nature portfolio

Peer Review File

Imaging foundation model for universal enhancement of non-ideal measurement CT

Corresponding Author: Dr Yuting He

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

A version of this paper was originally rejected for publication by Nature Communications, however that decision was reconsidered after appeal by the authors.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript presents TAMP, a foundation model for enhancing CT images acquired with non-ideal protocols (NICT). TAMP is pre-trained on 10.8 million synthetic image pairs (SimNICT) generated via physics-based simulation and can be adapted to specific use cases through parameter-efficient fine-tuning. Evaluated across 27 synthetic benchmark tasks varying in protocol, artifact severity, and anatomy, TAMP and its task-adapted variant (TAMP-S) outperform six specialized baselines and a prior foundation model (ProCT). The code and synthetic data will be made available. To the best of my knowledge, this is the first foundation model specifically designed for CT artifact correction. The study is timely and of potential significance to the field. However, several concerns and suggestions are outlined below.

Major Comments

1. Synthetic-to-real generalization gap

The primary evaluation focuses on synthetic data generated using the same simulator as the pretraining source. While the inclusion of real-world evaluation is appreciated, the real-data cohort is limited to only five volumes, and in several cases, part of the data is used for adaptation rather than held-out testing. Stronger evidence of generalization to clinical data would considerably strengthen the impact of the work.

2. Lack of task-specific or diagnostic relevance metrics

All evaluation metrics (PSNR, SSIM, RMSE, LPIPS) and radiologist scores (PBN, SQR, PCA) focus on image similarity and subjective image quality. While these are useful, they do not demonstrate improvements in diagnostic utility. Including task-specific metrics—e.g., nodule detection accuracy, segmentation performance, or radiomic feature preservation—would help link image quality improvements to downstream clinical value.

3. Evaluation reproducibility and dataset access

The authors mention plans to release the synthetic pretraining dataset, which is commendable. However, both the synthetic and real-world evaluation datasets are created in-house and are not publicly accessible. This limits the reproducibility of results and makes external validation difficult. The authors are encouraged to consider releasing anonymized versions of the real-data data.

4. Effect of large-scale pretraining

The manuscript highlights the use of 10.8 million pretraining image pairs and a transformer-based architecture. However, no ablation is provided on the effects of training data size or model size on downstream performance. Including such analysis would help clarify the scaling behavior and inform future work.

Minor Comments

1. Figure 1 – The phrase ‘time-cost model training’ may be unclear.
2. Terminology consistency – The manuscript inconsistently uses ‘physics-driven’ and ‘physical-driven’
3. Abbreviations – Some abbreviations such as PCA, PBN, SQR are used without clear definitions in the main text.

(Remarks on code availability)
The repo appears well-organized and clearly documented. Due to time constraints, I have not attempted to run the code and the model.

However, several key components necessary for full reproducibility are currently missing:
- Access to the pretraining synthetic dataset (SimNICT). Only a small subset is accessible.
- Code for running the pretraining pipeline
- Evaluation data splits for synthetic benchmarks

Reviewer #2

(Remarks to the Author)
This paper proposes TAMP, an imaging foundation model for universal enhancement of non-ideal measurement CT (NICT), including low-dose CT (LDCT), sparse-view CT (SVCT), and limited-angle CT (LACT). The authors construct a large physics-driven simulated dataset (SimNICT, 10.8M paired images), introduce a multi-scale transformer with dual-domain learning (image + projection), and apply parameter-efficient fine-tuning (LoRA) for rapid adaptation with few paired slices.

1. The Introduction argues that specialized models are costly to develop and fragile to protocol changes, motivating a universal FM across LDCT, SVCT, and LACT. However, in practice each scan uses a single, fixed NICT protocol, and this protocol is always known in advance. The manuscript should better clarify in what real clinical scenarios a universal enhancer is preferable to simply deploying the appropriate specialized model.

2. All experiments (e.g., Fig. 2) are conducted under fixed, single-scenario settings (LDCT-only, SVCT-only, LACT-only) using datasets where NICT is simulated from ICT. This mirrors conditions where specialized models already suffice and does not by itself prove universality. Moreover, since the test data are also simulated, it remains unclear how the model performs on measured public datasets. it would be valuable to evaluate on real NICT test data, such as the AAPM-Mayo LDCT dataset or other publicly available measured SVCT/LACT datasets, rather than only simulated counterparts.

3. TAMP is pretrained on the very large SimNICT dataset, whereas baselines are trained only on smaller task-specific datasets. Since pretraining is also training, much of the reported gain may come from data scale rather than architectural novelty. Ablation is needed to disentangle the contributions of pretraining, the multi-scale transformer backbone, and the dual-domain loss.

4. The framework combines multi-scale feature extraction, a transformer backbone, and dual-domain loss. These elements have each appeared in prior enhancement methods, and it is not clear what methodological improvement is offered here beyond re-assembly. The paper would benefit from a clearer discussion of how its design advances existing approaches.

5. Although the universal model motivation is emphasized, the evaluations again remain limited to single-scenario settings. This essentially reproduces conditions where specialized models are sufficient. To substantiate universality, the authors should provide pooled or mixed test cohorts (with LDCT, SVCT, and LACT cases together) or comparisons against an oracle ensemble of specialized models.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
In this manuscript, the authors propose a foundation model to enhance CT images affected by either high noise levels, limited-angle artifacts, or sparse-view artifacts. My detailed comments are as follows:

• The manuscript lacks a clear motivation and an intended clinical application. In particular, the authors confuse several x-ray–based imaging modalities such as clinical CT, cone-beam CT, and tomosynthesis, as detailed in the following points.

• Introduction: The authors introduce the considered artifacts, i.e., high image noise, limited-angle artifacts, and sparse-view artifacts. However, only high image noise is relevant to clinical CT. Limited-angle artifacts and sparse-view artifacts are more common in cone-beam CT or in dedicated modalities such as breast tomosynthesis. These modalities, however, differ significantly from clinical CT in terms of spatial resolution, noise, etc. It is unclear what the intended application of the foundation model might be, since it was trained solely on clinical CT data. Furthermore, such non-standard CT settings,

except for high image noise, are not widely used in clinical CT practice, contrary to the authors' claim.

• Introduction: Limited-angle and sparse-view artifacts are not of interest in clinical CT. In particular, limited-angle acquisitions due to a restricted angular range are not possible in clinical CT given the gantry-based setup. Moreover, sparse-view acquisitions would not increase scan speed as claimed by the authors, since the gantry rotation speed is constant. In addition, sparse-view artifacts would look very different from what is shown in this manuscript, as clinical CT data are typically acquired using a spiral trajectory. It seems the authors considered only circular trajectories often used in cone-beam CT.

• Introduction: Dose reduction via a reduction of tube voltage is usually not desired in clinical imaging. While reducing tube current results only in increased image noise, reducing tube voltage leads to an undesired change in image contrast.

• Data quantity: The authors state that data quantity is an issue. I disagree. The considered artifacts can easily be simulated using clinical data. This approach has been used for decades and is not new.

• Evaluation: All evaluation metrics used (PSNR, SSIM, RMSE, LPIPS) are clinically irrelevant, and clinical validation is missing. So far, the authors have only demonstrated that the proposed method performs "image prettification." Diagnostic quality was not assessed in the manuscript.

• Section 2.4: The data used for validation require further explanation. What systems were used? What trajectories were considered? How were sparse-view artifacts (circular vs. spiral trajectory) or limited-angle artifacts (what is the intended application given the angular ranges) simulated?

• Section 2.5: The validation by radiologists is, in my opinion, not very meaningful. Naturally, radiologists will rate an image with limited-angle artifacts lower compared to the same image without such artifacts. However, this does not prove that the corrected images preserve diagnostically relevant structures. It would be much more meaningful to investigate dedicated clinical applications. For example: using lung images with known nodules or lesions, are the lesions still visible if high noise is simulated and corrected for? Or in tomosynthesis, can microcalcifications still be detected after correction of limited-angle artifacts using the proposed model? What would happen in the latter case, given that the model was trained using clinical CT data and is applied to tomosynthesis data that has a much higher spatial resolution but a much lower dynamic range? The important part here is to demonstrate the clinical applicability.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you for the revision and rebuttal. I have reviewed the revised manuscript and responses.

All of my major concerns have been satisfactorily addressed, and the added experiments and clarifications significantly strengthen the work. I have no further comments and am happy with the manuscript in its current form.

(Remarks on code availability)
The repo appears well-organized and clearly documented. Due to time constraints, I have not attempted to run the code and the model.

Reviewer #2

(Remarks to the Author)
The reviewer thoroughly addressed my previous concerns and the manuscript was improved for publication. No more comments.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have adequately addressed all of my concerns, and I particularly appreciate the inclusion of clinically meaningful experiments.

(Remarks on code availability)

Major revisions we made include:

1. We have added three downstream clinical validation experiments in Sec. 5.6 of the Supplementary Materials (page 13, 14, 15), so that the diagnostic utility of TAMP has been demonstrated.
2. We have expanded the real-world validation by adding two new datasets in Sec. 2.4 of the main text (page 9), so that the synthetic-to-real generalization evidence has been substantially strengthened.
3. We have added validation experiments on an external cone-beam CT dataset in Sec. 5.3.2 of the Supplementary Materials (page 9, 10), thus demonstrating the cross-modality transfer capability.
4. We have added ablation studies on training data size in Sec. 5.5.2 of the Supplementary Materials (page 13), and model size in Sec. 5.5.3 (page 13), so that the scaling behavior of our foundation model has been clarified.
5. We have added component ablation studies in Sec. 5.5.1 of the Supplementary Materials (page 11, 12, 13) to disentangle the contributions of each component, so that the source of performance gains has been clearly demonstrated.
6. We have added mixed test cohort evaluation in Sec. 5.4.1 of the Supplementary Materials (page 10, 11), and oracle ensemble comparison in Sec. 5.4.2 (page 11), thus validating the universality of our foundation model.
7. We have clarified the target application scenarios and applicable modalities in the Introduction section (page 2), so that future readers will understand the intended use cases clearly.
8. We have added discussion on methodological improvements in the Method section (page 15), so that the novelty beyond existing approaches has been clarified.
9. We have released the complete SimNICT pretraining dataset in the Data Availability section, (page 16) so that future researchers can fully reproduce our pretraining.
10. We have released the pretraining code and evaluation data splits in the Code Availability section (page 16), thus ensuring full reproducibility of our results.
11. We have added detailed validation data information in Sec. 2.1 and Sec. 2.2 of the Supplementary Materials (page 3) with complete acquisition and reconstruction protocols, so that the experimental setup has been fully documented.
12. We have added definitions of radiologist evaluation metrics in Sec. 2.5 of the main text (page 11), thus avoiding potential confusion about these abbreviations.
13. We have re-checked the terminology and figures throughout the manuscript, so that the presentation has been more consistent and accurate.

All the comments and suggestions raised by the reviewers were carefully considered. Below are our itemized responses to all the points raised by the reviewers.

Comments from reviewer 1:

Comment#1: This manuscript presents TAMP, a foundation model for enhancing CT images acquired with non ideal protocols (NICT). TAMP is pre-trained on 10.8 million synthetic image pairs (SimNICT) generated via physics-based simulation and can be adapted to specific use cases through parameter-efficient fine-tuning. Evaluated across 27 synthetic benchmark tasks varying in protocol, artifact severity, and anatomy, TAMP and its task-adapted variant (TAMP-S) outperform six specialized baselines and a prior foundation model (ProCT). The code and synthetic data will be made available. To the best of my knowledge, this is the first foundation model specifically designed for CT artifact correction. The study is timely and of potential significance to the field. However, several concerns and suggestions are outlined below.

Response#1: We sincerely thank you for your recognition that TAMP represents "the first foundation model specifically designed for CT artifact correction" and for your thoughtful evaluation that our work "is timely and of potential significance to the field."

Our TAMP is designed as a foundation model for universal enhancement of non-ideal measurement CT (NICT), aiming to mitigate the limitations of specialized models in the NICT enhancement field, including high development costs and low generalizability for new protocol data. This work attempts to advance from task-specific solutions to a unified pre-training and fine-tuning foundation model framework, providing a potential solution for versatile CT acquisition scenarios.

In this version, we have further improved our paper according to your insightful suggestions, particularly strengthening the experimental validation and clarifying the methodological contributions.

Comment#2: 1. Synthetic-to-real generalization gap

The primary evaluation focuses on synthetic data generated using the same simulator as the pretraining source. While the inclusion of real-world evaluation is appreciated, the real-data cohort is limited to only five volumes, and in several cases, part of the data is used for adaptation rather than held-out testing. Stronger evidence of generalization to clinical data would considerably strengthen the impact of the work.

Response#2: We sincerely thank you for this valuable suggestion, which will help strengthen the evidence of generalization and clinical impact of our work. In this version, we have **1) expanded the real-world validation scale to 32 volumes** (6,775 slices) covering diverse clinical scenarios, and **2) validated TAMP's generalization to these extended clinical data** across three anatomical regions (chest, cardiac, abdomen) and three NICT types (LDCT, SVCT, LACT):

1. **Dataset expansion:** We expanded the real-world validation scale by collecting two additional datasets with 20 volumes (4,529 slices), resulting in 32 volumes (6,775 slices) in total (Table A). This scale exceeds existing NICT enhancement studies with real-world clinical data validation [1,2]. The expanded datasets include diverse clinical scenarios and imaging protocols, providing comprehensive and statistically reliable evidence of sim-to-real generalization.

Table A: Dataset expansion summary showing original, newly added, and total volumes/slices for real-world validation.

Dataset	Volumes		Slices		Clinical Application
	Train	Test	Train	Test	
Original datasets					
NJDTH-A (LDCT)	5	1	1,234	262	Routine chest CT
NJDTH-C (SVCT, LACT)	5	1	625	125	Cardiac imaging
Newly added datasets					
NJDTH-B (LDCT)	5	5	1,527	1,275	Coronary CTA
Mayo (SVCT, LACT)	5	5	900	737	Abdominal imaging
Total	**20**	**12**	**4,376**	**2,399**	

2. **Real-world validation:** We validated TAMP's generalization to clinical data on the expanded real-world datasets.
   a) **Experiment settings:** Six evaluation tasks were designed on these real-world datasets: **1) Routine chest LDCT** (LDCT on NJDTH-A) for thoracic screening imaging; **2) Coronary Low-dose CTA** (LDCT on NJDTH-B) for cardiac imaging with vascular structures; **3) Cardiac SVCT** (SVCT on NJDTH-C) for geometric artifacts in cardiac scans from angular undersampling; **4) Abdominal SVCT** (SVCT on Mayo) for abdominal imaging from sparse-view acquisition; **5) Cardiac LACT** (LACT on NJDTH-C) for geometric artifacts in cardiac scans from limited-angle acquisition; **6) Abdominal LACT** (LACT on Mayo) for abdominal imaging from limited-angle acquisition. These tasks validate TAMP's generalization to real-world clinical data across anatomical regions (chest, cardiac, abdomen) and clinical applications (routine screening, coronary angiography, cardiac imaging). Baseline methods (RED-CNN, FBPConvNet, ProCT) were trained from scratch on each training set, TAMP was evaluated zero-shot, and TAMP-S was adapted via parameter-efficient fine-tuning.
   b) **Experiment results:** TAMP and TAMP-S effectively enhance the quality of real-world NICT images across diverse clinical scenarios, based on two observations:

Table B: Real-world NICT enhancement results across six clinical scenarios (PSNR, dB, mean±std). All baseline methods were trained from scratch on task-specific training sets.

Method	LDCT		LACT		SVCT	
	NJDTH-A	NJDTH-B	NJDTH-C	Mayo	NJDTH-C	Mayo
RED-CNN	33.74±1.20	29.48±1.92	28.95±0.27	17.33±1.39	39.78±1.71	17.30±1.29
FBPConvNet	33.56±1.19	28.78±1.61	34.34±0.50	31.26±2.72	39.72±1.58	31.69±1.43
ProCT	32.83±0.97	28.99±1.74	36.98±0.97	31.26±3.86	40.98±1.90	32.85±1.60
TAMP	32.58±1.06	29.02±1.91	31.37±0.51	31.28±2.28	37.96±0.97	29.20±1.48
TAMP-S	34.02±1.24	30.12±1.86	37.09±0.96	35.65±1.68	41.15±2.03	33.41±1.57

   a) **TAMP demonstrates robust simulation-to-reality generalization across the newly added datasets without additional training.**

   TAMP achieves competitive zero-shot performance across all expanded real-world NICT tasks. As shown in Table B, on the newly added NJDTH-B LDCT dataset, TAMP achieves PSNR of 29.02 dB, comparable to baseline methods. For geometric artifact tasks on the newly added Mayo dataset, TAMP demonstrates effective zero-shot transfer with PSNR of 31.28 dB on LACT and

29.20 dB on SVCT, validating its learned universal representations.

b) **After parameter-efficient adaptation, TAMP-S achieves state-of-the-art performance across all real-world NICT tasks from diverse clinical scenarios.**

TAMP-S achieves the best PSNR performance across all 6 tasks (Table B). For the newly added NJDTH-B LDCT dataset, TAMP-S achieves PSNR of 30.12 dB, outperforming all baselines trained from scratch including RED-CNN (29.48 dB), FBPConvNet (28.78 dB), and ProCT (28.99 dB). For the newly added Mayo dataset, TAMP-S achieves PSNR of 35.65 dB (+4.39 dB vs FBPConvNet) on LACT tasks and 33.41 dB (+0.56 dB vs ProCT) on SVCT tasks. These consistent improvements validate TAMP's effective parameter-efficient adaptation capability for real-world NICT enhancement.

Manuscript changes: We have substantially revised the real-world validation section (Sec.2.4 in the main text and corresponding sections in Supplementary Materials) to incorporate the two newly added datasets (NJDTH-B and Mayo), thereby providing stronger and more statistically reliable evidence of sim-to-real generalization across diverse clinical scenarios and imaging protocols. Specifically:

"To evaluate TAMP's real-world enhancement capability, we collected real-world NICT data from diverse clinical scenarios and imaging protocols. The datasets include: 1) NJDTH-A: LDCT data from Nanjing Drum Tower Hospital Jiangbei (NJDTH) with 6 cases (1,496 slices)...; 2) NJDTH-B: LDCT data from NJDTH with 10 cases (2,802 slices) acquired at low dose (120 kVp, 50 mAs) and high dose (120 kVp, 200 mAs)...; 3) NJDTH-C: SVCT and LACT data from NJDTH with 6 cases (750 slices)...; 4) Mayo: SVCT and LACT data with 10 cases reconstructed from publicly available projection data of Mayo Clinic Low Dose CT Grand Challenge... These four datasets constitute 6 evaluation tasks."

"Quantitatively, TAMP achieves competitive zero-shot performance across all real-world NICT tasks. In LDCT tasks, TAMP achieves median LPIPS scores of 7.65% (NJDTH-A) and 16.54% (NJDTH-B), substantially outperforming compared methods. For geometric artifact tasks, TAMP demonstrates effective zero-shot transfer on LACT (median PSNR: 31.53 dB on NJDTH-C, 30.61 dB on Mayo) and SVCT (median SSIM: 94.35% on NJDTH-C, 63.37% on Mayo), validating its learned universal representations."

"TAMP-S achieves the best median performance in all 24 metric-task combinations (6 tasks × 4 metrics). Notably, for challenging geometric artifact reconstruction, TAMP-S demonstrates substantial improvements on LACT tasks with median PSNR of 37.01 dB (NJDTH-C, +0.26 dB vs ProCT) and 35.00 dB (Mayo, +3.26 dB vs FBPConvNet), and on SVCT tasks with median SSIM of 96.57% (NJDTH-C, +0.73% vs ProCT) and 77.97% (Mayo, +1.39% vs ProCT)."

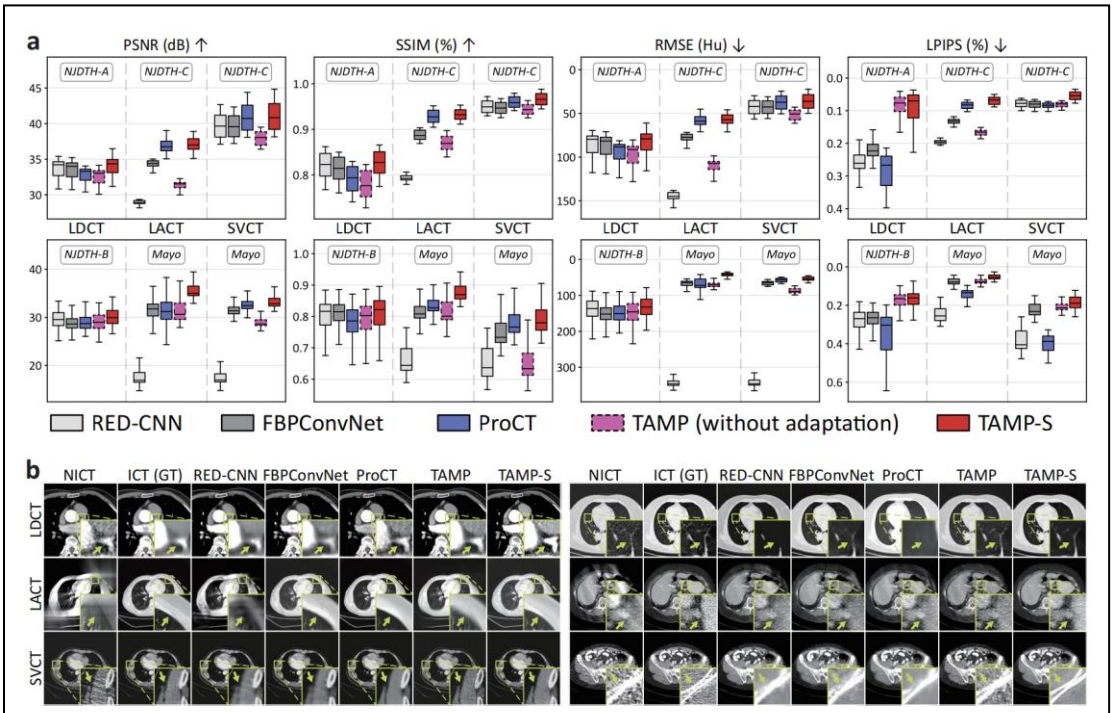

Fig. 4 Real-world validation demonstrates our great application potential in real-world NICT images. a) Our TAMP can directly enhance real-world NICT images and TAMP-S is adapted to specific NICT settings, achieving further improvement. b) Our TAMP has significant visual enhancement demonstrating its great clinical application potential in real-world NICT images.

[1] Meng M, Wang Y, Zhu M, et al. DDT-Net: Dose-Agnostic Dual-Task Transfer Network for Simultaneous Low-Dose CT Denoising and Simulation. IEEE Journal of Biomedical and Health Informatics. 2024;28(6):3298-3309. doi:10.1109/JBHI.2024.3376628.

[2] Lee J, Baek J. Iterative reconstruction for limited-angle CT using implicit neural representation. Physics in Medicine & Biology. 2024;69(8):085009. doi:10.1088/1361-6560/ad3c8e.

** The newly added NJDTH-B dataset was approved under the same ethics protocol as NJDTH-A and NJDTH-C (Ethics approval: AF/SC-08/03.0, 2025-0068-02).*

Comment#3: 2. Lack of task-specific or diagnostic relevance metrics

All evaluation metrics (PSNR, SSIM, RMSE, LPIPS) and radiologist scores (PBN, SQR, PCA) focus on image similarity and subjective image quality. While these are useful, they do not demonstrate improvements in diagnostic utility. Including task-specific metrics—e.g., nodule detection accuracy, segmentation performance, or radiomic feature preservation—would help link image quality improvements to downstream clinical value.

Response#3: We sincerely thank you for this insightful suggestion, which will help us demonstrate the clinical diagnostic value of our work. In this new version, we have conducted three downstream clinical validation experiments to validate that TAMP-enhanced images improve performance across critical diagnostic applications including **pulmonary nodule segmentation**, **pulmonary nodule malignancy diagnosis**, and **radiomics feature analysis**.

1. **Pulmonary nodule segmentation:** We have conducted pulmonary nodule segmentation

downstream validation experiments to evaluate how our TAMP enhancement of NICT images improves segmentation performance. Pulmonary nodule segmentation is a critical component in lung cancer diagnosis, as accurate delineation of nodule boundaries directly impacts malignancy assessment and treatment planning.

a) **Experiment setting:** We conducted nodule segmentation experiments on LIDC-IDRI dataset to measure task-specific diagnostic performance. We selected 8,553 training, 856 validation, and 2,139 test slices with expert consensus masks. ICT images were degraded into LDCT, SVCT, LACT and enhanced by TAMP. A U-Net [1] segmentation model trained on ICT was tested on NICT and TAMP-enhanced images, with Dice and IoU as task-specific metrics quantifying lesion delineation accuracy.

b) **Experiment results:** TAMP enhancement restores task-specific segmentation performance degraded by NICT. As shown in Table 15, Dice improved from 0.676, 0.527, 0.451 to 0.690, 0.662, 0.621 for LDCT, SVCT, LACT (+0.014, +0.135, +0.170). The substantial SVCT and LACT improvements demonstrate TAMP captures geometric artifacts and restores boundary information, directly improving diagnostic utility for nodule delineation in clinical workflows.

Table 15: Our enhancement improves pulmonary nodule segmentation on LIDC-IDRI (w/o vs. w/ our method).

Setting	Dice			IoU		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.676	0.690	+0.014	0.527	0.544	+0.017
SVCT	0.527	0.662	+0.135	0.380	0.513	+0.133
LACT	0.451	0.621	+0.170	0.314	0.468	+0.154

c) **Manuscript changes:** We have added the pulmonary nodule segmentation experiment in subsection “Downstream clinical validation” of the Supplementary Materials, so that readers can understand TAMP's clinical value in improving lesion delineation. Specifically:

“Pulmonary nodule segmentation is a critical component in lung cancer diagnosis, as accurate delineation of nodule boundaries directly impacts malignancy assessment and treatment planning. We conducted pulmonary nodule segmentation experiments on the LIDC-IDRI dataset to evaluate how TAMP enhancement improves automated lesion delineation accuracy.

We selected CT slices with expert consensus nodule masks, splitting them into 8,553, 856, 2,139 slices for training, validation, test respectively. ICT images were synthetically degraded into three NICT types (LDCT, SVCT, LACT) and subsequently enhanced by TAMP. A U-Net segmentation model was trained on ICT images with their corresponding masks, then tested on both NICT-degraded and TAMP-enhanced images to quantify segmentation accuracy improvements.

TAMP enhancement effectively restores segmentation accuracy degraded by NICT artifacts, validating improved diagnostic utility for lesion delineation. As shown in Table 15, TAMP improves Dice from 0.676, 0.527, 0.451 to 0.690, 0.662, 0.621 for LDCT, SVCT, LACT, corresponding to relative gains of +0.014, +0.135, +0.170. The substantial improvements in SVCT and LACT demonstrate that TAMP's learned

representations capture geometric artifact patterns and effectively restore boundary information corrupted by angular undersampling."

Table 15: Our enhancement improves pulmonary nodule segmentation on LIDC-IDRI (w/o vs. w/ our method).

Input	Dice			IoU		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.676	0.690	+0.014	0.527	0.544	+0.017
SVCT	0.527	0.662	+0.135	0.380	0.513	+0.133
LACT	0.451	0.621	+0.170	0.314	0.468	+0.154

2. **Pulmonary nodule malignancy diagnosis:** We have conducted pulmonary nodule malignancy diagnosis experiments to evaluate how TAMP enhancement improves malignancy assessment. Accurate diagnosis of nodule malignancy levels is critical for lung cancer screening, as it directly impacts clinical decision-making for follow-up and treatment.
   a) **Experiment setting:** We conducted malignancy classification experiments on the LIDC-IDRI dataset. We formulated a five-class malignancy classification task at the 2D slice level, where nodules are classified into expert consensus ratings: 1 (highly unlikely for cancer), 2 (moderately unlikely), 3 (indeterminate likelihood), 4 (moderately suspicious), and 5 (highly suspicious for cancer). A ResNet-18 classifier was trained on 8,458 ICT training slices with their malignancy labels, then tested on 2,139 test slices across seven scenarios: ICT baseline, three NICT types (LDCT, SVCT, LACT), and three TAMP-enhanced types (LDCT, SVCT, LACT) to quantify classification accuracy improvements.
   b) **Experiment results:** TAMP enhancement recovers task-specific diagnostic performance degraded by NICT. As shown in Table 16, accuracy improved from 0.817, 0.740, 0.758 to 0.831, 0.843, 0.839 for LDCT, SVCT, LACT (+0.014, +0.103, +0.081). Strong AUC $>$ 0.96 across all scenarios demonstrates TAMP restores diagnostically critical features for malignancy assessment, directly supporting clinical decision-making in lung cancer screening workflows.

Table 16: Our enhancement improves pulmonary nodule malignancy diagnosis on LIDC-IDRI (w/o vs. w/ our method).

Setting	**Accuracy**			**AUC**		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.817	0.831	+0.014	0.962	0.967	+0.005
SVCT	0.740	0.843	+0.103	0.932	0.966	+0.034
LACT	0.758	0.839	+0.081	0.939	0.964	+0.025

   c) **Manuscript changes:** We have added the pulmonary nodule malignancy diagnosis experiment in subsection "Downstream clinical validation" of the Supplementary Materials, so that readers can understand TAMP's clinical value in improving malignancy assessment. Specifically:

   "Accurate diagnosis of nodule malignancy levels is critical for lung cancer screening, as it directly impacts clinical decision-making for follow-up and treatment.

We conducted pulmonary nodule malignancy diagnosis experiments on the LIDC-IDRI dataset to evaluate how TAMP enhancement improves malignancy assessment.

We formulated a five-class malignancy classification task at the 2D slice level, where nodules are classified into expert consensus ratings: 1 (highly unlikely for cancer), 2 (moderately unlikely), 3 (indeterminate likelihood), 4 (moderately suspicious), and 5 (highly suspicious for cancer). A ResNet-18 classifier was trained on 8,458 ICT training slices with their malignancy labels, then tested on 2,139 test slices across seven scenarios: ICT baseline, three NICT types (LDCT, SVCT, LACT), and three TAMP-enhanced types to quantify classification accuracy improvements.

TAMP enhancement effectively recovers diagnostic accuracy degraded by NICT artifacts, validating improved clinical utility for malignancy assessment. As shown in Table 16, TAMP improves accuracy from 0.817, 0.740, 0.758 to 0.831, 0.843, 0.839 for LDCT, SVCT, LACT, corresponding to relative gains of +0.014, +0.103, +0.081. The substantial improvements in SVCT and LACT demonstrate that TAMP's learned representations effectively restore diagnostically critical features corrupted by geometric artifacts, maintaining strong discriminative ability (AUC $>$ 0.96) for reliable lung cancer screening."

Table 16: Our enhancement improves pulmonary nodule malignancy diagnosis on LIDC-IDRI (w/o vs. w/ our method).

Input	Accuracy			AUC		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.817	0.831	+0.014	0.962	0.967	+0.005
SVCT	0.740	0.843	+0.103	0.932	0.966	+0.034
LACT	0.758	0.839	+0.081	0.939	0.964	+0.025

3. **Radiomics feature analysis:** We have further evaluated how TAMP affects quantitative radiomics features. Radiomics analysis is highly sensitive to image quality, and feature preservation is critical for reliable downstream clinical applications.
   a) **Experiment setting:** To assess TAMP's clinical value in preserving quantitative imaging biomarkers, we evaluated radiomics feature stability on the same LIDC-IDRI lung nodules. We extracted 93 standard 2D radiomics features from ICT baseline, three NICT types (LDCT, SVCT, LACT), and TAMP-enhanced images using PyRadiomics. Feature stability was quantified by computing Pearson correlations between NICT (or TAMP) features and ICT reference features, where stable features achieving $r > 0.85$ are considered reliable for downstream clinical analysis.
   b) **Experiment results:** TAMP enhancement improves radiomics feature stability. As shown in Table 17, mean correlation improved from 0.646, 0.722, 0.721 to 0.685, 0.741, 0.750 for LDCT, SVCT, LACT (+0.039, +0.019, +0.029). Modest improvements compared to segmentation and diagnosis tasks indicate that while TAMP restores visual and diagnostic features effectively, quantitative radiomics remain sensitive to artifacts, requiring careful interpretation for clinical radiomics applications.

Table 17: Our enhancement improves radiomics feature preservation (w/o vs. w/ our method).

Setting	Mean Correlation (r)			Stable Rate (r > 8.5)		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ (%)
LDCT	0.646	0.685	+0.039	40.9%	41.9%	+1.0
SVCT	0.722	0.741	+0.019	51.7%	52.7%	+1.0
LACT	0.721	0.750	+0.029	48.4%	52.7%	+4.3

c) **Manuscript changes:** We have added the radiomics feature analysis experiment in subsection "Downstream clinical validation" of the Supplementary Materials, so that readers can understand TAMP's clinical value in preserving quantitative imaging biomarkers. Specifically:

"Radiomics analysis is highly sensitive to image quality, and feature preservation is critical for reliable downstream clinical applications. We evaluated how TAMP affects quantitative radiomics features to assess its clinical value in preserving quantitative imaging biomarkers.

We extracted 93 standard 2D radiomics features from ICT baseline, three NICT types (LDCT, SVCT, LACT), and TAMP-enhanced images using PyRadiomics on the same LIDC-IDRI lung nodules. Feature stability was quantified by computing Pearson correlations between NICT (or TAMP) features and ICT reference features, where stable features achieving $r > 0.85$ are considered reliable for downstream clinical analysis.

TAMP enhancement partially restores radiomics feature preservation degraded by NICT artifacts, validating improved quantitative imaging biomarker reliability. As shown in Table 17, TAMP improves mean correlation from 0.646, 0.722, 0.721 to 0.685, 0.741, 0.750 for LDCT, SVCT, LACT, corresponding to relative gains of +0.039, +0.019, +0.029. The modest improvements compared to segmentation and diagnosis tasks demonstrate that while TAMP's learned representations effectively restore visual quality and diagnostic features, quantitative radiomics features remain more sensitive to reconstruction artifacts, indicating the need for careful interpretation in radiomics-based clinical analysis."

Table 17: Our enhancement improves radiomics feature preservation (w/o vs. w/ our method)

Input	Mean Correlation (r)			Stable Rate ($r > 0.85$)		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ (%)
LDCT	0.646	0.685	+0.039	40.9%	41.9%	+1.0%
SVCT	0.722	0.741	+0.019	51.7%	52.7%	+1.0%
LACT	0.721	0.750	+0.029	48.4%	52.7%	+4.3%

[1] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015:234-241. doi:10.1007/978-3-319-24574-4_28.

Comment#4: 3. Evaluation reproducibility and dataset access

The authors mention plans to release the synthetic pretraining dataset, which is commendable. However, both the synthetic and real-world evaluation datasets are created in-house and are not publicly accessible. This limits the reproducibility of results and makes external validation difficult. The authors are encouraged to consider releasing anonymized versions of the real-data data.

Response#4: We sincerely thank you for this valuable suggestion about dataset accessibility, which will help us enhance the reproducibility and impact of our work. In this new version, we have made extensive datasets and resources publicly available to ensure reproducibility across three key areas:

Table B: Datasets and resources availability for reproducibility.

Dataset / Resource	Access Method	Platform / Location
Synthetic Pretraining Dataset		
SimNICT	Publicly released	HuggingFace, Internet Archive
SimNICT-AMOS-Sample	Publicly released	HuggingFace
NICT simulation code	Publicly released	HuggingFace
AutoPET, HECKTOR22	Restricted by license	Not available
Synthetic Evaluation Datasets		
Evaluation datasets	Publicly released	HuggingFace
Training / testing code	Publicly released	GitHub
Real-world Evaluation Datasets		
Private clinical datasets	Privacy-protected	Contact corresponding auther
Mayo Clinic LDCT dataset	Publicly available	AAPM Grand Challenge
Trained model weights	Publicly released	GitHub
Testing code	Publicly released	GitHub

1. **For synthetic pretraining dataset:** We have made the pretraining dataset SimNICT publicly available for full reproducibility. We have released 8 out of 10 datasets on HuggingFace (https://huggingface.co/datasets/YutingHe-list/SimNICT) from our SimNICT pretraining data. The data is hosted on Internet Archive (https://archive.org/search?query=creator%3A%22TAMP+Research+Group%22) with batch download scripts provided. The two unreleased datasets (AutoPET and HECKTOR22) are restricted by their original licensing terms. Additionally, a 78 MB sample dataset (SimNICT-AMOS-Sample, 55 volumes) is available on HuggingFace for quick prototyping. We have also released the physics-driven simulation code implementing the complete NICT generation pipeline using ODL and ASTRA Toolbox, enabling researchers to generate LDCT, SVCT, and LACT data with randomized degradation parameters.
2. **For synthetic evaluation datasets:** We have uploaded synthetic evaluation datasets on HuggingFace (https://huggingface.co/datasets/YutingHe-list/SimNICT/tree/main) in three folders: SimNICT-AMOS-Sample, SimNICT-APET-Sample, and SimNICT-COVID-Sample. Each folder contains train_nii_files and test_nii_files subfolders with clear training/testing splits.
3. **For real-world evaluation data:** 1) For our clinical data from Nanjing Drum Tower Hospital, due to patient privacy regulations, the raw data cannot be publicly released; however, we have provided trained model weights and complete testing code on GitHub

(https://github.com/YutingHe-list/TAMP) to enable researchers to verify our results. 2) For the Mayo Clinic LDCT dataset, this is a publicly available dataset from the AAPM Low Dose CT Grand Challenge (https://www.aapm.org/GrandChallenge/LowDoseCT/).

Manuscript changes: We have comprehensively revised the Data Availability section to provide detailed information about dataset accessibility. Specifically:

"The datasets and code used in this work are publicly available to ensure reproducibility. 1) SimNICT pretraining dataset: This work enrolled ten publicly available datasets, including COVID-19-NY-SBU, STOIC, MELA, LUNA, LNDb, HECKTOR22, CT_COLONOGRAPHY, AutoPET, AMOS22, and CT Images in COVID-19. 8 out of 10 datasets are released on HuggingFace with data hosted on Internet Archive; AutoPET and HECKTOR22 are excluded due to original licensing restrictions. 2) Synthetic evaluation datasets: Complete evaluation data with training/testing splits for all 27 benchmark tasks are available on HuggingFace. 3) Real-world evaluation data: Clinical data from Nanjing Drum Tower Hospital cannot be released due to patient privacy (Ethics approval: AF/SC-08/03.0, 2025-0068-02); however, trained model weights and testing code are provided on GitHub. Additionally, the Mayo Clinic LDCT dataset is publicly available."

Comment#5: 4. Effect of large-scale pretraining

The manuscript highlights the use of 10.8 million pretraining image pairs and a transformer-based architecture. However, no ablation is provided on the effects of training data size or model size on downstream performance. Including such analysis would help clarify the scaling behavior and inform future work.

Response#5: We sincerely thank you for this insightful suggestion, which will help us explore and analyze the effects of our pretraining approach. In this version, **training data size ablation** and **model size ablation** have been conducted, which will help clarify the scaling behavior and inform future work.

1. **Training data size ablation:** To validate the effect of training data size, we have conducted ablation studies using different subsets of our SimNICT dataset.
    a) **Experiment settings:** We pre-trained TAMP variants on 100k, 1M, and 10.8M image pairs respectively, and evaluated their performance across all 27 benchmark tasks. For each variant, we calculated the average PSNR across the three NICT categories (LDCT, LACT, SVCT) to comprehensively assess the scaling effect of pretraining data size.
    b) **Experiment results:** As shown in Table 12, the results demonstrate consistent and substantial improvement with increased data scale across all three NICT categories. Scaling from 100k to 1M image pairs yields significant improvements of 3.35 dB, 3.17 dB, and 2.83 dB for LDCT, LACT, and SVCT respectively. Further scaling to the full 10.8M dataset brings additional substantial gains of 4.11 dB, 4.08 dB, and 3.43 dB respectively, demonstrating that large-scale pretraining continues to improve performance without saturation. The full 10.8M dataset achieves optimal performance across all diverse NICT scenarios, with the most significant improvement observed in LACT (7.25 dB total improvement from 100k to 10.8M), where complex geometric artifacts benefit substantially from diverse training examples. This scaling behavior validates the importance of large-scale physics-driven pretraining for establishing

effective foundation models, aligning with foundation model principles demonstrated in natural language processing and general computer vision domains.

Table 12: Effect of pretraining data size on NICT enhancement performance. PSNR (dB) is reported as mean±std.

Training data size	APET-LDCT-High	APET-LACT-High	APET-SVCT-High
100k	44.16±0.76	38.43±0.82	44.00±0.75
1M	47.51±0.75	41.60±0.91	46.83±0.89
10.8M	51.62±1.07	45.68±0.98	50.26±1.04

c) **Manuscript changes:** We have added the training data size ablation experiment in subsection "Ablation studies" of the Supplementary Materials, so that readers can understand the effect of pretraining data size on model performance. Specifically:

"To validate the effect of training data size, we pre-trained TAMP variants on 100k, 1M, and 10.8M image pairs respectively, and evaluated their performance across all 27 benchmark tasks. As shown in Table 12, the results demonstrate consistent and substantial improvement with increased data scale across all three NICT categories. Scaling from 100k to 1M image pairs yields significant improvements of 3.35 dB, 3.17 dB, and 2.83 dB for LDCT, LACT, and SVCT respectively. Further scaling to the full 10.8M dataset brings additional substantial gains of 4.11 dB, 4.08 dB, and 3.43 dB respectively, demonstrating that large-scale pretraining continues to improve performance without saturation. This scaling behavior validates the importance of large-scale physics-driven pretraining for establishing effective foundation models."

Table 12: Effect of pretraining data size on NICT enhancement performance. PSNR (dB) is reported as mean±std.

Training Data Size	APET-LDCT-High	APET-LACT-High	APET-SVCT-High
100k	44.16±0.76	38.43±0.82	44.00±0.75
1M	47.51±0.75	41.60±0.91	46.83±0.89
10.8M	51.62±1.07	45.68±0.98	50.26±1.04

2. **Model size ablation:** We have investigated the impact of model capacity by training three variants with systematically reduced channel dimensions while maintaining the same network depth.

a) **Experiment settings:** Three model variants were created by systematically reducing channel dimensions while maintaining the same 60-layer transformer depth and multi-scale architecture. Table 13 summarizes their configurations. All variants were trained on the same 100k image pairs with identical hyperparameters and evaluated across three NICT scenarios (LDCT/SVCT/LACT) on APET, AMOS, and COVID datasets.

Table 13: Model size ablation: architectural configurations and computational costs.

Layer name	MITNet-Base	MITNet-Small	MITNet-Tiny
Embedding-1	Conv 3×3, stride 1, ch=2	Conv 3×3, stride 1, ch=1	Conv 3×3, stride 1, ch=1
DFE-1	[RSTB, dim=2, depth=[2], heads=[1]] × 1	[RSTB, dim=1, depth=[2], heads=[1]] × 1	[RSTB, dim=1, depth=[2], heads=[1]] × 1
Embedding-2	Conv 4×4, stride 2, ch=8	Conv 4×4, stride 2, ch=4	Conv 4×4, stride 2, ch=2
DFE-2	[RSTB, dim=8, depth=[4,4], heads=[2,2]] × 2	[RSTB, dim=4, depth=[4,4], heads=[2,2]] × 2	[RSTB, dim=2, depth=[4,4], heads=[1,1]] × 2
Embedding-4	Conv 6×6, stride 4, ch=32	Conv 6×6, stride 4, ch=16	Conv 6×6, stride 4, ch=8
DFE-4	[RSTB, dim=32, depth=[6,6,6], heads=[4,4,4]] × 3	[RSTB, dim=16, depth=[6,6,6], heads=[4,4,4]] × 3	[RSTB, dim=8, depth=[6,6,6], heads=[2,2,2]] × 3
Embedding-8	Conv 10×10, stride 8, ch=128	Conv 10×10, stride 8, ch=64	Conv 10×10, stride 8, ch=32
DFE-8	[RSTB, dim=128, depth=[8,8,8,8], heads=[8,8,8,8]] × 4	[RSTB, dim=64, depth=[8,8,8,8], heads=[8,8,8,8]] × 4	[RSTB, dim=32, depth=[8,8,8,8], heads=[4,4,4,4]] × 4
Fusion-4	Upsample + Conv 3×3, ch=80	Upsample + Conv 3×3, ch=40	Upsample + Conv 3×3, ch=20
Fusion-2	Upsample + Conv 3×3, ch=44	Upsample + Conv 3×3, ch=22	Upsample + Conv 3×3, ch=11
Fusion-1	Upsample + Conv 3×3, ch=96	Upsample + Conv 3×3, ch=48	Upsample + Conv 3×3, ch=24
Enhance	Conv 3×3, ch=1	Conv 3×3, ch=1	Conv 3×3, ch=1
Params (M)	7.48	1.95	0.52
FLOPs (G)	96.13	24.50	3.07

b) **Experiment results:** Model capacity scaling demonstrates significant performance improvements for NICT enhancement, validating the importance of sufficient model expressiveness for complex artifact correction. As shown in Table 14, the Base model (7.48M) achieves PSNR of 44.16, 38.43, 44.00 dB for LDCT, LACT, SVCT, substantially outperforming Tiny (0.52M) by 4.83, 1.16, 4.08 dB respectively. The substantial improvements in LDCT and SVCT (>4 dB) demonstrate that TAMP's transformer architecture effectively leverages increased capacity to learn complex geometric artifact patterns, while the smaller LACT improvement suggests limited-angle reconstruction is more constrained by the ill-posed inverse problem than model expressiveness.

Table 14: Model size ablation: performance comparison on APET test cases (PSNR in dB).

Model Size	APET-LDCT-High	APET-LACT-High	APET-SVCT-High
MITNet-Tiny	39.33±0.92	37.27±0.70	39.92±0.99
MITNet-Small	40.26±0.55	37.56±0.74	40.14±0.55
MITNet-Base	44.16±0.76	38.43±0.82	44.00±0.75

c) **Manuscript changes:**We have added the model size ablation experiment in subsection “Ablation studies” of the Supplementary Materials, so that readers can understand the impact of model capacity on NICT enhancement performance. Specifically:

“We investigated the impact of model capacity by training three variants with systematically reduced channel dimensions while maintaining the same network depth and multi-scale architecture. Table 13 shows the detailed architectural

configurations of each variant. All variants were trained on the same 100k image pairs with identical hyperparameters. As shown in Table 14, the Base model (7.48M) achieves PSNR of 44.16, 38.43, 44.00 dB for LDCT, LACT, SVCT, substantially outperforming Tiny (0.52M) by 4.83, 1.16, 4.08 dB respectively. The substantial improvements in LDCT and SVCT (>4 dB) demonstrate that TAMP's transformer architecture effectively leverages increased capacity to learn complex geometric artifact patterns."

Table 13: Detailed architectural configurations of three model variants for model size ablation.

Layer Name	MITNet-Base	MITNet-Small	MITNet-Tiny
Embedding-1	Conv 3×3, stride 1, ch=2	Conv 3×3, stride 1, ch=1	Conv 3×3, stride 1, ch=1
DFE-1	[RSTB, dim=2, depth=[2], heads=[1]] × 1	[RSTB, dim=1, depth=[2], heads=[1]] × 1	[RSTB, dim=1, depth=[2], heads=[1]] × 1
Embedding-2	Conv 4×4, stride 2, ch=8	Conv 4×4, stride 2, ch=4	Conv 4×4, stride 2, ch=2
DFE-2	[RSTB, dim=8, depth=[4,4], heads=[2,2]] × 2	[RSTB, dim=4, depth=[4,4], heads=[2,2]] × 2	[RSTB, dim=2, depth=[4,4], heads=[1,1]] × 2
Embedding-4	Conv 6×6, stride 4, ch=32	Conv 6×6, stride 4, ch=16	Conv 6×6, stride 4, ch=8
DFE-4	[RSTB, dim=32, depth=[6,6,6], heads=[4,4,4]] × 3	[RSTB, dim=16, depth=[6,6,6], heads=[4,4,4]] × 3	[RSTB, dim=8, depth=[6,6,6], heads=[2,2,2]] × 3
Embedding-8	Conv 10×10, stride 8, ch=128	Conv 10×10, stride 8, ch=64	Conv 10×10, stride 8, ch=32
DFE-8	[RSTB, dim=128, depth=[8,8,8,8], heads=[8,8,8,8]] × 4	[RSTB, dim=64, depth=[8,8,8,8], heads=[8,8,8,8]] × 4	[RSTB, dim=32, depth=[8,8,8,8], heads=[4,4,4,4]] × 4
Fusion-4	Upsample + Conv, ch=80	Upsample + Conv, ch=40	Upsample + Conv, ch=20
Fusion-2	Upsample + Conv, ch=44	Upsample + Conv, ch=22	Upsample + Conv, ch=11
Fusion-1	Upsample + Conv, ch=96	Upsample + Conv, ch=48	Upsample + Conv, ch=24
Enhance	Conv 3×3, ch=1	Conv 3×3, ch=1	Conv 3×3, ch=1
Params (M)	7.48	1.95	0.52
FLOPs (G)	96.1	24.5	6.4

Table 14: Model size ablation: performance comparison on APET test cases (PSNR in dB).

Model Size	LDCT-High	LACT-High	SVCT-High
MITNet-Tiny	39.33±0.92	37.27±0.70	39.92±0.99
MITNet-Small	40.26±0.55	37.56±0.74	40.14±0.55
MITNet-Base	44.16±0.76	38.43±0.82	44.00±0.75

Comment#6: Minor Comments

1. Figure 1 – The phrase 'time-cost model training' may be unclear.
2. Terminology consistency – The manuscript inconsistently uses 'physics-driven' and 'physical-driven'
3. Abbreviations – Some abbreviations such as PCA, PBN, SQR are used without clear definitions in the main text.

Response#6: We sincerely thank you for pointing out these issues regarding clarity and consistency, which will help improve the readability of our manuscript. We have carefully

revised all three concerns throughout the paper.

1. For the question "The phrase 'time-cost model training' may be unclear".

   We have revised the unclear phrase "time-cost model training" to "complex specialized model development" in Figure 1. This complex development involves multiple stages including model architecture design, training from scratch, iterative experimentation, and deployment optimization, requiring substantial time and computational resources. In contrast, our foundation model approach enables efficient adaptation through parameter-efficient fine-tuning, significantly reducing development overhead.

   Manuscript changes: We have revised the Introduction section to use "complex specialized model development" instead of the unclear phrase, and updated Figure 1 for terminology consistency, so that readers will understand the development challenges more clearly.

   "Complex specialized model development. Each new NICT protocol requires extensive dataset collection, model architecture design, and training from scratch, extending the development cycle of NICT imaging devices (Fig.1c)."

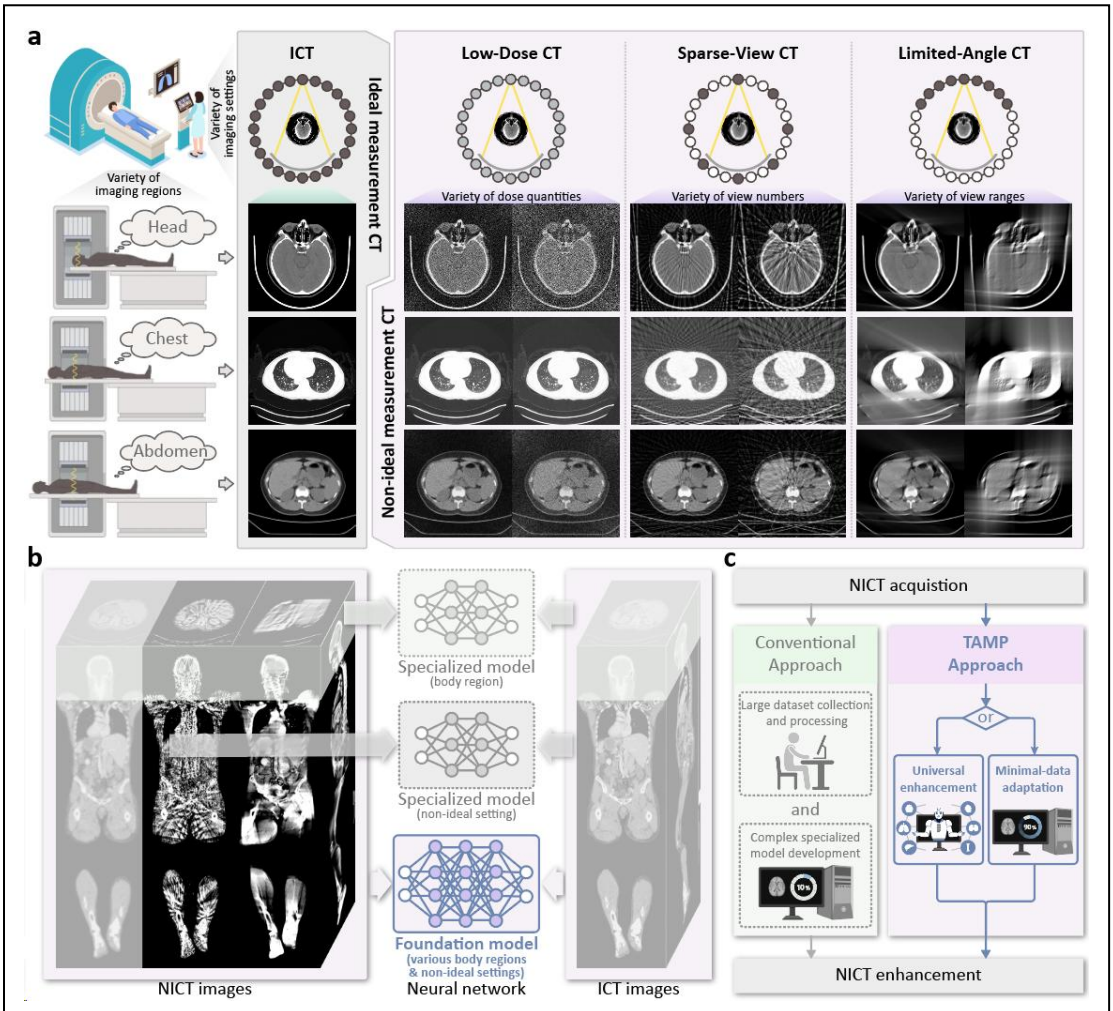

Fig. 1. Our TAMP is a universal non-ideal measurement computed tomography (NICT) enhancement foundation model that is able to enhance NICT images with various body regions and non-ideal settings and improve the efficiency of developing specialized NICT enhancement models. a) NICT expands the scope of CT applications with the advantages of radiation dose reduction, scanning acceleration, and adaptation of restricted scanning posture. However, the image quality of NICT is reduced, limiting its clinical effectiveness. b) Specialized NICT enhancement models that focus on specific body regions or non-ideal settings, are limited by the application scopes and construction costs. c) Our TAMP can directly enhance NICT images and adapt to specialized NICT enhancement tasks with low data and computational costs, improving imaging applications' development efficiency and effectiveness.

2. For the question "The manuscript inconsistently uses 'physics-driven' and 'physical-driven'":

We have unified all instances to "physics-driven" throughout the manuscript, so that our paper maintains consistent terminology.

Manuscript changes: We have revised the caption of Figure 6 to use "hysics-driven" instead of "physical-driven" for terminology consistency.

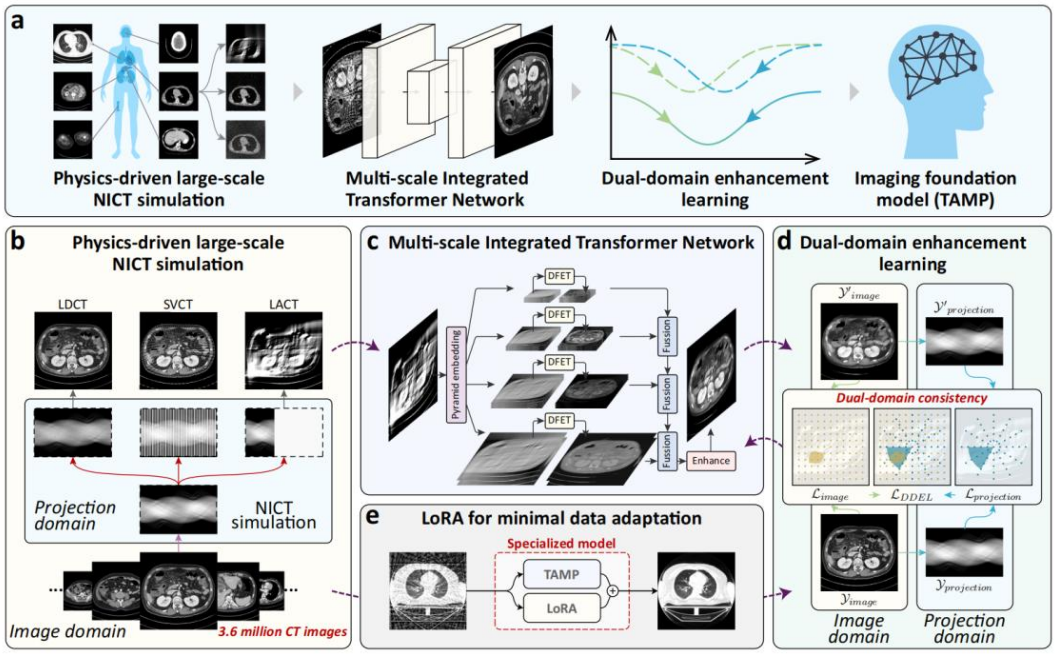

Fig. 6. The framework of our TAMP for universal enhancement of NICT images. a) It simulates large-scale NICT dataset (SimNICT), and trains our multi-scale integrated transformer network (MITNet) via our dual-domain enhancement learning (DDEL), thus achieving our TAMP model for universal NICT enhancement. b) By simulating the defects that meet the physical process of non-ideal measurement, NICT images are synthesized for our SimNICT. c) Our MITNet represents the multi-granularity defect features in the varied NICT data. d) Our DDEL trains the universal NICT enhancement both in image and projection domains. e) A parameter-efficient fine-tuning strategy, low-rank adaptation (LoRA), is employed for efficient adaptation.

3. For the question "Some abbreviations such as PCA, PBN, SQR are used without clear definitions in the main text":

We have provided detailed definitions and calculation formulas for PCA (Probability of Clinical Acceptability), PBN (Probability of Better than NICT), and SQR (Subjective Quality Ranking) directly in the main text, making these radiologist evaluation metrics clear and accessible to readers.

Manuscript changes: We have added the mathematical definitions of PBN, SQR, and PCA with explanations of key symbols in Section 2.4 (Radiologist validation) of the main text, so that readers can understand these metrics.

"These metrics are defined as: $PBN(x) = \frac{\sum_{s\in S_{NICT_x,NICT}}[R^s_{NICT_x} > R^s_{NICT}]}{|S_{NICT_x,NICT}|}$, $SQR(x) = \frac{\sum_{s\in S} R^s_{NICT_x}}{|S_{NICT_x}|}$, $PCA(x) = \frac{\sum_{s\in S}[A^s_{NICT_x}]}{|S_{NICT_x}|}$, where x denotes a model, $NICT_x$ represents the enhanced NICT image by model x, $S_{NICT_x,NICT}$ denotes the set of groups where both $NICT_x$ and NICT are selected for scoring, $R^s_{NICT_x}$ and R^s_{NICT} represent the subjective quality ratings, $A^s_{NICT_x}$ represents the clinical acceptability given by radiologists, and $[\cdot]$ denotes the Iverson bracket."

Comment#7: Remarks on code availability

The repo appears well-organized and clearly documented. Due to time constraints, I have not attempted to run the code and the model.

However, several key components necessary for full reproducibility are currently missing:

- Access to the pretraining synthetic dataset (SimNICT). Only a small subset is accessible.
- Code for running the pretraining pipeline
- Evaluation data splits for synthetic benchmarks

Response#7: We sincerely thank you for recognizing that our repo is "well-organized and clearly documented" and for pointing out these important reproducibility concerns, which will help ensure complete accessibility and reproducibility of our work. All three previously unavailable components have been provided to ensure full reproducibility of our results.

1. For the question "Access to the pretraining synthetic dataset (SimNICT). Only a small subset is accessible":

   We have released the complete SimNICT pretraining dataset at HuggingFace (https://huggingface.co/datasets/YutingHe-list/SimNICT) for full reproducibility. As shown in Table C, the dataset release includes:

Table C: SimNICT pretraining dataset release.

Component	Content	Platform
Preprocessed ICT data	8 datasets (7,742 volumes, ~823 GB)	Internet Archive
Batch download scripts	Automated download tools	HuggingFace
NICT simulation code	Generate LDCT, SVCT, LACT with ICT	HuggingFace

   The preprocessed ICT data is hosted on Internet Archive (https://archive.org/search?query=creator%3A%22TAMP+Research+Group%22) with batch download scripts provided. Researchers can download all datasets or select specific ones using the provided script. The two unreleased datasets (AutoPET and HECKTOR22) are restricted by their original licensing terms. The physics-driven simulation code implements the complete NICT generation pipeline, enabling researchers to generate LDCT, SVCT, and LACT data with randomized degradation parameters and reproduce the exact 10.8M image pairs used in our pretraining.

   Manuscript changes: We have added the complete dataset release information with specific hosting platforms and access URLs in the Data Availability section, so that researchers can readily access and reproduce our work.

   "Specifically, 8 out of 10 datasets containing 7,742 CT volumes (~823 GB) have been released on HuggingFace, with data hosted on Internet Archive. Unfortunately, 2 out of 10 datasets (AutoPET and HECKTOR22) are not released due to restrictions imposed by their original licensing terms. However, our NICT simulation code is publicly released to enable researchers to generate complete and extended SimNICT datasets with their own ICT data."

2. For the question "Code for running the pretraining pipeline":

   We have released the pretraining code at GitHub (https://github.com/YutingHe-list/TAMP/blob/main/utils/pretrain.py), so that researchers can readily reproduce our pretraining.

3. For the question "Evaluation data splits for synthetic benchmarks":

We have provided detailed evaluation data splits for all 27 synthetic benchmark tasks, including clear documentation of which data is used for fine-tuning versus testing. The splits are reproducible and ensure fair comparison with our reported results.

Manuscript changes: We have added detailed information about synthetic evaluation datasets in the Data Availability section. Specifically:

"Synthetic evaluation datasets: Complete evaluation data with training/testing splits for all 27 benchmark tasks are available on HuggingFace."

==========

Comments from reviewer 2:

Comment#1: This paper proposes TAMP, an imaging foundation model for universal enhancement of non-ideal measurement CT (NICT), including low-dose CT (LDCT), sparse-view CT (SVCT), and limited-angle CT (LACT). The authors construct a large physics-driven simulated dataset (SimNICT, 10.8M paired images), introduce a multi-scale transformer with dual-domain learning (image + projection), and apply parameter-efficient fine-tuning (LoRA) for rapid adaptation with few paired slices.

Response#1: We sincerely thank you for this clear summary of our work, particularly recognizing the large physics-driven simulated dataset, multi-scale transformer with dual-domain learning, and LoRA adaptation.

Our TAMP as a foundation model enables development of NICT enhancement models through parameter-efficient fine-tuning with minimal data, reducing development cycles from months to days. This capability addresses critical deployment needs including scenario-specific protocol adaptation, accelerated model availability for novel imaging devices, and efficient updates across multi-institutional protocol variations. The efficiently developed models advance clinical practice across diverse NICT modalities: LDCT models enable broader implementation of dose-reduction strategies in routine diagnostic workflows, while SVCT and LACT models address treatment CT applications where geometric constraints necessitate sparse-angle and limited-angle acquisitions, including cone-beam CT for intraoperative guidance and interventional imaging for real-time monitoring during minimally invasive procedures.

In this version, we have further improved our paper according to your insightful suggestions, particularly strengthening the motivation for universal enhancement approaches and clarifying the methodological contributions of our framework.

Comment#2: The Introduction argues that specialized models are costly to develop and fragile to protocol changes, motivating a universal FM across LDCT, SVCT, and LACT. However, in practice each scan uses a single, fixed NICT protocol, and this protocol is always known in advance. The manuscript should better clarify in what real clinical scenarios a universal enhancer is preferable to simply deploying the appropriate specialized model.

Response#2: Thank you for pointing out this potential misunderstanding, which will help clarify the target of our foundation model. 1) Our target is not to develop a single model for direct application across diverse real clinical scenarios, but to provide a foundation for the emergence [1,2] of specialized models with low-cost and high-performance. 2) Our foundation model approach is preferable to simply deploying the appropriate specialized model, by enabling rapid validation of emerging NICT protocols and facilitating low-cost development of high-performance models. Specifically:

1. **Target: Foundation for specialized model development**

   Our target is not to develop a single model for direct application across diverse real clinical scenarios, but to provide a foundation for the emergence [1,2] of specialized models with low-cost and high-performance. Because each clinical scan uses a single, fixed NICT protocol, this necessitates extensive protocol-specific model development efforts. Specifically, these efforts includes extensive paired data collection (typically

thousands of images), iterative neural network architecture design, and training from scratch, resulting in prolonged development cycles that delay clinical deployment. However, our foundation model approach provides a model development foundation through rich prior NICT knowledge learned from large-scale pretraining. It enables low-cost development of high-performance specialized models with only dozens of images, outperforming those trained from scratch while dramatically reducing development time and resource consumption.

2. **Advantages: Rapid validation and low-cost high-performance model development**

Our foundation model approach is preferable to simply deploying the appropriate specialized model, by enabling rapid validation of emerging NICT protocols and facilitating low-cost development of high-performance models.

a) **Rapid validation of emerging NICT protocols:** Our foundation model enables training-free enhancement of data from newly emerged NICT devices or protocols, allowing immediate assessment of data quality and feasibility for clinical deployment without any model development effort. This capability provides critical guidance for evaluating the potential of new protocols and determining optimal model development strategies.

b) **Low-cost and high-performance specialized model development:** Through parameter-efficient fine-tuning with only dozens of images, our foundation model enables rapid development of specialized models that outperform those trained from scratch. The rich prior knowledge from large-scale pretraining across diverse NICT scenarios enables effective disentanglement of artifact patterns from anatomical structures, achieving superior enhancement quality while dramatically reducing development time and resource consumption.

Manuscript changes: We have revised the Introduction section to elaborate on how foundation models address NICT enhancement challenges through universal enhancement and efficient adaptation, thereby clarifying the advantages of FM approach in rapid protocol validation and low-cost model development. Specifically:

"Foundation models (FMs) offer potential to address these challenges through two key capabilities: a) Universal enhancement enables rapid validation of emerging NICT protocols. FMs' generalization across diverse NICT settings supports training-free quality enhancement for a broad range of protocols, enabling rapid assessment of new device potential for clinical deployment. b) Efficient adaptation enables low-cost development of high-performance specialized models. FMs' transferable prior knowledge improves the emergence of specialized models with limited fine-tuning data, enabling low-cost deployment of models that surpass training-from-scratch approaches."

[1] Wei J, Tay Y, Bommasani R, et al. Emergent Abilities of Large Language Models. Transactions on Machine Learning Research. 2022. doi:10.48550/arXiv.2206.07682.
[2] Moor M, Banerjee O, Abad ZFH, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259-265. doi:10.1038/s41586-023-05881-4.

Comment#3: All experiments (e.g., Fig. 2) are conducted under fixed, single-scenario settings (LDCT-only, SVCT-only, LACT-only) using datasets where NICT is simulated from ICT. This

mirrors conditions where specialized models already suffice and does not by itself prove universality. Moreover, since the test data are also simulated, it remains unclear how the model performs on measured public datasets. it would be valuable to evaluate on real NICT test data, such as the AAPM-Mayo LDCT dataset or other publicly available measured SVCT/LACT datasets, rather than only simulated counterparts.

Response#3: We sincerely thank you for this insightful suggestion, which will help prove the universality of our foundation model and validate its performance on publicly abailable measured dataset. In this version, we have 1) clarified the conditions where our foundation model demonstrates universality, and 2) supplemented evaluation on AAPM-Mayo LDCT dataset, which is a publicly abailable measured dataset.

1. For the question "All experiments (e.g., Fig. 2) are conducted under fixed, single-scenario settings (LDCT-only, SVCT-only, LACT-only) using datasets where NICT is simulated from ICT. This mirrors conditions where specialized models already suffice and does not by itself prove universality.":

   Our foundation model approach provides a foundation for developing specialized NICT enhancement models, thereby promoting the exploration and adoption of emerging NICT devices and protocols. This foundation demonstrates two key capabilities: 1) direct enhancement without any training enabling rapid validation of emerging NICT protocols, and 2) providing a foundation for low-cost high-performance specialized model development for specific scenarios. Specifically:

   a) **Direct enhancement without any training enabling rapid validation of emerging NICT protocols**

      Our foundation model enables direct enhancement of data from newly emerged NICT devices or protocols without any training, through rich prior knowledge learned from large-scale pretraining, allowing immediate assessment of data quality and feasibility for clinical deployment without any model development effort. This capability provides critical guidance for evaluating the potential of new protocols and determining optimal model development strategies. In all 27 benchmark tasks, our model achieved comparable performance to specialized models without requiring protocol-specific training data, demonstrating effective rapid validation capability.

   b) **Provide a foundation for low-cost high-performance specialized model development for specific scenarios**

      Through parameter-efficient fine-tuning with only dozens of images, our foundation model enables rapid development of specialized models that outperform those trained from scratch. The rich prior knowledge from large-scale pretraining across diverse NICT scenarios enables effective disentanglement of artifact patterns from anatomical structures, achieving superior enhancement quality while dramatically reducing development time and resource consumption. Our evaluations demonstrate that specialized models fine-tuned with only 5 slices outperform those trained from scratch on 20 volumes, achieving state-of-the-art performance across all 27 diverse NICT scenarios.

   Manuscript changes: We have revised the Introduction section to elaborate on foundation models' universal enhancement and efficient adaptation capabilities for addressing NICT challenges, thereby proving the universality of our FM approach. Specifically:

"Foundation models (FMs) offer potential to address these challenges through two key capabilities: a) Universal enhancement enables rapid validation of emerging NICT protocols. FMs' generalization across diverse NICT settings supports training-free quality enhancement for a broad range of protocols, enabling rapid assessment of new device potential for clinical deployment. b) Efficient adaptation enables low-cost development of high-performance specialized models. FMs' transferable prior knowledge improves the emergence of specialized models with limited fine-tuning data, enabling low-cost deployment of models that surpass training-from-scratch approaches."

2. For the question "Moreover, since the test data are also simulated, it remains unclear how the model performs on measured public datasets. it would be valuable to evaluate on real NICT test data, such as the AAPM-Mayo LDCT dataset or other publicly available measured SVCT/LACT datasets, rather than only simulated counterparts.":

In this version, we have supplemented comprehensive evaluation on publicly available measured NICT datasets across multiple real-world scenarios, including the Mayo Clinic Low Dose CT Grand Challenge dataset, to demonstrate TAMP's performance on measured public datasets beyond simulated test data.

a) **Experiment settings:** Building upon the original real-world validation datasets (NJDTH-A for LDCT and NJDTH-C for SVCT/LACT), we supplemented two additional datasets including the publicly available Mayo dataset to validate TAMP across multiple real-world scenarios:

   i. **NJDTH-B (LDCT):** We collected an additional LDCT dataset with 10 cases (2,802 slices) from Nanjing Drum Tower Hospital, acquired using a current-reduction-only protocol. This dataset is split into 5 training cases (1,527 slices) and 5 test cases (1,275 slices). Combined with the original NJDTH-A dataset, we now have a total of 16 LDCT cases for real-world LDCT validation across multiple clinical protocols.

   ii. **Mayo (SVCT and LACT):** We collected SVCT and LACT data with 10 cases from the publicly available Mayo Clinic Low Dose CT Grand Challenge projection data by undersampling at sparse views and limited angles, split into 5 training cases (990 slices) and 5 test cases (737 slices). This publicly accessible dataset enables independent validation and reproducibility verification by the research community. Combined with the original NJDTH-C dataset, we now have a total of 16 cases for SVCT/LACT validation, providing cross-institutional validation across different CT scanners and reconstruction protocols.

   Together with the original datasets, these four datasets (NJDTH-A, NJDTH-B, NJDTH-C, Mayo) constitute 6 real-world evaluation tasks covering multiple NICT scenarios. Baseline methods (RED-CNN, FBPConvNet, ProCT) were trained from scratch on each training set, TAMP was evaluated zero-shot, and TAMP-S was adapted via parameter-efficient fine-tuning.

b) **Experiment results:** TAMP and TAMP-S effectively enhance the quality of real-world NICT images across multiple scenarios, based on two observations:

   i. **TAMP demonstrates robust simulation-to-reality generalization across multiple real-world NICT scenarios without additional training.**

   Quantitatively, TAMP achieves competitive zero-shot performance across

all 6 real-world evaluation tasks. In LDCT tasks, TAMP achieves median LPIPS of 16.54% on NJDTH-B, substantially outperforming compared methods. For geometric artifact tasks on the publicly available Mayo dataset, TAMP demonstrates effective zero-shot transfer with median PSNR of 30.61 dB on LACT and median SSIM of 63.37% on SVCT, validating its learned universal representations transfer to publicly accessible measured data. Qualitatively, TAMP enhances LDCT images by not only suppressing photon noise, but also producing clearly distinguishable edges of anatomical structures, resulting in smoother textures and enhanced structural clarity. For LACT images, TAMP reconstructed severely damaged chest edges and soft tissues where the original structures are difficult to infer from surrounding information.

ii. **After parameter-efficient adaptation, TAMP-S achieves state-of-the-art performance across all real-world NICT tasks from multiple clinical scenarios.**

Quantitatively, TAMP-S achieves the best median performance in all 24 metric-task combinations (6 tasks × 4 metrics). For the newly added NJDTH-B dataset, TAMP-S achieves PSNR of 30.09 dB and SSIM of 80.03%, outperforming all baselines trained from scratch. For the publicly available Mayo dataset, TAMP-S demonstrates substantial improvements on LACT tasks with median PSNR of 35.00 dB (+3.26 dB vs FBPConvNet) and on SVCT tasks with median SSIM of 77.97% (+1.39% vs ProCT). These consistent improvements across multiple scenarios validate TAMP's effective parameter-efficient adaptation capability for real-world NICT enhancement. Qualitatively, TAMP-S retains the visual advantages of TAMP while demonstrating more task-specific denoising and more accurate reconstruction of real-world NICT images across diverse clinical scenarios.

c) **Manuscript changes:** We have substantially revised the real-world validation section (Sec.2.4 in the main text and corresponding sections in Supplementary Materials) to incorporate the two newly added datasets (NJDTH-B and Mayo), thereby providing stronger and more statistically reliable evidence of sim-to-real generalization across diverse clinical scenarios and imaging protocols. Specifically:

"To evaluate TAMP's real-world enhancement capability, we collected real-world NICT data from diverse clinical scenarios and imaging protocols. The datasets include: 1) NJDTH-A: LDCT data from Nanjing Drum Tower Hospital Jiangbei (NJDTH) with 6 cases (1,496 slices)...; 2) NJDTH-B: LDCT data from NJDTH with 10 cases (2,802 slices) acquired at low dose (120 kVp, 50 mAs) and high dose (120 kVp, 200 mAs)...; 3) NJDTH-C: SVCT and LACT data from NJDTH with 6 cases (750 slices)...; 4) Mayo: SVCT and LACT data with 10 cases reconstructed from publicly available projection data of Mayo Clinic Low Dose CT Grand Challenge... These four datasets constitute 6 evaluation tasks."

"Quantitatively, TAMP achieves competitive zero-shot performance across all real-world NICT tasks. In LDCT tasks, TAMP achieves median LPIPS scores of 7.65% (NJDTH-A) and 16.54% (NJDTH-B), substantially outperforming compared methods. For geometric artifact tasks, TAMP demonstrates effective zero-shot

transfer on LACT (median PSNR: 31.53 dB on NJDTH-C, 30.61 dB on Mayo) and SVCT (median SSIM: 94.35% on NJDTH-C, 63.37% on Mayo), validating its learned universal representations."

"TAMP-S achieves the best median performance in all 24 metric-task combinations (6 tasks × 4 metrics). Notably, for challenging geometric artifact reconstruction, TAMP-S demonstrates substantial improvements on LACT tasks with median PSNR of 37.01 dB (NJDTH-C, +0.26 dB vs ProCT) and 35.00 dB (Mayo, +3.26 dB vs FBPConvNet), and on SVCT tasks with median SSIM of 96.57% (NJDTH-C, +0.73% vs ProCT) and 77.97% (Mayo, +1.39% vs ProCT)."

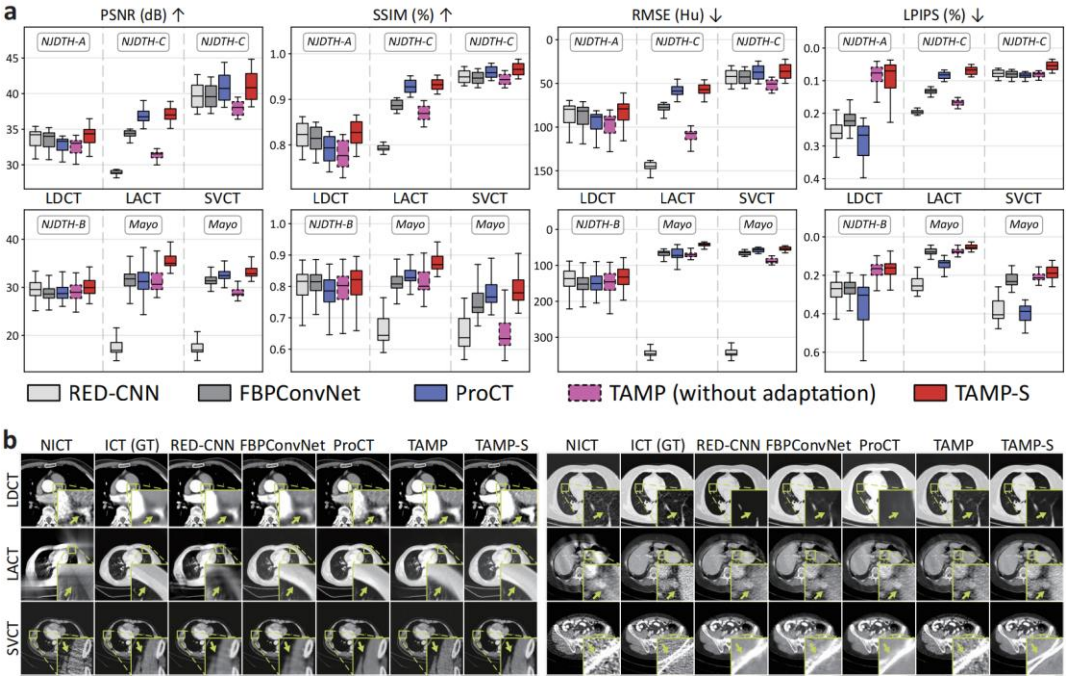

Fig. 4 Real-world validation demonstrates our great application potential in real-world NICT images. a) Our TAMP can directly enhance real-world NICT images and TAMP-S is adapted to specific NICT settings, achieving further improvement. b) Our TAMP has significant visual enhancement demonstrating its great clinical application potential in real-world NICT images.

Comment#4: TAMP is pretrained on the very large SimNICT dataset, whereas baselines are trained only on smaller task-specific datasets. Since pretraining is also training, much of the reported gain may come from data scale rather than architectural novelty. Ablation is needed to disentangle the contributions of pretraining, the multi-scale transformer backbone, and the dual-domain loss.

Response#4: We sincerely thank you for this insightful suggestion, which will help disentangle the contributions of different components. In this version, we have added ablation studies on the pretraining, the multi-scale transformer backbone (MITNet), and the dual-domain loss (DDEL). Specifically:

1. **Experiment settings:** Experiment settings: We conducted ablation studies by progressively incorporating each component (Table 11). Starting from the baseline RED-

CNN (single-scale convolutional network with MSE loss only), we first replaced the network with our MITNet. We then sequentially added the loss components of DDEL, beginning with SSIM loss, followed by VGG loss, and finally projection domain loss. After training all variants from scratch, we evaluated the effect of large-scale pretraining by initializing model weights with pre-trained TAMP parameters and applying LoRA-based fine-tuning. All experiments were conducted on three representative NICT tasks (APET-LDCT-High, APET-LACT-High, APET-SVCT-High) from the AutoPET dataset.

2. **Experiment results:** As shown in Table 11, each component contributes substantially to the final performance through progressive improvements. The baseline RED-CNN with MSE loss achieves 45.63, 37.51, and 43.96 dB PSNR for the three tasks respectively. Replacing RED-CNN with MITNet brings 0.04, 2.02, and 0.10 dB improvements, with the most significant gain observed in LACT where multi-scale features are critical for handling complex geometric artifacts. Adding SSIM loss provides the largest single improvement of 2.79, 2.46, and 3.95 dB, validating that structural similarity constraints are crucial for all NICT types. Subsequently incorporating VGG loss yields modest but consistent gains of 0.16, 0.25, and 0.28 dB for perceptual quality enhancement. The projection domain loss further boosts performance by 0.85, 2.03, and 0.50 dB, with particularly strong contribution to LACT reconstruction due to its geometric nature that benefits from physics-driven constraints. Finally, initializing with pre-trained TAMP weights and applying LoRA fine-tuning achieves 51.62, 45.68, and 50.26 dB, representing 2.15, 1.41, and 1.47 dB improvements over training from scratch with all components. This consistent 1.4-2.2 dB improvement across all tasks demonstrates the substantial benefits of large-scale physics-driven pretraining beyond architectural and loss function designs, validating that TAMP's superior performance results from the synergistic integration of all three innovations rather than data scale alone.

Table 11: Ablation study of TAMP. The first two rows compare the networks of RED-CNN and MITNet. The subsequent rows illustrate the impact of the loss components in DDEL. The final rows demonstrate the effect of initializing network weights with pre-trained TAMP weights. PSNR is used as the evaluation metric.

Pre-train	**MITNet**	**DDEL**				**Task (PSNR)**		
		Proj	VGG	SSIM	MSE	LDCT	LACT	SVCT
					√	45.63±1.20	37.51±0.80	43.96±1.13
	√				√	45.67±0.97	39.53±1.11	44.06±0.95
	√			√	√	48.46±1.11	41.99±1.04	48.01±1.04
	√		√	√	√	48.62±0.95	42.24±1.18	48.29±0.96
	√	√	√	√	√	49.47±0.94	44.27±1.01	48.79±1.02
√	√	√	√	√	√	51.62±1.07	45.68±0.98	50.26±1.04

3. **Manuscript changes:** We have conducted ablation studies and added comprehensive analysis in the Supplementary Materials to clearly demonstrate each component's individual contribution. Specifically:

"To verify the contributions of the MITNet architecture, model pretraining, DDEL training strategy, and the four loss functions in various NICT enhancement tasks, we conducted an ablation study, with the results shown in Table 11. The ablation study

systematically evaluates three key components:

Table 11: Ablation study of TAMP. The first two rows compare the networks of RED-CNN and MITNet. The subsequent rows illustrate the impact of the loss components in DDEL. The final rows demonstrate the effect of initializing network weights with pre-trained TAMP weights. PSNR is used as the evaluation metric.

Pre-train	MITNet	DDEL				Task		
		Proj loss	VGG loss	SSIM loss	MSE loss	APET-LDCT-High	APET-LACT-High	APET-SVCT-High
					✓	45.63±1.20	37.51±0.80	43.96±1.13
	✓				✓	45.67±0.97	39.53±1.11	44.06±0.95
	✓			✓	✓	48.46±1.11	41.99±1.04	48.01±1.04
	✓		✓	✓	✓	48.62±0.95	42.24±1.18	48.29±0.96
	✓	✓	✓	✓	✓	49.47±0.94	44.27±1.01	48.79±1.02
✓	✓	✓	✓	✓	✓	51.62±1.07	45.68±0.98	50.26±1.04

1) MITNet architecture contribution: By replacing RED-CNN's single-scale channel residual convolution network with our multi-scale integrated transformer architecture (MITNet), the model shows improvements across all three tasks, with particularly notable enhancement in LACT task (2.02 dB PSNR improvement from 37.51 to 39.53 dB), moderate improvement in SVCT task (0.10 dB improvement from 43.96 to 44.06 dB), and minimal improvement in LDCT task (0.04 dB improvement from 45.63 to 45.67 dB). This demonstrates that the multi-scale transformer architecture is especially effective for complex geometric artifacts (LACT) where large-scale features are critical, while providing consistent but modest improvements for noise-dominated tasks (LDCT and SVCT).

2) DDEL training strategy contribution: By gradually incorporating the four loss components of DDEL, the model achieves substantial PSNR improvements. The largest single improvement comes from adding SSIM loss, which provides 2.79 dB improvement for LDCT (from 45.67 to 48.46 dB), 2.46 dB for LACT (from 39.53 to 41.99 dB), and 3.95 dB for SVCT (from 44.06 to 48.01 dB), validating that structural similarity constraints are crucial for all NICT types. The VGG loss contributes modest but consistent improvements of 0.16 dB for LDCT, 0.25 dB for LACT, and 0.28 dB for SVCT. The projection domain loss provides the final boost with 0.85 dB for LDCT, 2.03 dB for LACT, and 0.50 dB for SVCT, with particularly strong contribution to LACT reconstruction due to its geometric nature.

3) Large-scale pretraining contribution: By initializing the model weights with pre-trained TAMP and using LoRA fine-tuning, we achieved final PSNR values of 51.62, 45.68, and 50.26 dB for the three tasks, representing improvements of 2.15 dB, 1.41 dB, and 1.47 dB respectively over training from scratch. This consistent 1.4-2.2 dB improvement across all tasks confirms the substantial benefits of large-scale physics-driven pretraining for universal NICT enhancement capabilities, with the largest benefit observed for LDCT tasks where the diverse noise patterns in pretraining data provide robust initialization."

Comment#5: The framework combines multi-scale feature extraction, a transformer backbone, and dual-domain loss. These elements have each appeared in prior enhancement methods, and it is not clear what methodological improvement is offered here beyond re-assembly. The paper would benefit from a clearer discussion of how its design advances existing approaches.

Response#5: We sincerely thank you for this insightful suggestion, which will help clarify how

our methodological improvements enable the foundation model paradigm for universal NICT enhancement. Our methodological improvements are reflected in two key designs: **MITNet** (multi-scale feature extraction with a transformer backbone) and **DDEL** (dual-domain loss). Specifically:

1. **MITNet** advances scale-independent multi-scale feature representation through a distinctive parallel-to-serial architecture, enabling unified encoding of fundamentally varied NICT artifacts across diverse protocols. NICT degradations from different protocols manifest at varying scales and shapes. Sequential multi-scale encoding introduces interference from preceding scale features, limiting adequate capture of diverse characteristics. Our parallel-to-serial architecture addresses this through: **1) Parallel multi-scale embedding:** Different patch scales and linear layers independently extract embedding features at multiple pyramid levels in parallel, isolating scale-specific information and enabling comprehensive encoding without cross-scale interference. **2) Serial progressive fusion:** Features are progressively fused from lowest to highest resolution, mimicking the coarse-to-fine refinement process of image reconstruction. This design enables artifact removal and detail reconstruction adapted to different NICT categories, establishing a unified foundation model capable of handling fundamentally varied degradation patterns.
2. **DDEL** advances dual-domain consistency constraints through bridging physical simulation and representation learning, enabling the learning of both anatomical structures and NICT artifact features (Fig. A). Foundation model development requires learning universal NICT representations authentically capturing both real anatomical features and diverse degradation patterns. Our DDEL achieves this through: 1) **Dual-domain constraints:** Image domain employs multi-granularity losses (MSE, SSIM, VGG) to constrain $Y'_{image} \approx Y_{image}$ for anatomical structure preservation, while projection domain employs MSE loss through differentiable forward projection (ODL) to constrain $Y'_{projection} \approx Y_{projection}$ for physical measurement consistency. 2) **Bridging through bidirectional mappings:** By constraining $Y' = f' \circ f(Y) \approx Y$, we bridge the physical simulation f and representation learning f', training f' to invert the forward mapping f. This ensures learned representations are grounded in both anatomical reality and physical degradation processes.

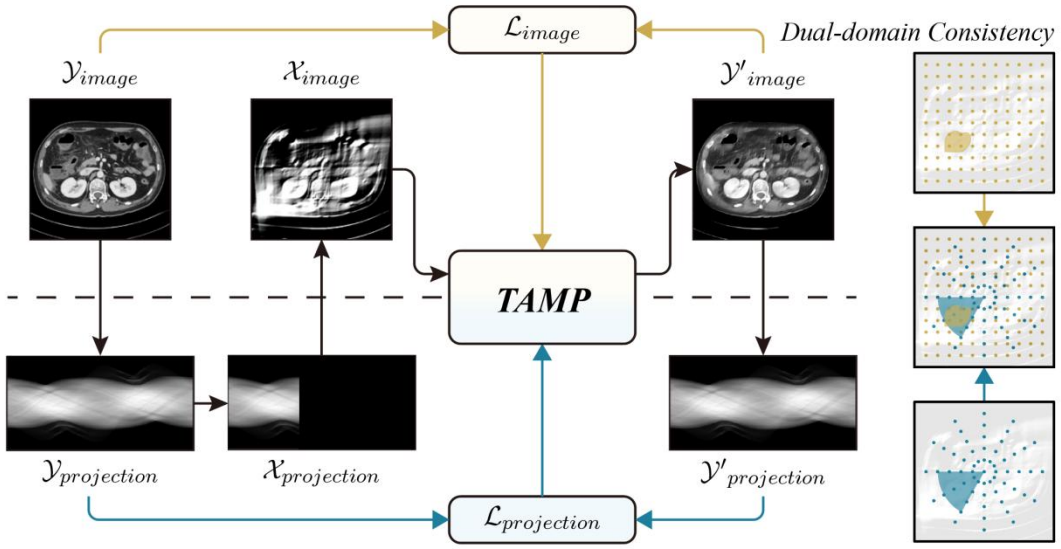

Fig. A. DDEL advances dual-domain consistency constraints through bridging physical simulation and representation learning, enabling the learning of both anatomical structures and NICT artifact features.

Manuscript changes: We have revised the Method section and added a technical discussion paragraph in the Discussion section to clarify how our design advances existing approaches for universal NICT enhancement. Specifically:

"Our MITNet advances scale-independent multi-scale feature representation through a distinctive parallel-to-serial transformer architecture, enabling unified encoding of varied NICT artifacts across diverse protocols. Since NICT degradations from different protocols manifest at varying scales and shapes, sequential multi-scale encoding introduces interference from preceding scale features. Our parallel-to-serial architecture addresses this through two aspects: 1) Parallel multi-scale embedding: Different patch scales and linear layers independently extract embedding features at multiple pyramid levels in parallel, isolating scale-specific information without cross-scale interference. 2) Serial progressive fusion: Features are progressively fused from lowest to highest resolution, mimicking the coarse-to-fine refinement process of CT iterative reconstruction. This design enables artifact removal and detail reconstruction adapted to different NICT categories, establishing a unified foundation model capable of handling varied degradation patterns.

We design a dual-domain enhancement learning (DDEL) that bridges physical simulation and representation learning through dual-domain consistency constraints. Our DDEL achieves this through two aspects: 1) Dual-domain constraints: In the image domain, multi-granularity losses including MSE loss, SSIM loss, and VGG loss learn real anatomical features. In the projection domain, the enhanced and target images are mapped via the Operator Discretization Library for sinogram maps, and the MSE loss learns diverse NICT degradation patterns. 2) Bridging through bidirectional mappings: By constraining $Y' = f' \circ f(Y) \approx Y$, we bridge the physical simulation f (SimNICT) and representation learning f' (MITNet), ensuring learned representations are grounded in both anatomical reality and physical degradation processes."

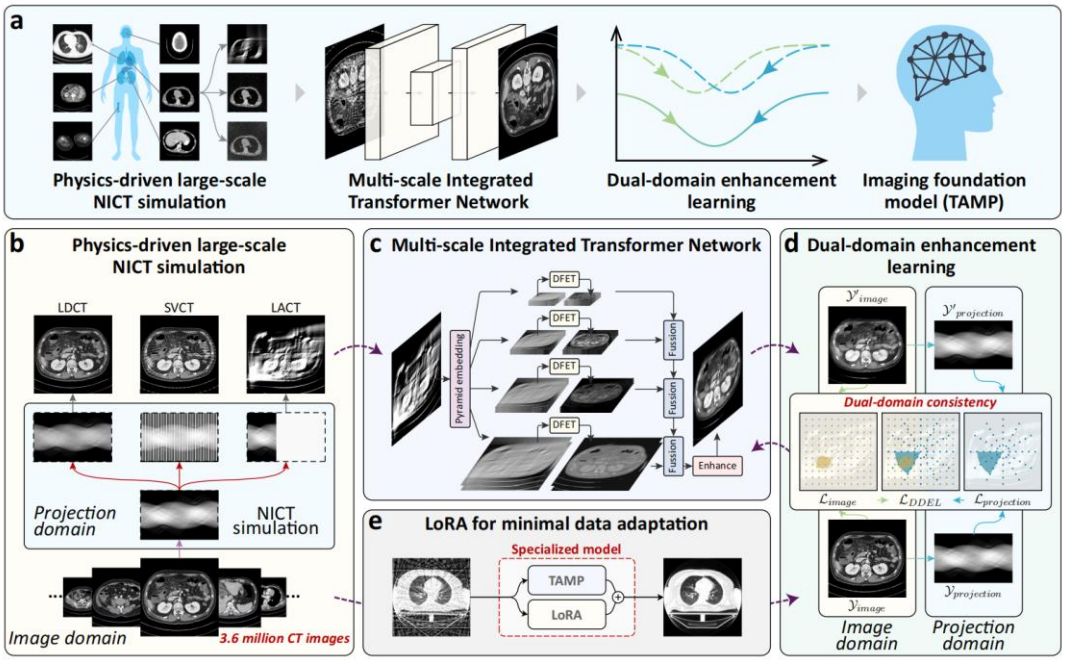

Fig. 6. The framework of our TAMP for universal enhancement of NICT images. a) It simulates large-scale NICT dataset (SimNICT), and trains our multi-scale integrated transformer network (MITNet) via our dual-domain enhancement learning (DDEL), thus achieving our TAMP model for universal NICT enhancement. b) By simulating the defects that meet the physical process of non-ideal measurement, NICT images are synthesized for our SimNICT. c) Our MITNet represents the multi-granularity defect features in the varied NICT data. d) Our DDEL trains the universal NICT enhancement both in image and projection domains. e) A parameter-efficient fine-tuning strategy, low-rank adaptation (LoRA), is employed for efficient adaptation.

"Technically, we designed MITNet and DDEL to encode diverse artifact patterns and learn universal representations for large-scale NICT pretraining. For network architecture, since NICT artifacts manifest at varied scales across protocols (noise in LDCT, streaks in SVCT, blurring in LACT), we designed scale-specific disentanglement and progressive integration mechanisms. MITNet employs parallel-to-serial transformers where parallel multi-scale embedding extracts scale-specific features, while serial progressive fusion enables coarse-to-fine refinement mimicking CT iterative reconstruction. For pretraining paradigm, foundation models require learning universal representations that capture both anatomical structures and degradation patterns. DDEL bridges image-domain structure learning and projection-domain physics-grounded degradation learning via dual-domain consistency, enabling the model to represent NICT degradation inversion. These designs establish the foundation for universal NICT enhancement beyond single-protocol models."

Comment#6: Although the universal model motivation is emphasized, the evaluations again remain limited to single-scenario settings. This essentially reproduces conditions where specialized models are sufficient. To substantiate universality, the authors should provide pooled or mixed test cohorts (with LDCT, SVCT, and LACT cases together) or comparisons

against an oracle ensemble of specialized models.

Response#6: We sincerely thank you for this insightful suggestion, which will help further demonstrate our foundation model's universal enhancement capabilities. In this version, we have conducted **1) mixed test cohort evaluation** and **2) oracle ensemble comparison** to validate the superior performance of our foundation model in multi-scenario deployment settings that represent real-world clinical practice.

1. **Mixed test cohort evaluation:** To validate our model's universal enhancement capability across diverse NICT scenarios simultaneously, we have conducted mixed test cohort evaluation where LDCT, SVCT, and LACT images are pooled together for comprehensive assessment.
    a) **Experiment settings:** We constructed a mixed test cohort by pooling all 27 diverse NICT enhancement scenarios (with 3 datasets, 3 NICT types, and 3 degradation degrees) from the main manuscript test sets. For baseline methods (RED-CNN, FBPConvNet, ProCT), we trained 27 specialized models, each targeting one specific scenario. To simulate real-world deployment where scene information is unavailable, we evaluated each test scenario by applying all 27 specialized models and averaging their performance metrics. In contrast, TAMP employs a single universal model and directly tested across all 27 mixed scenarios without adaptation, representing the zero-shot generalization capability of our foundation model.

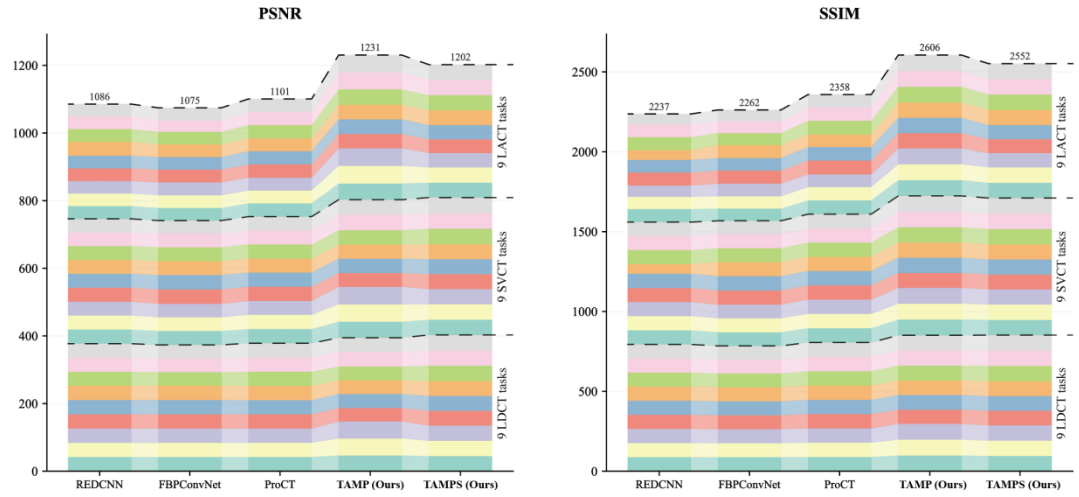

Fig. 21. Stacked bar chart showing PSNR and SSIM performance across 27 NICT enhancement scenarios. Each bar represents one method, with 27 segments stacked to show individual scenario contributions. TAMP (single universal model) achieves the highest cumulative performance without requiring scene-specific adaptation.

    b) **Experiment results:** TAMP demonstrates superior zero-shot generalization across diverse NICT scenarios, validating the foundation model's universal enhancement capability with a single model. As shown in Fig. 21, TAMP achieves cumulative PSNR of 1230.56 dB and SSIM of 2605.66 across all 27 scenarios, outperforming the best baseline ProCT by 11.79% and 10.48% respectively. The consistent performance superiority across heterogeneous scenarios demonstrates that TAMP's physics-driven pretraining learns universal representations that effectively capture diverse artifact

patterns, enabling robust generalization without requiring specialized models for each scenario combination.

c) **Manuscript changes:** We have added the mixed test cohort evaluation in subsection "Ablation studies" of the Supplementary Materials to substantiate the universality of our foundation model, through evaluating TAMP's zero-shot generalization across 27 pooled NICT scenarios. Specifically:

"We constructed a mixed test cohort by pooling all 27 diverse NICT enhancement scenarios. TAMP demonstrates superior zero-shot generalization across diverse NICT scenarios, validating the foundation model's universal enhancement capability with a single model. TAMP achieves cumulative PSNR of 1230.56 dB and SSIM of 2605.66 across all 27 scenarios, outperforming the best baseline ProCT by 11.79% and 10.48% respectively. The consistent performance superiority across heterogeneous scenarios demonstrates that TAMP's physics-driven pretraining learns universal representations that effectively capture diverse artifact patterns."

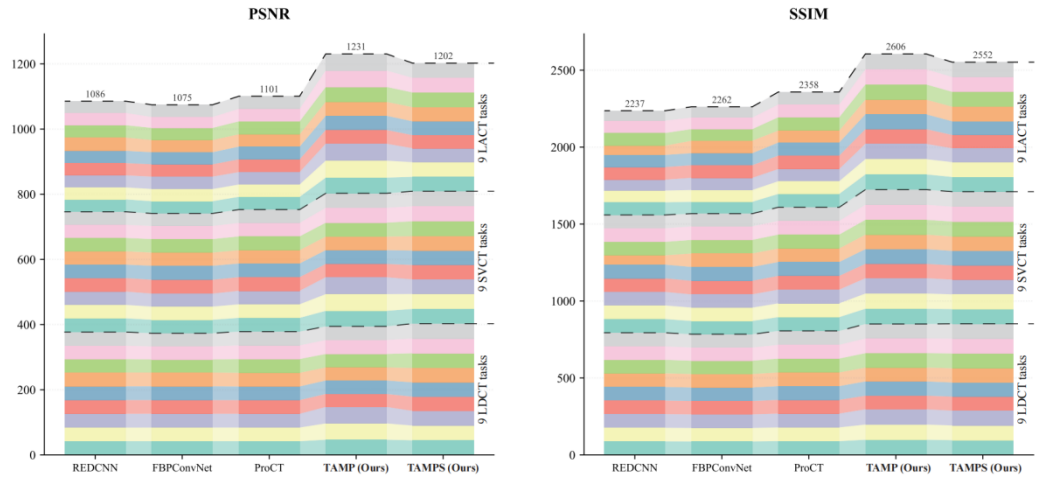

Fig. 21: Stacked bar chart showing PSNR and SSIM performance across 27 NICT enhancement scenarios. Each bar represents one method, with 27 segments stacked to show individual scenario contributions. TAMP (single universal model) achieves the highest cumulative performance without requiring scenespecific adaptation.

2. **Oracle ensemble comparison:** To validate our model's performance against the theoretical upper bound of specialized model approaches, we have conducted comparison against an oracle ensemble of specialized models.

a) **Experiment settings:** To compare TAMP against the theoretical upper bound of specialized model approaches, we constructed an oracle ensemble evaluation on the same 27 NICT enhancement scenarios. The oracle ensemble assumes perfect prior knowledge: for each test image, an oracle knows its exact (dataset, NICT type, degradation degree) combination and selects the corresponding perfectly-matched specialized model from the 27 trained models. This represents the theoretical performance upper bound of specialized approaches, which is unattainable in real-world deployment where scene information is typically unavailable. For baseline methods (RED-CNN, FBPConvNet, ProCT), we evaluated their oracle ensemble

performance under this ideal assumption. In contrast, TAMP employs a single universal model without any prior scene knowledge, directly tested across all scenarios.

b) **Experiment results:** TAMP achieves superior performance compared to oracle ensembles with perfect scene knowledge, validating the effectiveness of universal representations over specialized models. As shown in Table 10, TAMP achieves PSNR of 46.16 dB and SSIM of 97.35%, outperforming the best oracle baseline (FBPConvNet) by +1.67 dB and +2.81% respectively. The superior performance demonstrates that TAMP's physics-driven pretraining on 10.8M diverse SimNICT samples learns more generalizable enhancement patterns than scene-specific features from limited data, while achieving 25-61% lower standard deviations across scenarios for enhanced robustness.

Table 10: Oracle ensemble comparison: TAMP vs perfectly-matched specialized models.

Method	**PSNR (dB)↑**	**SSIM (%)↑**	**RMSE (Hu)↓**	**LPIPS (%)↑**
Oracle RED-CNN	43.90±4.80	91.75±11.20	30.33±16.51	9.73±9.58
Oracle FBPConvNet	44.49±5.02	94.54±5.95	28.64±16.91	7.86±5.56
Oracle ProCT	42.71±4.40	91.45±10.90	33.96±16.99	13.42±10.59
TAMP	**46.16±3.30**	**97.35±2.33**	**21.67±8.11**	**4.97±3.63**

c) **Manuscript changes:** We have added the oracle ensemble comparison in subsection "Ablation studies" of the Supplementary Materials to substantiate the universality of our foundation model, through comparing TAMP against the theoretical upper bound of perfectly-matched specialized models. Specifically:

"To compare TAMP against the theoretical upper bound of specialized model approaches, we constructed an oracle ensemble evaluation. The oracle ensemble assumes perfect prior knowledge: for each test image, an oracle knows its exact (dataset, NICT type, degradation degree) combination and selects the corresponding perfectly-matched specialized model from the 27 trained models. This represents the theoretical performance upper bound of specialized approaches, which is unattainable in real-world deployment where scene information is typically unavailable. As shown in Table 10, TAMP achieves PSNR of 46.16 dB and SSIM of 97.35%, outperforming the best oracle baseline (FBPConvNet) by +1.67 dB and +2.81% respectively. The superior performance demonstrates that TAMP's physics-driven pretraining on 10.8M diverse SimNICT samples learns more generalizable enhancement patterns than scene-specific features from limited data, while achieving 25-61% lower standard deviations across scenarios for enhanced robustness."

Table 10: Oracle ensemble comparison: TAMP vs perfectly-matched specialized models.

Method	PSNR (dB) ↑	SSIM (%) ↑	RMSE (HU) ↓	LPIPS (%) ↓
Oracle RED-CNN	43.90±4.80	91.75±11.20	30.33±16.51	9.73±9.58
Oracle FBPConvNet	44.49±5.02	94.54±5.95	28.64±16.91	7.86±5.56
Oracle ProCT	42.71±4.40	91.45±10.90	33.96±16.99	13.42±10.59
TAMP	**46.16±3.30**	**97.35±2.33**	**21.67±8.11**	**4.97±3.63**

Comments from reviewer 3:

Comment#1: In this manuscript, the authors propose a foundation model to enhance CT images affected by either high noise levels, limited-angle artifacts, or sparse-view artifacts. My detailed comments are as follows:

Response#1: We sincerely thank you for this concise and accurate summary of our work, which captures the core objective of our TAMP foundation model approach.

Our TAMP addresses NICT enhancement across axial CT imaging modalities including clinical CT and cone-beam CT through a unified foundation model framework. For clinical CT, dose reduction protocols (LDCT) produce noise artifacts through reduced tube current or voltage. For cone-beam CT, sparse-view or limited-angle acquisitions (SVCT and LACT) produce geometric artifacts due to angular sampling constraints in image-guided therapy. The foundation model enables effective enhancement across these axial CT modalities through learning universal anatomical and degradation priors via physics-driven simulation, achieving adaptation despite differences in acquisition systems and imaging protocols.

In this version, we have further clarified the clinical motivation and intended applications across diverse CT scenarios, demonstrating the practical advantages of our universal foundation model approach over traditional specialized methods.

Comment#2: My detailed comments are as follows: The manuscript lacks a clear motivation and an intended clinical application. In particular, the authors confuse several x-ray–based imaging modalities such as clinical CT, cone-beam CT, and tomosynthesis, as detailed in the following points.

Response#2: We sincerely thank you for these important concerns, which will help clarify the clinical motivation, intended applications, and applicable modalities of our TAMP framework.

1. For the question "My detailed comments are as follows: The manuscript lacks a clear motivation and an intended clinical application":
   a) **Clarify motivation:** We provide a foundation model to enable low-cost high-performance development of specialized models for emerging NICT enhancement scenarios. With the advancement of medical imaging technology, new NICT devices and protocols are continuously emerging to meet diverse clinical requirements. However, specialized NICT enhancement models face development challenges, requiring extensive data collection, architecture design, and training from scratch for each new protocol, which substantially delays clinical deployment. Foundation models have demonstrated potential to rapidly develop high-performance specialized models through parameter-efficient fine-tuning with minimal data, offering a promising solution to address this deployment bottleneck.
   b) **Intended clinical application:** Our TAMP foundation model enables two key clinical applications: 1) Direct enhancement without any training, enabling rapid validation of emerging NICT protocols. 2) Low-cost high-performance specialized model development for specific scenarios through parameter-efficient fine-tuning with minimal data.
   c) **Manuscript changes:** We have revised the Introduction section to clarify the clinical motivation and intended applications of foundation models for NICT enhancement,

through elaborating on FM's universal enhancement and efficient adaptation capabilities. Specifically:

"Foundation models (FMs) offer potential to address these challenges through two key capabilities: a) Universal enhancement enables rapid validation of emerging NICT protocols. FMs' generalization across diverse NICT settings supports training-free quality enhancement for a broad range of protocols, enabling rapid assessment of new device potential for clinical deployment. b) Efficient adaptation enables low-cost development of high-performance specialized models. FMs' transferable prior knowledge improves the emergence of specialized models with limited fine-tuning data, enabling low-cost deployment of models that surpass training-from-scratch approaches."

2. For the question "In particular, the authors confuse several x-ray-based imaging modalities such as clinical CT, cone-beam CT, and tomosynthesis":

Our TAMP is designed for axial non-ideal measurement CT (NICT), including both clinical CT and cone-beam CT. 1) Axial CT performs tomographic scanning along the axial direction and stores images as 3D volumes. Through large-scale pretraining on such axial NICT data, our foundation model learns universal representations of anatomical structures and artifact features manifesting in axial planes, thereby targeting clinical CT and cone-beam CT applications with axial imaging geometry. 2) Tomosynthesis (e.g., breast tomosynthesis) performs acquisition and reconstruction in coronal or sagittal planes, which is incompatible with our axial imaging geometry and 3D volume representation. Therefore, tomosynthesis is outside the current scope of TAMP.

Manuscript changes: We have revised the Introduction and Discussion sections to clarify that TAMP is designed for axial CT imaging systems and tomosynthesis is outside the current scope, through specifying the applicable modalities and limitations. Specifically:

"In this paper, we propose the multi-scale integrated Transformer AMPlifier (TAMP), an imaging FM for universal enhancement of NICT images in axial CT systems (e.g., diagnostic CT, cone-beam CT)."

"TAMP is designed for axial CT imaging systems where artifacts manifest in axial planes. Tomosynthesis (e.g., breast tomosynthesis), which performs acquisition and reconstruction in coronal or sagittal planes, is outside the current scope and will be explored in future work."

** We respond to your specific concerns about imaging modality applications and clinical scenarios in detail in Response#3 below.*

Comment#3: Introduction: The authors introduce the considered artifacts, i.e., high image noise, limited-angle artifacts, and sparse-view artifacts. However, only high image noise is relevant to clinical CT. Limited-angle artifacts and sparse-view artifacts are more common in cone-beam CT or in dedicated modalities such as breast tomosynthesis. These modalities, however, differ significantly from clinical CT in terms of spatial resolution, noise, etc. It is unclear what the intended application of the foundation model might be, since it was trained solely on clinical CT data. Furthermore, such non-standard CT settings, except for high image noise, are not widely used in clinical CT practice, contrary to the authors' claim.

Response#3: We sincerely thank you for this insightful point, which will help clarify the intended application and applicable modalities of TAMP. Specifically: 1) Axial CT is our target (including clinical CT and cone-beam CT). 2) It is the large-scale pretraining that enables TAMP to achieve universal NICT enhancement knowledge applicable across these modalities. 3) In this version, we have supplemented experiments on real cone-beam CT datasets to validate TAMP's superior enhancement performance.

1. **Axial CT is our target (including clinical CT and cone-beam CT)**

   TAMP is designed for axial CT imaging where reconstruction planes are perpendicular to the patient's longitudinal axis and artifacts manifest in axial (transverse) slices. This encompasses clinical CT with LDCT artifacts and cone-beam CT with SVCT and LACT artifacts. In contrast, tomosynthesis modalities (e.g., breast tomosynthesis) employ coronal or sagittal reconstruction planes parallel to the longitudinal axis, representing distinct imaging geometry outside TAMP's current scope.

2. **Why TAMP enables universal NICT enhancement across axial CT modalities**

   TAMP enables universal NICT enhancement capability across clinical CT and cone-beam CT through large-scale pretraining and adaptation:

   a) **Large-scale pretraining learns generalizable NICT enhancement knowledge.** Through pretraining on 10.8 million diverse axial NICT image pairs, TAMP learns universal knowledge transferable across axial CT modalities. First, anatomical structure representations across different body regions (head, chest, abdomen, lower-limbs) remain consistent across axial CT modalities. Second, fundamental CT degradation mechanisms (photon noise, electronic noise, projection deficiency) in axial reconstruction geometry learned through physics-driven simulation capture the reconstruction physics universal to axial CT systems. These generalizable priors enable direct enhancement for clinical LDCT and effective transfer to cone-beam CT with LACT and SVCT artifacts.

   b) **Parameter-efficient fine-tuning adapts to diverse axial CT modalities.** TAMP employs parameter-efficient fine-tuning that adapts pretrained knowledge to specific scenarios including clinical CT and cone-beam CT with different spatial resolutions, dynamic ranges, or acquisition trajectories. This adaptation strategy requires minimal paired data and ensures robust enhancement performance across diverse axial CT modalities.

3. **Supplemented experiments on cone-beam CT**

   In this version, we have supplemented experiments on real cone-beam CT datasets to validate TAMP's capability across axial CT modalities.

   a) **Experiment settings:** We collected and evaluated TAMP on an external cone-beam CT dataset with limited-angle artifacts, comprising 4 cases (1576 slices) from radiotherapy workflows. For baseline methods, we trained specialized models from scratch on the cone-beam CT training set. TAMP was evaluated in both zero-shot and fine-tuned (TAMP-S) settings.

   b) **Experiment results:** TAMP demonstrates effective knowledge transfer from clinical CT to cone-beam CT, achieving superior quantitative metrics and anatomical structure reconstruction quality through fine-tuning, which validates the foundation model's universal NICT enhancement capability across axial CT modalities.

i. **Quantitative evaluation:** TAMP demonstrates effective cross-modality transfer capability from diagnostic CT to cone-beam CT through parameter-efficient fine-tuning, validating the foundation model's adaptability across CT imaging modalities. As shown in Table D, TAMP-S achieves PSNR of 30.59 dB and SSIM of 84.33%, outperforming the best baseline ProCT by +0.43 dB and +1.38% respectively. This quantitative superiority demonstrates that TAMP's universal representations learned from large-scale diagnostic CT pretraining effectively transfer to cone-beam CT through efficient fine-tuning, enabling consistent improvement in pixel-level accuracy and structural similarity metrics despite domain shift challenges in spatial resolution and dynamic range differences.

Table D: Comparative validation on limited-angle cone-beam CT (LA-CBCT) dataset.

Methods	Training slice number	PSNR (dB)↑	SSIM (%)↑	RMSE (Hu)↓	LPIPS (%)↓
LA-CBCT (Input)	-	23.48±4.77	75.16±10.39	522.17±298.40	36.12±7.66
RED-CNN	2,758	21.67±6.35	62.46±8.94	599.97±259.62	40.78±13.07
FBPConvNet	2,758	26.64±4.77	79.75±9.21	369.38±226.09	*26.09±8.49*
ProCT	2,758	*30.16±5.54*	*82.95±8.51*	*252.77±170.80*	32.04±10.87
TAMP (Ours)	0	25.12±5.51	71.39±10.43	415.73±220.76	37.26±9.18
TAMP-S (Ours)	2,758	**30.59±5.87**	**84.33±7.86**	**244.97±173.44**	**25.10±8.69**

ii. **Qualitative evaluation:** TAMP achieves superior anatomical structure reconstruction with effective artifact suppression, validating the clinical utility of cross-modality transfer for diagnostic quality improvement. As shown in Fig. C, in the limited-angle CBCT input, vertebral bone structures are severely corrupted with radial wedge-shaped artifacts. The three baseline methods (RED-CNN, FBPConvNet, ProCT) fail to suppress these artifacts or reconstruct the damaged structures. TAMP effectively suppresses artifacts, rendering edge structures smooth and distinguishable, while TAMP-S further reconstructs the corrupted vertebral bone structures approaching the ground truth CBCT. The qualitative superiority demonstrates that TAMP's learned geometric artifact patterns enable robust limited-angle artifact suppression, and parameter-efficient adaptation further recovers fine-grained anatomical structure integrity critical for clinical interpretation.

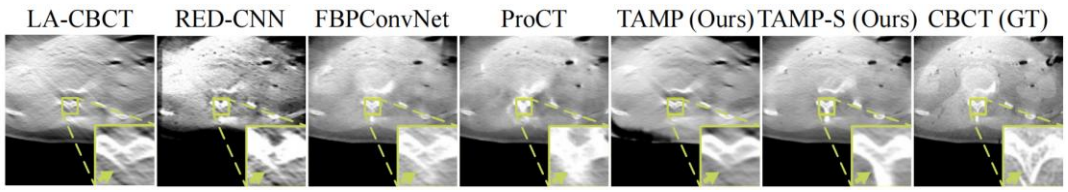

Fig. 19: Qualitative comparison on limited-angle cone-beam CT enhancement. TAMP effectively suppresses radial artifacts and TAMP-S further reconstructs corrupted anatomical details.

Manuscript changes: We have revised the Introduction and Discussion sections to clarify TAMP's target modalities and added cone-beam CT experiments in the Supplementary Materials to validate universal enhancement capability across axial CT systems, through

specifying the applicable modalities (diagnostic CT and cone-beam CT) and supplementing external validation. Specifically:

"In this paper, we propose the multi-scale integrated Transformer AMPlifier (TAMP), an imaging FM for universal enhancement of NICT images in axial CT systems (e.g., diagnostic CT, cone-beam CT)."

"TAMP is designed for axial CT imaging systems where artifacts manifest in axial planes. Tomosynthesis (e.g., breast tomosynthesis), which performs acquisition and reconstruction in coronal or sagittal planes, is outside the current scope and will be explored in future work."

"We evaluate TAMP on an external cone-beam CT dataset with limited-angle artifacts to validate its capability across axial CT modalities. The dataset comprises 4 cases (1576 slices) from radiotherapy workflows. For baseline methods (RED-CNN, FBPConvNet, ProCT), we trained specialized models from scratch on the cone-beam CT training set. TAMP was evaluated in both zero-shot and fine-tuned (TAMP-S) settings. As shown in Table 8 and Fig. 19, TAMP demonstrates effective knowledge transfer from diagnostic CT to cone-beam CT. TAMP-S achieves PSNR of 30.59 dB and SSIM of 84.33%, outperforming the best baseline ProCT by +0.11 dB and +1.60% respectively. Qualitatively, in the limited-angle CBCT input, vertebral bone structures are severely corrupted with radial wedge-shaped artifacts. The baseline methods fail to suppress these artifacts or reconstruct the damaged structures, while TAMP effectively suppresses artifacts and TAMP-S further reconstructs the corrupted vertebral bone structures approaching the ground truth."

Table 8: Quantitative evaluation on limited-angle cone-beam CT enhancement.

Method	PSNR (dB) ↑	SSIM (%) ↑	RMSE (Hu) ↓	LPIPS (%) ↓
LA-CBCT (Input)	23.48±4.77	75.16±10.39	522.17±298.40	36.12±7.66
RED-CNN	21.67±6.35	62.46±8.94	599.97±259.62	40.78±13.07
FBPConvNet	26.64±4.77	79.75±9.21	369.38±226.09	26.09±8.49
ProCT	30.16±5.54	82.95±8.51	252.77±170.80	32.04±10.87
TAMP	25.12±5.51	71.39±10.43	415.73±220.76	37.26±9.18
TAMP-S	**30.59±5.87**	**84.33±7.86**	**244.97±173.44**	**25.10±8.69**

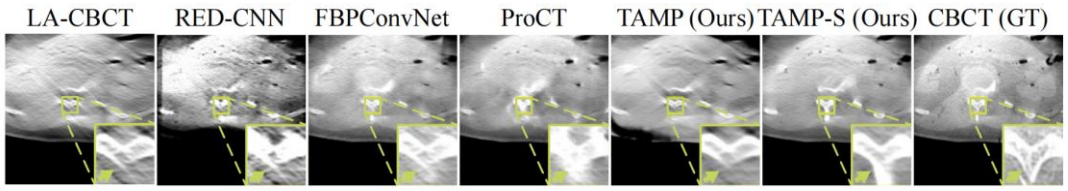

Fig. 19: Qualitative comparison on limited-angle cone-beam CT enhancement. TAMP effectively suppresses radial artifacts and TAMP-S further reconstructs corrupted anatomical details.

Comment#4: Introduction: Limited-angle and sparse-view artifacts are not of interest in clinical CT. In particular, limited-angle acquisitions due to a restricted angular range are not possible in clinical CT given the gantry-based setup. Moreover, sparse-view acquisitions would not increase scan speed as claimed by the authors, since the gantry rotation speed is constant. In addition, sparse-view artifacts would look very different from what is shown in this manuscript, as clinical CT data are typically acquired using a spiral trajectory. It seems the

authors considered only circular trajectories often used in cone-beam CT.

Response#4: We sincerely thank you for these insightful technical observations, which help clarify the precise application scenarios and technical scope of our TAMP foundation model approach.

1. For the question "Limited-angle and sparse-view artifacts are not of interest in clinical CT. In particular, limited-angle acquisitions due to a restricted angular range are not possible in clinical CT given the gantry-based setup.":

   Our study focuses on axial CT imaging, including both diagnostic CT and treatment cone-beam CT scenarios. TAMP is pretrained on clinical CT data with physics-driven simulation of axial imaging artifacts including LDCT, SVCT, and LACT. Through large-scale pretraining, TAMP learns universal representations of anatomical structures and artifact patterns in axial reconstruction geometry. This learned knowledge generalizes effectively to cone-beam CT, because both modalities share the same axial reconstruction geometry where artifacts manifest in transverse planes. As described in Response#3, TAMP effectively transfers to cone-beam CT through parameter-efficient fine-tuning with minimal data.

   Manuscript changes: We have revised the Introduction and Discussion sections to clarify that TAMP targets axial CT imaging systems including diagnostic CT and cone-beam CT, through specifying the applicable modalities and limitations. Specifically:

   "In this paper, we propose the multi-scale integrated Transformer AMPlifier (TAMP), an imaging FM for universal enhancement of NICT images in axial CT systems (e.g., diagnostic CT, cone-beam CT)."

   "TAMP is designed for axial CT imaging systems where artifacts manifest in axial planes. Tomosynthesis (e.g., breast tomosynthesis), which performs acquisition and reconstruction in coronal or sagittal planes, is outside the current scope and will be explored in future work."

2. For the question "Moreover, sparse-view acquisitions would not increase scan speed as claimed by the authors, since the gantry rotation speed is constant.":

   We sincerely thank you for pointing out this potential misunderstanding. The primary benefit of sparse-view acquisitions is radiation dose reduction through fewer projection acquisitions, which is particularly valuable for cone-beam CT applications requiring repeated imaging in image-guided therapy. In this version, we have revised the manuscript to remove the incorrect claim about scan speed.

   Manuscript changes: We have revised the Introduction section to accurately describe the benefits of sparse-view acquisitions as radiation dose reduction rather than scan speed improvement. Specifically:

   "..., SVCT reduces radiation dose through sparse angle sampling, ..."

3. For the question "In addition, sparse-view artifacts would look very different from what is shown in this manuscript, as clinical CT data are typically acquired using a spiral trajectory. It seems the authors considered only circular trajectories often used in cone-beam CT.":

   We thank you for this important observation. 1) Our foundation model approach enables adaptation to cone-beam CT with different acquisition trajectories (both circular

and spiral), through the learned universal priors from axial CT reconstruction geometry and parameter-efficient fine-tuning. 2) We have validated TAMP on an external cone-beam CT dataset with limited-angle artifacts from radiotherapy workflows using circular trajectory acquisition.

a) **Experiment settings:** We evaluated TAMP on cone-beam CT with circular trajectory acquisition (4 cases, 1576 slices from radiotherapy workflows, same dataset as Response#3 point 3). TAMP was evaluated in zero-shot and fine-tuned settings to assess trajectory-adaptive capability.

b) **Experiment results:** TAMP-S achieved effective adaptation from helical diagnostic CT to circular trajectory CBCT (PSNR: 30.59 dB, SSIM: 84.33%), effectively suppressing radial wedge-shaped artifacts characteristic of circular acquisition (see Fig 19 and Table D). This validates that universal priors from axial reconstruction geometry enable trajectory-adaptive enhancement across different acquisition trajectories.

Table D: Comparative validation on limited-angle cone-beam CT (LA-CBCT) dataset.

Methods	**Training slice number**	**PSNR (dB)↑**	**SSIM (%)↑**	**RMSE (Hu)↓**	**LPIPS (%)↓**
LA-CBCT (Input)	-	23.48±4.77	75.16±10.39	522.17±298.40	36.12±7.66
RED-CNN	2,758	21.67±6.35	62.46±8.94	599.97±259.62	40.78±13.07
FBPConvNet	2,758	26.64±4.77	79.75±9.21	369.38±226.09	*26.09±8.49*
ProCT	2,758	*30.16±5.54*	*82.95±8.51*	*252.77±170.80*	32.04±10.87
TAMP (Ours)	0	25.12±5.51	71.39±10.43	415.73±220.76	37.26±9.18
TAMP-S (Ours)	2,758	**30.59±5.87**	**84.33±7.86**	**244.97±173.44**	**25.10±8.69**

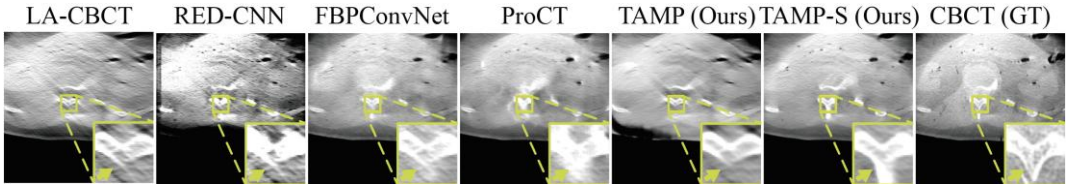

Fig. 19: Qualitative comparison on limited-angle cone-beam CT enhancement. TAMP effectively suppresses radial artifacts and TAMP-S further reconstructs corrupted anatomical details.

c) **Manuscript changes:** We have added the limited-angle CBCT experiment in subsection "External-scene adaptation evaluation results" of the Supplementary Materials to validate TAMP's adaptability across axial CT modalities with circular trajectory acquisition. Specifically:

"We evaluate TAMP on an external cone-beam CT dataset with limited-angle artifacts to validate its capability across axial CT modalities. The dataset comprises 4 cases (1576 slices) from radiotherapy workflows. For baseline methods (RED-CNN, FBPConvNet, ProCT), we trained specialized models from scratch on the cone-beam CT training set. TAMP was evaluated in both zero-shot and fine-tuned (TAMP-S) settings. As shown in Table 8 and Fig. 19, TAMP demonstrates effective knowledge transfer from diagnostic CT to cone-beam CT. TAMP-S achieves PSNR of 30.59 dB and SSIM of 84.33%, outperforming the best baseline ProCT by +0.11 dB and +1.60%

respectively. Qualitatively, in the limited-angle CBCT input, vertebral bone structures are severely corrupted with radial wedge-shaped artifacts. The baseline methods fail to suppress these artifacts or reconstruct the damaged structures, while TAMP effectively suppresses artifacts and TAMP-S further reconstructs the corrupted vertebral bone structures approaching the ground truth."

Table 8: Quantitative evaluation on limited-angle cone-beam CT enhancement.

Method	PSNR (dB) ↑	SSIM (%) ↑	RMSE (Hu) ↓	LPIPS (%) ↓
LA-CBCT (Input)	23.48±4.77	75.16±10.39	522.17±298.40	36.12±7.66
RED-CNN	21.67±6.35	62.46±8.94	599.97±259.62	40.78±13.07
FBPConvNet	26.64±4.77	79.75±9.21	369.38±226.09	26.09±8.49
ProCT	30.16±5.54	82.95±8.51	252.77±170.80	32.04±10.87
TAMP	25.12±5.51	71.39±10.43	415.73±220.76	37.26±9.18
TAMP-S	**30.59±5.87**	**84.33±7.86**	**244.97±173.44**	**25.10±8.69**

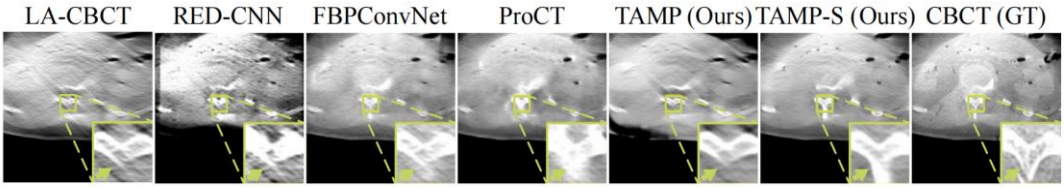

Fig. 19: Qualitative comparison on limited-angle cone-beam CT enhancement. TAMP effectively suppresses radial artifacts and TAMP-S further reconstructs corrupted anatomical details.

Comment#5: Introduction: Dose reduction via a reduction of tube voltage is usually not desired in clinical imaging. While reducing tube current results only in increased image noise, reducing tube voltage leads to an undesired change in image contrast.

Response#5: We sincerely thank you for this technical clarification. 1) In principle, tube voltage reduction for dose reduction is a clinically acceptable practice, and research advances have demonstrated its feasibility across multiple clinical scenarios [1-3], which has been widely recognized. 2) In addition, we have supplemented experimental validation of our method's enhancement performance on LDCT through solely reducing tube current.

1. **Tube voltage reduction for dose reduction has been recognized as a clinically acceptable practice in CT imaging.** A search of prominent radiology journals (e.g., Radiographics [1], Radiology [2], American Journal of Roentgenology [3]) reveals that controlled tube voltage reduction has been successfully implemented and validated in multiple clinical scenarios including pediatric imaging, abdominal contrast-enhanced CT, and pulmonary CT angiography. This established clinical practice provides justification for including tube voltage reduction in our LDCT simulation protocol.

[1] Nagayama Y, Oda S, Nakaura T, et al. Radiation Dose Reduction at Pediatric CT: Use of Low Tube Voltage and Iterative Reconstruction. Radiographics. 2018;38(5):1421-1440. doi:10.1148/rg.2018180041.

[2] Nakayama Y, Awai K, Funama Y, et al. Abdominal CT with low tube voltage: preliminary observations about radiation dose, contrast enhancement, image quality, and noise. Radiology. 2005;237(3):945-951. doi:10.1148/RADIOL.2373041655.

[3] Szucs-Farkas Z, Schibler F, Cullmann J, et al. Diagnostic accuracy of pulmonary CT angiography at low tube voltage: intraindividual comparison of a normal-dose protocol at 120 kVp and a low-dose protocol at 80 kVp using reduced amount of contrast medium in a simulation study. AJR Am J Roentgenol. 2011;197(5):W852-9. doi:10.2214/AJR.11.6750.

2. **We supplemented experimental validation of our method's enhancement performance on LDCT through solely reducing tube current.**
   a) **Experiment settings:** We collected a new real-world LDCT dataset using a current-reduction-only protocol: tube voltage fixed at 120 kVp, tube current reduced from 200 mAs to 50 mAs (Table E). This dataset comprises 10 patients with paired LDCT-ICT acquisitions (5 for training, 5 for testing; 1527 training slices, 1275 test slices). For baseline methods (RED-CNN, FBPConvNet, ProCT), we trained specialized models from scratch on the training set. TAMP was evaluated in both zero-shot and fine-tuned (TAMP-S) settings.

Table E: Comparison of current-reduction-only LDCT dataset (added) and original LDCT dataset.

Parameters	Original LDCT / ICT	Added LDCT / ICT
Tube voltage (kVp)	80 / 120 (Reduction)	120 / 120 (Same)
Tube current (mAs)	30 / 90 (Reduction)	50 / 200 (Reduction)

   b) **Experiment results:** TAMP demonstrates effective enhancement performance on current-reduction-only LDCT, achieving superior quantitative metrics and structural preservation quality, which validates its applicability to the clinically preferred dose reduction protocol.
      i. **Quantitative evaluation:** TAMP demonstrates effective enhancement performance on current-reduction-only LDCT, validating its applicability to the clinically preferred dose reduction protocol. As shown in Table F, TAMP-S achieves PSNR of 29.39 dB, SSIM of 79.49%, and RMSE of 142.52 HU, outperforming all baselines across all metrics. The superior performance demonstrates that TAMP's learned noise suppression mechanisms effectively handle the pure noise degradation in current-reduction-only LDCT without contrast changes, while maintaining exceptional perceptual quality (LPIPS 17.83%) through physics-driven pretraining on diverse noise patterns.

Table F: Real-world LDCT enhancement results, with 10 cases (2,464 slices) for training and 6 cases (1,527 slices) for testing.

Methods	Training slice number	PSNR (dB)↑	SSIM (%)↑	RMSE (Hu)↓	LPIPS (%)↓
LDCT (Input)	-	28.57±1.84	73.59±9.13	156.13±33.61	23.20±6.17
RED-CNN	2,464	*30.01±2.04*	79.27±4.05	*134.98±29.95*	28.61±3.37
FBPConvNet	2,464	29.43±2.11	*79.49±3.84*	143.43±30.70	27.20±3.50
ProCT	2,464	29.46±1.81	76.82±4.42	142.42±27.19	35.86±7.21
TAMP (Ours)	0	29.44±1.86	77.51±4.29	143.28±30.07	*17.00±4.64*
TAMP-S (Ours)	2,464	**30.09±1.94**	**80.03±4.08**	**133.19±27.88**	**16.54±4.18**

      ii. **Qualitative evaluation:** TAMP achieves superior structural preservation with

effective noise suppression, validating clinical utility for current-reduction-only LDCT enhancement. As shown in Fig. D, zoomed-in regions highlight cardiac structures. Baseline methods produce over-smoothed results with blurred edges and reduced contrast, rendering structures difficult to discern. TAMP preserves clear edges while suppressing noise, and TAMP-S further improves noise suppression with well-defined structures. This superiority demonstrates that baseline methods trained on limited real paired data with spatial misalignment learn suboptimal representations, while TAMP's large-scale pretrained priors enable effective separation of anatomical structures from noise patterns, achieving robust enhancement through efficient adaptation.

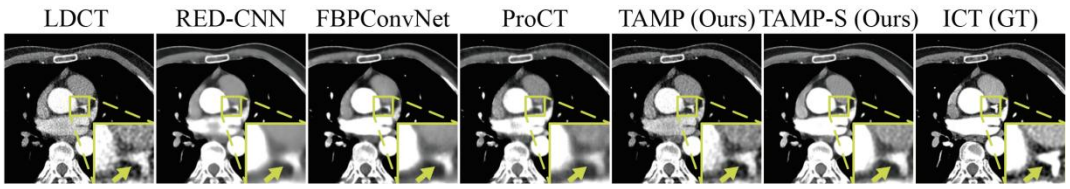

Fig. D. Qualitative comparison on current-reduction-only LDCT enhancement. Zoomed-in regions highlight cardiac structures, where baseline methods produce over-smoothed results while TAMP and TAMP-S preserve structural clarity with effective noise suppression.

** The added LDCT dataset was approved under the same ethics protocol as the original datasets (Ethics approval: AF/SC-08/03.0, 2025-0068-02).*

Comment#6: Data quantity: The authors state that data quantity is an issue. I disagree. The considered artifacts can easily be simulated using clinical data. This approach has been used for decades and is not new.

Response#6: We sincerely thank you for this important perspective, which helps us clarify the data quantity issue. 1) NICT enhancement faces issues that artifact simulation does not solve: complex development and limited generalizability. Foundation models provide a promising solution. 2) Data quantity is the issue in foundation model development, and we address it with artifact simulation. Specifically:

1. **NICT enhancement faces issues that artifact simulation does not solve: complex development and limited generalizability.**

   NICT enhancement faces complex development and limited generalizability issues that simulation methods cannot address.Each new imaging protocol requires extensive development cycles including data collection, architecture design, and training from scratch, resulting in prolonged deployment and high costs [1][2]. Moreover, specialized models trained on one protocol exhibit substantial performance degradation on different settings, limiting their generalizability and necessitating protocol-specific development [3]. These issues are inherent to the specialized model paradigm and cannot be resolved by simulation methods alone. Foundation models offer potential to address these issues by enabling rapid adaptation to new protocols through fine-tuning with minimal data while achieving superior performance through universal representations learned from large-scale pretraining.

2. **Data quantity is the issue in foundation model development, and we address it with artifact simulation.**

Foundation model pretraining requires large-scale paired datasets (NICT-ICT pairs) to learn universal representations across diverse scenarios [4][5]. However, acquiring clinical paired data at the required scale faces critical challenges. For LDCT, dual-acquisition protocols raise ethical concerns due to additional radiation exposure. For SVCT and LACT, accessing raw projection data is limited by proprietary system configurations across different vendors and institutions. We address this data quantity issue with artifact simulation. Traditional NICT simulation methods generate limited synthetic data (typically fewer than 100,000 pairs) for specialized model training [6][7][8]. Our SimNICT generates 10.8 million diverse paired images covering various anatomies, protocols, artifact types, and degradation levels by simulating projection-domain defects following established physics principles [9]. This physics-driven design ensures realistic artifact patterns enabling effective generalization to clinical data, as validated in our real-world evaluations.

Manuscript changes: We have revised the Introduction section to establish the necessity of foundation models for addressing NICT enhancement challenges and clarify the data quantity and variation issues that motivate our physics-driven simulation approach. Specifically:

"Although numerous studies have demonstrated that deep learning can enhance NICT image quality in specific scenarios, significant challenges remain. 1) Complex specialized model development. Each new NICT protocol requires extensive dataset collection, model architecture design, and training from scratch, extending the development cycle of NICT imaging devices. 2) Limited transfer learning capability. These studies typically target specific body regions (e.g., head, chest, abdomen) and NICT settings (LDCT, SVCT, LACT). When emerging NICT devices or updated protocols are introduced in clinical practice, these specialized models cannot effectively transfer their learned knowledge to new scenarios with limited available data. Overall, this specialized model development paradigm hinders the widespread and rapid clinical deployment of AI-integrated NICT technologies.

Foundation models (FMs) offer potential to address these challenges through two key capabilities: a) Universal enhancement enables rapid validation of emerging NICT protocols. FMs' generalization across diverse NICT settings supports training-free quality enhancement for a broad range of protocols, enabling rapid assessment of new device potential for clinical deployment. b) Efficient adaptation enables low-cost development of high-performance specialized models. FMs' transferable prior knowledge improves the emergence of specialized models with limited fine-tuning data, enabling low-cost deployment of models that surpass training-from-scratch approaches.

However, two main challenges have so far hindered their success in this domain: a) Data quantity. Ethical concerns restrict the creation of large datasets for NICT FM training. The inherent radiation risks of CT scanning make it unethical to repeatedly scan individuals solely for data collection. This limitation prevents the direct acquisition of large NICT datasets, thereby compromising the model's ability to generalize in universal scenarios. b) Data variation. Different physical processes in NICT settings lead to highly varied defect patterns. For example, LDCT images are characterized by fine-grained noise, whereas LACT images exhibit pronounced angular defects. This variability poses a challenge for universal NICT enhancement models, which struggle to accommodate a wide spectrum of defect patterns. Additionally,

existing specialized NICT enhancement models focus on specific defect types within particular NICT settings, limiting their capacity for universal representation and learning during FM training."

[1] De Biase et al. Clinical adoption of deep learning target auto-segmentation for radiation therapy: challenges, clinical risks, and mitigation strategies. BJR|Artificial Intelligence, 2024.
[2] Herington et al. Ethical Considerations for Artificial Intelligence in Medical Imaging: Data Collection, Development, and Evaluation. Journal of Nuclear Medicine, 2023.
[3] Szumel et al. The Impact of Scanner Domain Shift on Deep Learning Performance in Medical Imaging: an Experimental Study. arXiv, 2024.
[4] Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258, 2021.
[5] Xu H, Usuyama N, Bagga J, et al. A whole-slide foundation model for digital pathology from real-world data. Nature. 2024;630(8015):181-188. doi:10.1038/s41586-024-07441-w.
[6] Leuschner J, Schmidt M, Baguer DO, et al. LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction. Scientific Data. 2021;8(1):109. doi:10.1038/s41597-021-00893-z.
[7] Chen H, Zhang Y, Kalra MK, et al. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. IEEE Transactions on Medical Imaging. 2017;36(12):2524-2535. doi:10.1109/TMI.2017.2715284.
[8] Zhang Z, Liang X, Dong X, et al. A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution. IEEE Transactions on Medical Imaging. 2018;37(6):1407-1417. doi:10.1109/TMI.2018.2823338.
[9] Gao C, Killeen B, Hu Y, et al. Synthetic data accelerates the development of generalizable learning-based algorithms for X-ray image analysis. Nature Machine Intelligence. 2023;5(5):294-308. doi:10.1038/s42256-023-00629-1.

Comment#7: Evaluation: All evaluation metrics used (PSNR, SSIM, RMSE, LPIPS) are clinically irrelevant, and clinical validation is missing. So far, the authors have only demonstrated that the proposed method performs "image prettification." Diagnostic quality was not assessed in the manuscript.

Response#7: We sincerely thank you for this insightful suggestion, which will help us demonstrate the clinical diagnostic value of our work. In this new version, we have conducted three downstream clinical validation experiments to validate that TAMP-enhanced images improve performance across critical diagnostic applications including **pulmonary nodule segmentation**, **pulmonary nodule malignancy diagnosis**, **and radiomics feature analysis.**

1. **Pulmonary nodule segmentation:** We have conducted pulmonary nodule segmentation downstream validation experiments to evaluate how our TAMP enhancement of NICT images improves segmentation performance. Pulmonary nodule segmentation is a critical component in lung cancer diagnosis, as accurate delineation of nodule boundaries directly impacts malignancy assessment and treatment planning.
    a) **Experiment setting:** We evaluated automated lesion delineation performance on LIDC-IDRI [1] with 8,553 training, 856 validation, and 2,139 test slices. We degraded ICT images into LDCT, SVCT, LACT, applied TAMP enhancement, and tested a U-

Net [2] trained on ICT images to assess whether TAMP-enhanced images restore diagnostic-quality segmentation.

b) **Experiment results:** TAMP restores diagnostic-quality segmentation degraded by NICT. Table 15 shows NICT reduces Dice from ICT baseline (0.744) to 0.676, 0.527, 0.451, while TAMP recovers to 0.690, 0.662, 0.621 (+0.014, +0.135, +0.170), validating that enhanced images achieve clinically acceptable nodule delineation for treatment planning.

Table 15: Our enhancement improves pulmonary nodule segmentation on LIDC-IDRI (w/o vs. w/ our method).

Setting	**Dice**			**IoU**		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.676	0.690	+0.014	0.527	0.544	+0.017
SVCT	0.527	0.662	+0.135	0.380	0.513	+0.133
LACT	0.451	0.621	+0.170	0.314	0.468	+0.154

c) **Manuscript changes:** We have added the pulmonary nodule segmentation experiment in subsection "Downstream clinical validation" of the Supplementary Materials to demonstrate clinical diagnostic quality. Specifically:

"Pulmonary nodule segmentation is a critical component in lung cancer diagnosis, as accurate delineation of nodule boundaries directly impacts malignancy assessment and treatment planning. We conducted pulmonary nodule segmentation experiments on the LIDC-IDRI dataset to evaluate how TAMP enhancement improves automated lesion delineation accuracy.

We selected CT slices with expert consensus nodule masks, splitting them into 8,553, 856, 2,139 slices for training, validation, test respectively. ICT images were synthetically degraded into three NICT types (LDCT, SVCT, LACT) and subsequently enhanced by TAMP. A U-Net segmentation model was trained on ICT images with their corresponding masks, then tested on both NICT-degraded and TAMP-enhanced images to quantify segmentation accuracy improvements.

TAMP enhancement effectively restores segmentation accuracy degraded by NICT artifacts, validating improved diagnostic utility for lesion delineation. As shown in Table, TAMP improves Dice from 0.676, 0.527, 0.451 to 0.690, 0.662, 0.621 for LDCT, SVCT, LACT, corresponding to relative gains of +0.014, +0.135, +0.170. The substantial improvements in SVCT and LACT demonstrate that TAMP's learned representations capture geometric artifact patterns and effectively restore boundary information corrupted by angular undersampling."

Table 15: Our enhancement improves pulmonary nodule segmentation on LIDC-IDRI (w/o vs. w/ our method).

Input	Dice			IoU		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.676	0.690	+0.014	0.527	0.544	+0.017
SVCT	0.527	0.662	+0.135	0.380	0.513	+0.133
LACT	0.451	0.621	+0.170	0.314	0.468	+0.154

2. **Pulmonary nodule malignancy diagnosis:** We have conducted pulmonary nodule malignancy diagnosis experiments to evaluate how TAMP enhancement improves malignancy assessment. Accurate diagnosis of nodule malignancy levels is critical for lung cancer screening, as it directly impacts clinical decision-making for follow-up and treatment.

   a) **Experiment setting:** We conducted malignancy classification experiments on the LIDC-IDRI dataset. We formulated a five-class malignancy classification task at the 2D slice level, where nodules are classified into expert consensus ratings: 1 (highly unlikely for cancer), 2 (moderately unlikely), 3 (indeterminate likelihood), 4 (moderately suspicious), and 5 (highly suspicious for cancer). A ResNet-18 classifier was trained on 8,458 ICT training slices with their malignancy labels, then tested on 2,139 test slices across seven scenarios: ICT baseline, three NICT types (LDCT, SVCT, LACT), and three TAMP-enhanced types (LDCT, SVCT, LACT) to quantify classification accuracy improvements.

   b) **Experiment results:** TAMP enhancement effectively recovers diagnostic accuracy degraded by NICT artifacts, validating improved clinical utility for malignancy assessment. As shown in Table 16, TAMP improves accuracy from 0.817, 0.740, 0.758 to 0.831, 0.843, 0.839 for LDCT, SVCT, LACT, corresponding to relative gains of +0.014, +0.103, +0.081. The substantial improvements in SVCT and LACT demonstrate that TAMP's learned representations effectively restore diagnostically critical features corrupted by geometric artifacts, maintaining strong discriminative ability (AUC $>$ 0.96) for reliable lung cancer screening.

Table 16: Our enhancement improves pulmonary nodule malignancy diagnosis on LIDC-IDRI (w/o vs. w/ our method).

Setting	Accuracy			AUC		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.817	0.831	+0.014	0.962	0.967	+0.005
SVCT	0.740	0.843	+0.103	0.932	0.966	+0.034
LACT	0.758	0.839	+0.081	0.939	0.964	+0.025

   c) **Manuscript changes:** We have added the pulmonary nodule malignancy diagnosis experiment in subsection "Downstream clinical validation" of the Supplementary Materials to demonstrate clinical diagnostic quality. Specifically:

   "Accurate diagnosis of nodule malignancy levels is critical for lung cancer screening, as it directly impacts clinical decision-making for follow-up and treatment. We conducted pulmonary nodule malignancy diagnosis experiments on the LIDC-IDRI dataset to evaluate how TAMP enhancement improves malignancy assessment.

   We formulated a five-class malignancy classification task at the 2D slice level, where nodules are classified into expert consensus ratings: 1 (highly unlikely for cancer), 2 (moderately unlikely), 3 (indeterminate likelihood), 4 (moderately suspicious), and 5 (highly suspicious for cancer). A ResNet-18 classifier was trained on 8,458 ICT training slices with their malignancy labels, then tested on 2,139 test slices across seven scenarios: ICT baseline, three NICT types (LDCT, SVCT, LACT), and three TAMP-enhanced types to quantify classification accuracy improvements.

TAMP enhancement effectively recovers diagnostic accuracy degraded by NICT artifacts, validating improved clinical utility for malignancy assessment. As shown in Table, TAMP improves accuracy from 0.817, 0.740, 0.758 to 0.831, 0.843, 0.839 for LDCT, SVCT, LACT, corresponding to relative gains of +0.014, +0.103, +0.081. The substantial improvements in SVCT and LACT demonstrate that TAMP's learned representations effectively restore diagnostically critical features corrupted by geometric artifacts, maintaining strong discriminative ability (AUC $>$ 0.96) for reliable lung cancer screening."

Table 16: Our enhancement improves pulmonary nodule malignancy diagnosis on LIDC-IDRI (w/o vs. w/ our method).

Input	Accuracy			AUC		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.817	0.831	+0.014	0.962	0.967	+0.005
SVCT	0.740	0.843	+0.103	0.932	0.966	+0.034
LACT	0.758	0.839	+0.081	0.939	0.964	+0.025

3. **Radiomics feature analysis:** We have further evaluated how TAMP affects quantitative radiomics features. Radiomics analysis is highly sensitive to image quality, and feature preservation is critical for reliable downstream clinical applications.
   a) **Experiment setting:** To assess TAMP's clinical value in preserving quantitative imaging biomarkers, we evaluated radiomics feature stability on the same LIDC-IDRI lung nodules. We extracted 93 standard 2D radiomics features from ICT baseline, three NICT types (LDCT, SVCT, LACT), and TAMP-enhanced images using PyRadiomics. Feature stability was quantified by computing Pearson correlations between NICT (or TAMP) features and ICT reference features, where stable features achieving $r > 0.85$ are considered reliable for downstream clinical analysis.
   b) **Experiment results:** TAMP enhancement partially restores radiomics feature preservation degraded by NICT artifacts, validating improved quantitative imaging biomarker reliability. As shown in Table 17, TAMP improves mean correlation from 0.646, 0.722, 0.721 to 0.685, 0.741, 0.750 for LDCT, SVCT, LACT, corresponding to relative gains of +0.039, +0.019, +0.029. The modest improvements compared to segmentation and diagnosis tasks demonstrate that while TAMP's learned representations effectively restore visual quality and diagnostic features, quantitative radiomics features remain more sensitive to reconstruction artifacts, indicating the need for careful interpretation in radiomics-based clinical analysis.

Table 17: Our enhancement improves radiomics feature preservation (w/o vs. w/ our method).

Setting	**Mean Correlation (r)**			**Stable Rate (r > 8.5)**		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ (%)
LDCT	0.646	0.685	+0.039	40.9%	41.9%	+1.0
SVCT	0.722	0.741	+0.019	51.7%	52.7%	+1.0
LACT	0.721	0.750	+0.029	48.4%	52.7%	+4.3

   c) **Manuscript changes:** We have added the radiomics feature analysis experiment in

subsection “Downstream clinical validation” of the Supplementary Materials to demonstrate quantitative biomarker quality. Specifically:

`Radiomics analysis is highly sensitive to image quality, and feature preservation is critical for reliable downstream clinical applications. We evaluated how TAMP affects quantitative radiomics features to assess its clinical value in preserving quantitative imaging biomarkers.

We extracted 93 standard 2D radiomics features from ICT baseline, three NICT types (LDCT, SVCT, LACT), and TAMP-enhanced images using PyRadiomics on the same LIDC-IDRI lung nodules. Feature stability was quantified by computing Pearson correlations between NICT (or TAMP) features and ICT reference features, where stable features achieving $r > 0.85$ are considered reliable for downstream clinical analysis.

TAMP enhancement partially restores radiomics feature preservation degraded by NICT artifacts, validating improved quantitative imaging biomarker reliability. As shown in Table 17, TAMP improves mean correlation from 0.646, 0.722, 0.721 to 0.685, 0.741, 0.750 for LDCT, SVCT, LACT, corresponding to relative gains of +0.039, +0.019, +0.029. The modest improvements compared to segmentation and diagnosis tasks demonstrate that while TAMP's learned representations effectively restore visual quality and diagnostic features, quantitative radiomics features remain more sensitive to reconstruction artifacts, indicating the need for careful interpretation in radiomics-based clinical analysis.”

Table 17: Our enhancement improves radiomics feature preservation (w/o vs. w/ our method)

Input	Mean Correlation (r)			Stable Rate ($r > 0.85$)		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ (%)
LDCT	0.646	0.685	+0.039	40.9%	41.9%	+1.0%
SVCT	0.722	0.741	+0.019	51.7%	52.7%	+1.0%
LACT	0.721	0.750	+0.029	48.4%	52.7%	+4.3%

[1] Armato SG, McLennan G, Bidaut L, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical Physics. 2011;38(2):915-931. doi:10.1118/1.3528204.

[2] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Confernce on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015:234-241. doi:10.1007/978-3-319-24574-4_28.

Comment#8: Section 2.4: The data used for validation require further explanation. What systems were used? What trajectories were considered? How were sparse-view artifacts (circular vs. spiral trajectory) or limited-angle artifacts (what is the intended application given the angular ranges)

Response#8: We sincerely thank you for this important feedback requesting more detailed information about the validation data, which helps provide clearer technical specifications of our experimental setup.

1. For the question "What systems were used?"

   We collected real-world data from Nanjing Drum Tower Hospital using the uCT 960+ scanner from United Imaging Healthcare (UIH). For LDCT-ICT paired data, six patients (1 male and 5 females, aged 49-79 years) underwent scans at 120 kVp and 80 kVp tube voltages. For raw projection data used to reconstruct SVCT and LACT, six patients (5 males and 1 female, aged 43-79 years) underwent scans at 39 mA tube current and 120 kVp tube voltage. All data acquisition followed standard clinical protocols to ensure clinical relevance of our validation experiments.

2. For the question "What trajectories were considered?"

   We employed two distinct acquisition trajectories in our validation experiments. For LDCT data acquisition, the scan protocol was set to Helical mode with a reconstruction kernel of B_VSHARP_C, which represents the standard clinical helical trajectory used in routine diagnostic CT imaging. For SVCT and LACT data acquisition, the scan protocol was set to AXIAL mode with a reconstruction kernel of C_SOFT_BA, enabling the collection of raw projection data necessary for subsequent sparse-view and limited-angle reconstruction experiments.

3. For the question "How were sparse-view artifacts (circular vs. spiral trajectory) or limited-angle artifacts (what is the intended application given the angular ranges) simulated?"

   We reconstructed SVCT and LACT from real-world projection data using the TIGRE toolbox in MATLAB, following a systematic downsampling protocol. For SVCT reconstruction, we used projection angles to simulate sparse-view conditions, while for LACT reconstruction, we used projection angles to simulate limited-angle conditions. These specific angular ranges were selected to represent clinically relevant scenarios: SVCT with 240 projections simulates accelerated scanning protocols used for time-sensitive examinations, while LACT with 120 projections simulates imaging constraints encountered in interventional procedures or when patient positioning limits the available angular range. The reconstruction process utilized real clinical projection data rather than purely synthetic simulation, ensuring that the validation reflects authentic clinical imaging conditions and artifact patterns.

Manuscript changes: We have added detailed system specifications, trajectory information, and reconstruction parameters in Section "More details of real-world NICT" of the Supplementary Materials, so that readers can fully understand the technical details of our validation data acquisition and processing protocols. Specifically:

"The real-world dataset includes paired LDCT-ICT images and raw projection data for reconstructing LACT and SVCT. All data were collected from Nanjing Drum Tower Hospital using the uCT 960+ scanner from United Imaging Healthcare (UIH) with the following specifications. For the LDCT-ICT paired data, six patients (1 male and 5 females, aged 49-79 years) underwent two scans at 120 kVp and 80 kVp tube voltages. The scan protocol was set to Helical mode with a reconstruction kernel of B_VSHARP_C, which represents the standard clinical helical trajectory used in routine diagnostic CT imaging. For the raw projection data, six patients (5 males and 1 female, aged 43-79 years) underwent scans at the same 39 mA tube current and 120 kVp tube voltage. The scan protocol was set to AXIAL mode with a reconstruction kernel of C_SOFT_BA, enabling the collection of raw projection data necessary for subsequent sparse-view and limited-angle reconstruction experiments."

"For SVCT and LACT, using the TIGRE toolbox in MATLAB, we reconstructed SVCT and LACT from real-world projection data following the downsampling protocol with N_SV = 240 and N_LA = 120."

Comment#9: Section 2.5: The validation by radiologists is, in my opinion, not very meaningful. Naturally, radiologists will rate an image with limited-angle artifacts lower compared to the same image without such artifacts. However, this does not prove that the corrected images preserve diagnostically relevant structures. It would be much more meaningful to investigate dedicated clinical applications. For example: using lung images with known nodules or lesions, are the lesions still visible if high noise is simulated and corrected for? Or in tomosynthesis, can microcalcifications still be detected after correction of limited-angle artifacts using the proposed model? What would happen in the latter case, given that the model was trained using clinical CT data and is applied to tomosynthesis data that has a much higher spatial resolution but a much lower dynamic range? The important part here is to demonstrate the clinical applicability.

Response#9: We sincerely thank you for this valuable suggestion regarding clinical validation methodology, which provides important guidance for demonstrating the diagnostic utility of our work.

1. For the question "Section 2.5: The validation by radiologists is, in my opinion, not very meaningful. Naturally, radiologists will rate an image with limited-angle artifacts lower compared to the same image without such artifacts. However, this does not prove that the corrected images preserve diagnostically relevant structures."

   Our radiologist validation demonstrates clinical value through three aspects: 1) LACT's lower clinical acceptability illustrates the necessity of enhancement methods; 2) TAMP's substantial improvements validate effective preservation of diagnostically relevant structures; 3) TAMP's superior performance over baseline methods confirms its clinical utility. Specifically:

   a) **Our validation demonstrates image quality's impact on clinical acceptability, with LACT's lower acceptability illustrating the necessity of enhancement.** Limited-angle CT reconstructions inherently suffer from severe geometric artifacts due to incomplete angular sampling, which compromises image quality and clinical acceptability [1][2]. However, the clinical value of LACT lies in intraoperative guidance during surgical procedures and radiation therapy [3][4], where real-time imaging requirements necessitate limited angular sampling despite the resulting artifacts. This clinical reality motivates the need for effective image enhancement methods to improve the diagnostic utility of LACT images.

   b) **TAMP significantly improves clinical acceptability, validating effective preservation of diagnostically relevant structures.** To validate that TAMP preserves diagnostically relevant structures while removing artifacts, we conducted radiologist evaluation on LACT images using three clinical assessment metrics: PBN (Probability of Better than NICT), SQR (Subjective Quality Ranking), and PCA (Probability of Clinical Acceptability). As shown in Table G, TAMP achieves 100% PBN and improves SQR from 1.18 (NICT) to 2.59, while TAMP-S further improves SQR to 3.03. Most importantly, TAMP-S improves PCA by 13.89% compared to

uncorrected LACT, demonstrating that radiologists found the corrected images substantially more acceptable for clinical use. These substantial improvements across all three metrics directly validate that TAMP effectively preserves diagnostically relevant structures while removing artifacts.

Table G. The radiologist validation illustrates our superior improvement of clinical acceptance for LACT images, using metrics of Probability of better than NICT (PBN), Subjective quality ranking (SQR), and Probability of clinical acceptance (PCA).

Method	PBN (%)	SQR	PCA (%)
NICT	-	1.18	0
RedCNN	60.00	1.31	1.47
FBPConvNet	100.00	2.45	4.69
TAMP (Ours)	100.00	2.59	4.69
TAMP-S (Ours)	100.00	3.03	13.89

c) **TAMP outperforms baseline methods, validating its superior enhancement effectiveness.** TAMP and TAMP-S substantially outperform baseline methods across all radiologist assessment metrics for LACT images. Specifically, TAMP-S achieves the highest SQR of 3.03 compared to RED-CNN (1.31) and FBPConvNet (2.45), and most importantly, TAMP-S improves PCA by 9.4 times over RED-CNN (13.89% vs. 1.47%) and 2.96 times over FBPConvNet (13.89% vs. 4.69%). This superior performance across all three metrics validates that TAMP effectively enhances both subjective quality and clinical diagnostic utility for LACT images.

Manuscript changes: We have revised the radiologist validation results section to emphasize TAMP's clinical value in addressing NICT's inherently limited acceptability. Specifically, in the subsection "Radiologist validation: TAMP improves the clinical acceptability of NICT images", we have added:

"As shown in Fig.5d, despite NICT's inherently limited clinical acceptability, TAMP and TAMP-S efficiently improve the PCA by 54.60% and 56.73% respectively, demonstrating significant improvements that are recognized by radiologists."

[1] Xu, J., Zhao, Y., Li, H., & Zhang, P. (2019). An image reconstruction model regularized by edge-preserving diffusion and smoothing for limited-angle computed tomography. Inverse Problems, 35(8), 085004. doi:10.1088/1361-6420/ab08f9

[2] Wu, P., Tersol Montserrat, A., Clackdoyle, R., Boone, J., & Siewerdsen, J. (2023). Cone-beam CT sampling incompleteness: analytical and empirical studies of emerging systems and source-detector orbits. Journal of Medical Imaging, 10(3), 033503. doi:10.1117/1.JMI.10.3.033503

[3] Orth, R. C., et al. (2008). C-arm cone-beam CT: general principles and technical considerations for use in interventional radiology. Journal of Vascular and Interventional Radiology, 19(6), 814-820.

[4] Jaffray, D. A., et al. (2002). Flat-panel cone-beam computed tomography for image-guided radiation therapy. International Journal of Radiation Oncology Biology Physics, 53(5), 1337-1349.

2. For the question "It would be much more meaningful to investigate dedicated clinical applications. For example: using lung images with known nodules or lesions, are the lesions still visible if high noise is simulated and corrected for?"

We sincerely thank you for this insightful suggestion. To validate whether lesions remain visible after TAMP enhancement, we have conducted pulmonary nodule segmentation validation on the LIDC-IDRI dataset.

a) **Experiment settings:** We conducted pulmonary nodule segmentation experiments on the LIDC-IDRI dataset. We selected CT slices with expert consensus nodule masks, splitting them into 8,553, 856, 2,139 slices for training, validation, test respectively. ICT images were synthetically degraded into three NICT types (LDCT, SVCT, LACT) and subsequently enhanced by TAMP. A U-Net segmentation model was trained on ICT images with their corresponding masks, then tested on both NICT-degraded and TAMP-enhanced images to quantify segmentation accuracy improvements.

b) **Experiment results:** TAMP enhancement effectively restores segmentation accuracy degraded by NICT artifacts, validating improved diagnostic utility for lesion delineation. As shown in Table 15, TAMP improves Dice from 0.676, 0.527, 0.451 to 0.690, 0.662, 0.621 for LDCT, SVCT, LACT, corresponding to relative gains of +0.014, +0.135, +0.170. The substantial improvements in SVCT and LACT demonstrate that TAMP's learned representations capture geometric artifact patterns and effectively restore boundary information corrupted by angular undersampling.

Table 15: Our enhancement improves pulmonary nodule segmentation on LIDC-IDRI (w/o vs. w/ our method).

Setting	Dice			IoU		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.676	0.690	+0.014	0.527	0.544	+0.017
SVCT	0.527	0.662	+0.135	0.380	0.513	+0.133
LACT	0.451	0.621	+0.170	0.314	0.468	+0.154

c) **Manuscript changes:** We have added the pulmonary nodule segmentation experiment in subsection "Downstream clinical validation experiments" of the Supplementary Materials to validate TAMP's preservation of diagnostically relevant structures for lesion detection. Specifically:

"Pulmonary nodule segmentation is a critical component in lung cancer diagnosis, as accurate delineation of nodule boundaries directly impacts malignancy assessment and treatment planning. We conducted pulmonary nodule segmentation experiments on the LIDC-IDRI dataset to evaluate how TAMP enhancement improves automated lesion delineation accuracy.

We selected CT slices with expert consensus nodule masks, splitting them into 8,553, 856, 2,139 slices for training, validation, test respectively. ICT images were synthetically degraded into three NICT types (LDCT, SVCT, LACT) and subsequently enhanced by TAMP. A U-Net segmentation model was trained on ICT images with their corresponding masks, then tested on both NICT-degraded and TAMP-enhanced images to quantify segmentation accuracy improvements.

TAMP enhancement effectively restores segmentation accuracy degraded by

NICT artifacts, validating improved diagnostic utility for lesion delineation. As shown in Table 15, TAMP improves Dice from 0.676, 0.527, 0.451 to 0.690, 0.662, 0.621 for LDCT, SVCT, LACT, corresponding to relative gains of +0.014, +0.135, +0.170. The substantial improvements in SVCT and LACT demonstrate that TAMP's learned representations capture geometric artifact patterns and effectively restore boundary information corrupted by angular undersampling."

Table 15: Our enhancement improves pulmonary nodule segmentation on LIDC-IDRI (w/o vs. w/ our method).

Input	Dice			IoU		
	w/o Enhance	w/ Ours	Δ	w/o Enhance	w/ Ours	Δ
LDCT	0.676	0.690	+0.014	0.527	0.544	+0.017
SVCT	0.527	0.662	+0.135	0.380	0.513	+0.133
LACT	0.451	0.621	+0.170	0.314	0.468	+0.154

3. For the question "Or in tomosynthesis, can microcalcifications still be detected after correction of limited-angle artifacts using the proposed model? What would happen in the latter case, given that the model was trained using clinical CT data and is applied to tomosynthesis data that has a much higher spatial resolution but a much lower dynamic range? The important part here is to demonstrate the clinical applicability."

We sincerely thank you for this important observation. TAMP is designed for axial CT imaging systems (axial acquisition/reconstruction, 3D volume representation), which is incompatible with tomosynthesis that performs coronal/sagittal acquisition. Therefore, tomosynthesis is outside TAMP's scope. However, to demonstrate clinical applicability, we evaluated TAMP on cone-beam CT with limited-angle artifacts from radiotherapy workflows.

a) **Experiment settings:** We evaluated TAMP on clinical cone-beam CT with limited-angle artifacts (4 cases, 1576 slices from radiotherapy workflows, same dataset as Respons\#3 point 3). TAMP was evaluated in zero-shot and fine-tuned settings to assess clinical transferability across imaging modalities.

b) **Experiment results:** TAMP-S achieved clinically viable cross-modality transfer (PSNR: 30.59 dB, SSIM: 84.33%), effectively reconstructing corrupted vertebral structures for radiotherapy imaging and demonstrating practical clinical applicability beyond the training modality (see Fig. 19 and Table D in Response#3). This validates that parameter-efficient fine-tuning overcomes domain-specific differences in spatial resolution and dynamic range for clinical applications.

Table D: Comparative validation on limited-angle cone-beam CT (LA-CBCT) dataset.

Methods	**Training slice number**	**PSNR (dB)↑**	**SSIM (%)↑**	**RMSE (Hu)↓**	**LPIPS (%)↓**
LA-CBCT (Input)	-	23.48±4.77	75.16±10.39	522.17±298.40	36.12±7.66
RED-CNN	2,758	21.67±6.35	62.46±8.94	599.97±259.62	40.78±13.07
FBPConvNet	2,758	26.64±4.77	79.75±9.21	369.38±226.09	*26.09±8.49*
ProCT	2,758	*30.16±5.54*	*82.95±8.51*	*252.77±170.80*	32.04±10.87
TAMP (Ours)	0	25.12±5.51	71.39±10.43	415.73±220.76	37.26±9.18
TAMP-S (Ours)	2,758	**30.59±5.87**	**84.33±7.86**	**244.97±173.44**	**25.10±8.69**

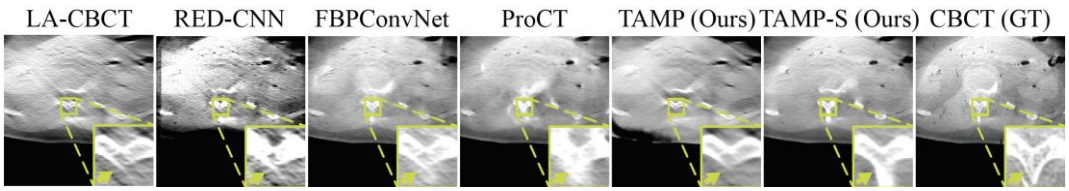

Fig. 19: Qualitative comparison on limited-angle cone-beam CT enhancement. TAMP effectively suppresses radial artifacts and TAMP-S further reconstructs corrupted anatomical details.

c) **Manuscript changes:** We have added the limited-angle cone-beam CT experiment in subsection "External-scene adaptation evaluation results" of the Supplementary Materials to demonstrate clinical applicability across axial CT modalities through cross-modality transfer validation. Specifically:

"We evaluate TAMP on an external cone-beam CT dataset with limited-angle artifacts to validate its capability across axial CT modalities. The dataset comprises 4 cases (1576 slices) from radiotherapy workflows. For baseline methods (RED-CNN, FBPConvNet, ProCT), we trained specialized models from scratch on the cone-beam CT training set. TAMP was evaluated in both zero-shot and fine-tuned (TAMP-S) settings.

As shown in Table 8 and Fig. 19, TAMP demonstrates effective knowledge transfer from diagnostic CT to cone-beam CT. TAMP-S achieves PSNR of 30.59 dB and SSIM of 84.33%, outperforming the best baseline ProCT by +0.11 dB and +1.60% respectively. Qualitatively, in the limited-angle CBCT input, vertebral bone structures are severely corrupted with radial wedge-shaped artifacts. The baseline methods fail to suppress these artifacts or reconstruct the damaged structures, while TAMP effectively suppresses artifacts and TAMP-S further reconstructs the corrupted vertebral bone structures approaching the ground truth."

Table 8: Quantitative evaluation on limited-angle cone-beam CT enhancement.

Method	PSNR (dB) ↑	SSIM (%) ↑	RMSE (Hu) ↓	LPIPS (%) ↓
LA-CBCT (Input)	23.48±4.77	75.16±10.39	522.17±298.40	36.12±7.66
RED-CNN	21.67±6.35	62.46±8.94	599.97±259.62	40.78±13.07
FBPConvNet	26.64±4.77	79.75±9.21	369.38±226.09	26.09±8.49
ProCT	30.16±5.54	82.95±8.51	252.77±170.80	32.04±10.87
TAMP	25.12±5.51	71.39±10.43	415.73±220.76	37.26±9.18
TAMP-S	**30.59±5.87**	**84.33±7.86**	**244.97±173.44**	**25.10±8.69**

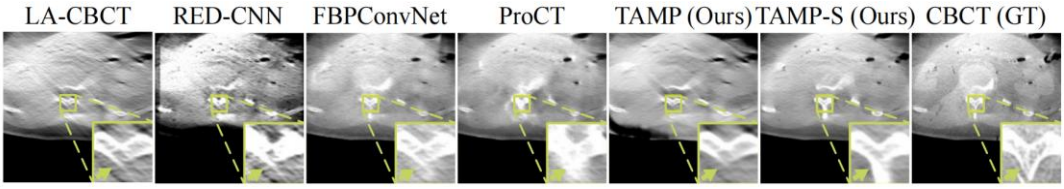

Fig. 19: Qualitative comparison on limited-angle cone-beam CT enhancement. TAMP effectively suppresses radial artifacts and TAMP-S further reconstructs corrupted anatomical details.

Comments from reviewer 1:

Comment#1: Thank you for the revision and rebuttal. I have reviewed the revised manuscript and responses. All of my major concerns have been satisfactorily addressed, and the added experiments and clarifications significantly strengthen the work. I have no further comments and am happy with the manuscript in its current form.

Response#1: We sincerely thank you for the thorough and constructive review throughout the revision process. We are glad that the added experiments and clarifications have addressed all major concerns and strengthened the work.

Comment#2: The repo appears well-organized and clearly documented. Due to time constraints, I have not attempted to run the code and the model.

Response#2: We appreciate your positive feedback on our code repository. The complete codebase, pretrained models, and documentation remain publicly available to facilitate reproducibility.

Comments from reviewer 2:

Comment#1: The reviewer thoroughly addressed my previous concerns and the manuscript was improved for publication. No more comments.

Response#1: We sincerely thank you for the valuable suggestions during the review process. We are pleased that the revisions have fully addressed the previous concerns and improved the manuscript to publication standard.

Comments from reviewer 3:

Comment#1: The authors have adequately addressed all of my concerns, and I particularly appreciate the inclusion of clinically meaningful experiments.

Response#1: We sincerely thank you for the insightful feedback. We are glad that the inclusion of clinically meaningful experiments has been well received and has strengthened the clinical relevance of our work.