

Patient-Level Anatomy Meets Scanning-Level Physics: Personalized Federated Low-Dose CT Denoising Empowered by Large Language Model

Ziyuan Yang^{1,5}, Yingyu Chen^{1,5}, Zhiwen Wang², Hongming Shan³, Yang Chen⁴, Yi Zhang^{1,5,*}

¹Sichuan University ²Southwest Petroleum University ³Fudan University

⁴Southeast University ⁵Tianfu Jiangxi Laboratory

{czyuanyang, cyy262511, zwwang1228}@gmail.com, hmshan@fudan.edu.cn

chenyang.list@seu.edu.cn, yzhang@scu.edu.cn

Abstract

*Reducing radiation doses benefits patients, but the resultant low-dose computed tomography (LDCT) images often suffer from clinically unacceptable noise and artifacts. While deep learning (DL) has shown promise in LDCT reconstruction, it requires large-scale data collection from multiple clients, raising privacy concerns. Federated learning (FL) has been introduced to mitigate these privacy concerns; however, current methods are typically tailored to specific scanning protocols, which limits their generalizability and makes them less effective for unseen protocols. To address these issues, we propose **SCAN-PhysFed**, a novel **SC**anning- and **AN**atomy-level personalized **Physics-Driven Federated learning** paradigm for LDCT reconstruction. Since the noise distribution in LDCT data is closely tied to scanning protocols and anatomical structures, we propose a dual-level physics-informed way to address these challenges. Specifically, we incorporate physical and anatomical prompts into our physics-informed hypernetworks to capture scanning- and anatomy-specific information, enabling dual-level physics-driven personalization of imaging features. These prompts are derived from the scanning protocol and the radiology report generated by a medical large language model (MLLM). Subsequently, client-specific decoders project these dual-level personalized imaging features back into the image domain. Besides, to tackle the challenge of unseen data, we introduce a novel protocol vector-quantization strategy (PVQS), which ensures consistent performance across new clients by quantifying unseen scanning codes to the closest match in the scanning codebook. Extensive experimental results demonstrate the superior performance of SCAN-PhysFed on pub-*

*lic datasets*¹.

1. Introduction

Computed tomography (CT) is a key diagnostic tool that non-invasively visualizes anatomical structures within the human body. Despite its clinical benefits, CT scans pose radiation concerns, which potentially lead to genetic, cancerous, and other health diseases. Lowering CT radiation dose—either by decreasing incident photons or limiting sampling views—has emerged as a promising solution to reduce radiation risks and accelerate scanning. However, reconstructed images under these conditions suffer from significant quality degradation, compromising clinical applicability [27].

Recently, deep learning (DL)-based methods have shown great promise for low-dose CT (LDCT) imaging [28]. However, most DL-based LDCT reconstruction methods are condition-specific and overlook privacy concerns. Recently, researchers have attempted to introduce federated learning (FL), a privacy-preserving distributed learning paradigm, for LDCT reconstruction to alleviate privacy concerns. Although these methods aim to enhance generalization, they still face significant challenges arising from data heterogeneity, caused by diverse noise distributions in CT imaging across different protocols and anatomical structures. We believe that the further advancement of FL in LDCT reconstruction urgently requires a new framework capable of leveraging comprehensive physical information from the scanning process, thereby breaking through current bottlenecks to achieve better results.

To tackle this issue, we aim to leverage physical principles of the imaging process. During the projection process, the measurement is influenced by the anatomical structures the X-ray passed through, such as fat, bone, and fluid, each

*Corresponding author.

¹ Our code has been released at <https://github.com/Zi-YuanYang/SCAN-PhysFed>.

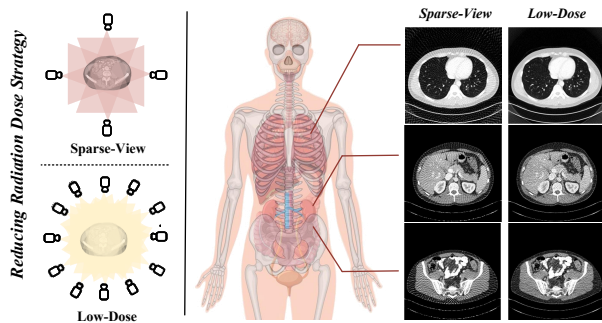


Figure 1. Illustration of different radiation reduction strategies. ‘Sparse-View’ refers to scanning with a reduced number of projection views, while ‘Low-Dose’ represents scanning with a lower number of incident photons.

with distinct attenuation coefficients. Additionally, the system matrix is significantly affected by scanning protocols, including tube current, voltage settings, and scanning angles. These factors collectively impact the quality and characteristics of reconstructed CT images during the backprojection phase, resulting in varying levels of noise and artifacts. An example is provided in Figure 1 to illustrate this concept more clearly.

By leveraging the physics-specific knowledge discussed earlier, we reimagine LDCT reconstruction within the federated learning framework. Our approach adopts a dual-level strategy to alleviate heterogeneity, addressing variations at both the scanning and patient levels. In practice, scanning protocols, which serve as scanner hyperparameters, are readily available. However, generating anatomical information has traditionally been challenging due to patient variability and the complexity of accurately modeling different tissue types. Recent advances in large language models (LLMs) offer new possibilities for automatically generating anatomical information, potentially improving the reconstruction.

Building on this, we propose **SCAN-PhysFed**, a novel **SC**anning- and **AN**atomy-level personalized **PH**ysics-Driven **F**ederated learning paradigm for LDCT reconstruction. Specifically, we address the data heterogeneous using a dual-level strategy. A detailed radiology report is generated using a pretrained LLM. Then, a patient-level anatomy-informed hypernetwork is designed to generate a modulation map from the radiology report, which seamlessly integrates with imaging features and aligns them with each patient’s unique anatomical structures. Simultaneously, scanning features are integrated via a scanning-informed hypernetwork to personalize imaging features based on the protocol. To ensure distinguishability of scanning feature vectors, we impose an orthogonal loss during training to reinforce personalization. This dual-modulation strategy facilitates personalized CT imaging at both patient and scanning levels, guided by the principles of physical imaging.

Additionally, existing personalized FL CT imaging methods do not address the challenge posed by unseen clients with varying scanning protocols. In this paper, we propose a **Protocol Vector Quantization Strategy (PVQS)** to quantize unseen scanning protocol by matching it with the closest existing vectors, thereby selecting the corresponding client-specific decoder. The main contributions of this paper are summarized as follows:

- We introduce LLM into LDCT reconstruction within the FL framework, proposing SCAN-PhysFed. To the best of our knowledge, this is the first work to apply LLM in this field.
- To address the challenge of data heterogeneity, we draw inspiration from physical imaging principles and propose a dual-level approach, achieving personalized CT imaging at both the scanning and patient levels.
- We introduce a novel strategy, PVQS, to ensure our method remains effective for unseen clients with diverse scanning protocols—a critical challenge often overlooked in previous works.

2. Related Works

LDCT Imaging. Traditional reconstruction methods primarily rely on image sparsity [25]. However, these methods are generally time-consuming. Recently, deep learning-based methods have shown faster processing and promising performance, gaining significant attention in this field. For example, Chen *et al.* [4] introduced residual structure into an autoencoder and proposed the encoder-decoder convolutional neural network (RED-CNN) for LDCT denoising. You *et al.* [39] extended RED-CNN by incorporating a generative adversarial network (GAN) for super resolution. Additionally, researchers also explored transformer-based architectures to enhance robustness, as seen in Uformer [31], TransCT [40], and RegFormer [34]. Recently, researchers have proposed diffusion-based CT reconstruction methods [9, 21, 33], which show promising results but incur substantial computational costs. Moreover, the methods mentioned above are tailored to specific scanning protocols and depend on data collection from multiple clients, raising significant privacy concerns.

Federated Learning. In recent years, FL has gained increasing attention as it offers a decentralized approach to DL while preserving data privacy [38]. FedAvg [23], the earliest FL method, introduced a server-client framework to learn a global-shared model by averaging local models from different parties. However, this method suffers from model shift due to serious statistic heterogeneity [11]. To alleviate this concern, recent works have proposed methods to obtain a more robust global-shared model, such as FedProx [16], MOON [15], and FedNova [29]. For serious non-iid data distributions, a single global-shared model is insufficient to address the variability across different datasets. As

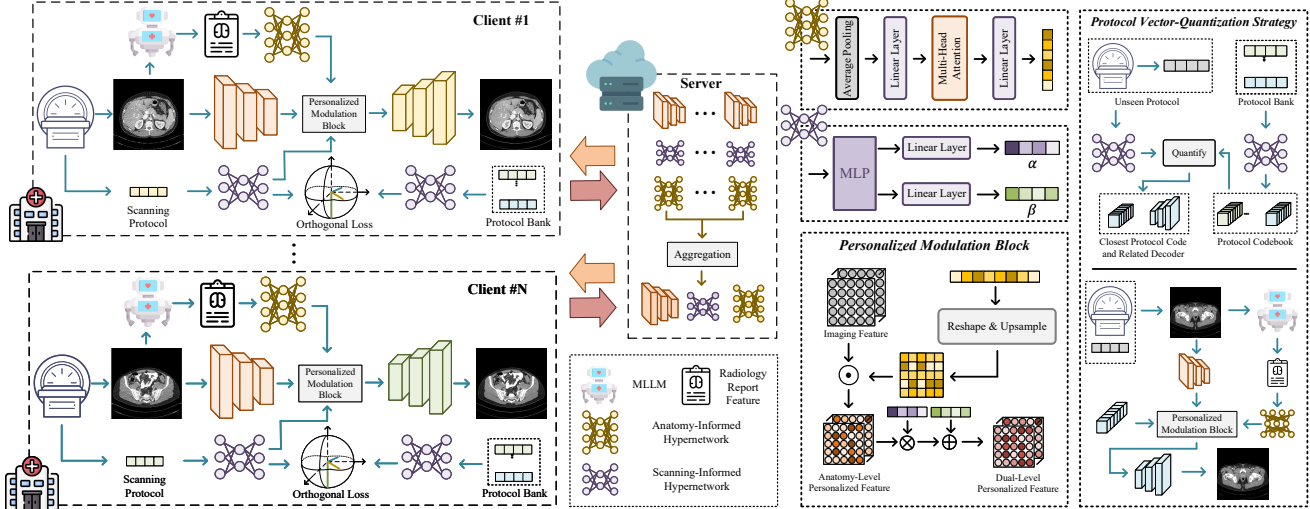


Figure 2. The overall learning paradigm of our proposed SCAN-PhysFed.

a result, several personalized FL methods have been proposed [3, 7, 17, 35].

FL has recently been introduced in healthcare applications due to its privacy-preserving nature [10, 32, 36]. Several studies have focused on its applications in medical imaging [8, 14]. For example, Chen *et al.* [5] proposed a cross-domain learning approach that combines individual-client sinogram learning and cross-client image reconstruction learning. Yang *et al.* [37] proposed a client-specific hypernetwork to modulate the global-shared data. Chen *et al.* [6] utilized separate networks to process high- and low-frequency components of CT data. While these methods have achieved promising performance, they still face significant challenges from data heterogeneity and fail to address the issue of unseen protocols.

Large Language Model. The emergence of large language models (LLMs), such as GPT [1] and Llama [26], has ushered in a new era of artificial intelligence. Recently, large-scale medical LLMs (MLLMs) have been developed specifically for healthcare applications [18, 20, 41]. These models demonstrate significant potential in understanding and reasoning over medical data, excelling in tasks such as visual question answering (VQA) and text generation [24]. For instance, Alkhaldi *et al.* [2] developed Minigtpt-med, which supports multiple tasks, including VQA, medical report generation, and disease identification from medical images. Although MLLMs have been applied to various downstream tasks [13], they have not yet been explored for mitigating domain shift in CT reconstruction within FL settings.

3. Methodology

3.1. Overview

This paper aims to address generalization challenges in FL for LDCT reconstruction, with a focus on alleviating data

heterogeneity among existing clients and maintaining performance for new clients. Figure 2 illustrates the overall learning paradigm of the proposed SCAN-PhysFed framework. Specifically, robust imaging representations are obtained using a collaboratively learned shared encoder across clients. To tackle the heterogeneity issue, two physics-informed hypernetworks are designed to capture scanning- and anatomy-specific information, personalizing the imaging features via a specialized modulation block. Subsequently, client-specific decoders project the personalized imaging features back to the image domain. For the i -th client, the reconstruction process is formulated as follows:

$$\hat{Y} = D_i(P(f_{\mathbf{X}}, H_s(g; \theta_{H_s}), H_a(f_t; \theta_{H_a})); \theta_{D_i}), \quad (1)$$

where $f_{\mathbf{X}}$ denotes the imaging feature of \mathbf{X} extracted from the shared encoder E , and \hat{Y} is the corresponding normal-dose CT image. θ_{H_s} and θ_{H_a} are the scanning- and anatomy-informed hypernetworks, parameterized by θ_s and θ_a , respectively. g and f_t represent the scanning protocol vector and the radiology report feature derived from the MLLM. P denotes the personalized modulation block, and D_i denotes the i -th client decoder, parameterized by θ_{D_i} .

Additionally, to handle unseen data, we propose a novel protocol vector-quantization strategy (PVQS), which ensures robust performance across previously unseen clients by quantizing the unseen scanning code to the closest match in the scanning codebook.

3.2. Personalized Modulation Block

Anatomy-Level Personalization. The patient’s unique anatomy significantly affects noise distribution during the imaging process, making it an essential factor to consider. To address this issue, we leverage MLLM, which is trained on large-scale datasets and demonstrates strong performance in radiology report generation. Specifically,

we adapt miniGPT-Med [2] to generate radiology reports from CT images, using the text encoder’s feature output $f_t \in \mathbb{R}^{1 \times d}$ as a prompt for our anatomy-informed hyper-network, where d denotes the feature dimension.

Given the high dimensionality of f_t , we first apply a pooling operation to reduce its size. The process is formulated as follows:

$$f_{tl} = \text{Linear}(\text{AvgPool}(f_t)), \quad (2)$$

where AvgPool denotes an average pooling operation that reduces the feature dimension by a factor of 1/4. A linear layer is then applied to map the feature to $f_{tl} \in \mathbb{R}^{1 \times h}$, where h denotes the hidden dimension.

Next, we capture long-range dependencies within f_{tl} using the multi-head attention mechanism, followed by a linear layer to extract the anatomy feature f_{an} . This process is formulated as:

$$f_{an} = \text{Linear}(\text{Multi-Head}(f_{tl})). \quad (3)$$

After this stage, f_{an} encapsulates anatomy-specific information tailored to different CT data. Since this feature encodes anatomy-specific details, we use it to spatially personalize the imaging feature $f_{\mathbf{X}}$. The process is formulated as:

$$f_{ana} = f_{\mathbf{X}} \odot \text{Reshape}(f_{an}), \quad (4)$$

where $f_{\mathbf{X}}$ denotes the imaging feature, Reshape represents the reshape operation for dimensional alignment, and “ \odot ” denotes the Hadamard product. f_{ana} is the anatomy-level personalized feature.

Scanning-Level Personalization. In addition to anatomical variations affecting noise distribution, the physical imaging process shows that noise distribution is primarily governed by the scanning protocol. This protocol comprises several key physical parameters in CT scanners, including the number of views, the number of detector bins, pixel length, detector bin length, the distance between the source and rotation center, the distance between the detector and rotation center, and the photon number of incident X-rays.

Then, we utilize g , a vector encoding these parameters, to construct our scanning-level prompt, which captures the latent relationship between the scanning protocol and noise distribution for the imaging network. A detailed description of g is provided in Appendix B.

Given the low dimensionality of the physical parameters, we employ a Multi-Layer Perceptron (MLP) with three linear layers to capture and represent the underlying physical information. Two modulation parameter vectors are then derived to encode scanning-level personalization knowledge. This process can be formulated as follows:

$$\alpha, \beta = H_s(g; \theta_s), \quad (5)$$

where α and β are the modulation parameter vectors, which are used to inject scanning-level information to personalize the imaging feature as:

$$f_{per} = \alpha \otimes f_{ana} + \beta, \quad (6)$$

where f_{per} is the dual-level personalized imaging feature, and \otimes denotes the multiplication operation along the channel dimension. This approach applies the scanning prior to modulate the imaging feature.

Our proposed personalization block is architecture-agnostic, allowing the personalization parameters to be fine-tuned accordingly.

3.3. Protocol Vector-Quantization Strategy

Existing FL-based CT imaging methods have not addressed the challenge of unseen data, resulting in degraded performance in new domains. To overcome this limitation and ensure robust performance for unseen protocols, we introduce PVQS. While patient anatomy remains largely consistent, the primary source of noise variation between unseen and existing clients stems from the low-dose strategy. Therefore, PVQS is designed to enable SCAN-PhysFed to maintain consistent performance across diverse and unseen low-dose protocols.

Specifically, in PVQS, protocol codes for different existing scanning protocols, generated by the MLP in H_s , are stored in a protocol codebook. For an unseen protocol, we first quantize it to the closest code vector in the protocol codebook. The process is defined as follows:

$$i^* = \arg \min_i \left(1 - \frac{\mathbf{c}_{un} \cdot \mathbf{c}_i}{\|\mathbf{c}_{un}\| \|\mathbf{c}_i\|} \right), \quad (7)$$

where \mathbf{c}_{un} and \mathbf{c}_i denote the protocol code of the unseen protocol and that of the i -th client, respectively. i^* indicates the index of the closest code vector.

Instead of using the unseen protocol as input to H_s , we directly replace it with \mathbf{c}_{i^*} to personalize the imaging feature and use the corresponding decoder D_{i^*} to project the personalized features back to the image domain.

Thus, PVQS effectively avoids potential distribution shifts in unseen data and preserves a consistent feature space across domains, ensuring reliable and robust performance for unseen protocols.

3.4. Learning Paradigm

Since the noise distribution varies across protocols, it is challenging to use a single shared imaging network to accommodate all variations. In FL-based LDCT imaging, as assumed in previous works [6, 37], the optimization challenge consists of two main components: imaging feature extraction and personalized projection. Specifically, for feature extraction, we propose aggregating client-side encoders to capture robust, comprehensive imaging features.

Table 1. Quantitative Results (PSNR (dB) and SSIM (%)) for CNN-based LDCT Imaging Method.

	Client #1		Client #2		Client #3		Client #4		Client #5		Client #6		Client #7		Client #8		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Generic FL Methods																		
FedAvg [23]	29.99	82.20	32.82	78.23	34.68	86.99	32.30	80.63	36.87	92.06	37.36	89.02	32.73	80.66	35.46	87.65	34.03	84.70
FedProx [16]	29.94	83.12	32.69	75.46	36.95	91.58	32.32	80.91	38.95	95.34	37.58	88.13	32.94	81.62	35.86	88.65	34.65	85.60
FedNova [29]	29.95	82.29	32.57	80.57	37.77	92.16	32.09	78.98	34.42	88.69	36.05	89.57	32.81	81.41	35.89	88.71	33.94	85.30
MOON [15]	32.50	75.61	35.55	85.41	35.72	80.30	35.69	82.10	35.38	78.47	36.31	84.35	35.70	81.17	35.62	80.40	35.31	80.98
FedDG [19]	33.37	81.40	34.01	82.39	33.56	76.37	34.43	81.49	34.06	76.70	33.97	80.17	34.75	80.87	34.25	79.23	34.05	79.83
FedKD [32]	29.74	83.11	32.17	71.65	36.65	90.93	32.10	79.19	40.58	96.71	37.07	82.55	32.88	81.30	36.28	89.08	34.68	84.31
Personalized FL Methods																		
FedPer [3]	33.23	88.20	35.06	84.89	38.79	92.70	35.58	86.12	41.76	96.97	36.85	89.96	35.69	86.30	38.48	91.69	36.93	89.60
FedBN [17]	32.55	84.59	34.76	84.15	38.18	90.30	34.77	83.30	41.24	96.11	38.66	91.19	34.74	83.35	37.89	90.35	36.60	87.92
FedMRI [8]	33.28	83.43	34.51	85.81	38.62	92.47	35.86	87.21	41.49	96.27	38.22	91.32	35.87	86.34	38.09	91.94	36.99	89.35
HyperFed [37]	33.74	88.57	34.83	85.05	38.87	92.78	35.44	85.40	42.94	97.48	39.15	92.96	34.99	85.19	38.12	91.29	37.26	89.84
FedFDD [6]	36.18	93.67	38.43	94.19	39.72	96.47	38.51	95.30	45.56	98.73	42.72	97.61	40.74	96.01	38.54	94.39	40.05	95.79
SCAN-PhysFed	34.49	96.40	39.21	96.09	40.45	97.93	40.51	97.61	47.81	99.14	44.08	98.51	42.54	97.86	40.02	97.37	41.14	97.62

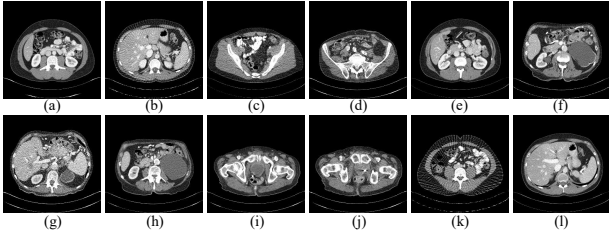


Figure 3. Simulated examples under different protocols. (a)-(h) show examples from training clients with known protocols, while (i)-(l) present examples from unseen protocols.

To fully leverage the diversity of client data, we globally share the hypernetworks in SCAN-PhysFed. However, projecting imaging features from different domains into a consistent, noise-free representation is challenging with a single shared decoder. Therefore, we propose protocol-specific decoders to effectively adapt to variations in scanning protocols across clients.

In PVQS, the unseen protocol code is quantized based on the nearest entry in the protocol codebook. Therefore, it is essential to ensure the protocol code is discriminative and informative for personalization. To achieve this, we introduce an orthogonal loss $\mathcal{L}_{\text{orth}}$ for the i -th client, defined as follows:

$$\mathcal{L}_{\text{orth}} = \sum_{j=1}^K |\mathbf{c}_i \cdot \mathbf{c}_j|^2 \quad \text{s.t.} \quad i \neq j, \quad (8)$$

where K is the number of clients (protocols), and j denotes the client index.

The Mean Squared Error (MSE) is used as the imaging loss, and the total loss is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \tau \mathcal{L}_{\text{orth}}, \quad (9)$$

where \mathbf{Y} represents the reconstructed image, and τ denotes the temperature weight of $\mathcal{L}_{\text{orth}}$.

To facilitate understanding of the implementation details, we provide the pseudocode of our method in Appendix A.

4. Experiments

Experimental Setting. The proposed method was implemented using PyTorch and optimized by Adam [12] at a learning rate of 0.001. The number of communication rounds was set to 200 and the batch size was 20. The experiments were conducted on a system equipped with an AMD Ryzen 7 5800X CPU and a single NVIDIA GTX 3080 Ti GPU.

Dataset. The ‘‘2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge’’ dataset [22], which contains 5,936 full-dose CT images from 10 patients, was used to evaluate the proposed method. In this study, 8 patients were used for training, while the remaining 2 were reserved for testing. To simulate a realistic scenario, we ensured that the patients for training did not overlap across different clients. The two patients for testing were simulated using different protocols to form the testing set.

Simulated examples are shown in Figure 3, where noticeable noise heterogeneity among clients can be observed, primarily due to differences in anatomy and scanning protocols. Detailed dataset preparation steps and scanning protocols are provided in Appendix B.

Baselines. We compared our SCAN-PhysFed approach with both generic and personalized FL methods. For generic FL methods, we include FedAvg [23], FedProx [16], FedNova [29], MOON [15], FedDG [19], and FedKD [32]. For personalized FL methods, we evaluated FedPer [3], FedBN [17], FedMRI [8], HyperFed [37], and FedFDD [6]. To ensure fairness, we kept the training settings consistent across all methods.

Evaluation Metrics. We use the peak signal-to-noise ratio (PSNR) to evaluate pixel-wise accuracy and the structural similarity (SSIM) to assess perceived visual quality [30]. For both metrics, higher values indicate better performance.

4.1. Comparison with other methods

We evaluate our method alongside other FL methods. To ensure fairness, we use the same imaging network, RED-

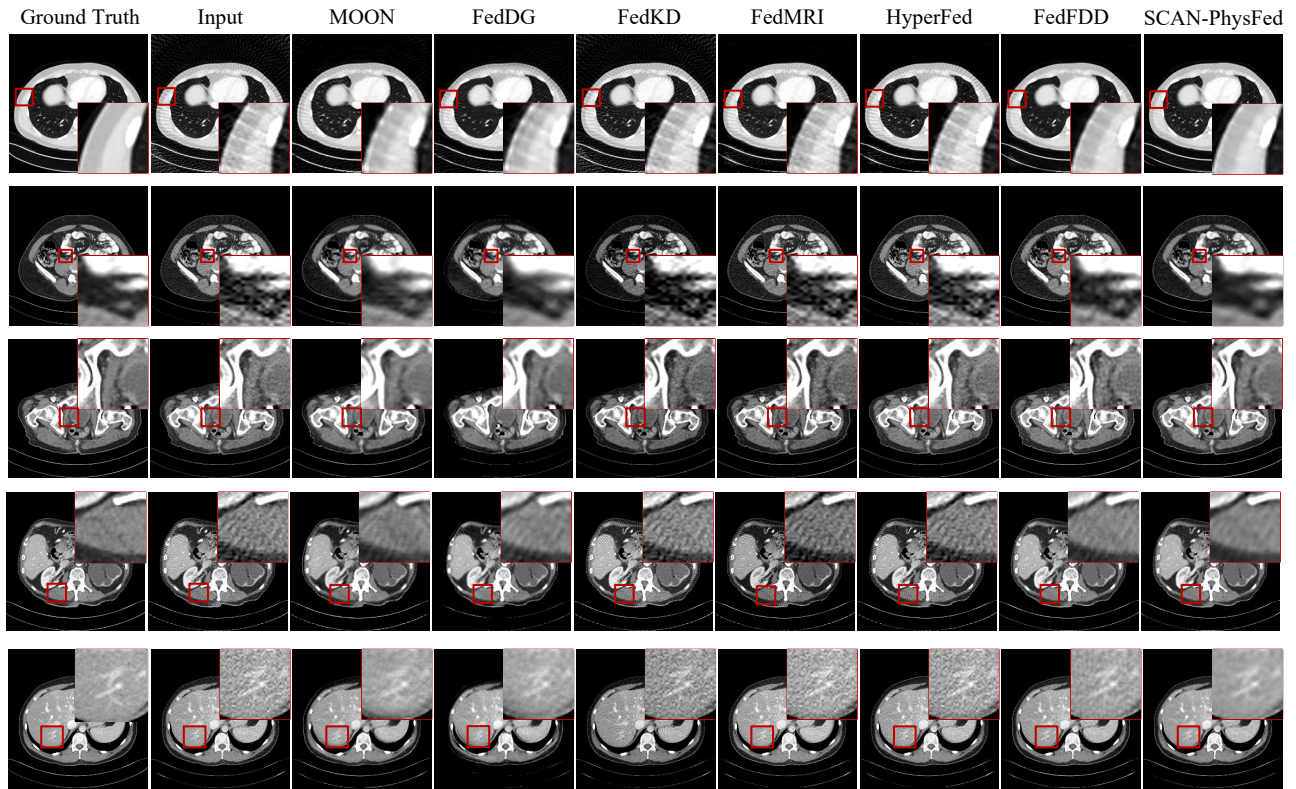


Figure 4. Qualitative results of six selected comparison methods and our method across different clients using the classical convolutional-based LDCT imaging network. Rows one to five correspond to Clients #2, #3, #5, #6, and #7, respectively. The display window for the first row is $[-1024, 200]$ HU, while for the remaining rows, it is $[-160, 240]$ HU.

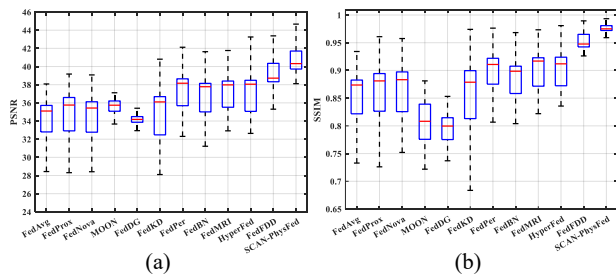


Figure 5. Boxplots of the average results across all clients. (a) and (b) represent PSNR and SSIM, respectively.

CNN [4], a state-of-the-art model in CT imaging. Quantitative results are provided in Table 1. Our method achieves the best performance across most clients, outperforming other generic and personalized FL methods in terms of average PSNR and SSIM. Furthermore, unlike other personalized methods that show instability across clients, such as HyperFed in Client #2 / #7 and FedFDD in Client #4, our method maintains consistent performance across varying protocols. We also observe that sparse-view protocols (Clients #2, #4, and #6) pose challenges for both generic and personalized FL methods, resulting in significant performance gaps compared to other protocols. How-

ever, by incorporating dual-level physics-informed prior, our method maintains robust performance.

Representative slices produced by different methods are shown in Figure 4. We compared the top three methods from both generic and personalized FL approaches. Representative clients were selected based on different radiation reduction strategies: sparse-view (Clients #2 and #6) and low-dose (Clients #3, #5, and #7) reconstructions. It can be observed that other methods exhibit noticeable noise or artifacts in sparse-view protocols, and fail to preserve details in low-dose protocols. In contrast, our method effectively reduces noise and artifacts in sparse-view protocols while preserving fine details in low-dose protocols. SCAN-PhysFed achieves strong performance across slices from different anatomical regions, benefiting from the introduction of anatomy priors. Both quantitative and qualitative results demonstrate the effectiveness of incorporating physics-informed priors.

Additionally, we present boxplots of the results across all clients for different methods in Figure 5 to validate their stability. It can be seen that SCAN-PhysFed outperforms all other methods in both PSNR and SSIM scores. In both plots, SCAN-PhysFed demonstrates a higher median and

Table 2. Quantitative Results of PSNR and SSIM for Transformer-based LDCT Imaging Methods.

	Client #1		Client #2		Client #3		Client #4		Client #5		Client #6		Client #7		Client #8		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Generic FL Methods																		
FedAvg [23]	34.31	87.88	38.06	90.54	43.51	97.56	41.05	95.43	45.24	98.47	43.15	97.09	41.45	95.95	42.88	96.16	41.21	94.89
FedProx [16]	34.32	87.01	37.74	88.89	43.73	97.95	41.12	93.92	45.39	98.61	42.99	97.08	41.38	94.31	42.10	94.55	41.10	94.04
FedNova [29]	29.55	78.44	31.38	68.13	37.07	82.47	31.79	73.45	40.61	88.80	35.46	79.77	32.38	73.66	35.77	82.22	34.25	78.37
MOON [15]	34.28	93.20	38.09	91.17	43.64	97.77	41.09	95.66	45.25	98.43	42.82	96.91	41.42	95.97	42.20	96.18	41.10	95.66
FedKD [32]	35.92	89.93	38.70	91.21	44.26	98.13	42.30	95.99	45.93	98.75	43.57	97.41	42.54	96.22	43.51	96.23	42.09	95.48
Personalized FL Methods																		
FedPer [3]	35.37	93.18	39.03	93.55	42.75	97.13	40.88	96.20	46.55	98.83	42.87	96.68	41.87	96.70	40.79	94.78	41.26	95.82
FedMRI [8]	34.60	93.67	38.04	94.55	42.94	97.44	39.34	95.78	46.79	98.91	42.82	97.30	40.81	95.32	40.85	95.70	40.78	96.01
HyperFed [37]	34.60	87.59	38.62	91.08	44.13	97.61	42.25	96.24	45.92	98.59	43.71	97.37	42.44	96.66	43.56	96.63	41.90	95.22
FedFDD [6]	37.35	96.20	39.67	95.18	44.31	98.26	39.00	95.21	47.89	99.20	43.81	98.06	42.17	97.40	40.50	95.49	41.84	96.89
SCAN-PhysFed	38.55	94.06	41.74	96.20	45.52	98.52	42.48	96.55	49.50	99.39	45.32	97.93	43.35	96.73	42.71	96.79	43.65	97.03

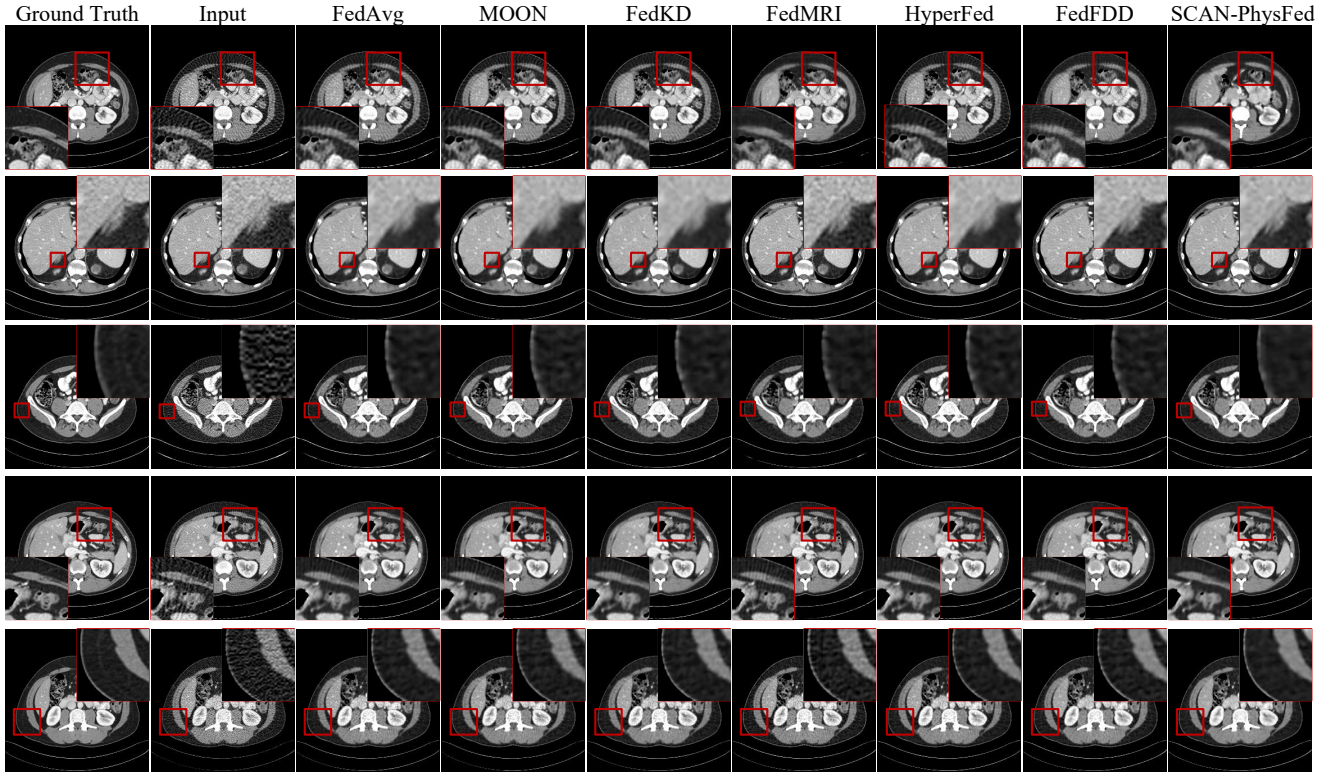


Figure 6. Qualitative results of the compared methods across different clients based on a transformer-based imaging network. Rows one to five correspond to Clients #2, #3, #5, #6, and #7, respectively. The display window is set to [-160, 240] HU.

a narrower interquartile range, indicating strong stability in restoring LDCT data with varying noise distributions across different clients. Individual boxplots for each client are provided in Appendix C.

4.2. Generalization Evaluation

4.2.1 Backbone

In addition to using RED-CNN as the imaging network in previous experiments, we evaluate our method and other FL methods with an alternative imaging network, Uformer [31], a transformer-based model for LDCT imaging. This comparison further validates the generalization capability of different FL methods.

The qualitative and quantitative results are shown in Figure 6 and Table 2, respectively. Consistent with previous experiments, our method achieves the best performance compared to other methods across most clients, with significantly higher average PSNR and SSIM values. This demonstrates the strong generalization capability of our method when integrated with different imaging networks. Among competing methods, FedFDD serves as a strong baseline. While FedFDD’s SSIM is slightly higher than ours in some clients, our method achieves a significantly better PSNR, demonstrating superior detail preservation. Additionally, unlike FedFDD, which requires two imaging networks, our method achieves high performance with-

Table 3. The Quantitative Results of PSNR and SSIM for the Unseen Clients.

	Unseen Client #1		Unseen Client #2		Unseen Client #3		Unseen Client #4		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Generic FL Methods										
FedAvg [23]	35.57	88.68	34.00	85.26	32.52	76.71	36.86	91.44	32.89	81.35
FedProx [16]	36.69	91.16	34.30	86.39	32.58	72.13	39.13	94.87	34.74	85.52
FedNova [29]	35.45	88.83	34.23	86.29	32.39	69.20	35.40	89.99	34.38	83.58
MOON [15]	35.21	78.19	35.30	79.81	34.55	83.07	35.56	78.94	35.16	80.00
FedDG [19]	34.86	79.54	34.79	80.91	33.15	80.31	34.49	77.42	34.32	79.55
FedKD [32]	36.82	91.60	34.27	86.37	32.49	70.54	40.73	95.90	36.08	86.10
Personalized FL Methods										
FedPer [3]	36.86	88.58	35.46	87.02	31.96	71.64	38.56	90.90	35.71	84.53
FedBN [17]	37.84	89.95	36.16	88.46	33.67	81.63	40.21	94.89	36.97	88.73
FedMRI [8]	31.12	82.42	31.95	79.95	31.72	71.96	32.04	81.63	31.71	78.99
HyperFed [37]	36.18	89.60	33.47	82.49	31.94	65.52	41.70	94.85	35.90	83.12
FedFDD [6]	37.80	92.33	37.06	91.77	33.71	80.26	40.53	95.12	37.27	89.87
SCAN-PhysFed	40.77	97.69	38.77	95.75	34.38	81.80	40.27	96.92	38.55	93.04

Table 4. Ablation on different components in SCAN-PhysFed.

Scanning	Component			Average	
	Anatomy	Paradigm	Loss	PSNR	SSIM
\times	\times	Generic	\mathcal{L}_{MSE}	34.03	84.70
\checkmark	\times	Personalized	\mathcal{L}_{MSE}	36.82	88.78
\times	\checkmark	Personalized	\mathcal{L}_{MSE}	38.04	94.14
\checkmark	\checkmark	Personalized	\mathcal{L}_{MSE}	40.67	97.27
\checkmark	\checkmark	Personalized	$\mathcal{L}_{MSE} \& \mathcal{L}_{orth}$	41.14	97.62

out relying on an additional network. Similar to the previous experiment, SCAN-PhysFed shows better qualitative performance and lower variation across different protocols, indicating more consistent results irrespective of scanning protocols or imaging networks.

4.2.2 Unseen Protocol

Existing FL methods in medical imaging struggle to maintain performance when encountering unseen noise distributions from new clients. Here, we evaluate the effectiveness of the proposed PVQS across four unseen protocols. Detailed protocol settings can be found in Appendix B. For generic FL methods, we directly evaluate the shared model on unseen clients. However, personalized methods mostly ignore how to operate in the unseen domain, so we average the performance of all local models to establish a vanilla baseline in this paper.

The results of different methods on unseen clients are presented in Table 3. We observe that the performance of generic FL methods shows only minor degradation on unseen clients, whereas personalized FL methods experience significant declines. This is because local optimization in personalized FL methods focuses heavily on each client’s specific data distribution, underfitting other potential distributions and leading to greater performance drops on unseen clients. In PVQS, we prevent the network from directly mapping the unseen protocol to modulate the imaging network, as this could cause network collapse if the unseen protocol significantly differs from known ones. Instead, we quantize the unseen protocol using a protocol codebook, al-

lowing SCAN-PhysFed to leverage prior knowledge for better adaptation to unseen distributions.

4.3. Ablation Study

We evaluate the effectiveness of various components in SCAN-PhysFed, and the results are presented in Table 4. Compared to the baseline results, which exclude scanning and anatomy information, introducing either significantly improves performance, highlighting the importance of integrating physical prior in LDCT imaging. This approach substantially reduces learning complexity and improves the overall model performance. Moreover, introducing both types of physical information simultaneously is complementary rather than redundant, leading to further performance gains. Building on this, we use \mathcal{L}_{orth} to enhance the discriminative ability of the scanning code, thereby personalizing imaging features and effectively mitigating data heterogeneity. In the above studies, “Personalized” indicates that the decoder is tailored to each client. An ablation study on the personalized components is provided in Appendix D.

5. Conclusion

To preserve privacy and address heterogeneity in CT imaging, we propose a dual-level, physics-informed approach that utilizes both scanning- and anatomy-level prompts to enable personalized CT imaging with the assistance of an MLLM. This approach facilitates both patient- and scanning-level personalization, with extensive results demonstrating that incorporating physical prior significantly improves performance and effectively mitigates heterogeneity. Additionally, we propose a novel strategy, PVQS, to quantize protocol codes and maintain robust performance on unseen protocols. PVQS leverages known protocol codes to prevent latent collapse when encountering significantly different unseen protocols. In future work, we aim to develop a general FL framework for medical imaging that can support both CT and MRI denoising, which we believe holds promising potential.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62271335, in part by the Sichuan Science and Technology Program under Grant 2025ZNSFSC0470, and in part by Sichuan University “From 0 to 1” Innovative Research Program under Grant 2022SCUH0016.

References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. [3](#)
- [2] Asma Alkhalidi, Raneem Alnajim, Layan Alabdullatef, Rawan Alyahya, Jun Chen, Deyao Zhu, Ahmed Alsinan, and Mohamed Elhoseiny. Minigpt-med: Large language model as a general interface for radiology diagnosis. *arXiv preprint arXiv:2407.04106*, 2024. [3, 4](#)
- [3] Manoj Ghuhana Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, et al. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019. [3, 5, 7, 8](#)
- [4] Hu Chen, Yi Zhang, Mannudeep K Kalra, et al. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Trans. Med. Imaging*, 36(12):2524–2535, 2017. [2, 6](#)
- [5] Shixuan Chen, Boxuan Cao, Yinda Du, Yaoduo Zhang, Ji He, Zhaoying Bian, Dong Zeng, and Jianhua Ma. Federated condition generalization on low-dose ct reconstruction via cross-domain learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 47–56. Springer, 2023. [3](#)
- [6] Xuhang Chen, Zeju Li, Zikun Xu, Cheng Ouyang, Chen Qin, et al. Fedfdd: Federated learning with frequency domain decomposition for low-dose ct denoising. In *Medical Imaging with Deep Learning*, 2024. [3, 4, 5, 7, 8](#)
- [7] Dongshang Deng, Xuanguo Wu, Tao Zhang, Xiangyun Tang, Hongyang Du, Jiawen Kang, Jiqiang Liu, and Dusit Niyato. Fedasa: A personalized federated learning with adaptive model aggregation for heterogeneous mobile edge computing. *IEEE Transactions on Mobile Computing*, 2024. [3](#)
- [8] Chun-Mei Feng, Yunlu Yan, Shanshan Wang, Yong Xu, Ling Shao, and Huazhu Fu. Specificity-preserving federated learning for mr image reconstruction. *IEEE Transactions on Medical Imaging*, 42(7):2010–2021, 2022. [3, 5, 7, 8](#)
- [9] Qi Gao, Zilong Li, Junping Zhang, Yi Zhang, and Hongming Shan. Corediff: Contextual error-modulated generalized diffusion model for low-dose ct denoising and generalization. *IEEE Transactions on Medical Imaging*, 2023. [2](#)
- [10] Hao Guan, Pew-Thian Yap, Andrea Bozoki, and Mingxia Liu. Federated learning for medical image analysis: A survey. *Pattern Recognition*, page 110424, 2024. [3](#)
- [11] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020. [2](#)
- [12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [5](#)
- [13] Gurucharan Marthi Krishna Kumar, Aman Chadha, Janine Mendola, and Amir Shmuel. Medvisionllama: Leveraging pre-trained large language model layers to enhance medical image segmentation. *arXiv preprint arXiv:2410.02458*, 2024. [3](#)
- [14] Danyang Li, Hao Wang, Jingyi Liao, Mingqiang Meng, Cuidie Zeng, Yuting Wang, Dong Zeng, and Jianhua Ma. Semi-centralized federated learning network for low-dose ct imaging. In *Medical Imaging 2023: Physics of Medical Imaging*, pages 1024–1028. SPIE, 2023. [3](#)
- [15] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 10713–10722, 2021. [2, 5, 7, 8](#)
- [16] Tian Li, Anit Kumar Sahu, Manzil Zaheer, et al. Federated optimization in heterogeneous networks. In *Proc. Mach. Learn. Syst. (MLSys)*, pages 429–450, 2020. [2, 5, 7, 8](#)
- [17] Xiaoxiao Li, Meirui JIANG, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. In *International Conference on Learning Representations*, 2021. [3, 5, 8](#)
- [18] Fenglin Liu, Tingting Zhu, Xian Wu, Bang Yang, Chenyu You, Chenyang Wang, Lei Lu, Zhangdaihong Liu, Yefeng Zheng, Xu Sun, et al. A medical multimodal large language model for future pandemics. *NPJ Digital Medicine*, 6(1): 226, 2023. [3](#)
- [19] Quande Liu, Cheng Chen, Jing Qin, Qi Dou, and Pheng-Ann Heng. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1013–1023, 2021. [5, 8](#)
- [20] Qidong Liu, Xian Wu, Xiangyu Zhao, Yuanshao Zhu, Derong Xu, Feng Tian, and Yefeng Zheng. When moe meets llms: Parameter efficient fine-tuning for multi-task medical applications. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1104–1114, 2024. [3](#)
- [21] Zexin Lu, Qi Gao, Tao Wang, Ziyuan Yang, Zhiwen Wang, Hui Yu, Hu Chen, Jiliu Zhou, Hongming Shan, and Yi Zhang. Pridediff: Physics-regularized generalized diffusion model for ct reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2024. [2](#)
- [22] C McCollough. Tu-fg-207a-04: overview of the low dose ct grand challenge. *Med. Phys.*, 43(6):3759–3760, 2016. [5](#)
- [23] Brendan McMahan, Eider Moore, Daniel Ramage, et al. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282. PMLR, 2017. [2, 5, 7, 8](#)
- [24] Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M Krumholz, Jure Leskovec, Eric J Topol, and Pranav Rajpurkar. Foundation models for generalist medi-

- cal artificial intelligence. *Nature*, 616(7956):259–265, 2023. 3
- [25] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1): 86–109, 2019. 2
- [26] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023. 3
- [27] Ge Wang, Hengyong Yu, and Bruno De Man. An outlook on x-ray ct research and development. *Medical physics*, 35(3): 1051–1064, 2008. 1
- [28] Ge Wang, Jong Chul Ye, and Bruno De Man. Deep learning for tomographic image reconstruction. *Nature machine intelligence*, 2(12):737–748, 2020. 1
- [29] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*, 33:7611–7623, 2020. 2, 5, 7, 8
- [30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5
- [31] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022. 2, 7
- [32] Chuhan Wu, Fangzhao Wu, Lingjuan Lyu, Yongfeng Huang, and Xing Xie. Communication-efficient federated learning via knowledge distillation. *Nature Communications*, 13(1): 2032, 2022. 3, 5, 7, 8
- [33] Wenjun Xia, Qing Lyu, and Ge Wang. Low-dose ct using denoising diffusion probabilistic model for 20x speedup. *arXiv preprint arXiv:2209.15136*, 2022. 2
- [34] Wenjun Xia, Ziyuan Yang, Zexin Lu, Zhongxian Wang, and Yi Zhang. Regformer: A local–nonlocal regularization-based model for sparse-view ct reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 8(2): 184–194, 2023. 2
- [35] Xiyuan Yang, Wenke Huang, and Mang Ye. Fedas: Bridging inconsistency in personalized federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11986–11995, 2024. 3
- [36] Ziyuan Yang, Yingyu Chen, Huijie Huangfu, Maosong Ran, Hui Wang, Xiaoxiao Li, and Yi Zhang. Dynamic corrected split federated learning with homomorphic encryption for u-shaped medical image networks. *IEEE Journal of Biomedical and Health Informatics*, 27(12):5946–5957, 2023. 3
- [37] Ziyuan Yang, Wenjun Xia, Zexin Lu, Yingyu Chen, Xiaoxiao Li, and Yi Zhang. Hypernetwork-based physics-driven personalized federated learning for ct imaging. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 3, 4, 5, 7, 8
- [38] Mang Ye, Xiuwen Fang, Bo Du, Pong C Yuen, and Dacheng Tao. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, 56(3):1–44, 2023. 2
- [39] Chenyu You, Guang Li, Yi Zhang, et al. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE Trans. Med. Imaging*, 39(1): 188–203, 2019. 2
- [40] Zhicheng Zhang, Lequan Yu, Xiaokun Liang, Wei Zhao, and Lei Xing. Transct: dual-path transformer for low dose computed tomography. In *Medical Image Computing and Computer Assisted Intervention*, pages 55–64. Springer, 2021. 2
- [41] Lanyun Zhu, Tianrun Chen, Deyi Ji, Jieping Ye, and Jun Liu. Llaf: When large language models meet few-shot segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3065–3075, 2024. 3