

TAT: Task-Adaptive Transformer for All-in-One Medical Image Restoration

Zhiwen Yang¹, Jiaju Zhang¹, Yang Yi¹, Jian Liang², Bingzheng Wei³, and Yan Xu¹ (✉)

¹ School of Biological Science and Medical Engineering, State Key Laboratory of Software Development Environment, Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education, Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China
xuyan04@gmail.com

² Sinovision Technologies (Beijing) Co., Ltd., Beijing 101102, China

³ ByteDance Inc., Beijing 100098, China

Abstract. Medical image restoration (MedIR) aims to recover high-quality medical images from their low-quality counterparts. Recent advancements in MedIR have focused on All-in-One models capable of simultaneously addressing multiple different MedIR tasks. However, due to significant differences in both modality and degradation types, using a shared model for these diverse tasks requires careful consideration of two critical inter-task relationships: *task interference*, which occurs when conflicting gradient update directions arise across tasks on the same parameter, and *task imbalance*, which refers to uneven optimization caused by varying learning difficulties inherent to each task. To address these challenges, we propose a task-adaptive Transformer (TAT), a novel framework that dynamically adapts to different tasks through two key innovations. First, a task-adaptive weight generation strategy is introduced to mitigate *task interference* by generating task-specific weight parameters for each task, thereby eliminating potential gradient conflicts on shared weight parameters. Second, a task-adaptive loss balancing strategy is introduced to dynamically adjust loss weights based on task-specific learning difficulties, preventing task domination or undertraining. Extensive experiments demonstrate that our proposed TAT achieves state-of-the-art performance in three MedIR tasks—PET synthesis, CT denoising, and MRI super-resolution—both in task-specific and All-in-One settings. Code is available at [this https URL](#).

Keywords: All-in-One · Task Interference · Task Imbalance.

1 Introduction

Medical image restoration (MedIR) is a fundamental task in medical imaging, focused on reconstructing high-quality (HQ) images from their low-quality (LQ)

Equal contribution – Z. Yang, J. Zhang, and Y. Yi.

counterparts. LQ medical images often suffer from substantial diagnostic quality degradation due to suboptimal imaging conditions, such as reduced radiation exposure time and insufficient radiation intensity, which are employed to minimize potential health risks to patients. Significant progress has been made in recent years for many specific MedIR tasks, including PET synthesis [1,2,3,4,5,6,7], CT denoising [8,9,10,11,12], and MRI super-resolution [13,14,15,16].

Despite their success in specific scenarios, task-specific MedIR models face critical limitations that hinder their clinical adaptability. **(1) Limited Generalization.** In complex multimodal imaging workflows (e.g., PET/CT or PET/MRI), multiple MedIR tasks often coexist. However, due to inherent differences in imaging modalities and degradation types, task-specific models trained for one MedIR task struggle to adapt to others, leading to significant performance drops. **(2) Inefficiency and Redundancy.** Task-specific models require redundant development efforts and resource allocation, as each task demands separate architectures, training pipelines, storage solutions, and computational resources. This fragmented approach escalates costs and complicates clinical deployment. **(3) Data Scarcity and Isolation.** Task-specific models are particularly vulnerable to data scarcity, as they rely on narrow, task-specific datasets that are often limited in medical imaging. This constraint not only increases their susceptibility to model overfitting but also isolates them from potential cross-task and cross-modal synergies due to their task-specific training. In summary, task-specific models face limitations in generalization, efficiency, and data availability, which collectively hinder their scalability and real-world applicability.

To overcome the limitations of task-specific models, recent efforts [17,18,19,20,21] have focused on developing an All-in-One model capable of handling multiple tasks simultaneously. This approach directly addresses the three key limitations of task-specific models: (1) improving cross-task generalization through multi-task training, (2) eliminating redundancies by consolidating workflows into a single model, and (3) mitigating data scarcity by leveraging both multitask and multimodal data. To handle various restoration tasks, recent All-in-One methods in natural image restoration use advanced techniques, such as contrastive learning [17] and visual prompting [20], to learn task-discriminative representations that guide the model’s adaptation to different tasks. In medical image restoration, the pioneering work of AMIR [21] introduces a task routing strategy that allocates different tasks to separate network paths. These approaches have significantly contributed to the development of effective All-in-One models.

However, due to the substantial differences between MedIR tasks in terms of both modality and degradation types, using a shared model for such diverse tasks requires careful consideration of two crucial inter-task relationships: *task interference*, which occurs when conflicting gradient update directions arise on the same parameter between tasks [21], and *task imbalance*, which refers to uneven optimization caused by varying learning difficulties inherent to each task [22]. Regarding *task interference*, while several methods have attempted to address this [18,21], they still lack the adaptability to handle the complexity and diversity across different MedIR tasks. This is because most of them still rely

on fixed parameters shared across tasks, which prevents them from adapting to the specific needs of each task. When gradient conflicts arise in these shared parameters, they inevitably lead to suboptimal performance. On the other hand, *task imbalance* has been largely overlooked in existing studies. These methods fail to recognize that different MedIR tasks have varying levels of learning difficulty, and using a uniform weight for the loss across tasks often results in some tasks dominating while others remain undertrained. Therefore, there is a pressing need for novel All-in-One approaches that effectively address the inter-task relationships of *task interference* and *task imbalance* in MedIR tasks.

In this paper, we introduce a novel task-adaptive transformer (TAT) that effectively addresses two crucial inter-task relationships—*task interference* and *task imbalance*—for All-in-One medical image restoration. This is accomplished through two key innovations: a task-adaptive weight generation strategy and a task-adaptive loss balancing strategy. Specifically, **(1)** to mitigate *task interference* between distinct MedIR tasks, we propose a task-adaptive weight generation strategy that dynamically generates task-specific weight parameters for processing, thereby eliminating potential conflicts in weight updates. **(2)** To properly address *task imbalance* and prevent task domination or undertraining, we introduce a task-adaptive loss balancing strategy that dynamically adjusts the loss weights for different tasks during training, ensuring the most effective optimization path. Extensive experiments demonstrate that our proposed TAT achieves state-of-the-art performance on tasks such as PET synthesis, CT denoising, MRI super-resolution, and All-in-One medical image restoration.

2 Method

In this section, we provide a detailed introduction to our proposed task-adaptive transformer (TAT). We first describe the overall architecture of TAT in Subsection 2.1. Then, in Subsections 2.2 and 2.3, we present the two key innovations: the task-adaptive weight generation strategy, which addresses *task interference*, and the task-adaptive loss balancing strategy, which addresses *task imbalance*.

2.1 Overall Network Architecture

As shown in Fig. 1, the proposed TAT features a multi-level U-shaped architecture consisting of a three-stage encoder with Transformer blocks [23] and a four-stage decoder incorporating weight-adaptive Transformer blocks (WATBs). To address the issue of *task interference*, a task representation extraction network (TREN) is employed to extract task-specific representations that guide the generation of weights in the WATBs. The TAT begins by extracting initial features, denoted as $I^{IF} \in \mathbb{R}^{H \times W \times C}$, from an input low-quality (LQ) medical image $I^{LQ} \in \mathbb{R}^{H \times W \times 1}$ via a 3×3 convolutional layer, where H , W , and C represent the height, width, and channel dimensions, respectively. These features are then encoded into latent representations I^{LF} through the Transformer-based encoder. The pipeline subsequently splits into two branches: the first processes

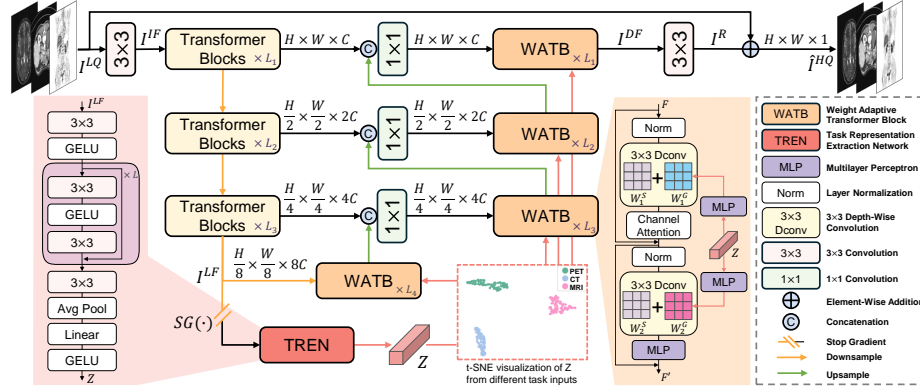


Fig. 1. Overview of the proposed task-adaptive transformer (TAT) network.

I^{LF} through the decoder to produce deep features I^{DF} , while the second passes a gradient-detached copy of I^{LF} through TREN to extract a task-specific representation $Z \in \mathbb{R}^d$. The WATBs utilize Z to generate task-adaptive weights, enabling specialized feature refinement during the decoding process. Finally, a 3×3 convolutional layer transforms I^{DF} into a residual image $I^R \in \mathbb{R}^{H \times W \times 1}$, which is added to the original LQ image I^{LQ} to yield the restored high-quality (HQ) output, $\hat{I}^{HQ} = I^{LQ} + I^R$.

2.2 Task-Adaptive Weight Generation Strategy

Most existing All-in-One models share a common limitation: they rely on a single model with fixed parameters to handle multiple tasks. This one-size-fits-all approach often results in suboptimal performance due to *task interference*, where conflicting gradient updates from different tasks impede effective parameter optimization. As a result, the weight parameters are not specialized for any particular task, leading to diminished performance across broad. To address this, we propose a novel task-adaptive weight generation strategy that dynamically generates task-specific parameters for specialized processing, thereby eliminating potential interference. We introduce this strategy from two perspectives: task-specific representation extraction and task-adaptive weight generation.

Task-Specific Representation Extraction. Previous All-in-One models for natural images often utilize advanced techniques, such as contrastive learning [17] or auxiliary classification tasks [18], to learn task-specific representations as guidance. However, we argue that these methods are unnecessary in the context of medical image restoration. Due to the significant semantic differences between various medical imaging modalities, the latent features encoded with semantic information inherently exhibit task-specific variations. Consequently, even straightforward feature extraction from the latent feature enables distinctions across tasks, eliminating the need for complex representation learning. Building upon this insight, we propose a simple task representation extraction network

(TREN), consisting of sequential convolutional blocks that directly extract task-specific representations $Z \in \mathbb{R}^d$ from the latent features I^{LF} , as expressed below:

$$Z = \text{TREN}(\text{SG}(I^{LF})), \quad (1)$$

where $\text{SG}(\cdot)$ denotes the stop-gradient operator, which decouples the extraction of latent features I^{LF} from the extraction of task representations Z , thereby preventing potential interference between the two processes with distinct objectives. The resulting extracted task representations Z are specific to each task, as evidenced by the t-SNE visualization in Fig. 1.

Task-Adaptive Weight Generation. *Task interference* occurs when different tasks conflict in their update directions for the same weight parameters. To address this, we propose generating task-specific parameters for each task. Using the task-specific representation Z , we use multi-layer perceptrons (MLPs) to estimate weight parameters for each decoding Transformer block. However, traditional approaches that generate weights for linear layers or standard convolutions face scalability issues: their parameter counts grow quadratically with channel dimension C (i.e., $\mathcal{O}(C^2)$), leading to computational inefficiency and unreliable parameter estimation. To mitigate this, we shift focus to depth-wise convolutions, a lightweight alternative with only $k \times k \times C$ parameters (where k is the kernel size), scaling linearly with C (i.e., $\mathcal{O}(C)$) since $k \ll C$. This choice is motivated by two key advantages. First, depth-wise convolutions preserve local spatial information while complementing global attention mechanisms—a synergy shown to enhance performance in vision Transformers [24,23]. Second, their parameter efficiency enables accurate and compact weight generation. Consequently, our weight generation process can be formulated as follows:

$$W^G = \text{Reshape}(\text{MLP}(Z)), \quad (2)$$

where W^G denotes the dynamically generated weight for depth-wise convolution, which is obtained by first transforming Z through an MLP and then reshaping it into the target kernel shape. Finally, the generated task-specific weight W^G is summed with the previously shared weight W^S as follows:

$$W = W^S + \lambda W^G, \quad (3)$$

where W is the final weight of the depth-wise convolution and λ is a learnable parameter. By incorporating the generated task-specific weight into the Transformer block, we transform the original Transformer block [23] into a weight-adaptive Transformer block (WATB), as illustrated in Fig. 1.

2.3 Task-Adaptive Loss Balancing Strategy

Existing All-in-One methods ignore the challenge of *task imbalance*, where different tasks present varying learning difficulties, leading to some tasks dominating while others remain undertrained. This concern has been addressed in the field

of multi-task learning by using a loss balancing strategy [22] that dynamically allocates task-specific weights during training. A common formulation is:

$$Loss = \sum_{t=1}^T \left(\frac{1}{2\sigma_t^2} L^t + \log \sigma_t \right), \quad (4)$$

where T denotes the total number of tasks. L^t denotes the loss for the t -th task. $\sigma_t \in \mathbb{R}^1$ is a learnable parameter. Here, $\frac{1}{2\sigma_t^2}$ dynamically scales the loss weight of each task, while $\log \sigma_t$ regularizes the scaling. When the loss L^t is large and tends to dominate the overall loss, σ_t increases to suppress its weight, and vice versa. This mechanism autonomously balances task contributions, ensuring equitable training without manual intervention. However, while effective for task-level balancing, this approach lacks sample-level adaptability and struggles with implementation in task-specific models. To address these limitations, we propose a novel task-adaptive balancing strategy that achieves sample-level balancing by redefining the derivation of $\sigma \in \mathbb{R}^1$:

$$Loss = \frac{1}{2\sigma^2} L_1(\hat{I}^{HQ}, I^{HQ}) + \log \sigma, \quad (5)$$

$$\sigma = \text{MLP}(\text{SG}([L_1(I^{LQ}, I^{HQ}), L_1(I^{LQ}, \hat{I}^{HQ}), L_1(\hat{I}^{HQ}, I^{HQ})])), \quad (6)$$

where $L_1(\cdot)$ denotes the L1 distance, and $\text{SG}(\cdot)$ denotes the stop gradient operation that decouples loss balancing and model optimization. The three terms— $L_1(I^{LQ}, I^{HQ})$, $L_1(I^{LQ}, \hat{I}^{HQ})$, and $L_1(\hat{I}^{HQ}, I^{HQ})$ —encodes sample-specific training dynamics. By feeding their concatenated values into an MLP, we estimate σ adaptively for each sample, enabling fine-grained balancing. While this strategy shifts the derivation of σ from task-index-conditioned (σ_t in Eq. 4) to sample-loss-conditioned, the core mechanism—dynamic weighting via $\frac{1}{2\sigma^2}$ and regularization via $\log \sigma$ —remains aligned with the original theory, ensuring autonomous loss balancing while extending flexibility to the sample level.

3 Experiments and Results

3.1 Dataset

We utilize the dataset provided by the paper [21], which includes three distinct datasets from three different tasks: PET synthesis, CT denoising, and MRI super-resolution. (1) The PET synthesis dataset consists of paired low-dose LQ images, with a dose reduction factor of 12, and corresponding full-dose HQ images. Each image has a size of 400×92 . The dataset includes 8,350 PET images for training, 684 for validation, and 2,044 for testing. (2) The CT denoising dataset consists of paired quarter-dose LQ images and corresponding standard-dose HQ images. Each image has a size of 512×512 . The dataset includes 2,039 CT images for training, 128 for validation, and 211 for testing. (3) The MRI super-resolution dataset consists of paired $4\times$ downsampled LQ images and corresponding HQ images. Each image has a size of 256×256 . The dataset includes 40,500 MRI images for training, 5,828 for validation, and 11,400 for testing.

Table 1. Task-specific medical image restoration results. The best results are **bolded**, and the second-best results are underlined. * Denotes results that are significantly different from the best results paired t-test at $p < 0.05$.

Method	PET Synthesis			Method	CT Denoising			Method	MRI Super-Resolution		
	PSNR \uparrow	SSIM \uparrow	RMSE \downarrow		PSNR \uparrow	SSIM \uparrow	RMSE \downarrow		PSNR \uparrow	SSIM \uparrow	RMSE \downarrow
Xiang's [1]	35.93*	0.9167*	0.0980*	REDCNN [8]	33.19*	0.9113*	8.9427*	DAGAN [13]	30.55*	0.9189*	34.0866*
DCNN [2]	36.27*	0.9243*	0.0954*	EDCNN [9]	33.41*	0.9155*	8.7401*	SwinMR [14]	30.93*	0.9253*	32.7339*
CycleWGAN [3]	36.62*	0.9290*	0.0910*	Eformer [10]	33.35*	0.9175*	8.8030*	SDAUT [15]	30.96*	0.9257*	32.5928*
ARGAN [4]	36.73*	0.9406*	0.0902*	CTformer [11]	33.25*	0.9134*	8.8974*	F-UNet [16]	31.26*	0.9314*	31.5675*
DRMC [5]	36.00*	0.9352*	0.0998*	DenoMamba [12]	33.53*	0.9149*	8.6115*	MambaIR [25]	31.77*	0.9369*	29.8372*
TAT	37.31	0.9482	0.0851	TAT	33.78	0.9199	8.3799	TAT	32.13	0.9408	28.8921

Table 2. All-in-One medical image restoration results.

Method	PET Synthesis			CT Denoising			MRI Super-Resolution			Avg.		
	PSNR \uparrow	SSIM \uparrow	RMSE \downarrow	PSNR \uparrow	SSIM \uparrow	RMSE \downarrow	PSNR \uparrow	SSIM \uparrow	RMSE \downarrow	PSNR \uparrow	SSIM \uparrow	RMSE \downarrow
ARGAN [4]	36.75*	0.9389*	0.0907*	32.92*	0.9111*	9.2110*	30.08*	0.9083*	35.7999*	33.25*	0.9194*	15.0339*
DenoMamba [12]	36.81*	0.9367*	0.0895*	33.18*	0.9115*	8.9512*	30.32*	0.9091*	34.6972*	33.44*	0.9191*	14.5793*
MambaIR [25]	37.17*	0.9458*	0.0864*	33.50*	0.9165*	8.6345*	31.31*	0.9305*	31.3150*	33.99*	0.9309*	13.3453*
AirNet [17]	37.17*	0.9451*	0.0864*	33.62*	0.9176*	8.5226*	31.39*	0.9316*	31.1141*	34.06*	0.9314*	13.2410*
AMIR [21]	37.12*	0.9475*	0.0876*	33.70*	0.9182*	8.4520*	32.03*	0.9396*	29.0988*	34.28*	0.9351*	12.5461*
TAT	37.28	0.9480	0.0856	33.80	0.9192	8.3642	32.10	0.9402	28.9145	34.39	0.9358	12.4548

Table 3. Ablation study results of TAT.

Method		Params (M)	PET Synthesis			CT Denoising			MRI Super-Resolution		
			PSNR \uparrow	SSIM \uparrow	RMSE \downarrow	PSNR \uparrow	SSIM \uparrow	RMSE \downarrow	PSNR \uparrow	SSIM \uparrow	RMSE \downarrow
TAT		41.69	37.28	0.9480	0.0856	33.80	0.9192	8.3642	32.10	0.9402	28.9145
Weight Generation	w/o	26.12	37.16*	0.9472*	0.0871*	33.64*	0.9181*	8.4845*	31.85*	0.9374*	29.7689*
	Generate All Params	663.14	37.20*	0.9455*	0.0868*	33.69*	0.9183*	8.4611*	31.89*	0.9382*	29.5431*
	w/o Stop Gradient	41.69	37.25	0.9469*	0.0860	33.69*	0.9188*	8.4559*	32.07*	0.9401	29.0291*
Loss Balancing	w/o	41.69	37.11*	0.9474*	0.0877*	33.72*	0.9183*	8.4338*	32.01*	0.9397*	29.1954*
	Task-Level [22]	41.69	37.22*	0.9475*	0.0865*	33.76*	0.9191	8.3845*	32.05*	0.9398*	29.0858*

3.2 Implementation

For model architecture, the numbers of feature extraction blocks in TAT are $L_1 = 4$, $L_2 = L_3 = 6$, and $L_4 = 8$. The residual block number in TREN is $L = 2$. The task-specific representation Z has a dimensionality of $d = 256$. For model training, we use a total batch size of 12 (4 samples per dataset) and a patch size of 128×128 . The model is optimized using the AdamW optimizer with a learning rate of 2×10^{-4} trained for 4×10^5 iterations. For model evaluation, the restoration performance is evaluated using PSNR, SSIM, and RMSE metrics.

3.3 Comparative Experiment

For task-specific MedIR, we compare TAT with five state-of-the-art (SOTA) methods for each individual task. For All-in-One MedIR, we compare TAT with five methods: two SOTA methods (AirNet [17] and AMIR [21]) designed for All-in-One restoration, and three methods (ARGAN [4], DenoMamba [12], and MambaIR [25]) that perform the best in their respective specific tasks.

Task-Specific Medical Image Restoration. Although TAT is not specifically designed for task-specific MedIR, it still achieves the best performance as shown in Table. 1. This suggests that the two proposed task-adaptive strategies are extensible and effective for individual tasks as well.

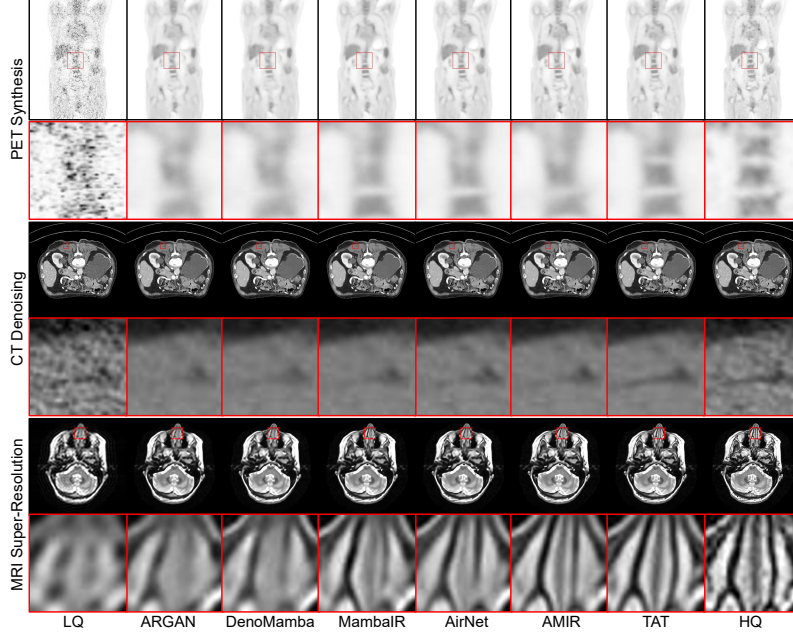


Fig. 2. Visual comparisons on All-in-One medical image restoration.

All-in-One Medical Image Restoration. The results of All-in-One MedIR are presented in Table. 2, where TAT significantly ($p < 0.05$) outperforms all comparison methods across all three tasks. Especially, it outperforms AMIR which is the current SOTA method in All-in-One MedIR. This is because TAT best deals with the inter-task relationships of *task interference* and *task imbalance*. In fact, our proposed TAT even achieves a comparable performance to task-specific models present in Table. 1. This indicates that TAT has strong model capacity and adaptability to deal with diverse tasks. The visual comparisons in Fig. 2 further highlight TAT’s superior ability to consistently restore finer structures and details of images across different tasks.

3.4 Ablation Study

We conduct comprehensive ablation studies by systematically removing or modifying individual components. The experimental results are summarized in Table. 3. For the task-adaptive weight generation, we introduce model variants: (1) without the weight generation strategy, (2) with the generation of all parameters in the Transformer block, and (3) without the stop-gradient operation. All variants exhibit consistent performance degradation, highlighting the validity of our design choices. For the task-adaptive loss balancing strategy, we evaluate two model variants: (1) without loss balancing and (2) with a conventional task-level balancing strategy [22]. Both alternatives lead to significant performance degradation, further validating the superiority of our loss balancing strategy.

4 Conclusion

In this paper, we introduce a novel task-adaptive transformer (TAT) to address the challenges of *task interference* and *task imbalance* in All-in-One medical image restoration. To mitigate *task interference*, we propose a task-adaptive weight generation strategy that produces task-specific weight parameters, thereby reducing conflicts during weight updates. To tackle *task imbalance*, we introduce a task-adaptive loss balancing strategy that dynamically adjusts the loss weights according to the learning difficulty of each task, ensuring the most effective optimization path. Experimental results demonstrate that our approach achieves state-of-the-art performance across different tasks. Considering that the two proposed task-adaptive strategies of TAT are architecture-agnostic, we plan to explore applying them to more modern architectures [26,27,25] in future work.

Acknowledgments. This work is supported by the Beijing Natural Science Foundation under Grant QY24144, the National Natural Science Foundation in China under Grant 62371016 and U23B2063, the Beijing Natural Science Foundation Haidian District Joint Fund in China under Grant L222032, the Fundamental Research Funds for the Central University of China from the State Key Laboratory of Software Development Environment in Beihang University in China, and the 111 Project in China under Grant B13003, and the high performance computing resources at Beihang University.

Disclosure of Interests. We have no conflicts of interest to disclose.

References

1. Xiang, L., Qiao, Y., Nie, D., An, L., Lin, W., Wang, Q., Shen, D.: Deep auto-context convolutional neural networks for standard-dose pet image estimation from low-dose pet/mri. *Neurocomputing* **267**, 406–416 (2017)
2. Chan, C., Zhou, J., Yang, L., Qi, W., Kolthammer, J., Asma, E.: Noise adaptive deep convolutional neural network for whole-body pet denoising. In: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC). pp. 1–4. IEEE (2018)
3. Zhou, L., Schaefferkoetter, J.D., Tham, I.W., Huang, G., Yan, J.: Supervised learning with cyclegan for low-dose fdg pet image denoising. *Medical image analysis* **65**, 101770 (2020)
4. Luo, Y., Zhou, L., Zhan, B., Fei, Y., Zhou, J., Wang, Y., Shen, D.: Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis. *Medical Image Analysis* **77**, 102335 (2022)
5. Yang, Z., Zhou, Y., Zhang, H., Wei, B., Fan, Y., Xu, Y.: Drmc: A generalist model with dynamic routing for multi-center pet image synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 36–46. Springer (2023)
6. Zhou, Y., Yang, Z., Zhang, H., Eric, I., Chang, C., Fan, Y., Xu, Y.: 3d segmentation guided style-based generative adversarial networks for pet synthesis. *IEEE Transactions on Medical Imaging* **41**(8), 2092–2104 (2022)
7. Gong, K., Johnson, K., El Fakhri, G., Li, Q., Pan, T.: Pet image denoising based on denoising diffusion probabilistic model. *European Journal of Nuclear Medicine and Molecular Imaging* **51**(2), 358–368 (2024)

8. Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G.: Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging* **36**(12), 2524–2535 (2017)
9. Liang, T., Jin, Y., Li, Y., Wang, T.: Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising. In: 2020 15th IEEE International conference on signal processing (ICSP). vol. 1, pp. 193–198. IEEE (2020)
10. Luthra, A., Sulakhe, H., Mittal, T., Iyer, A., Yadav, S.: Eformer: Edge enhancement based transformer for medical image denoising. *arXiv preprint arXiv:2109.08044* (2021)
11. Wang, D., Fan, F., Wu, Z., Liu, R., Wang, F., Yu, H.: Ctformer: convolution-free token2token dilated vision transformer for low-dose ct denoising. *Physics in Medicine & Biology* **68**(6), 065012 (2023)
12. Öztürk, Ş., Duran, O.C., Çukur, T.: Denomamba: A fused state-space model for low-dose ct denoising. *arXiv preprint arXiv:2409.13094* (2024)
13. Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., et al.: Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging* **37**(6), 1310–1321 (2017)
14. Huang, J., Fang, Y., Wu, Y., Wu, H., Gao, Z., Li, Y., Del Ser, J., Xia, J., Yang, G.: Swin transformer for fast mri. *Neurocomputing* **493**, 281–304 (2022)
15. Huang, J., Xing, X., Gao, Z., Yang, G.: Swin deformable attention u-net transformer (sdaut) for explainable fast mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 538–548. Springer (2022)
16. Sun, H., Li, Y., Li, Z., Yang, R., Xu, Z., Dou, J., Qi, H., Chen, H.: Fourier convolution block with global receptive field for mri reconstruction. *Medical Image Analysis* **99**, 103349 (2025)
17. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-in-one image restoration for unknown corruption. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17452–17462 (2022)
18. Park, D., Lee, B.H., Chun, S.Y.: All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5815–5824. IEEE (2023)
19. Ma, C., Li, Z., He, J., Zhang, J., Zhang, Y., Shan, H.: Prompted contextual transformer for incomplete-view ct reconstruction. *arXiv preprint arXiv:2312.07846* (2023)
20. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv:2306.13090* (2023)
21. Yang, Z., Chen, H., Qian, Z., Yi, Y., Zhang, H., Zhao, D., Wei, B., Xu, Y.: All-in-one medical image restoration via task-adaptive routing. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 67–77. Springer (2024)
22. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7482–7491 (2018)
23. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)

24. Guo, J., Han, K., Wu, H., Tang, Y., Chen, X., Wang, Y., Xu, C.: Cmt: Convolutional neural networks meet vision transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12175–12185 (2022)
25. Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., Xia, S.T.: Mambair: A simple baseline for image restoration with state-space model. In: European Conference on Computer Vision. pp. 222–241. Springer (2025)
26. Yang, Z., Chen, H., Qian, Z., Zhou, Y., Zhang, H., Zhao, D., Wei, B., Xu, Y.: Region attention transformer for medical image restoration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 603–613. Springer (2024)
27. Yang, Z., Li, J., Zhang, H., Zhao, D., Wei, B., Xu, Y.: Restore-rwkv: Efficient and effective medical image restoration with rwkv. arXiv preprint arXiv:2407.11087 (2024)