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

Supplementary Material

A. Implementation

Different from previous works [2, 3], which only takes the image as the input while ignoring the information contained in scanning. This paradigm can be formulated as:

$$\hat{\mathbf{Y}} = D(E(\mathbf{X}; \theta_E); \theta_D), \tag{A.1}$$

where E and D denote the encoder and decoder, respectively.

Unlike previous approaches, our method simultaneously considers multiple physical factors influencing noise in the image domain, such as scanning protocols and anatomical structures. Different anatomical structures have varying attenuation coefficients, which result in diverse noise distributions [7], such as fat, blood, and bone.

To leverage anatomical information, we utilize the generation capabilities of the MLLM, miniGPT-Med [1], pretrained on multiple large-scale datasets. This method can effectively generate radiology reports for CT data. Specifically, we use the prompt, "Please provide a radiology report of the CT slice." The feature extracted from the generated report by miniGPT-Med's text encoder is then used to personalize imaging features at the anatomy level.

To help readers understand the entire process of the proposed SCAN-PhysFed, we provide a pseudocode-style overview of our method in Algorithm 1, where $\mathrm{PMB}(\cdot)$ denotes our proposed personalized modulation block.

Algorithm 1 Main steps of SCAN-PhysFed.

Input: $\mathcal{D} \triangleq U_{k\in K}\mathcal{D}^k$, data from K institutions; $\mathcal{G} \triangleq \{G_1, ..., G_K\}$; $C \triangleq \{c1, ..., c_K\}$, the protocol set; T, the number of local training epoch; $Promp$, the prompt of MLLM, $\mathcal{M}$.

Output: The global shared parameters θ_E, θ_{H_s}, θ_{H_a} for shared encoder E, scanning-informed and anatomy-informed hypernetworks H_s and H_a, respectively. Additionally, client-specified decoders $\{D_1, ..., D_K\}$ for each of K clients.

```
Server Executes:
Initialize $\theta_E, \theta_{H_s}, \theta_{H_a}$ and $\theta_D$ and deliver them to each client.
for $t = 1, 2, ..., N$ do
    for $k = 1, 2, ..., K$ in parallel do
        send $\theta_E^t$, $\theta_{H_s}^t$, $\theta_{H_a}^t$ to $k$-th institution
        $\theta_E^{k,t}, \theta_{H_s}^{k,t}, \theta_{H_a}^{k,t} \leftarrow$ Local Train($\theta_E^t, \theta_{H_s}^t, \theta_{H_a}^t$)
    end for
    $\theta_E^{t+1}, \theta_{H_s}^{t+1}, \theta_{H_a}^{t+1} \leftarrow \sum_{k=1}^{K} \frac{|\mathcal{D}^k|}{|\mathcal{D}|}(\theta_E^{k,t}, \theta_{H_s}^{k,t}, \theta_{H_a}^{k,t})$
end for
Local Train:
$\theta_E^{t,1}, \theta_{H_s}^{t,1}, \theta_{H_a}^{t,1} \leftarrow \theta_E^t, \theta_{H_s}^t, \theta_{H_a}^t$
for $e = 1, 2, ..., E$ do
    for $(\mathbf{X}, \hat{\mathbf{Y}})$ in $\mathcal{D}^k$ do
        $f_t \leftarrow E(\mathbf{X})$
        $\alpha, \beta, c_k \leftarrow H_s(g_k)$
        $f_{an} \leftarrow H_a(\mathcal{M}(\mathbf{X}, Promp))$
        $f_{per} \leftarrow \mathrm{PMB}(f_{\mathbf{X}}, f_{an}, \alpha, \beta)$
        $\mathbf{Y} \leftarrow D_k(f_{per})$
        $\mathcal{L}_{total} \leftarrow \mathcal{L}_{\text{MSE}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \tau\mathcal{L}_{\text{orth}}(c_k, C)$
        $\theta_E^{t,e}, \theta_{H_s}^{t,e}, \theta_{H_a}^{t,e} \leftarrow \mathrm{BackProjection}$
    end for
end for
return $\theta_E^{t,E}, \theta_{H_s}^{t,E}, \theta_{H_a}^{t,E}$ to the server
```

B. Detailed Experimental Setup

In CT imaging, the noise distribution is closely related to the scanning protocol, which primarily consists of seven parameters [6]. These parameters include the number of views (NV), the number of detector bins (NDB), pixel length (PL), detector bin length (DBL), the distance between the source and rotation center (DSR), the distance between the detector and rotation center (DDR), as well as the photon number of incident X-rays (PN). The units for DBL, DSR, and DDR are all in millimeters. To assist readers in re-simulating our data, we present the scanning protocols in Table A. Besides, the scanning protocols of unseen clients can be found in Table B. It can be observed that the protocol differences between known clients are quite different, and the protocols of unseen clients differ significantly from those of known clients. Following the previous simulation process [5], we generate various LDCT data by introducing Poisson and electronic noise into the measured projection data to replicate low-dose conditions. The simulated function can be formulated as:

$$\mathbf{Y}_{si} = \ln \frac{I_0}{\text{Poisson}\left(I_0 \exp(-\hat{\mathbf{Y}}_{si})\right) + \text{Normal}\left(0, \sigma_e^2\right)}, \tag{A.2}$$

where $\mathbf{Y}_{si}$ and $\hat{\mathbf{Y}}_{si}$ denote the clean projection and noised projection, respectively. σ_2^2 represents the electronic noise variance, which is set to 10 in this paper following [5].

Since the parameters within g have different value

Table A. Geometry Parameters and Dose Levels in Different Known Clients.

	Client #1	Client #2	Client #3	Client #4	Client #5	Client #6	Client #7	Client #8
NV	1024	128	512	384	712	200	560	368
NDB	512	768	768	600	720	730	755	500
PL	0.66	0.78	1.00	1.40	0.60	0.88	1.20	1.00
DBL	0.72	0.58	1.20	1.50	0.82	0.78	1.30	1.30
DSR	250	350	500	350	300	350	300	350
DDR	250	300	400	300	350	280	400	350
PN	1×10^5	1×10^6	5×10^4	1.25×10^5	1.3×10^5	0.9×10^6	4.5×10^4	1.45×10^5

Table B. Geometry Parameters and Dose Levels in Different Unseen Clients.

	Unseen Client #1	Unseen Client #2	Unseen Client #3	Unseen Client #4
NV	768	428	100	896
NDB	550	590	768	730
PL	0.57	1.10	0.50	0.70
DBL	0.83	1.10	0.60	0.93
DSR	200	350	200	250
DDR	300	300	250	400
PN	1.3×10^5	1.4×10^5	1.1×10^6	9×10^4

Table C. Quantitative Results of PSNR and SSIM for the Post-processing Task.

	Average	
Personalized Component	PSNR	SSIM
Decoder E	40.67	97.27
Scanning-Informed Hypernetwork H_s	40.38	97.22
Decoder E & Scanning-Informed Hypernetwork H_s	40.63	97.25

ranges, we normalized all the parameters in g as:

$$\hat{g_j^i} = \frac{g_j^i - \min(g_j)}{\max(g_j) - \min(g_j)}, \tag{A.3}$$

where g_j^i is j-th element in i-th protocol vector g. $\hat{g_j^i}$ denotes the normalized feature. $\max(g_j)$ and $\min(g_j)$ represent the maximum and minimum values of the j-th element across all g of different strategies.

In our experiments, the validation set was exclusively used for model selection during testing and was not involved in training or fine-tuning.

C. Stability Experiment

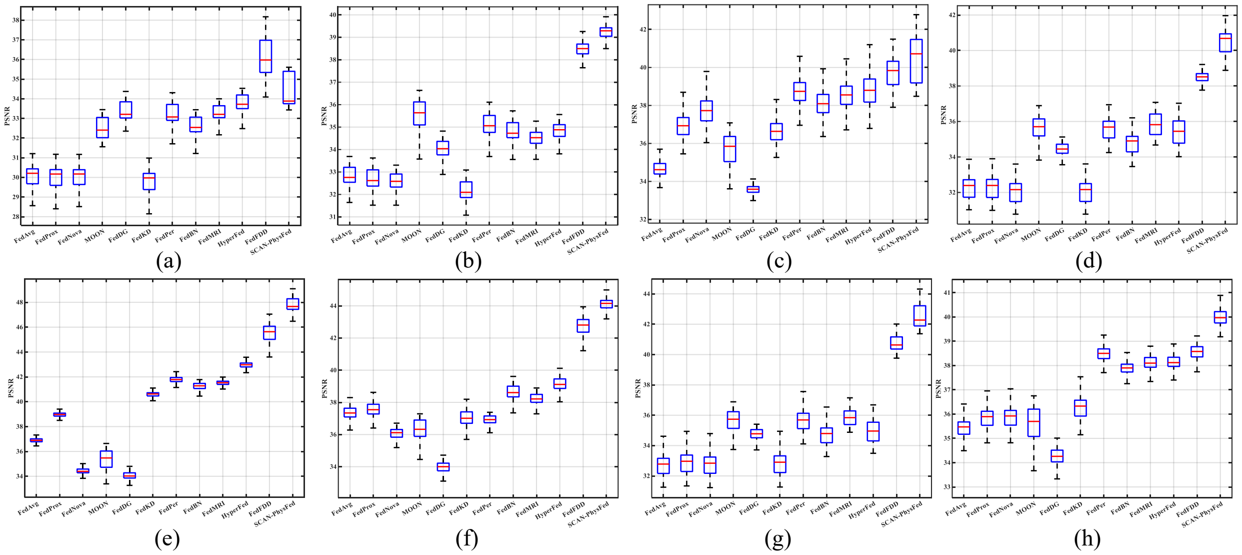
Figures A and B illustrate the boxplots of both PSNR and SSIM for different FL methods in different clients. As seen in the figures, our SCAN-PhysFed not only achieves a consistently high median but also exhibits a narrower interquartile range across all clients, indicating low variability and minimal outliers. This stability underscores the robustness of my method in comparison to others. In contrast, other methods often show wider interquartile ranges and more extreme outliers, reflecting considerable performance fluctuations among clients. Specifically, generic FL methods can maintain reasonable performance in certain clients, yet they often collapse in clients with different noise distributions. While personalized FL methods partially address this issue, they still exhibit poor performance in some clients because they lack sufficient knowledge to effectively personalize the imaging process.

Our method, SCAN-PhysFed, adopts a physics-driven approach, leveraging physical scanning parameters and anatomy priors to personalize the imaging process. This approach effectively reduces learning complexity and enhances overall model performance. The results strongly support this, with SCAN-PhysFed displaying consistent boxplot characteristics across all subplots, underscoring its robustness and reliability in maintaining steady performance with minimal variation.

D. Personalization Experiment

Additionally, we evaluate the performance with different personalized components. In these experiments, we use RED-CNN as the imaging backbone and optimize the network using only MSE loss, with other settings consistent with previous experiments. Attempts to only personalize the encoder resulted in non-converging training, suggesting that the encoder may not be suitable as a personalized component. Meanwhile, it is challenging to use a shared decoder to project varying imaging features into the image domain. Furthermore, since anatomy prior is independent of the scanning protocol and remains similar across different patients, we choose to leverage data from multiple clients to train a shared, robust anatomy feature extractor.

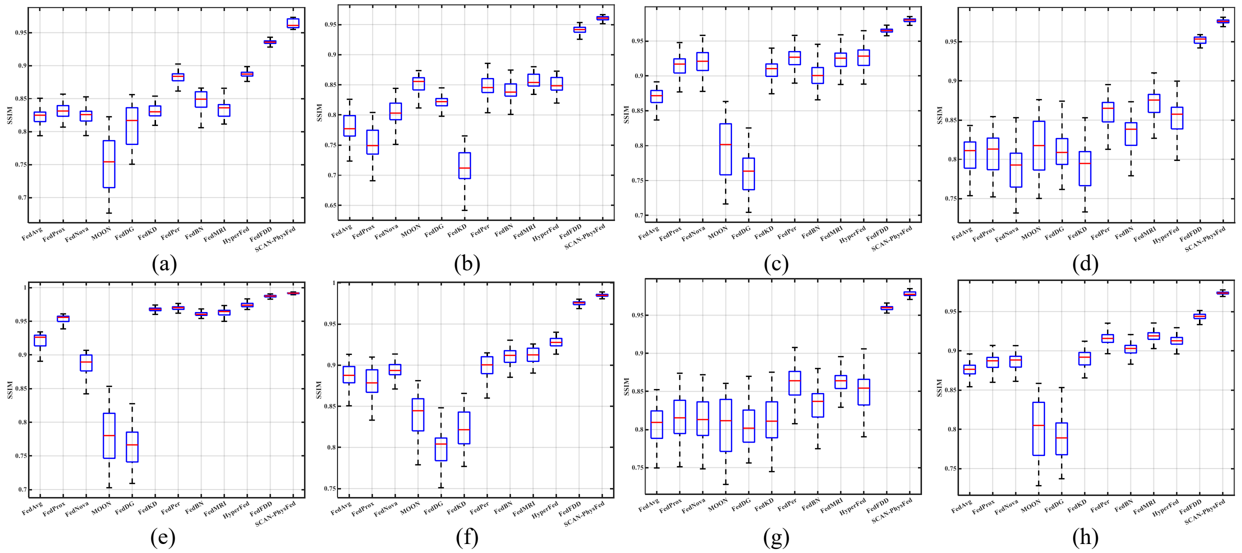
Therefore, in this study, we focus on personalizing the



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure A. Boxplots of PSNR in all clients based on different FL methods. (a)-(h) indicate the PSNR boxplots of Client #1 to Client #8, respectively.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure B. Boxplots of SSIM in all clients based on different FL methods. (a)-(h) indicate the SSIM boxplots of Client #1 to Client #8, respectively.

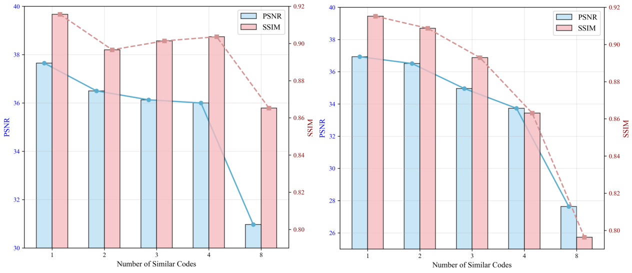
decoder and the scanning-informed hypernetwork. Results in Table C show minimal performance differences between these methods, indicating that both personalized decoder and hypernetwork approaches yield satisfactory results. However, to handle unseen clients, our PVQS strategy requires quantizing the scanning code using a protocol codebook, which necessitates extracting these scanning codes with a standardized extractor. As a result, we adopt client-specific decoders in this paper to project personalized features into the image domain.

E. Robustness Experiment

In this subsection, we expanded our evaluation using a larger dataset [4] with 40 training and 10 testing patients, divided into eight clients (four with head CT and four with torso CT). Meanwhile, we introduce a clinical metric MAHE to assess reconstruction accuracy. The results in Table D show that our method achieves promising performance across different data types, surpassing strong baselines like FedALA [8] and approaching the central-

Table D. Quantitative Results in the Large Dataset.

	Head			Torso			Average			Cross-Dataset Experiment	Head → Torso		
	PSNR↑	SSIM↑	MAHE↓	PSNR↑	SSIM↑	MAHE↓	PSNR↑	SSIM↑	MAHE↓		PSNR↑	SSIM↑	MAHE↓
CL	42.11	96.05	13.41	42.70	97.06	13.28	42.21	96.55	13.34		36.36	91.73	15.16
FedAvg	29.91	66.00	73.34	28.55	69.51	92.30	29.23	67.76	82.44		29.18	79.19	77.15
FedProx	30.83	68.80	64.42	30.12	72.97	75.87	30.47	70.88	70.80		30.37	80.63	66.14
HyperFed	34.06	75.53	37.52	35.44	80.15	37.62	34.75	77.84	38.82		34.75	85.90	32.14
FedFDD	37.96	90.13	23.48	39.24	91.94	22.17	38.50	91.03	22.82		35.60	91.24	24.63
FedALA	36.95	89.10	26.99	37.67	88.87	25.29	37.31	88.99	26.14		36.77	91.28	24.63
Ours†	39.72	93.90	17.27	**41.40**	96.00	**15.19**	40.56	94.95	16.23		**39.18**	**94.87**	**17.81**
Ours	**41.74**	**94.38**	**16.35**	41.09	**96.05**	15.66	**41.42**	**95.21**	**16.01**		37.24	94.47	19.15



(a) Head

(b) Torso

Figure C. Ablation Study about PVQS.

ized learning (CL) upper bound. This improvement stems from incorporating anatomical prior information, which enhances reconstruction quality and adaptability. Furthermore, in a cross-dataset and cross-part setting, we used models trained on head CT from [4] to validate torso data from the Mayo2016 dataset used in our main text. Our method demonstrates significant generalization improvements compared to other methods.

Besides, we combined our method with FedALA, referred to as "Ours†," with the results presented in Table D. Its performance is comparable to our standalone method. Additionally, it demonstrates robust generalization performance in the cross-dataset experiment. Compared to vanilla FedALA, "Ours" and "Ours†" achieve significantly better performance, highlighting the effectiveness of integrating anatomical and scanning information in addressing the heterogeneous issue. While personalization is not the primary focus of this paper, we believe future works could explore novel personalization approaches based on the proposed framework to further improve performance.

F. Ablation Study about PVQS

We compared the performance of aggregating multiple codes and the results are shown in Figure C. Our findings indicate that the nearest-neighbor strategy is generally more effective. Using more codes may result in negative outcomes due to the heterogeneous issue. However, we believe this may vary depending on specific cases, and we will discuss it by simulating more data. Specifically, when scanning parameters or anatomical regions are similar, aggregation may yield more robust results. This inspires us to consider developing a quantization method to determine the optimal number of codes in our future work. Overall, we suggest to chose the nearest code as the quantization.

Table E. The Computational Time.

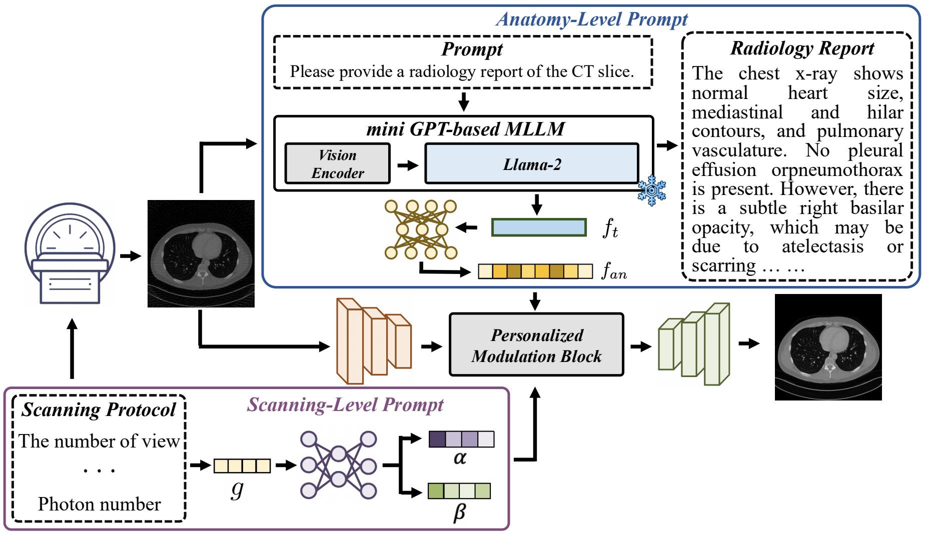
	MLLM	Imaging	Total
Time (s)	0.386	0.024	0.410

G. Prompt Example

To help readers understand our pipeline, we give a prompt example in this section. As shown in Figure D, anatomy-level prompts are generated by the MLLM from CT slices to extract textual features f_t for personalization. For scanning-level prompts, we utilize physical scanning protocol g (details can be found in the previous section) to further personalize the imaging features. These prompts are integrated via the Personalized Modulation Block to adaptively optimize the reconstruction process based on anatomical and scanning information.

H. Limitation

Although the proposed SCAN-PhysFed demonstrates satisfactory generalizability across different protocols, body parts, and imaging networks, there are still areas for improvement in future work. ***i)*** This method is closely aligned with the physical principles of CT imaging; therefore, adapting it to other imaging modalities would require adjustments based on the specific physical characteristics of those modalities. A promising direction for future work is to extend this approach to other medical imaging modalities, such as MRI or ultrasound, by incorporating relevant modality-specific physical knowledge. ***ii)*** While introducing anatomical information significantly enhances performance, it also raises potential security concerns. For instance, an attacker could compromise the MLLM and inject a backdoor into the feature representation, potentially



Please provide a radiology report of the CT slice.

The chest x-ray shows normal heart size, mediastinal and hilar contours, and pulmonary vasculature. No pleural effusion orpneumothorax is present. However, there is a subtle right basilar opacity, which may be due to atelectasis or scarring … …

Figure D. The Illustration of the Prompt Working.

deceiving downstream tasks and leading to inaccurate or manipulated results. This underscores the need for robust security measures to protect the model against such vulnerabilities.

A significant underlying concern pertains to the substantial increase in computational complexity that would inevitably accompany the implementation of MLLM. To mitigate potential computational complexity, we designed our method using a miniGPT-based MLLM, achieving relatively low computational cost. As shown in Table E, the total computation time is 0.410 seconds, with the MLLM module taking 0.386 seconds. In medical imaging tasks, real-time performance is not a strict requirement. We believe the current computation speed is adequate for practical applications.

References

[1] Asma Alkhaldi, 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. 1

[2] 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. 1

[3] 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. 1

[4] Taylor R Moen et al. Low-dose ct image and projection dataset. *Medical Physics*, 48(2):902–911, 2021. 3, 4

[5] Shanzhou Niu, Yang Gao, Zhaoying Bian, Jing Huang, Wufan Chen, Gaohang Yu, Zhengrong Liang, and Jianhua Ma. Sparse-view x-ray ct reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, 59(12): 2997, 2014. 1

[6] Ge Wang, Yi Zhang, Xiaojing Ye, and Xuanqin Mou. *Machine learning for tomographic imaging*. IOP Publishing, 2019. 1

[7] Habib Zaidi and Bruce Hasegawa. Determination of the attenuation map in emission tomography. *Journal of Nuclear Medicine*, 44(2):291–315, 2003. 1

[8] Jianqing Zhang et al. Fedala: Adaptive local aggregation for personalized federated learning. In *AAAI*, 2023. 3