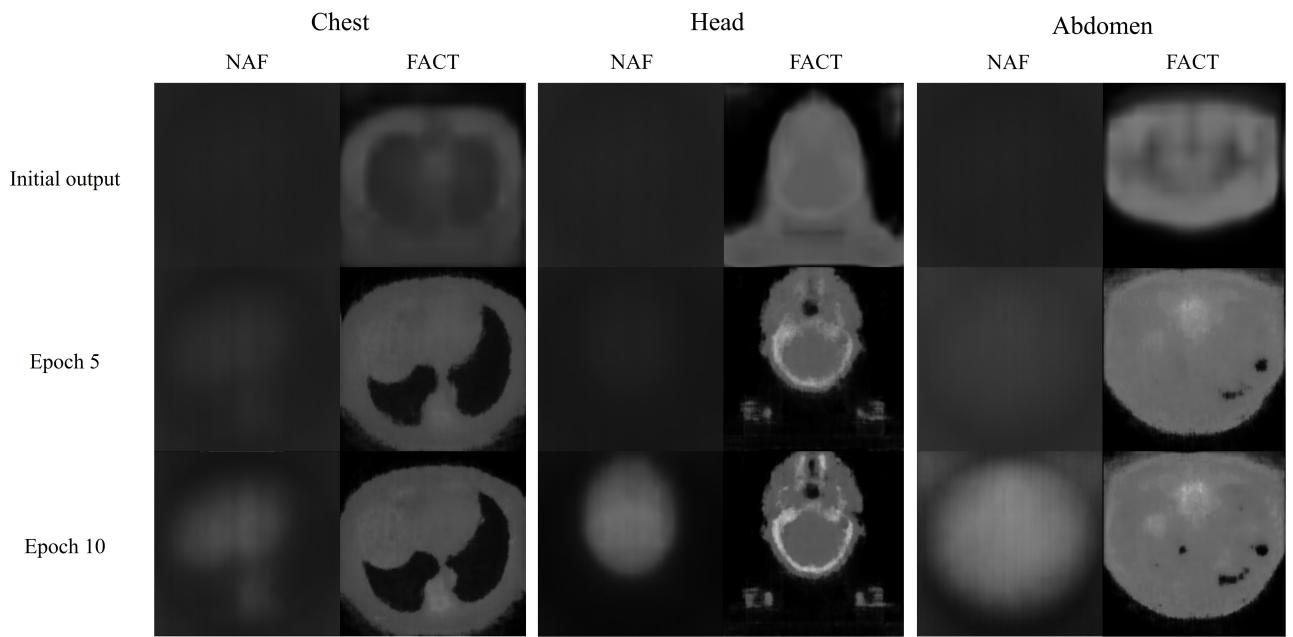
- Dual-domain residual-based optimization network for sparse-view ct reconstruction. *IEEE Transactions on Medical Imaging* 40, 3002–3014.
- Yang, J., Pavone, M., Wang, Y., 2023. Freenerf: Improving few-shot neural rendering with free frequency regularization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8254–8263.
- Yu, A., Ye, V., Tancik, M., Kanazawa, A., 2021. pixelnerf: Neural radiance fields from one or few images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4578–4587.
- Zang, G., Idoughi, R., Li, R., Wonka, P., Heidrich, W., 2021. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1960–1970.
- Zha, R., Zhang, Y., Li, H., 2022. Naf: neural attenuation fields for sparse-view cbct reconstruction, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 442–452.
- Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G., 2018. A sparse-view ct reconstruction method based on combination of densenet and deconvolution. *IEEE transactions on medical imaging* 37, 1407–1417.
- Zhou, B., Chen, X., Zhou, S.K., Duncan, J.S., Liu, C., 2022. Dudodr-net: Dual-domain data consistent recurrent network for simultaneous sparse view and metal artifact reduction in computed tomography. *Medical Image Analysis* 75, 102289.



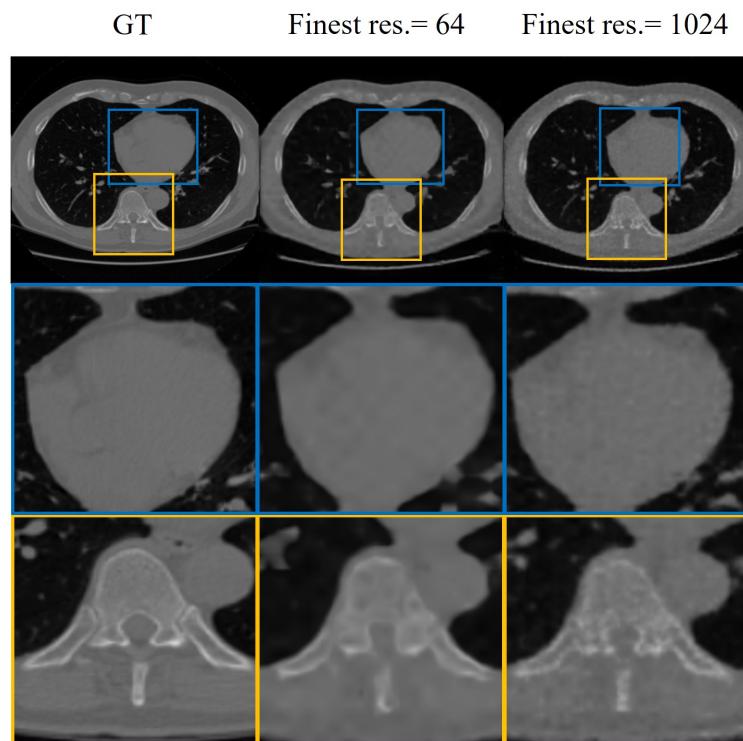
Supplementary Figure 1. Example CBCT reconstruction results of 30 views for the three CT vendors (Siemens, Phillips, and GE).



Supplementary Figure 2. CBCT optimization graphs (3D SSIM or 3D PSNR vs. optimization epochs) between the NAF and FACT methods for the head and abdomen CBCTs.



Supplementary Figure 3. Intermediate CBCT reconstruction results of the different body parts (chest, head, and abdomen) for the NAF and FACT methods at the very early stages of optimization (5 and 10 epochs).



Supplementary Figure 4. Example CBCT reconstruction results of the NAF method when the finest resolution was 1,024 and 64. The results with the finest resolution of 1,024 reveal notable noises compared to the results with the finest resolution of 64.