

Imaging foundation model for universal enhancement of non-ideal measurement CT (*Supplementary Material*)

Contents

1	More details of SimNICT dataset	2		
1.1	Simulation of NICT for pre-training	2		
1.2	Simulation of NICT for validation experiments	3		
2	More details of real-world NICT	3		
2.1	Detail information of real-world data	3		
2.2	Reconstruction of real-world data	3		
3	More details of Experiment	4		
3.1	Radiologist validation application	4		
3.2	Evaluation metrics	4		
3.2.1	NICT image quality	4		
3.2.2	Model's Clinical Application Value	5		
4	More technical details of TAMP	6		
4.1	Loss function in DDEL	6		
4.2	More details of TAMP pre-training	7		
4.3	More details of TAMP adaptation	8		
4.4	More details of MITNet architecture	8		
5	More experiment results	8		
5.1	Additional qualitative results	8		
5.2	Additional quantitative results	8		
5.3	External-scene adaptation evaluation results	8		
5.3.1	Sparse-view CBCT	9		
5.3.2	Limited-angle CBCT	10		
5.3.3	Low-dose Micro-CT	10		
5.4	Universal enhancement validation	11		
5.4.1	Mixed test cohort evaluation	11		
5.4.2	Oracle ensemble comparison	12		
5.5	Ablation study	13		
5.5.1	Component ablation	13		
5.5.2	Training data size ablation	13		
5.5.3	Model size ablation	14		
5.6	Downstream clinical validation	14		
			5.6.1	Pulmonary nodule segmentation
				14
			5.6.2	Pulmonary nodule malignancy diagnosis
				15
			5.6.3	Radiomics feature analysis
				16
			5.7	TAMP model analysis
				16

1 More details of SimNICT dataset

SimNICT is the dataset used for TAMP pre-training, derived from CT images across ten publicly available CT datasets to simulate a diverse range of NICT images. These source datasets cover various body regions, including the head, chest, abdomen, and lower limbs. By filtering out low-quality data, SimNICT ultimately includes 9,513 volumes with over 3.6 million slices for TAMP pre-training, while an additional 125 volumes with over 30 thousand slices from the AMOS22, AutoPET, and CT Images in COVID-19 datasets are used for downstream validation experiments. Detailed information is provided in Table 1.

The SimNICT pre-training data (8 out of 10 datasets, excluding AutoPET and HECKTOR22 due to original licensing restrictions) and complete evaluation data for all 27 benchmark tasks are available on HuggingFace (<https://huggingface.co/datasets/YutingHe-list/SimNICT>), with data hosted on Internet Archive (https://archive.org/details/@tamp_research_group). The Mayo Clinic LDCT dataset is publicly available at <https://www.aapm.org/GrandChallenge/LowDoseCT/>.

1.1 Simulation of NICT for pre-training

The SimNICT dataset used for TAMP pre-training simulates NICT images by introducing defects during the CT image projection and reconstruction processes (corresponding to Section Result1 of the paper). This process models a fan-beam CT with a full 720-degree angular view, implemented using code from The Operator Discretization Library (ODL, v1.0.0, <https://github.com/odlgroup/odl>), as represented by:

$$P_{\text{ICT}} = P_{\text{proj}}(I_{\text{ICT}}, \theta_{\text{Full}}), \quad \theta_{\text{Full}} = \{1, 2, \dots, 720\},$$

where I_{ICT} represents the ideal measurement CT image from SimNICT, and P_{ICT} denotes the simulated projection data. The process of introducing defects and reconstruction for the NICT images of three settings is presented below.

1) **LDCT**. The simulation method for LDCT images follows the approach used in the 2016

NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge ^{1, 2}. In this noise simulation process, quantum noise x is modeled based on photon counting statistics, following a Poisson distribution. For high photon counts, this is approximated by a Gaussian distribution. Noise x is introduced as a normally distributed random variable applied to adjust the initial measurement P_{ICT} for simulating lower dose conditions, expressed as:

$$P_{\text{LD}} = P_{\text{ICT}} + \sqrt{\frac{1-a}{a} \cdot \frac{\exp(P_{\text{ICT}})}{N_0}} \cdot x,$$

where $a = \frac{N_{\text{LD}}}{100}$, and $a \in (0, 1)$ is the scaling factor that adjusts the dose reduction through the parameter N_{LD} , and $N_0 = 1 \times 10^6$ is the initial incident photon count. The data is then reconstructed into the LDCT image as:

$$I_{\text{LD}} = P^{-1}(P_{\text{LD}}, \theta_{\text{Full}}).$$

2) **SVCT**. The simulation of SVCT is achieved by reconstructing the P_{ICT} projection data, which is evenly sampled based on a specified number of angles through the parameter N_{SV} , as follows:

$$I_{\text{SV}} = P^{-1}(P_{\text{ICT}}, \theta_{\text{SV}}),$$

$$\theta_{\text{SV}} = \left\{ \theta_i \mid \theta_i = \left\lfloor \frac{720 \cdot i}{N_{\text{SV}}} \right\rfloor, i = 1, 2, \dots, N_{\text{SV}} \right\}.$$

3) **LACT**. The simulation of LACT is achieved by reconstructing the P_{ICT} projection data, which is continuously sampled based on a range of angles through the parameter N_{LA} , as follows:

$$I_{\text{LA}} = P^{-1}(P_{\text{ICT}}, \theta_{\text{LA}}),$$

$$\theta_{\text{LA}} = \{1, 2, \dots, N_{\text{LA}}\}.$$

During the construction of SimNICT, each source CT slice is simulated into one LDCT, one SVCT, and one LACT image. In this simulation process, the parameters N_{LD} , N_{SV} , and N_{LA} are randomly sampled within a range to enhance the diversity of the SimNICT data, thereby improving the model's adaptability to different defect levels, with the range settings presented in Supplementary Table 2.

Supplementary Table 1: Source Datasets for SimNICT Pre-training and Validation Experiments.

Dataset name	Body regions	Original volumes	Pre-train/Downstream volumes	link	license
COVID-19-NY-SBU	Chest	1,384	459/0	✓	CC BY 4.0
STOIC	Chest	2,000	2,000/0	✓	CC BY-NC 4.0
MELA	Chest	1,100	1,100/0	✓	CC BY 4.0
LUNA	Chest	888	888/0	✓	CC BY 4.0
LNDb	Chest	294	294/0	✓	CC BY-NC-ND 4.0
HECKTOR22	Head, neck	882	882/0	✓	CC BY-NC-ND 4.0
CT_COLONOGRAPHY	Abdomen	1,730	1,730/0	✓	CC BY 4.0
CT Images in COVID-19	Chest	771	736/35	✓	CC BY 4.0
AutoPET	Head, chest, abdomen, lower-limbs	1,014	979/35	✓	TCIA Restricted
AMOS22	Abdomen	500	445/55	✓	CC BY 4.0

Supplementary Table 2: The settings of N_{LD} , N_{SV} , and N_{LA} for NICT simulation in the SimNICT dataset and validation experiment with three types of defect degrees.

Condition	Defect degree	N_{LD}	N_{SV}	N_{LA}
SimNICT	random	[5, 75]	[15, 360]	[75, 270]
Validation experiment	low	60	300	150
	mid	40	120	120
	high	20	60	90

1.2 Simulation of NICT for validation experiments

In the validation experiments, to specifically evaluate the capability of various methods on different types of NICT enhancement tasks, we configured three defect degrees (Low, Mid, and High) for each NICT setting (LDCT, SVCT, and LACT) in the simulation by adjusting N_{LD} , N_{SV} , and N_{LA} , as detailed in Supplementary Table 2.

2 More details of real-world NICT

We collect and reconstruct real-world NICT from diverse clinical sources including Nanjing Drum Tower Hospital Jiangbei and the Mayo Clinic Low Dose CT Grand Challenge to evaluate the adaptation of TAMP to real-world NICT. The details of the relevant information and reconstruction parameters are as follows.

2.1 Detail information of real-world data

The real-world validation dataset comprises four sources from diverse clinical scenarios and imaging protocols: NJDTH-A, NJDTH-B, and NJDTH-C from Nanjing Drum Tower Hospital Jiangbei

(NJDTH), and Mayo from the Mayo Clinic Low Dose CT Grand Challenge.

NJDTH-A: LDCT paired data from six patients (1 male and 5 females, aged 49-79 years) with 1,496 slices. Patients underwent dual scans at low dose (80 kVp, 30 mAs) and high dose (120 kVp, 90 mAs). 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.

NJDTH-B: LDCT paired data from ten patients with 2,802 slices. Patients underwent dual scans at low dose (120 kVp, 50 mAs) and high dose (120 kVp, 200 mAs), providing validation for current-reduction-based LDCT protocols.

NJDTH-C: Raw projection data from six patients (5 males and 1 female, aged 43-79 years) with 750 slices acquired at 120 kVp and 39 mA. 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.

Mayo: Publicly available projection data from the Mayo Clinic Low Dose CT Grand Challenge with ten patients. SVCT and LACT variants were reconstructed by undersampling the projection data at sparse views and limited angles, providing independent validation across different institutions.

2.2 Reconstruction of real-world data

NJDTH-A and NJDTH-B: For LDCT paired data, to address respiratory-induced misalignment between LDCT and ICT, we performed image registration using Elastix (v5.1.0) ^{3, 4}. The registration employed lung-specific parameters from

ElastixModelZoo¹, using ICT as fixed and LDCT as moving images, followed by cropping of non-overlapping regions.

NJDTH-C: For SVCT and LACT, using the TIGRE toolbox (v2.6) ⁵ in MATLAB (R2022b), we reconstructed SVCT and LACT from raw projection data following the downsampling protocol in Section 1.1 with $N_{SV} = 240$ and $N_{LA} = 120$.

Mayo: For SVCT and LACT, we first converted the helical CT projection data to fan-beam projections using helix2fan conversion, then performed iterative reconstruction using the ASTRA Toolbox (v2.1.2, <https://astra-toolbox.com/index.html>) with the CGLS_CUDA algorithm for 50 iterations, with $N_{SV} = 180$ and $N_{LA} = 140$.

3 More details of Experiment

3.1 Radiologist validation application

In the radiologist validation experiment, the enhancement results of each method on NICT images are presented to radiologists for scoring. To facilitate the experts' observation and scoring of the CT images, we developed an application with an interactive website to enable a more convenient and accurate evaluation, as shown in Supplementary Figure 1. We use the Flask Toolbox (<https://github.com/pallets/flask>) to load and present CT images in an interactive webpage. The webpage displays a group of four CT images at a time, allowing radiologists to rank the images based on quality and assess their clinical acceptability. Radiologists can also easily adjust window width and level, and zoom in on specific regions for detailed observation. Moreover, the detail of the process of our radiologist validation are shown in Supplementary Figure 2.

3.2 Evaluation metrics

3.2.1 NICT image quality

To evaluate the enhanced NICT image quality, we utilize four image quality assessment metrics: root mean square error (RMSE) ⁶, peak signal-to-noise ratio (PSNR) ⁶, structural similarity index

measure (SSIM) ⁷, and learned perceptual image patch similarity (LPIPS) ⁸. These metrics are defined as follows:

1) **RMSE** calculates the average difference between the pixel intensities of the ground truth image x and the predicted image $\hat{x}$, representing the pixel-wise error in the image, defined by:

$$\text{RMSE}(x, \hat{x}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x(i) - \hat{x}(i))^2},$$

where N is the total number of pixels in the image.

2) **PSNR** calculates the ratio of the maximum possible pixel value to the noise (RMSE), representing the peak signal-to-noise ratio and indicating the overall quality of the image, defined by:

$$\text{PSNR}(x, \hat{x}) = 20 \log_{10} \left(\frac{M - 1}{\text{RMSE}(x, \hat{x})} \right),$$

where $M = 4096$ for the image with units of Hu.

3) **SSIM** calculates the similarity between the luminance, contrast, and structure of the two images, representing the perceptual quality and structural similarity between the ground truth and predicted images, defined by:

$$\text{SSIM}(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)},$$

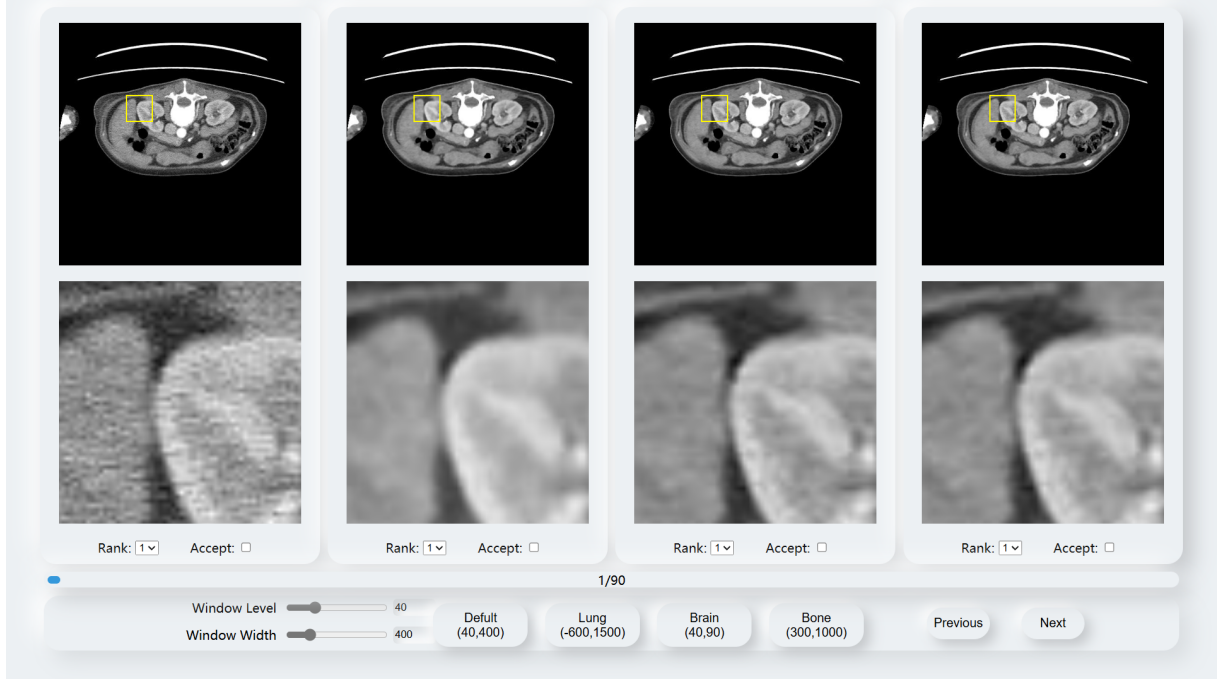
where μ_x and $\mu_{\hat{x}}$ are the mean values of x and $\hat{x}$ respectively, σ_x^2 and $\sigma_{\hat{x}}^2$ are the variances of x and $\hat{x}$, and $\sigma_{x\hat{x}}$ is the covariance between x and $\hat{x}$. The constants $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$, with $k_1 = 0.01$ and $k_2 = 0.03$ by default.

4) **LPIPS** calculates the perceptual distance between image patches using deep neural network features, representing the perceptual similarity based on learned feature spaces, defined by:

$$\text{LPIPS}(x, \hat{x}) = \sum_{l=1}^L w_l \cdot \|\phi_l(x) - \phi_l(\hat{x})\|_2,$$

where L is the total number of layers in the network, w_l is the weight assigned to the l -th layer, while $\phi_l(x)$ and $\phi_l(\hat{x})$ are the feature maps from the l -th layer for images x and $\hat{x}$, respectively, and $\|\cdot\|_2$ denotes the L2 norm.

¹Parameters: <https://github.com/SuperElastix/ElastixModelZoo/blob/master/models/Par0004>



Supplementary Figure 1: The scoring website for the radiologist validation.

3.2.2 Model's Clinical Application Value

In the radiologist validation experiment, to provide an intuitive assessment of each model's clinical application value, we designed three metrics to statistically analyze the scoring data from radiologists: probability of better than NICT (PBN), subjective quality ranking (SQR), and probability of clinical acceptability (PCA). These metrics are defined as follows:

1) **PBN** calculates the probability that the model's enhanced image is rated higher than the original NICT image, describing the significance of the model in improving the quality of NICT, defined by:

$$\text{PBN}(x) = \frac{1}{|S_{\text{NICT}_x, \text{NICT}}|} \sum_{s \in S_{\text{NICT}_x, \text{NICT}}} [R_{\text{NICT}_x}^s > R_{\text{NICT}}^s],$$

where x denotes a model, NICT_x represents the NICT image enhanced by model x , $S_{\text{NICT}_x, \text{NICT}}$ denotes the set of groups where both NICT_x and NICT are selected for scoring, $|\cdot|$ indicates the cardinality of the groups, and $R_{\text{NICT}_x}^s$ and R_{NICT}^s represent the subjective quality ratings of the NICT_x and NICT, respectively, in group s .

2) **SQR** calculates the average subjective quality rating of the model's enhanced images, describing the efficiency of the model in improving the

quality of NICT, defined by:

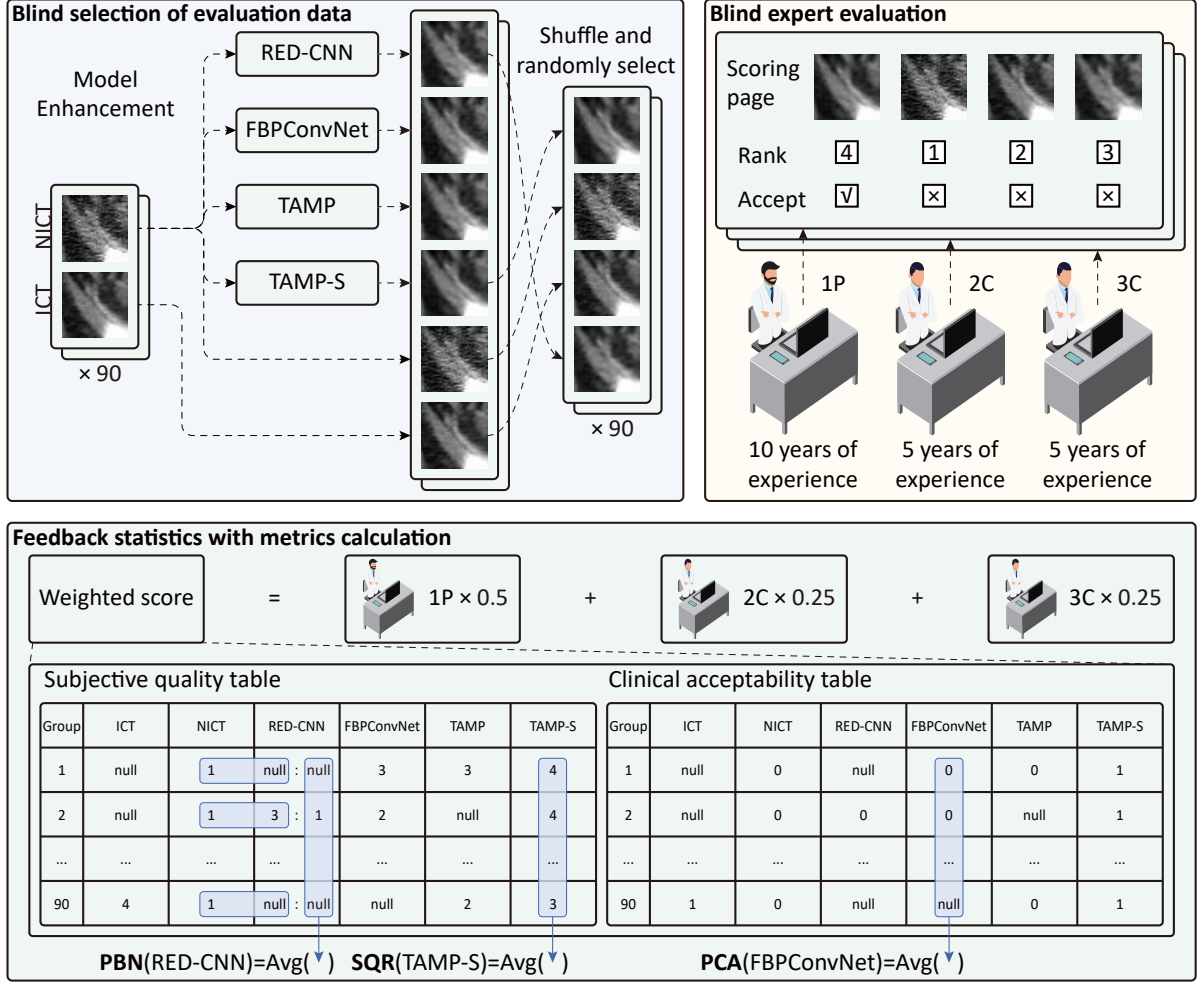
$$\text{SQR}(x) = \frac{1}{|S_{\text{NICT}_x}|} \sum_{s \in S} R_{\text{NICT}_x}^s,$$

where S_{NICT_x} denotes the set of groups where NICT_x is selected for scoring.

3) **PCA** calculates the probability of clinical acceptability of the model's enhanced images, describing the capability of the model in making NICT images more acceptable for clinical use, defined by:

$$\text{PCA}(x) = \frac{1}{|S_{\text{NICT}_x}|} \sum_{s \in S} [A_{\text{NICT}_x}^s],$$

where $A_{\text{NICT}_x}^s$ represents the clinical acceptability of the model's enhanced CT image in group s , given by radiologists. $[\cdot]$ denotes the Iverson bracket, which is 1 if the statement is true and 0 if false.



Supplementary Figure 2: The process of our radiologist validation. CT images are sourced from AMOS22²⁰, and the cartoon icons are adapted from Freepik (https://www.freepik.com/free-vector/medical-people-health_5669117.htm and https://www.freepik.com/free-vector/bundle-professional-medical-staff-icons_5720502.htm).

4 More technical details of TAMP

4.1 Loss function in DDEL

Dual-domain enhancement learning (DDEL) is designed with composite loss functions to enable TAMP to learn both image and projection domain features of NICT images, expressed as:

$$\mathcal{L}_{\text{train}} = w_1 \mathcal{L}_{\text{MSE}}^I + w_2 \mathcal{L}_{\text{SSIM}}^I + w_3 \mathcal{L}_{\text{VGG}}^I + w_4 \mathcal{L}_{\text{MSE}}^P,$$

where w_1, w_2, w_3, w_4 are set to 1.0, 5.0×10^{-3} , 1.0×10^{-4} , and 5.0×10^{-4} , respectively. The four loss terms are defined as follows:

1) $\mathcal{L}_{\text{MSE}}^I$ is commonly used in training self-supervised image denoising models⁹ and aims to minimize pixel-wise differences between the predicted and ground truth images, serving as a basic measure of enhancement quality. It is defined as:

$$\mathcal{L}_{\text{MSE}}^I = \frac{1}{N_I} \sum_{i=1}^{N_I} (I_{\text{Pred}}(i) - I_{\text{ICT}}(i))^2,$$

where I_{Pred} is the predicted image, I_{ICT} is the ground truth image, and N_I is the total number of pixels in the image.

2) $\mathcal{L}_{\text{SSIM}}^I$ calculates the structural similarity index, which assesses image similarity based on luminance, contrast, and structure¹⁰. This loss function improves perceptual image quality by aligning structural features. It is defined as:

$$\mathcal{L}_{\text{SSIM}}^I = 1 - \text{SSIM}(I_{\text{Pred}}, I_{\text{ICT}}),$$

where $\text{SSIM}(I_{\text{Pred}}, I_{\text{ICT}})$, as defined in Section 3.2.1, measures the structural similarity between the predicted and ground truth images.

3) $\mathcal{L}_{\text{VGG}}^I$ computes the difference between the feature maps extracted from a pre-trained VGG19 network, encouraging the model to preserve high-level perceptual features¹¹. This loss helps to improve the perceptual quality of the enhanced images. It is defined as:

$$\mathcal{L}_{\text{VGG}}^I = \frac{1}{N_\phi} \sum_{i=1}^{N_\phi} \|\phi(I_{\text{Pred}})(i) - \phi(I_{\text{ICT}})(i)\|_1,$$

where ϕ represents the feature maps extracted from the 35th layer of a pre-trained VGG19 network, and N_ϕ is the number of features in the layer.

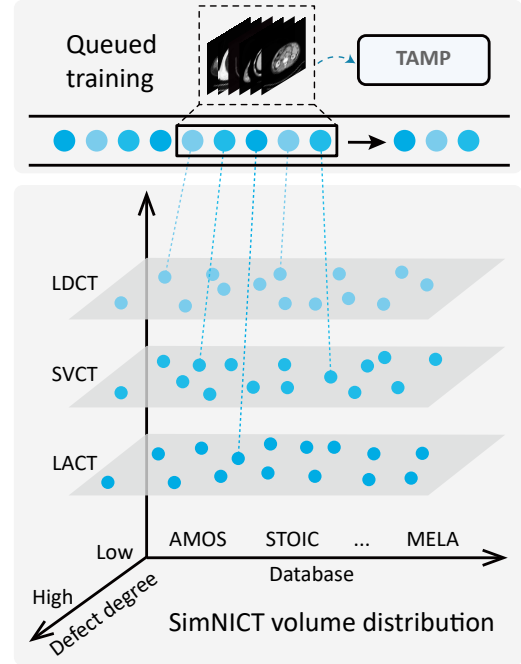
4) $\mathcal{L}_{\text{MSE}}^P$ calculates the pixel-wise mean squared error in the projection domain, ensuring that the model's predicted projections are consistent with the original projections. This loss ensures proper alignment in the projection domain. It is defined as:

$$\mathcal{L}_{\text{MSE}}^P = \frac{1}{N_P} \sum_{i=1}^{N_P} (P(I_{\text{Pred}})(i) - P(I_{\text{ICT}})(i))^2,$$

where P denotes the projection operation, and N_P is the number of projections. The projection operation uses the same simulation environment and toolkit (ODL) as in Section 1.1, and is executed on CUDA, allowing the computed loss to be backpropagated through the model for weight updates.

4.2 More details of TAMP pre-training

TAMP is implemented in PyTorch (v2.10.0, <https://pytorch.org/>) with Python 3.10.13 and CUDA 12.8. It is pre-trained on SimNICT datasets to achieve universal NICT enhancement capability. Firstly, to reduce the time consumption of inputting and outputting (IO) large-scale data on disk during training, a queued training process is designed. As shown in Supplementary Figure 3, it loads N (we set $N = 5$) NICT volumes into memory as a queue and shuffles the slices for learning. To avoid over-fitting for a fixed defect pattern in the training process, we ensure that all NICT settings exist in the queue. Once all the slices have been iterated, the oldest volume is removed from the queue, and a new NICT volume is loaded, thus effectively reducing the IO cost. Secondly, to improve model convergence, we use warm-up at the beginning of the TAMP pre-training. The hyperparameter settings are provided in Supplementary Table 3.



Supplementary Figure 3: The queued training process designed for TAMP pre-training (CT images sourced from AutoPET²¹).

Supplementary Table 3: Hyperparameter settings used in the TAMP pre-training.

Hyperparameter	Value
Optimizer	Adan ¹²
Warm-up epochs	10
Full epochs	28,539
Initial learning rate at warm-up	0.00005
Learning rate after warm-up	0.0005
Learning betas	[0.98, 0.92, 0.99]
Epochs per learning rate decay	100
Learning rate decay rate	0.95
Batch size on each GPU	5
Training environment	2× NVIDIA RTX 3090 GPUs

4.3 More details of TAMP adaptation

We employ LoRA ¹³, a parameter-efficient fine-tuning method, to adapt our TAMP to specialized NICT enhancement tasks with few training data and computation. LoRA tunes only a small number of parameters, reducing the risk of overfitting and enabling professional performance with limited data. Specifically, we add low-rank matrix bypasses to the linear and convolutional layers in the network, tuning only the parameters in the LoRA bypasses while keeping the original TAMP parameters fixed. The same optimizer, learning rate, learning betas, batch size, and training environment are used as in pre-training after the warm-up phase, with the learning rate halved every 10 epochs.

4.4 More details of MITNet architecture

Our MITNet constructs a transformer ¹⁴ architecture that is compatible with multi-scale features to represent the varied defect patterns in different NICT images. As shown in Fig. 4, MITNet first uses convolutions with varying kernel sizes to extract features from NICT images at four scales. It then employs multiple Deep Feature Extraction Transformer (DFET) blocks ¹⁵, with Swin Transformer layers and convolutional operations alternated across multiple levels, to refine defects at each scale. This is followed by a step-by-step fusion process, restoring features to the original NICT image scale, and culminating in an

enhancement phase that produces the final output image.

5 More experiment results

5.1 Additional qualitative results

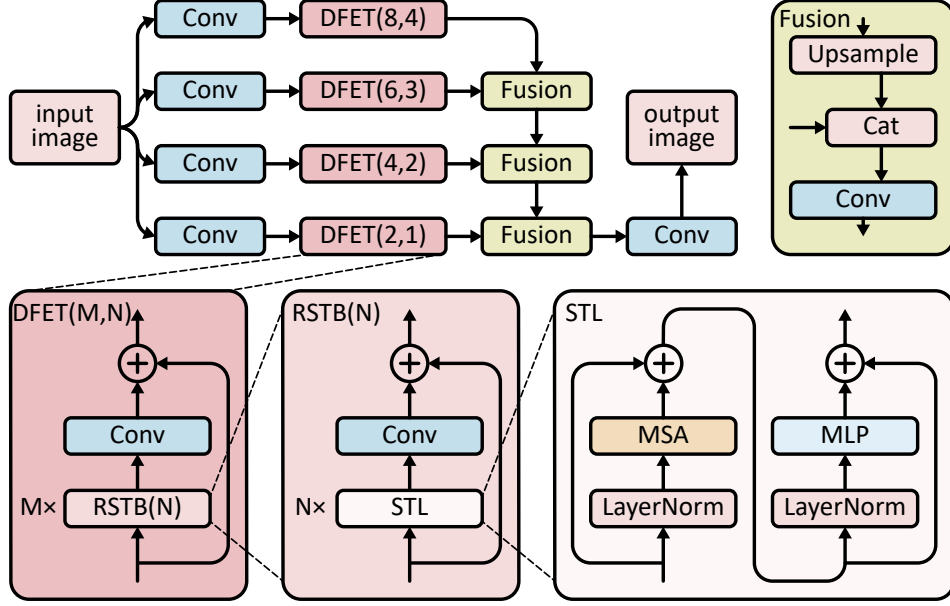
Our TAMP achieves superior image quality improvement in diverse NICT enhancement tasks across different NICT settings (LDCT, LACT, SVCT) and body regions (head, chest, abdomen, lower-limbs), as shown in Supplementary Figures 5–16. Moreover, TAMP demonstrates equally effective enhancement for NICTs with lesions, highlighting its significant potential in assisting medical diagnosis across a wide range of clinical scenarios, as shown in Supplementary Figure 17.

5.2 Additional quantitative results

The universal NICT enhancement performance of TAMP has been quantitatively evaluated across 27 tasks, including three datasets (AMOS22, AutoPET, COVID-19), three NICT settings (LDCT, LACT, SVCT), and three defect degrees (High, Mid, Low), using four evaluation metrics, i.e., PSNR, SSIM, LPIPS, and RMSE, as shown in Supplementary Tables 4–7.

5.3 External-scene adaptation evaluation results

We evaluated the performance of TAMP in adapting to sparse-view cone beam CT (CBCT) and low-dose Micro-CT enhancement tasks to



Supplementary Figure 4: Network architecture of MITNet.

Supplementary Table 4: Quantitative evaluation of PSNR (dB) for different methods across 27 tasks. The best results are highlighted in **bold**, and the second-best results are underlined.

Task	RedCNN	WGAN-VGG	FBPConvNet	SPECIAL	TransCT	MITNet	ProCT	TAMP	TAMP-S
AMOS-LDCT-High	48.96±1.59	47.78±1.40	50.07±1.90	49.50±1.83	46.30±1.08	46.49±1.24	47.79±1.25	50.49±1.95	50.65±1.99
AMOS-LDCT-Mid	50.13±1.71	44.99±1.01	51.43±2.06	47.29±0.98	47.70±1.30	46.92±1.33	48.38±1.37	51.87±2.15	52.03±2.19
AMOS-LDCT-Low	51.38±1.91	45.18±1.10	<u>52.40±2.19</u>	52.11±2.26	48.44±1.44	47.53±1.43	48.61±1.44	51.94±1.88	53.05±2.31
AMOS-LACT-High	37.46±0.74	38.30±0.76	38.84±0.71	39.81±0.96	38.59±0.66	39.37±0.65	35.53±0.72	40.30±0.81	42.48±0.98
AMOS-LACT-Mid	37.46±1.00	37.91±0.84	40.89±1.02	41.49±1.46	39.40±0.77	40.29±0.83	36.54±0.52	41.67±0.88	44.59±1.23
AMOS-LACT-Low	40.21±1.40	38.85±1.03	41.81±1.45	<u>44.03±1.63</u>	40.96±1.17	41.54±0.96	37.83±0.64	42.78±1.15	47.49±1.56
AMOS-SVCT-High	41.50±1.04	39.23±0.84	40.74±0.78	41.65±0.88	40.51±1.00	41.62±0.92	<u>41.84±0.74</u>	41.69±1.14	44.20±1.17
AMOS-SVCT-Mid	43.49±1.25	41.31±1.02	43.30±1.19	<u>44.72±1.39</u>	41.54±1.07	42.82±0.97	43.63±0.91	44.09±1.54	47.21±1.58
AMOS-SVCT-Low	49.02±1.85	43.24±0.96	48.59±1.76	49.10±1.97	45.44±1.40	45.25±1.21	46.78±1.07	<u>50.67±1.79</u>	51.85±2.27
APET-LDCT-High	49.13±1.21	41.33±0.90	50.75±1.58	50.38±1.46	47.69±1.11	48.20±1.13	48.00±0.95	<u>50.25±1.42</u>	50.87±1.64
APET-LDCT-Mid	50.31±1.18	40.29±0.92	<u>51.91±1.61</u>	51.82±1.47	47.48±1.03	48.45±1.03	48.66±0.92	51.43±1.32	52.13±1.61
APET-LDCT-Low	50.75±0.88	40.30±0.80	52.61±1.43	<u>52.74±1.41</u>	50.02±1.03	49.28±1.00	48.70±0.91	52.29±1.20	53.10±1.54
APET-LACT-High	36.95±0.91	38.32±0.47	40.69±1.06	41.06±0.90	39.28±0.83	40.35±1.01	37.00±0.82	41.25±0.85	42.68±1.07
APET-LACT-Mid	38.82±0.82	37.91±0.56	41.71±1.01	<u>42.72±0.99</u>	40.13±0.94	41.75±0.95	37.36±0.63	42.40±0.86	44.56±1.12
APET-LACT-Low	39.80±1.00	38.39±0.71	44.22±1.10	<u>45.71±1.12</u>	42.07±1.16	42.30±1.15	38.41±0.67	43.44±1.29	47.50±1.15
APET-SVCT-High	41.31±0.76	39.66±0.72	40.06±1.36	42.74±1.02	41.35±0.81	42.37±0.90	40.39±0.66	43.15±0.94	44.31±1.19
APET-SVCT-Mid	43.85±1.07	40.53±0.85	<u>47.16±1.26</u>	46.43±1.24	43.11±1.07	44.42±1.07	43.12±0.80	46.04±1.13	47.18±1.37
APET-SVCT-Low	46.33±1.52	43.63±0.87	51.93±1.26	50.51±1.14	47.07±1.16	47.18±1.03	46.98±0.83	50.76±1.06	<u>51.60±1.26</u>
COVID-LDCT-High	44.46±0.65	40.18±0.56	42.44±0.82	45.37±0.82	42.33±0.91	41.82±0.76	42.97±1.09	46.55±0.78	46.90±0.82
COVID-LDCT-Mid	45.02±0.75	40.13±0.52	43.25±1.09	46.87±1.06	43.06±1.07	42.36±0.81	43.12±1.16	<u>47.54±0.88</u>	47.99±0.92
COVID-LDCT-Low	45.27±0.96	47.15±5.96	43.56±1.13	46.95±1.01	42.92±1.15	42.04±0.84	43.17±1.18	<u>48.12±0.97</u>	48.67±1.01
COVID-LACT-High	37.37±0.56	38.80±0.78	38.41±0.67	38.92±0.50	38.40±0.39	38.59±0.52	35.67±0.42	<u>39.86±0.65</u>	40.99±0.77
COVID-LACT-Mid	38.01±0.63	38.06±0.45	33.28±0.41	39.82±0.64	38.41±0.41	38.85±0.53	36.94±0.52	<u>40.65±0.77</u>	42.48±0.89
COVID-LACT-Low	38.37±0.96	37.77±0.77	39.14±0.86	41.24±0.79	39.02±0.68	38.94±0.48	37.45±0.58	<u>42.15±0.81</u>	44.60±1.07
COVID-SVCT-High	39.51±0.68	38.09±0.51	38.27±0.57	40.13±0.67	38.74±0.63	39.39±0.51	37.25±0.89	41.04±0.70	41.30±0.77
COVID-SVCT-Mid	39.93±0.97	37.93±0.41	39.82±0.92	41.17±0.95	39.25±0.80	40.38±0.72	39.31±1.17	<u>42.48±0.99</u>	42.84±1.02
COVID-SVCT-Low	42.44±1.02	39.35±0.60	43.25±1.35	43.92±1.26	41.23±1.13	41.16±0.79	42.31±1.18	<u>45.66±1.10</u>	46.04±1.10

assess its generalization capability as a foundation model.

5.3.1 Sparse-view CBCT

We evaluate the adaptability of TAMP to the sparse-view CBCT enhancement task. The clinical Varian and Elekta datasets, sourced from the

Supplementary Table 5: Quantitative evaluation of SSIM (%) for different methods across 27 tasks. The best results are highlighted in **bold**, and the second-best results are underlined.

Task	RedCNN	WGAN-VGG	FBPConvNet	SPECIAL	TransCT	MITNet	ProCT	TAMP	TAMP-S
AMOS-LDCT-High	99.03±0.50	98.82±0.53	99.17±0.50	99.09±0.52	98.38±0.55	98.51±0.57	98.86±0.52	99.23±0.48	99.24±0.49
AMOS-LDCT-Mid	99.25±0.42	98.40±0.59	99.38±0.43	98.34±0.53	98.81±0.47	98.55±0.55	99.03±0.49	<u>99.42±0.42</u>	99.43±0.42
AMOS-LDCT-Low	99.43±0.39	97.83±0.60	<u>99.50±0.38</u>	99.47±0.42	98.96±0.45	98.90±0.51	99.10±0.48	<u>99.46±0.37</u>	99.54±0.37
AMOS-LACT-High	64.64±7.96	59.07±5.60	83.49±2.91	85.74±4.06	79.31±3.89	82.08±2.82	67.24±4.17	<u>90.32±2.35</u>	93.41±1.57
AMOS-LACT-Mid	78.79±5.82	71.86±5.48	89.80±2.94	91.17±4.18	85.04±3.68	87.67±2.87	78.67±4.02	<u>94.50±1.47</u>	96.57±1.00
AMOS-LACT-Low	89.19±4.16	85.42±5.04	90.83±3.52	<u>96.06±1.45</u>	91.31±2.65	92.14±1.86	85.57±2.99	<u>95.70±1.54</u>	98.35±0.60
AMOS-SVCT-High	93.12±1.78	88.47±2.81	90.90±1.70	<u>93.29±1.51</u>	90.93±2.21	92.49±1.73	93.86±1.14	<u>93.67±1.66</u>	96.60±0.88
AMOS-SVCT-Mid	96.43±1.13	94.30±1.68	96.10±1.13	97.24±0.98	94.52±1.53	95.44±1.12	97.20±0.73	<u>97.73±0.71</u>	98.49±0.57
AMOS-SVCT-Low	99.01±0.53	95.15±0.91	99.01±0.55	98.98±0.55	97.95±0.68	97.91±0.69	98.94±0.51	<u>99.33±0.41</u>	99.43±0.44
APET-LDCT-High	99.05±0.36	90.79±2.26	<u>99.28±0.30</u>	99.22±0.30	98.72±0.44	98.83±0.39	98.88±0.33	99.19±0.32	99.27±0.30
APET-LDCT-Mid	99.30±0.24	86.88±3.23	<u>99.45±0.24</u>	99.43±0.23	98.86±0.25	98.60±0.46	99.05±0.25	99.40±0.23	99.45±0.23
APET-LDCT-Low	99.34±0.22	86.99±3.47	<u>99.55±0.19</u>	99.54±0.17	99.27±0.20	99.12±0.23	99.06±0.25	99.52±0.18	99.56±0.18
APET-LACT-High	68.36±8.23	65.95±8.27	90.34±3.36	91.57±2.34	84.34±3.92	88.41±3.24	70.37±7.45	<u>92.40±1.75</u>	94.79±1.56
APET-LACT-Mid	86.49±3.58	76.60±6.00	93.32±2.05	94.76±1.59	88.97±3.14	92.68±1.99	77.00±6.58	<u>95.48±1.32</u>	97.04±0.89
APET-LACT-Low	90.68±2.78	85.51±4.26	97.23±0.98	<u>97.62±0.76</u>	93.61±1.95	94.18±1.59	87.25±3.74	<u>96.00±2.16</u>	98.53±0.44
APET-SVCT-High	93.59±1.47	89.36±2.37	94.99±1.20	95.16±1.27	93.19±1.56	94.58±1.34	94.33±1.16	<u>96.11±0.98</u>	96.79±0.90
APET-SVCT-Mid	96.79±1.74	90.87±2.65	<u>98.41±0.42</u>	98.17±0.49	96.28±0.87	97.06±0.69	97.19±0.57	<u>98.11±0.49</u>	98.42±0.45
APET-SVCT-Low	98.10±0.74	96.22±1.02	99.44±0.17	99.27±0.18	98.53±0.34	98.63±0.28	98.86±0.29	99.33±0.18	<u>99.42±0.17</u>
COVID-LDCT-High	97.69±0.50	92.73±0.82	95.76±0.94	98.15±0.39	96.09±0.68	95.31±0.69	97.41±0.43	<u>98.46±0.30</u>	98.59±0.27
COVID-LDCT-Mid	97.97±0.51	91.44±1.09	96.74±0.88	98.61±0.50	96.86±0.69	95.98±0.54	97.48±0.43	<u>98.78±0.24</u>	98.89±0.22
COVID-LDCT-Low	98.19±0.56	71.86±3.41	97.07±0.91	98.64±0.44	96.71±0.75	95.69±0.77	97.50±0.44	<u>98.94±0.21</u>	99.06±0.19
COVID-LACT-High	65.55±6.65	56.99±5.49	66.16±3.69	83.03±2.39	75.35±2.69	74.65±2.63	69.19±4.21	<u>88.25±1.84</u>	91.48±1.57
COVID-LACT-Mid	80.94±2.73	73.05±5.39	78.12±3.62	89.35±1.92	81.82±2.47	82.36±2.12	76.26±3.62	<u>93.22±1.38</u>	95.33±0.96
COVID-LACT-Low	85.90±3.49	72.88±5.37	85.45±3.20	93.08±1.43	86.66±2.25	83.85±1.88	82.32±3.10	<u>95.49±0.86</u>	97.40±0.60
COVID-SVCT-High	89.11±2.46	79.87±2.87	85.75±2.21	90.43±2.35	85.62±2.96	86.55±1.94	91.16±1.39	<u>93.20±1.16</u>	93.87±1.14
COVID-SVCT-Mid	92.70±1.89	78.70±3.25	92.32±2.10	94.28±2.14	90.34±2.68	90.87±1.85	94.88±0.98	<u>96.17±0.72</u>	96.51±0.69
COVID-SVCT-Low	95.76±1.07	87.82±1.53	97.01±1.04	97.35±0.96	94.93±1.53	94.29±1.07	97.30±0.48	<u>98.25±0.33</u>	98.37±0.32

SPARE Challenge¹⁶, are used for model training and testing. Specifically, each dataset contains multiple scans from five patients, with one scan used for training and the remaining scans used for testing. For each single scan, the CBCT FDK reconstruction is used as the reference for model inference, while one phase bin from the ten-phase sparse-view CBCT FDK reconstruction is used as the input for model inference. The TAMP-S is adapted, and FBPConvNet is trained for comparison. The test results of FBPConvNet and TAMP-S are visually presented, as shown in Supplementary Figure 18. Compared with FBPConvNet, the sparse-view CBCT with TAMP-S demonstrates clearer and more accurate enhancement, showing smoother skeletal lines and clearer vascular structures.

5.3.2 Limited-angle CBCT

We conducted limited-angle cone-beam CT enhancement experiments on a private First-Imaging CBCT (**FI-CBCT**) dataset to validate the capability of TAMP across axial CT modalities. The FI-CBCT dataset contains 4 cases (1576 slices) of cone-beam CT acquired on a C-arm CBCT system manufactured by Jiangsu First-Imaging Medical Equipment Co., Ltd. (Nantong,

China), and is an institution-internal dataset that is not publicly available. 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 Supplementary Table 8 and Supplementary Figure 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 (30.16 dB, 82.95%) by +0.43 dB and +1.38% 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.

5.3.3 Low-dose Micro-CT

We conducted low-dose Micro-CT enhancement experiments on a private MARS Bioimaging

Supplementary Table 6: Quantitative evaluation of LPIPS (%) for different methods across 27 tasks. The best results are highlighted in **bold**, and the second-best results are underlined.

Task	RedCNN	WGAN-VGG	FBPConvNet	SPECIAL	TransCT	MITNet	ProCT	TAMP	TAMP-S
AMOS-LDCT-High	5.37±1.72	5.09±1.70	5.50±1.88	5.90±1.99	6.63±2.10	6.28±1.78	4.71±1.98	2.29±2.45	3.52±1.44
AMOS-LDCT-Mid	3.41±1.22	9.83±2.68	3.92±1.40	4.57±1.57	4.70±1.43	5.66±1.51	5.42±1.97	1.67±1.93	2.34±0.92
AMOS-LDCT-Low	2.48±0.93	5.16±1.31	2.76±1.04	3.87±1.46	3.64±1.06	5.96±1.79	6.29±2.15	<u>1.74±0.65</u>	1.66±0.63
AMOS-LACT-High	31.54±4.24	34.48±4.82	22.06±4.13	15.39±2.89	24.62±3.87	22.59±3.68	37.51±4.79	9.04±2.38	9.88±2.25
AMOS-LACT-Mid	17.42±3.78	21.49±3.63	11.61±2.53	8.23±2.04	16.56±2.97	15.53±2.94	19.64±2.99	4.78±2.01	5.39±1.20
AMOS-LACT-Low	8.16±1.99	12.70±2.44	5.73±1.53	4.67±1.25	10.45±2.24	10.29±1.93	13.88±3.08	2.83±1.89	<u>3.02±0.72</u>
AMOS-SVCT-High	15.94±3.58	20.79±3.85	17.78±3.73	15.26±3.15	18.81±3.49	16.87±3.34	15.21±3.11	<u>15.24±4.13</u>	8.00±1.77
AMOS-SVCT-Mid	12.60±2.99	14.78±2.98	13.53±2.96	10.20±2.69	13.94±2.88	15.14±3.15	9.50±2.67	<u>5.90±1.14</u>	5.02±1.32
AMOS-SVCT-Low	4.44±1.89	8.04±1.76	5.82±2.37	4.26±1.52	6.25±1.95	7.53±1.89	6.35±2.22	1.86±1.98	<u>2.41±0.81</u>
APET-LDCT-High	5.99±3.10	8.28±2.55	7.12±3.77	8.14±4.03	6.06±3.64	4.72±2.58	4.21±3.03	<u>3.64±2.16</u>	3.62±2.52
APET-LDCT-Mid	3.93±2.00	10.13±2.98	3.62±2.10	3.75±2.01	8.35±4.46	3.62±1.93	5.76±4.41	<u>2.61±1.63</u>	2.52±1.80
APET-LDCT-Low	<u>2.79±1.49</u>	9.54±2.78	4.28±2.63	4.31±2.35	3.16±1.76	4.53±2.44	6.94±5.46	<u>3.05±2.28</u>	1.68±1.21
APET-LACT-High	31.95±7.85	29.59±7.24	13.08±4.80	10.64±3.82	20.82±5.44	16.48±6.23	27.01±17.59	<u>10.19±2.23</u>	8.85±3.79
APET-LACT-Mid	12.94±4.12	18.20±4.73	8.69±3.18	6.88±2.29	13.56±4.23	11.93±4.47	19.84±8.96	<u>6.67±2.08</u>	5.28±2.14
APET-LACT-Low	7.63±2.12	12.36±3.16	4.02±1.35	3.97±1.46	8.26±2.74	11.10±4.99	14.78±5.34	<u>3.69±1.44</u>	3.19±1.35
APET-SVCT-High	12.99±4.84	14.31±4.36	11.69±4.47	12.36±5.01	14.66±5.23	13.41±5.20	11.38±7.19	<u>7.59±2.33</u>	7.31±2.61
APET-SVCT-Mid	10.40±4.11	11.64±3.76	6.96±2.88	7.61±3.22	10.89±3.84	9.78±3.80	8.92±6.14	<u>4.85±1.64</u>	4.54±1.77
APET-SVCT-Low	4.19±1.47	6.31±1.94	3.40±1.37	4.27±2.04	5.38±2.05	7.32±3.29	7.03±5.62	1.81±0.84	<u>2.04±0.94</u>
COVID-LDCT-High	2.76±1.14	8.33±1.25	3.95±1.61	4.17±1.41	5.92±2.00	13.49±3.11	10.12±3.98	3.19±2.34	2.55±1.24
COVID-LDCT-Mid	<u>2.07±0.80</u>	7.61±1.35	2.90±0.88	2.47±0.96	4.81±1.58	9.62±2.41	11.18±4.55	2.33±1.55	1.65±0.79
COVID-LDCT-Low	<u>1.85±0.67</u>	21.14±3.83	2.76±0.75	2.01±0.85	4.97±1.63	12.28±2.84	11.63±4.81	3.15±1.36	1.36±0.60
COVID-LACT-High	29.90±4.26	35.12±4.27	32.76±3.48	15.95±2.58	25.36±3.82	32.43±5.09	35.05±7.71	<u>12.42±3.37</u>	9.96±2.32
COVID-LACT-Mid	15.95±2.59	19.55±2.99	15.29±2.27	9.50±1.98	17.64±3.29	22.24±3.82	27.12±4.92	<u>6.64±3.54</u>	5.73±1.53
COVID-LACT-Low	9.07±1.87	18.46±2.89	8.17±1.35	5.71±1.22	12.36±2.74	18.20±3.61	19.68±6.55	2.82±2.02	<u>3.37±0.92</u>
COVID-SVCT-High	16.47±4.38	19.55±4.29	19.31±4.59	14.40±3.57	19.94±4.01	20.60±4.47	12.63±4.77	8.75±1.55	<u>9.06±1.94</u>
COVID-SVCT-Mid	11.17±3.13	15.78±3.23	12.76±3.04	11.86±3.15	16.02±3.52	18.72±4.03	10.54±4.23	5.87±1.22	6.31±1.46
COVID-SVCT-Low	5.17±1.94	11.93±2.45	6.33±2.43	6.40±2.20	8.04±2.37	12.60±3.29	10.24±4.30	2.41±1.85	<u>3.26±0.94</u>

micro-CT (**MARS-Micro-CT**) dataset to evaluate the adaptability of TAMP. In the MARS-Micro-CT dataset, a chicken drumstick was scanned using both low-dose (80 kVp) and normal-dose (120 kVp) settings on a MARS spectral CT scanner (MARS Bioimaging Ltd., Christchurch, New Zealand), and the projections were reconstructed into Micro-CT image volumes. The MARS-Micro-CT dataset is an institution-internal dataset that is not publicly available. After registration and cropping, both low-dose and normal-dose Micro-CT volumes contain 5 bins of $280 \times 1216 \times 1216$ pixels. The first bin's 5 slices are used for training, while the remaining slices are reserved for testing. The TAMP-S is adapted, and FBPConvNet is trained for comparison. The test results of FBPConvNet and TAMP-S are visually presented, as shown in Supplementary Figure 20. Compared with FBPConvNet, the low-dose Micro-CT with TAMP-S demonstrates clearer and more accurate enhancement, showing sharper edges, and the ring artifacts are more thoroughly suppressed.

5.4 Universal enhancement validation

To validate TAMP's universal enhancement capability in multi-scenario deployment settings that represent real-world clinical practice, we conducted mixed test cohort evaluation and oracle ensemble comparison.

5.4.1 Mixed test cohort evaluation

We constructed a mixed test cohort by pooling all 27 diverse NICT enhancement scenarios (3 datasets $\times$ 3 NICT types $\times$ 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.

As shown in Supplementary Table 9 and Supplementary Figure 21, TAMP achieves cumulative

Supplementary Table 7: Quantitative evaluation of RMSE (Hu) for different methods across 27 tasks. The best results are highlighted in **bold**, and the second-best results are underlined.

Task	RedCNN	WGAN-VGG	FBPConvNet	SPECIAL	TransCT	MITNet	ProCT	TAMP	TAMP-S
AMOS-LDCT-High	14.86±2.92	16.96±2.87	13.18±3.20	14.05±3.25	19.98±2.56	19.61±2.93	16.88±2.63	12.58±3.15	12.37±3.19
AMOS-LDCT-Mid	13.03±2.81	23.22±2.76	11.33±3.08	17.81±2.20	17.08±2.69	18.69±3.01	15.82±2.70	<u>10.80±3.05</u>	10.61±3.09
AMOS-LDCT-Low	11.35±2.83	22.75±2.96	<u>10.18±2.99</u>	10.55±3.24	15.73±2.77	17.45±3.04	15.41±2.77	<u>10.64±2.70</u>	9.49±2.97
AMOS-LACT-High	55.05±4.59	49.98±4.25	46.95±3.76	42.11±4.68	48.33±3.60	44.18±3.24	68.80±5.71	<u>39.72±3.72</u>	31.00±3.55
AMOS-LACT-Mid	55.22±6.10	52.36±4.92	37.24±4.41	35.00±6.11	44.07±3.96	39.78±3.78	61.14±3.63	<u>33.96±3.50</u>	24.38±3.55
AMOS-LACT-Low	40.52±6.82	47.12±5.80	33.71±5.61	<u>26.23±5.13</u>	37.00±5.23	34.53±3.90	52.70±3.87	<u>30.00±4.05</u>	17.58±3.36
AMOS-SVCT-High	34.72±4.29	44.97±4.47	37.75±3.42	34.03±3.54	38.89±4.58	34.20±3.75	33.27±2.86	34.00±4.43	25.47±3.49
AMOS-SVCT-Mid	27.69±4.14	35.46±4.29	28.28±4.00	<u>24.10±4.05</u>	34.56±4.39	29.80±3.45	27.10±2.86	25.97±4.54	18.16±3.45
AMOS-SVCT-Low	14.84±3.40	28.37±3.20	15.57±3.41	<u>14.76±3.61</u>	22.19±3.60	22.60±3.24	18.91±2.48	<u>12.27±2.86</u>	10.87±3.27
APET-LDCT-High	14.46±1.97	35.35±3.56	<u>12.07±2.10</u>	12.57±2.03	17.03±2.09	16.07±1.99	16.41±1.75	12.74±2.00	11.92±2.15
APET-LDCT-Mid	12.60±1.65	39.85±4.10	<u>10.57±1.92</u>	10.65±1.75	17.44±1.99	15.58±1.76	15.20±1.55	11.11±1.67	10.30±1.85
APET-LDCT-Low	11.94±1.20	39.74±3.56	<u>9.72±1.58</u>	<u>9.57±1.51</u>	13.01±1.47	14.17±1.56	15.12±1.54	10.04±1.40	9.21±1.59
APET-LACT-High	58.55±6.46	49.76±2.74	38.11±4.67	36.45±3.71	44.69±4.19	39.59±4.40	58.15±5.54	35.65±3.48	30.31±3.61
APET-LACT-Mid	47.13±4.50	52.21±3.55	33.86±3.92	<u>30.16±3.38</u>	40.61±4.37	33.70±3.60	55.63±4.03	31.22±3.08	24.42±3.08
APET-LACT-Low	42.19±5.00	49.44±4.11	25.42±3.28	<u>21.42±2.84</u>	32.58±4.48	31.71±4.20	49.36±3.76	27.87±4.39	17.43±2.30
APET-SVCT-High	35.37±3.09	42.74±3.57	41.17±6.55	30.10±3.46	35.21±3.26	31.33±3.17	39.28±3.01	28.68±2.98	25.18±3.32
APET-SVCT-Mid	26.49±3.32	38.69±3.65	18.14±2.60	19.73±2.80	28.84±3.61	24.82±3.01	28.72±2.65	20.61±2.61	<u>18.14±2.78</u>
APET-SVCT-Low	20.07±3.73	27.12±2.68	10.48±1.47	12.32±1.57	18.31±2.45	18.05±2.10	18.42±1.70	11.96±1.42	<u>10.89±1.49</u>
COVID-LDCT-High	24.58±1.89	40.21±2.53	31.06±2.93	22.16±2.20	31.49±3.29	33.32±2.84	29.34±3.64	19.35±1.80	18.59±1.82
COVID-LDCT-Mid	23.07±2.12	40.41±2.37	28.41±3.57	18.71±2.55	29.00±3.58	31.34±2.85	28.84±3.78	17.28±1.85	16.43±1.84
COVID-LDCT-Low	22.48±2.66	21.07±8.95	27.42±3.61	18.54±2.43	29.52±3.89	32.53±3.06	28.69±3.83	<u>16.18±1.92</u>	15.20±1.91
COVID-LACT-High	55.57±3.59	47.20±4.17	49.36±3.75	46.47±2.68	49.27±2.19	48.25±2.96	67.51±3.25	41.73±3.08	36.69±3.25
COVID-LACT-Mid	51.64±3.64	51.26±2.66	88.84±4.25	41.95±3.05	49.27±2.31	46.84±2.87	58.39±3.41	<u>38.16±3.37</u>	30.96±3.15
COVID-LACT-Low	49.69±5.28	53.19±4.76	45.45±4.31	35.68±3.23	45.97±3.44	46.34±2.49	55.04±3.61	<u>32.10±3.00</u>	24.31±2.99
COVID-SVCT-High	43.49±3.40	51.14±2.97	50.09±3.23	40.46±3.14	47.49±3.37	44.01±2.59	56.51±5.87	<u>36.45±2.91</u>	35.39±3.12
COVID-SVCT-Mid	41.54±4.62	52.01±2.39	42.04±4.44	36.04±4.03	44.82±4.07	39.35±3.28	44.76±5.95	<u>30.99±3.52</u>	29.74±3.50
COVID-SVCT-Low	31.14±3.77	44.22±3.01	28.52±4.49	26.38±3.99	35.84±4.59	35.97±3.19	31.67±4.26	<u>21.53±2.80</u>	20.60±2.71

Supplementary 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

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.

5.4.2 Oracle ensemble comparison

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 Supplementary 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

Supplementary Table 9: Mixed test cohort evaluation: cumulative performance across 27 NICT enhancement scenarios.

Method	Cumulative PSNR (dB) $\uparrow$	Cumulative SSIM (%) $\uparrow$
RED-CNN	1087.53	2312.45
FBPConvNet	1100.82	2359.21
ProCT	1100.73	2359.02
TAMP	1230.56	2605.66

features from limited data, while achieving 25-61% lower standard deviations across scenarios for enhanced robustness.

5.5 Ablation study

5.5.1 Component ablation

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 Supplementary Table 11. The ablation study systematically evaluates three key components:

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.

5.5.2 Training data size ablation

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 Supplementary 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.

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

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

Supplementary 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
					$\checkmark$	45.63 $\pm$ 1.20	37.51 $\pm$ 0.80	43.96 $\pm$ 1.13
	$\checkmark$				$\checkmark$	45.67 $\pm$ 0.97	39.53 $\pm$ 1.11	44.06 $\pm$ 0.95
	$\checkmark$			$\checkmark$	$\checkmark$	48.46 $\pm$ 1.11	41.99 $\pm$ 1.04	48.01 $\pm$ 1.04
	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$	48.62 $\pm$ 0.95	42.24 $\pm$ 1.18	48.29 $\pm$ 0.96
	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	49.47 $\pm$ 0.94	44.27 $\pm$ 1.01	48.79 $\pm$ 1.02
$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	51.62 $\pm$ 1.07	45.68 $\pm$ 0.98	50.26 $\pm$ 1.04

5.5.3 Model size ablation

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. Supplementary 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 Supplementary 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.

5.6 Downstream clinical validation

We 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.

5.6.1 Pulmonary nodule segmentation

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 ¹⁷ 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 ¹⁸ 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 Supplementary 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 absolute gains of +0.014, +0.135,

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

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

Supplementary 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 $\times$ 3, stride 1, ch=2	Conv 3 $\times$ 3, stride 1, ch=1	Conv 3 $\times$ 3, stride 1, ch=1
DFE-1	$\begin{bmatrix} \text{RSTB, dim=2} \\ \text{depth}=[2] \\ \text{heads}=[1] \end{bmatrix} \times 1$	$\begin{bmatrix} \text{RSTB, dim=1} \\ \text{depth}=[2] \\ \text{heads}=[1] \end{bmatrix} \times 1$	$\begin{bmatrix} \text{RSTB, dim=1} \\ \text{depth}=[2] \\ \text{heads}=[1] \end{bmatrix} \times 1$
Embedding-2	Conv 4 $\times$ 4, stride 2, ch=8	Conv 4 $\times$ 4, stride 2, ch=4	Conv 4 $\times$ 4, stride 2, ch=2
DFE-2	$\begin{bmatrix} \text{RSTB, dim=8} \\ \text{depth}=[4,4] \\ \text{heads}=[2,2] \end{bmatrix} \times 2$	$\begin{bmatrix} \text{RSTB, dim=4} \\ \text{depth}=[4,4] \\ \text{heads}=[2,2] \end{bmatrix} \times 2$	$\begin{bmatrix} \text{RSTB, dim=2} \\ \text{depth}=[4,4] \\ \text{heads}=[1,1] \end{bmatrix} \times 2$
Embedding-4	Conv 6 $\times$ 6, stride 4, ch=32	Conv 6 $\times$ 6, stride 4, ch=16	Conv 6 $\times$ 6, stride 4, ch=8
DFE-4	$\begin{bmatrix} \text{RSTB, dim=32} \\ \text{depth}=[6,6,6] \\ \text{heads}=[4,4,4] \end{bmatrix} \times 3$	$\begin{bmatrix} \text{RSTB, dim=16} \\ \text{depth}=[6,6,6] \\ \text{heads}=[4,4,4] \end{bmatrix} \times 3$	$\begin{bmatrix} \text{RSTB, dim=8} \\ \text{depth}=[6,6,6] \\ \text{heads}=[2,2,2] \end{bmatrix} \times 3$
Embedding-8	Conv 10 $\times$ 10, stride 8, ch=128	Conv 10 $\times$ 10, stride 8, ch=64	Conv 10 $\times$ 10, stride 8, ch=32
DFE-8	$\begin{bmatrix} \text{RSTB, dim=128} \\ \text{depth}=[8,8,8,8] \\ \text{heads}=[8,8,8,8] \end{bmatrix} \times 4$	$\begin{bmatrix} \text{RSTB, dim=64} \\ \text{depth}=[8,8,8,8] \\ \text{heads}=[8,8,8,8] \end{bmatrix} \times 4$	$\begin{bmatrix} \text{RSTB, dim=32} \\ \text{depth}=[8,8,8,8] \\ \text{heads}=[4,4,4,4] \end{bmatrix} \times 4$
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 $\times$ 3, ch=1	Conv 3 $\times$ 3, ch=1	Conv 3 $\times$ 3, ch=1
Params (M)	7.48	1.95	0.52
FLOPs (G)	96.1	24.5	6.4

+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.

5.6.2 Pulmonary nodule malignancy diagnosis

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,

Supplementary 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 $\pm$ 0.92	37.27 $\pm$ 0.70	39.92 $\pm$ 0.99
MITNet-Small	40.26 $\pm$ 0.55	37.56 $\pm$ 0.74	40.14 $\pm$ 0.55
MITNet-Base	44.16 $\pm$ 0.76	38.43 $\pm$ 0.82	44.00 $\pm$ 0.75

Supplementary 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

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 Supplementary 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 absolute 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.

5.6.3 Radiomics feature analysis

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 Supplementary 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 absolute 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.

5.7 TAMP model analysis

TAMP-S effectively extracts artifact features of various scales and shapes with its multi-scale integrated transformer network through pre-training and adaptation. As shown in Supplementary Figure 22, we use Effective Receptive Fields (ERF) to visually indicate the input image regions that receive focused attention during specific processing units, enabling a detailed analysis of model behavior and feature extraction characteristics.

Supplementary 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

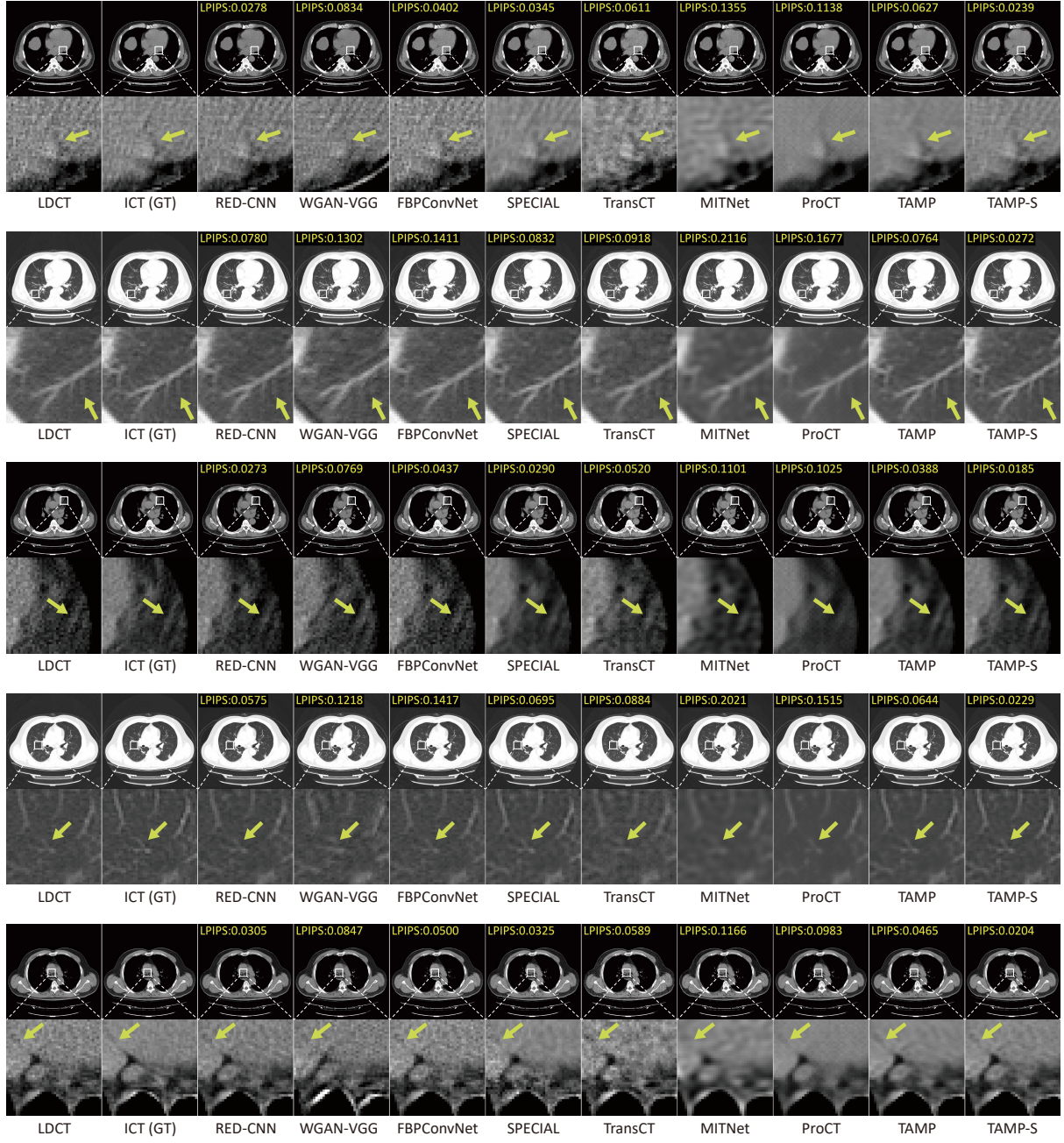
Supplementary Table 17: Our enhancement improves radiomics feature preservation (w/o vs. w/ our method).

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%

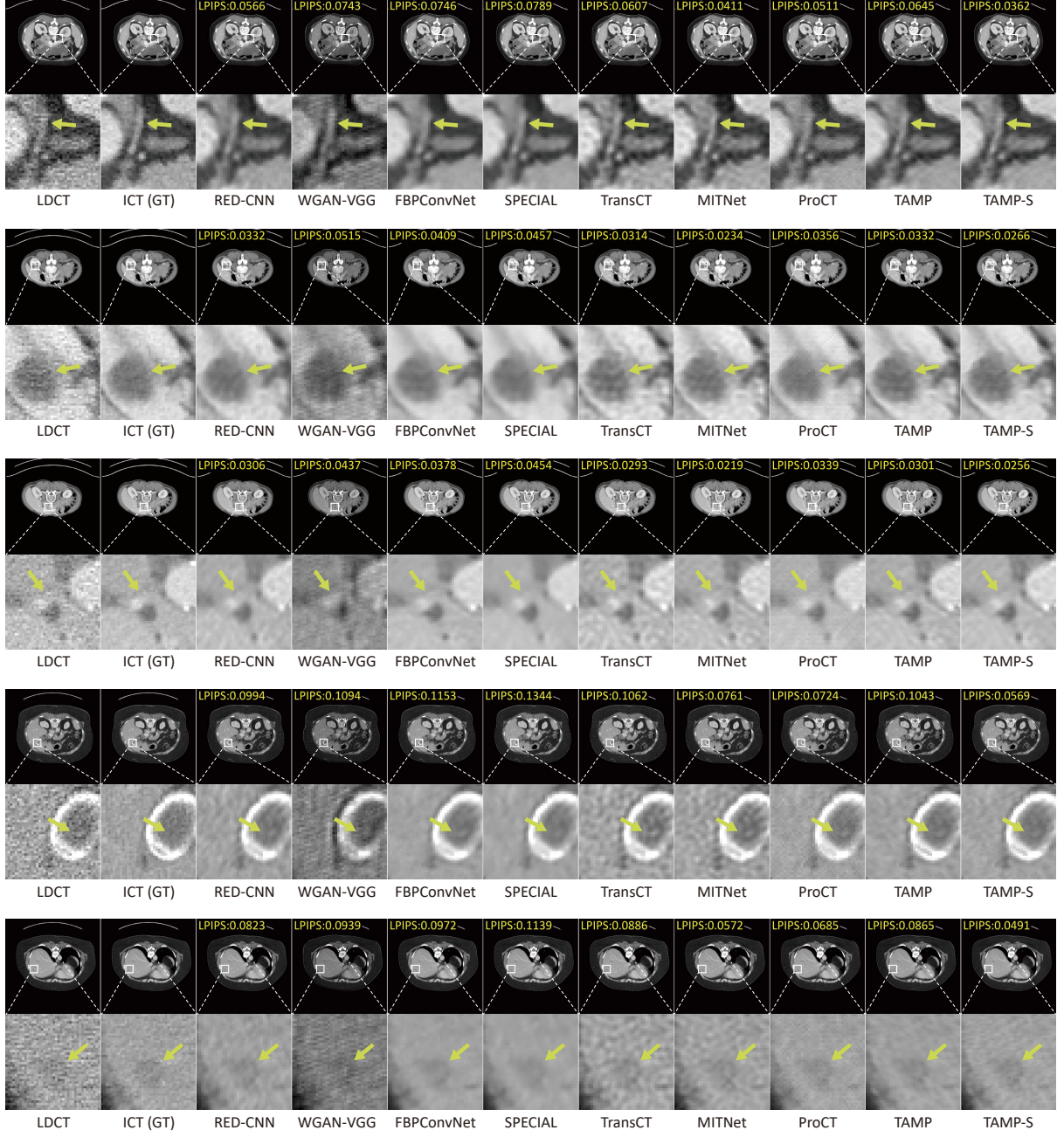
Our TAMP-S demonstrates precise ERF coverage of relevant regions, such as the concentrated ERF of the detailed region in LDCT and the wedge-shaped artifact in LACT, allowing it to accurately characterize a wide range of NICT defects and capture the most effective features for image enhancement. In contrast, single-scale convolutional networks (e.g., RED-CNN and WGAN-VGG) exhibit uniformly small-scale ERFs that fail to represent large-scale artifact structures. Although multi-scale U-Nets (e.g., FBPCovNet and SPECIAL) adapt to defects at different scales, their ERFs are approximately elliptical, indicating limited adaptability to precise artifact shapes. Furthermore, although multi-scale transformer networks (e.g., TransCT and ProCT) can accommodate artifacts of varying scales and shapes, their ERF coverage does not align well with the areas of interest, as evidenced by the inclusion of irrelevant regions in LACT.



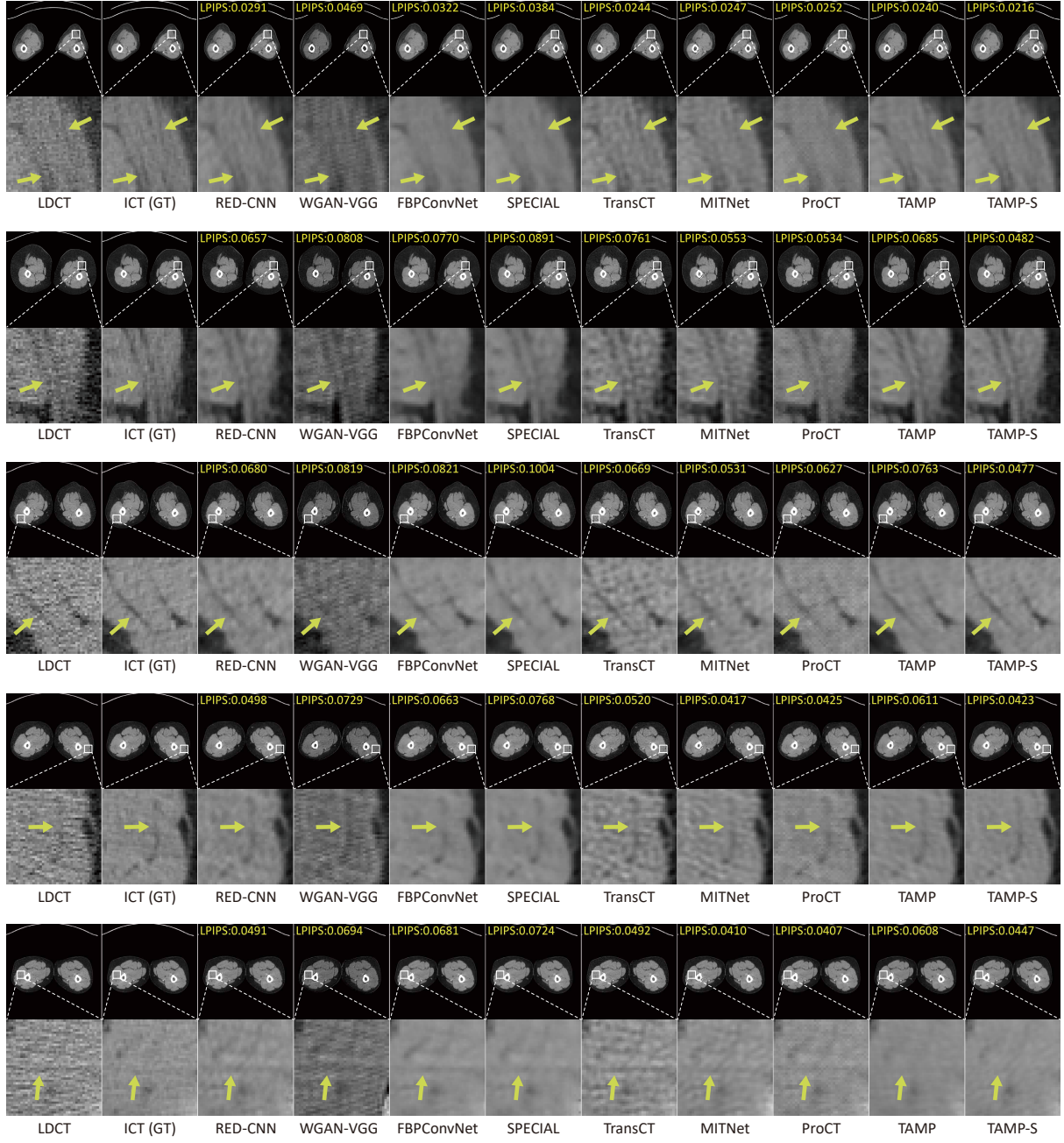
Supplementary Figure 5: Our TAMP has powerful universal enhancement capabilities, using the head LDCT as an example (CT images sourced from AutoPET [21](#)).



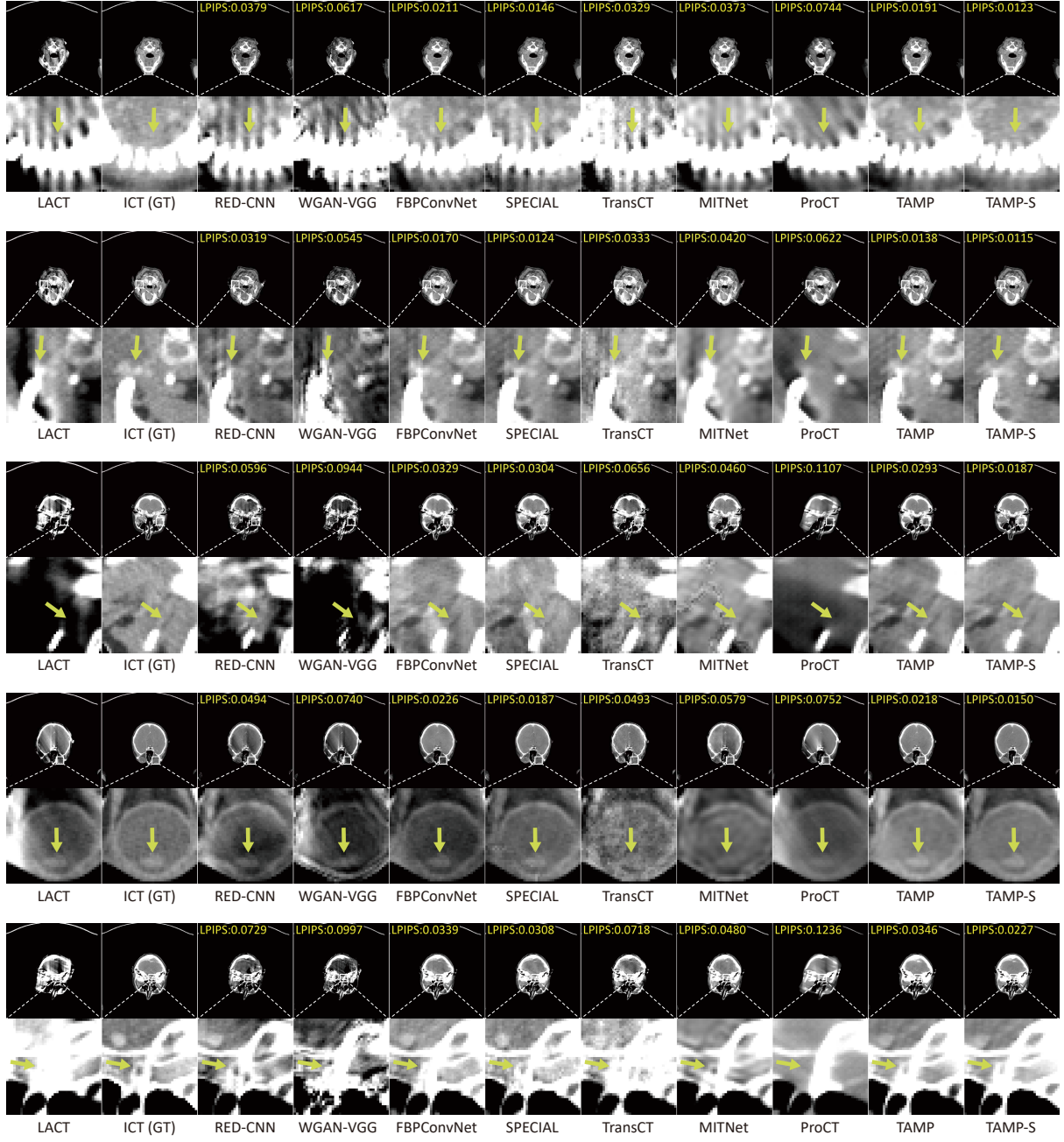
Supplementary Figure 6: Our TAMP has powerful universal enhancement capabilities, using the chest LDCT as an example (CT images sourced from COVID-19 [19](#)).



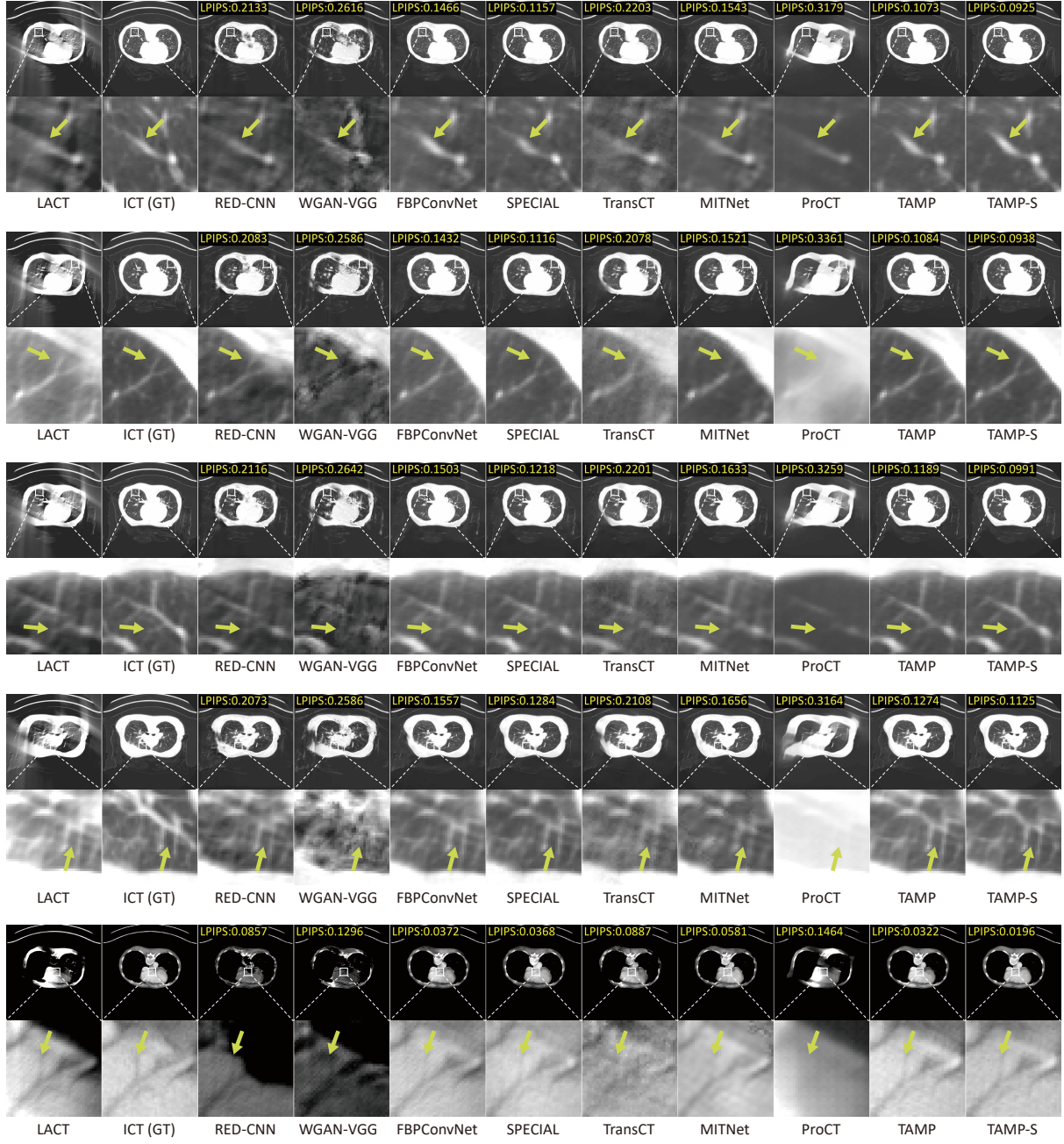
Supplementary Figure 7: Our TAMP has powerful universal enhancement capabilities, using the abdomen LDCT as an example (CT images sourced from AMOS22 [20](#)).



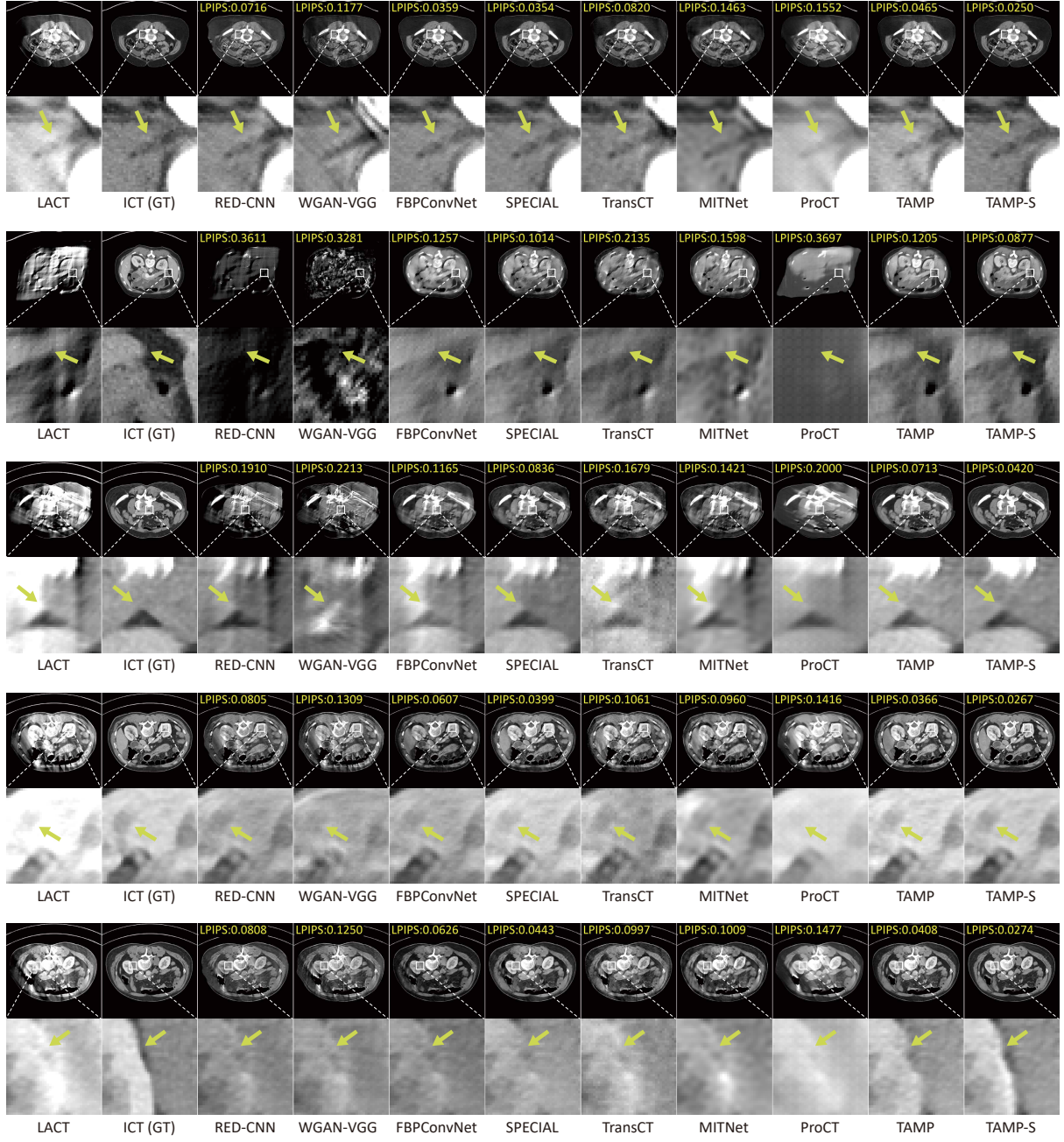
Supplementary Figure 8: Our TAMP has powerful universal enhancement capabilities, using the lower-limbs LDCT as an example (CT images sourced from AutoPET ²¹).



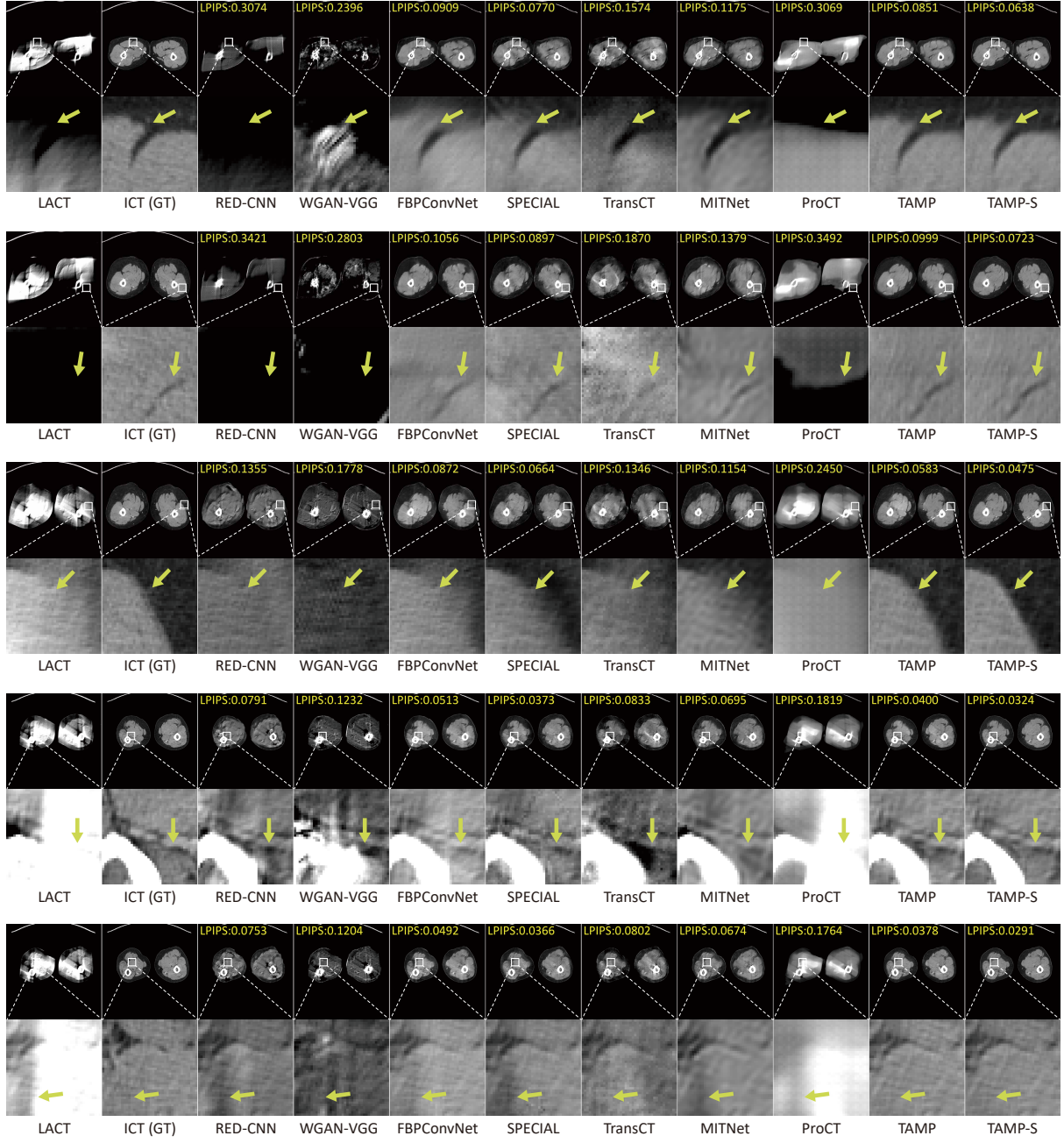
Supplementary Figure 9: Our TAMP has powerful universal enhancement capabilities, using the head LACT as an example (CT images sourced from AutoPET [21](#)).



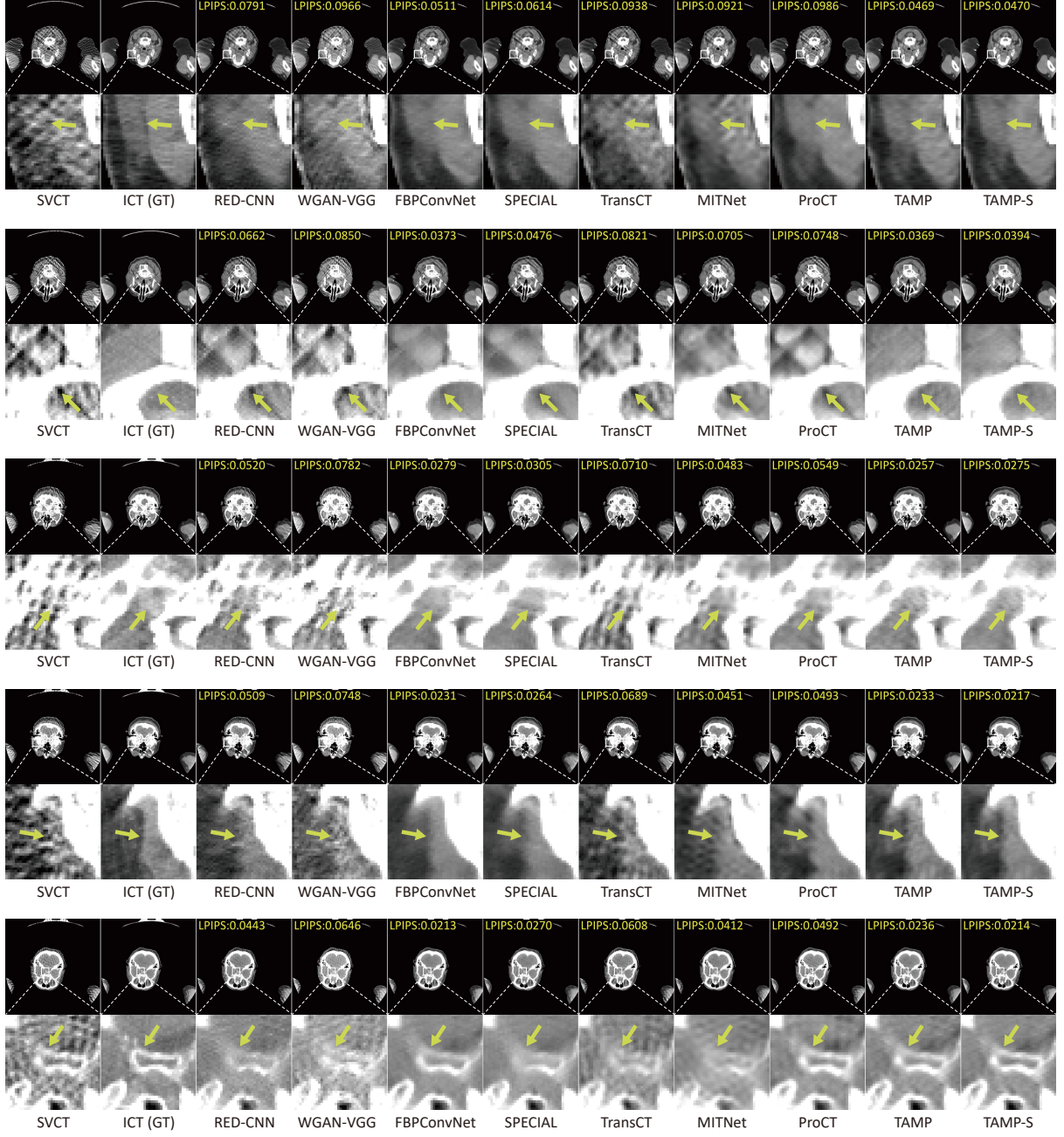
Supplementary Figure 10: Our TAMP has powerful universal enhancement capabilities, using the chest LACT as an example (CT images sourced from COVID-19 ¹⁹).



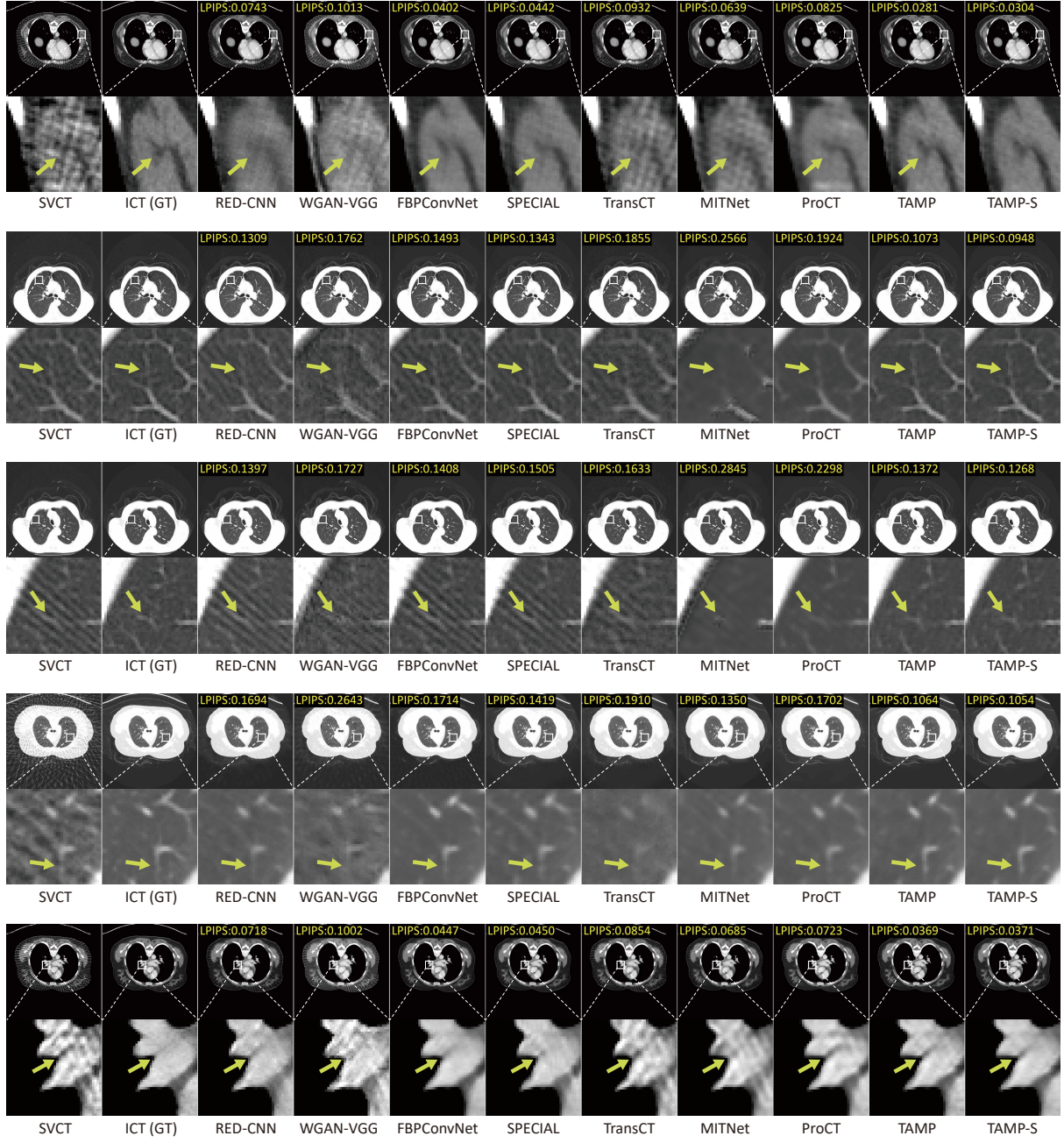
Supplementary Figure 11: Our TAMP has powerful universal enhancement capabilities, using the abdomen LACT as an example (CT images sourced from AMOS22 [20](#)).



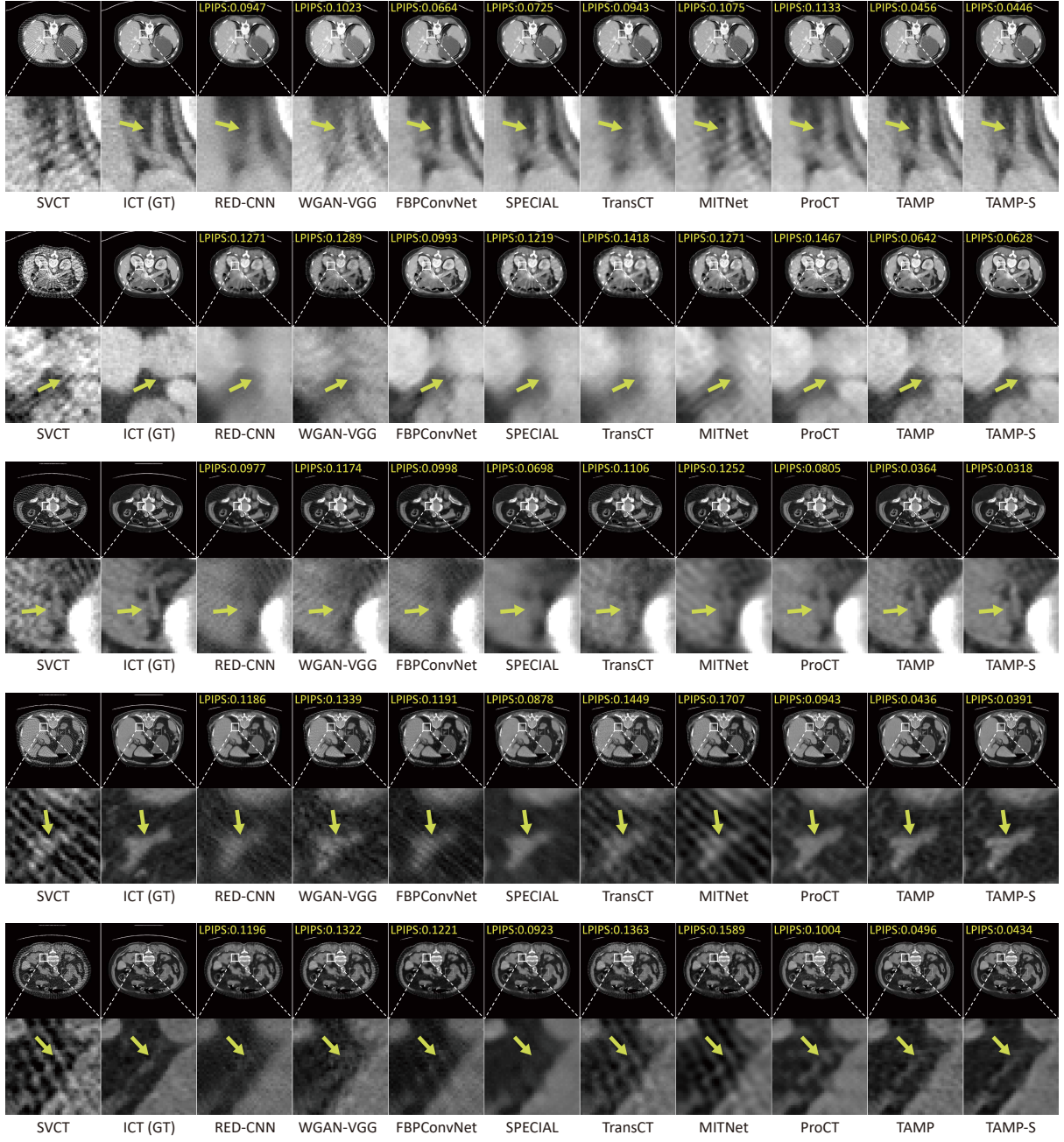
Supplementary Figure 12: Our TAMP has powerful universal enhancement capabilities, using the lower-limbs LACT as an example (CT images sourced from AutoPET ²¹).



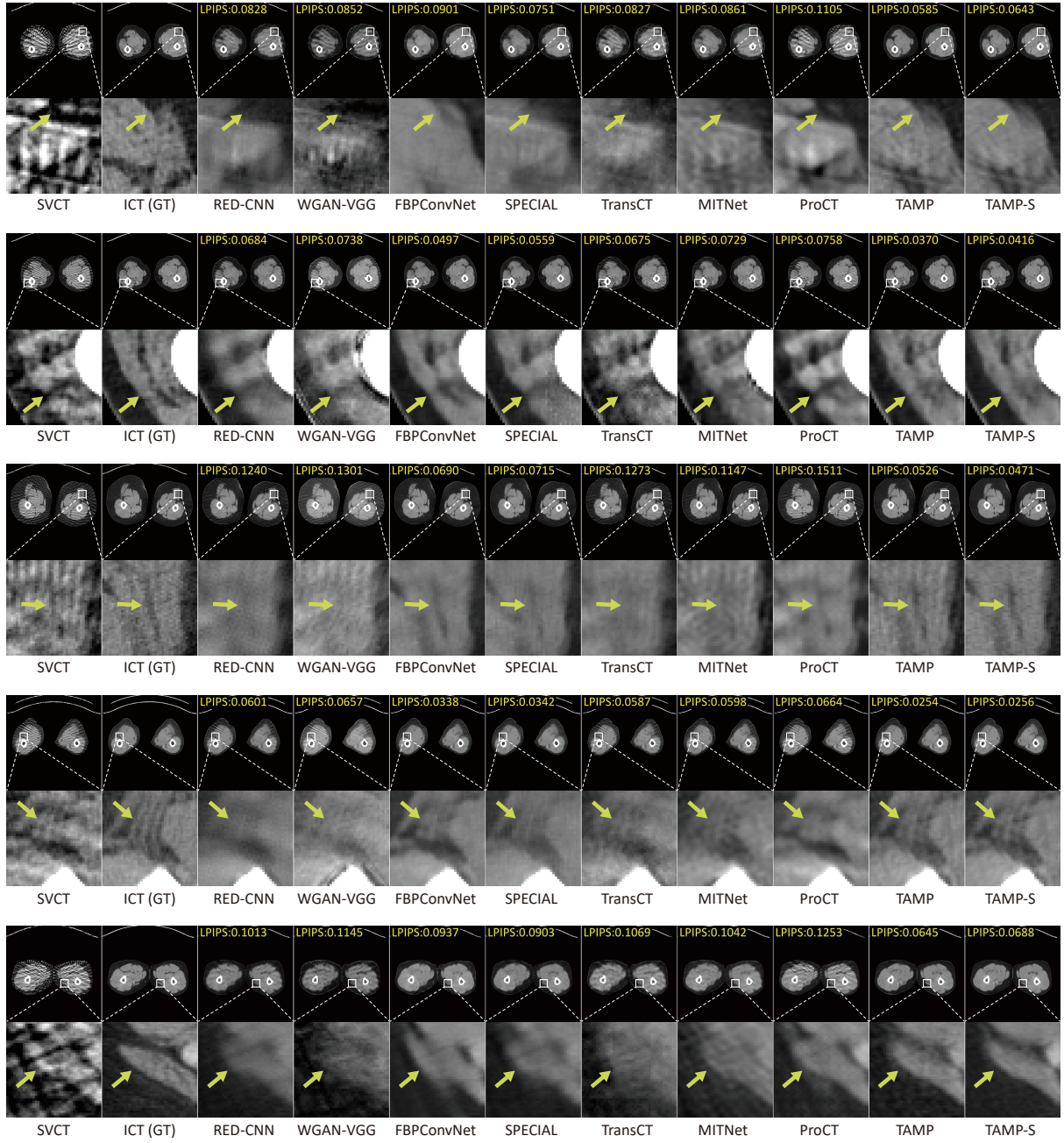
Supplementary Figure 13: Our TAMP has powerful universal enhancement capabilities, using the head SVCT as an example (CT images sourced from AutoPET ²¹).



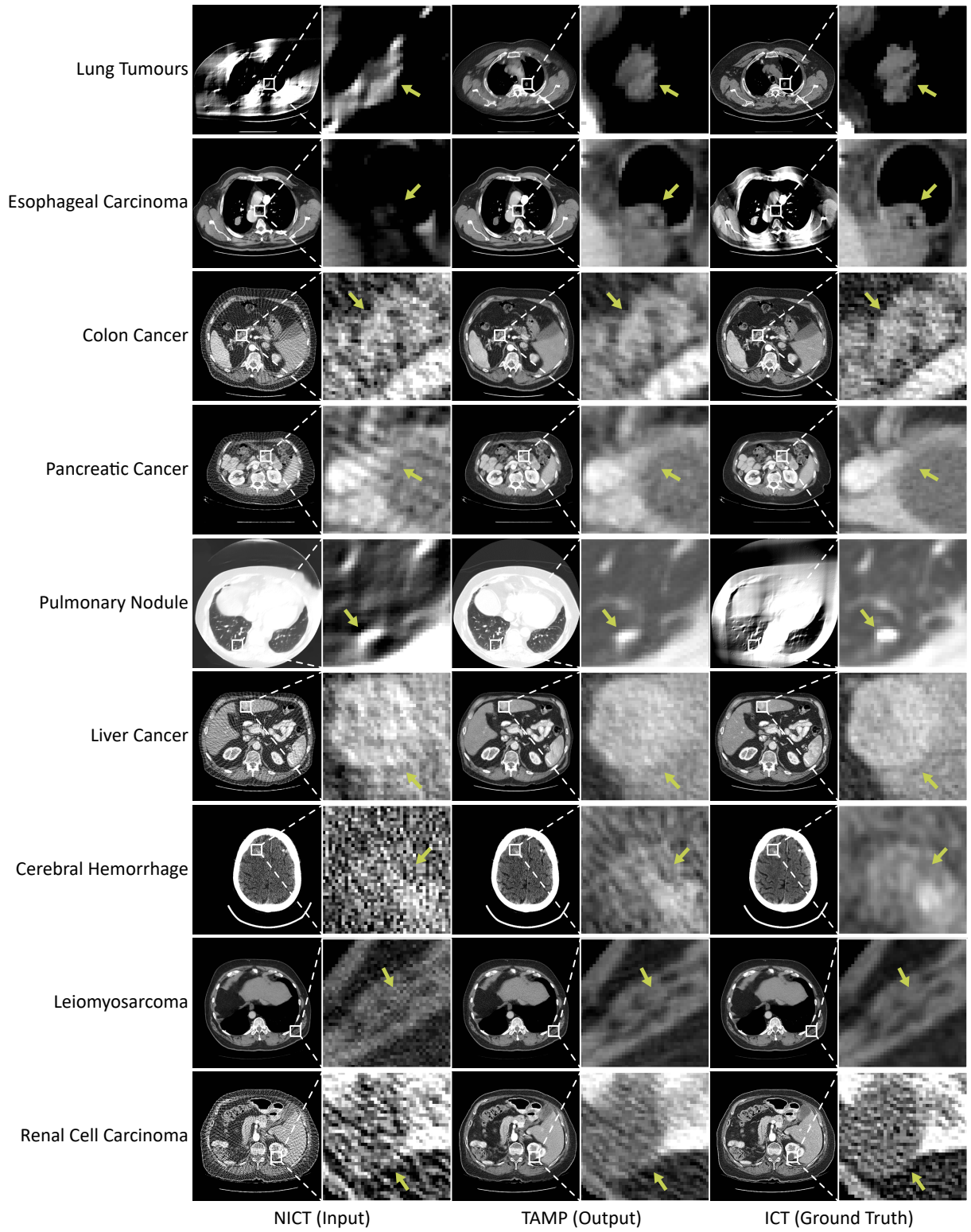
Supplementary Figure 14: Our TAMP has powerful universal enhancement capabilities, using the chest SVCT as an example (CT images sourced from COVID-19 ¹⁹).



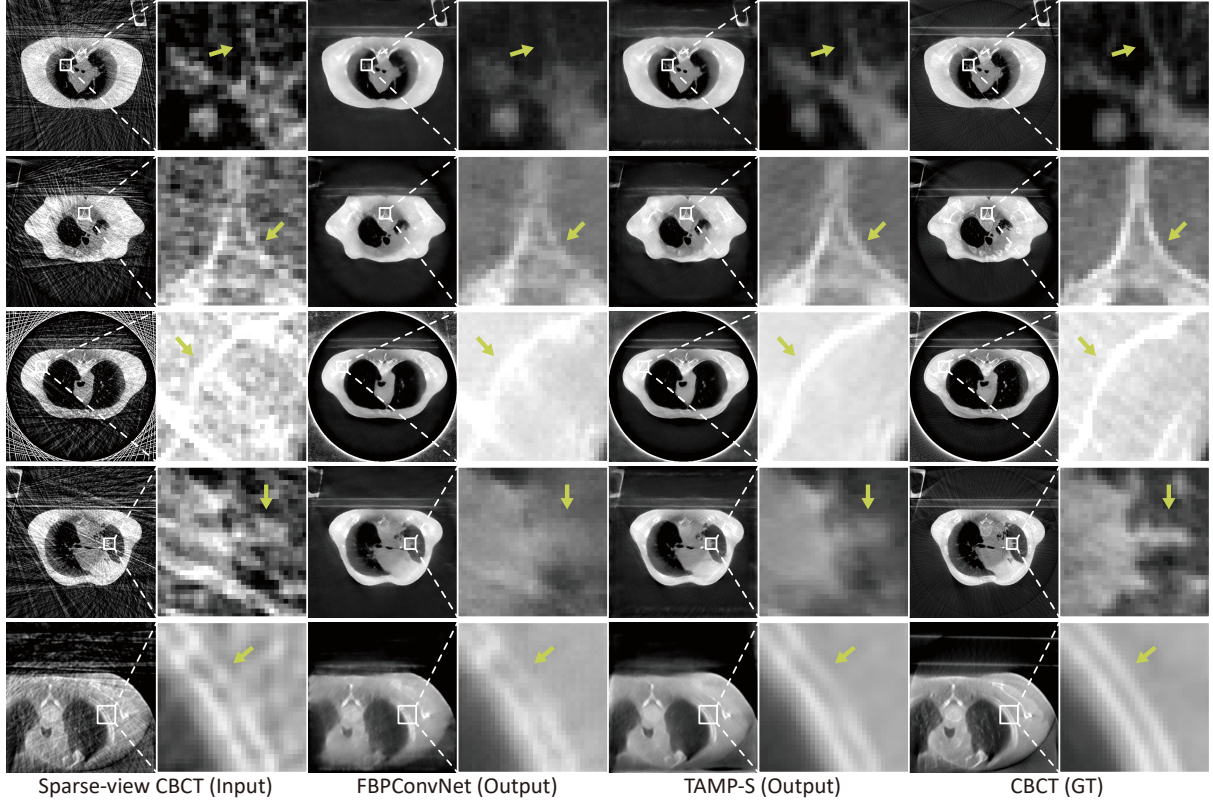
Supplementary Figure 15: Our TAMP has powerful universal enhancement capabilities, using the abdomen SVCT as an example (CT images sourced from AMOS22 ²⁰).



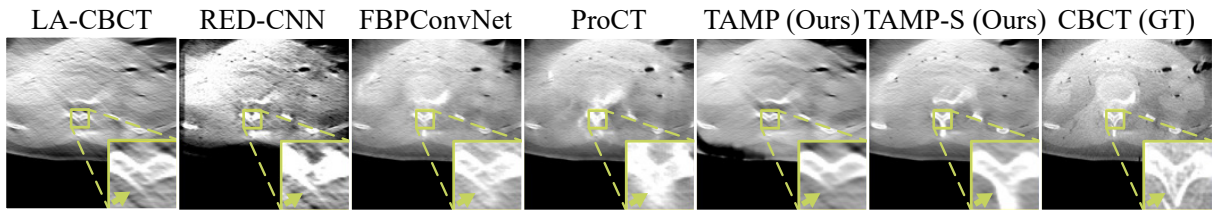
Supplementary Figure 16: Our TAMP has powerful universal enhancement capabilities, using the lower-limbs SVCT as an example (CT images sourced from AutoPET ²¹).



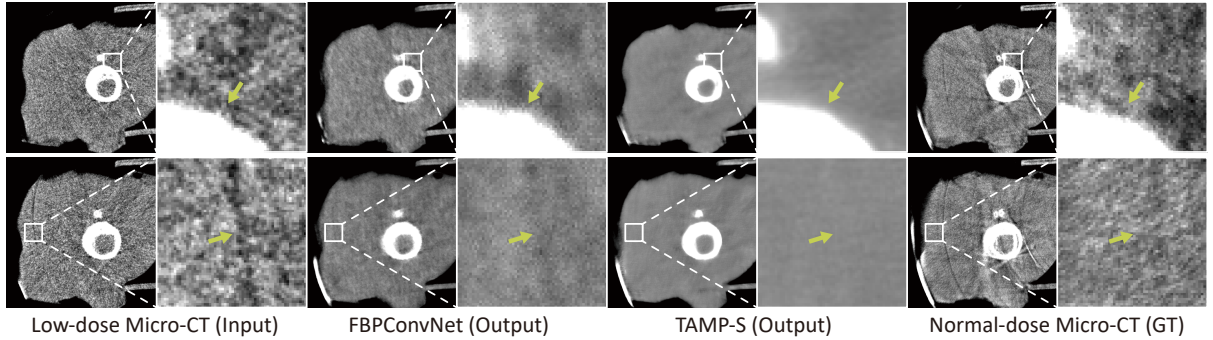
Supplementary Figure 17: Our TAMP demonstrates effective enhancement for NICTs with different lesions, highlighting its significant potential in assisting medical diagnosis across a wide range of clinical scenarios (CT images sourced from LIDC-IDRI [17](#)).



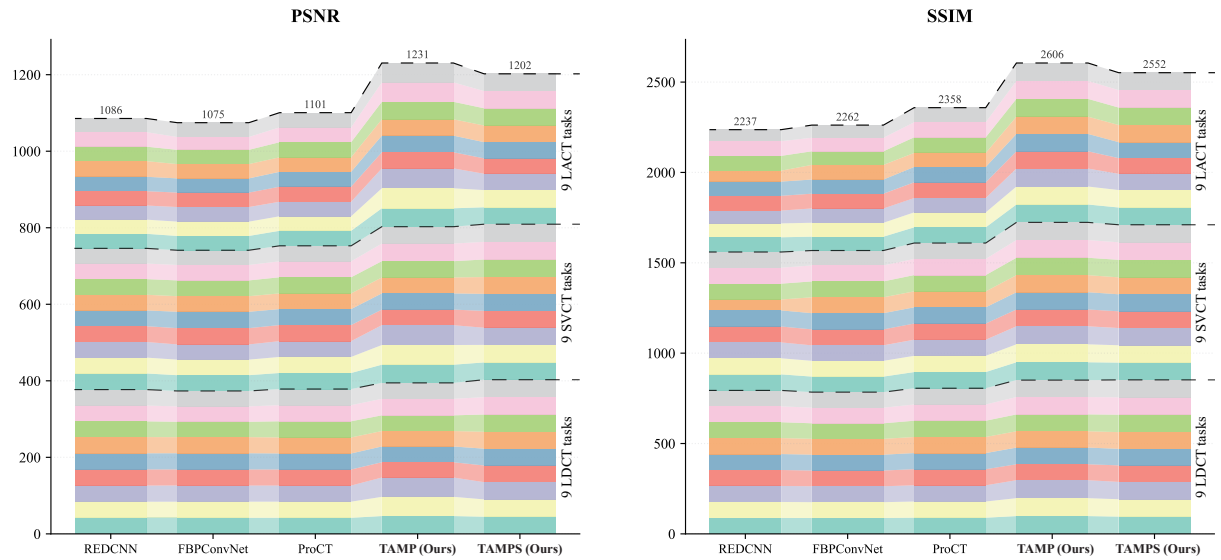
Supplementary Figure 18: Our TAMP demonstrates effective adaptation capability in enhancing sparse-view CBCT, highlighting its significant potential in widespread clinical scenarios. CBCT images are sourced from the SPARE Challenge Varian and Elekta datasets ¹⁶.



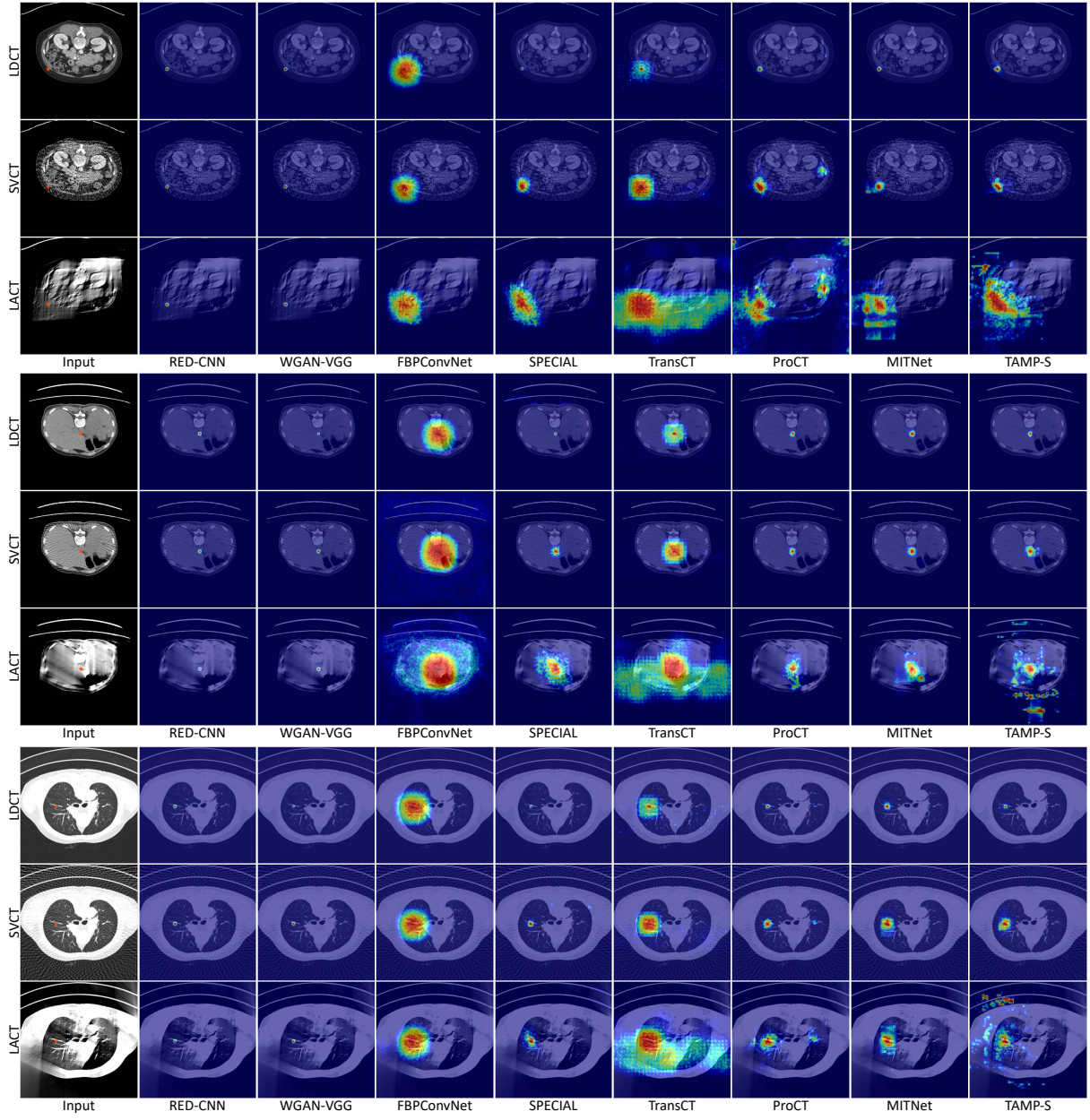
Supplementary Figure 19: Qualitative comparison on limited-angle cone-beam CT enhancement. TAMP effectively suppresses radial artifacts and TAMP-S further reconstructs corrupted anatomical details (CBCT images sourced from the FI-CBCT dataset).



Supplementary Figure 20: Our TAMP demonstrates effective adaptation capability in enhancing low-dose Micro-CT, highlighting its significant potential in widespread clinical scenarios (Micro-CT images sourced from the MARS-Micro-CT dataset).



Supplementary Figure 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.



Supplementary Figure 22: The effective receptive fields of different methods for inferring various types of NICT images. TAMP-S effectively extracts artifact features of different scales and shapes.

References

References

1. NIH, AAPM, Clinic, M.: Low Dose CT Grand Challenge. <https://www.aapm.org/GrandChallenge/LowDoseCT/>. Accessed: 2024-11-10 (2016)
2. Yu, L., Shiung, M., Jondal, D., McCollough, C.H.: Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols. *Journal of computer assisted tomography* **36**(4), 477–487 (2012)
3. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging* **29**(1), 196–205 (2009)
4. Shamonin, D.P., Bron, E.E., Lelieveldt, B.P., Smits, M., Klein, S., Staring, M., Initiative, A.D.N.: Fast parallel image registration on cpu and gpu for diagnostic classification of alzheimer’s disease. *Frontiers in neuroinformatics* **7**, 50 (2014)
5. Biguri, A., Dosanjh, M., Hancock, S., Soleimani, M.: Tigre: a matlab-gpu toolbox for cbct image reconstruction. *Biomedical Physics & Engineering Express* **2**(5), 055010 (2016)
6. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing* **15**(11), 3440–3451 (2006)
7. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
8. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: *CVPR* (2018)
9. Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep cnn denoiser prior for image restoration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3929–3938 (2017)
10. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging* **3**(1), 47–57 (2016)
11. Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M.K., Zhang, Y., Sun, L., Wang, G.: Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging* **37**(6), 1348–1357 (2018)
12. Xie, X., Zhou, P., Li, H., Lin, Z., Yan, S.: Adan: Adaptive nesterov momentum algorithm for faster optimizing deep models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024)
13. Hu, E.J., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. In: *International Conference on Learning Representations*
14. Vaswani, A.: Attention is all you need. *Advances in Neural Information Processing Systems* (2017)
15. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1833–1844 (2021)
16. Shieh, C.-C., Gonzalez, Y., Li, B., Jia, X., Rit, S., Mory, C., Riblett, M., Hugo, G., Zhang, Y., Jiang, Z., et al.: Spare: Sparse-view reconstruction challenge for 4d cone-beam ct from a 1-min scan. *Medical physics* **46**(9), 3799–3811 (2019)
17. Armato, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke,

- C.I., Hoffman, E.A., *et al.*: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical Physics* **38**(2), 915–931 (2011) <https://doi.org/10.1118/1.3528204>
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241 (2015). Springer
 19. An, P., Xu, S., Harmon, S.A., Turkbey, E.B., Sanford, T.H., Amalou, A., Kassin, M., Varble, N., Blain, M., Anderson, V., *et al.*: Ct images in covid-19 [data set]. *The Cancer Imaging Archive* **10**, 32 (2020)
 20. Ji, Y., Bia, H., Ge, C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., *et al.*: Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *Advances in neural information processing systems* **35**, 36722–36732 (2022)
 21. Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberger, C., Schölkopf, B., Küstner, T., Cyran, C., Rubin, D.: A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. *Scientific Data* **9**(1), 601 (2022)