Supporting Information

1. Statistical analysis

TABLE S-1 *p*-value between the quantitative metrics of denoising methods and P-HTC-net.

Dataset	Method	SNRs = (8, 15)			SNRs = (5, 8)		
		PSNR	SSIM	Dice	PSNR	SSIM	Dice
HP ^{129}Xe MRI dataset	Noisy	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	NLM	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	UNLM	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	DnCNN	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	WDNN	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	SwinIR	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	HTC-net	0.027	0.882	0.002	<0.001	<0.001	<0.045
IXI dataset	Noisy	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	NLM	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	UNLM	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	DnCNN	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	WDNN	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	SwinIR	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	HTC-net	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Statistical analysis using a paired t-test

TABLE S-2 *p*-value between the quantitative metrics of various network variants and P-HTC-net.

Dataset	Method	SNRs = (8, 15)		SNRs = (5, 8)	
		PSNR	SSIM	PSNR	SSIM
HP ^{129}Xe MRI dataset	u	<0.001	<0.001	<0.001	<0.001
	d	<0.001	<0.001	<0.001	<0.001
	u+d	<0.001	<0.001	<0.001	<0.001
	u+d+t	<0.001	<0.001	<0.001	<0.001
	u+d+t+f	0.027	0.882	<0.001	<0.001
IXI dataset	u	<0.001	<0.001	<0.001	<0.001
	d	<0.001	<0.001	<0.001	<0.001
	u+d	<0.001	<0.001	<0.001	<0.001
	u+d+t	<0.001	<0.001	<0.001	<0.001
	u+d+t+f	<0.001	<0.001	<0.001	<0.001

Statistical analysis using a paired t-test

TABLE S-3 *p*-value for the metrics of networks trained with different methods.

Dataset	Method	SNRs = (8, 15)		SNRs = (5, 8)	
		PSNR	SSIM	PSNR	SSIM
HP ^{129}Xe MRI dataset	w/o self-supervised pre-training	<0.001	0.046	<0.001	<0.001
	self-supervised pre-training (847 samples)	0.147	0.013	0.112	0.283
IXI dataset	w/o self-supervised pre-training	<0.001	<0.001	<0.001	<0.001
	self-supervised pre-training (4000 samples)	<0.001	<0.001	<0.001	<0.001

Statistical analysis using a paired t-test

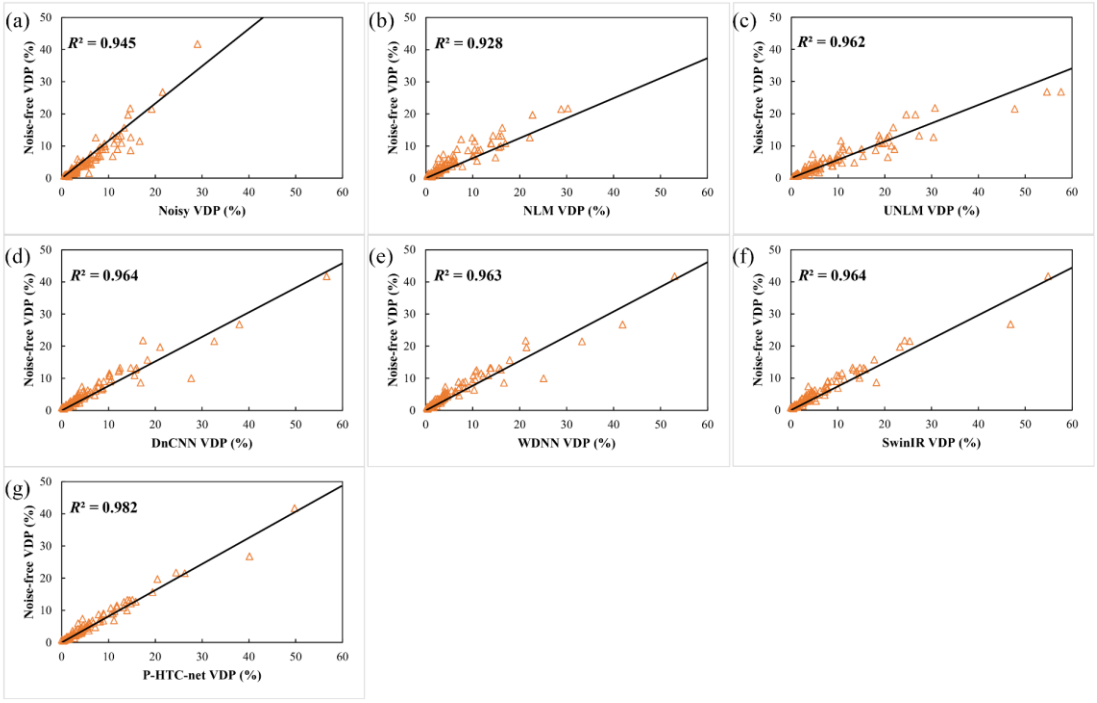

FIGURE S-1 VDP comparisons of noise-free images and different methods (SNR ≥ 8). (a) Scatter plots of VDP calculated from noisy images and noise-free images. (b) Scatter plots of VDP calculated from NLM and noise-free images. (c) Scatter plots of VDP calculated from UNLM and noise-free images. (d) Scatter plots of VDP calculated from DnCNN and noise-free images. (e) Scatter plots of VDP calculated from WDNN and noise-free images. (f) Scatter plots of VDP calculated from SwinIR and noise-free images. (g) Scatter plots of VDP calculated from P-HTC-net and noise-free images. R-squared values are listed in the scatter plots.

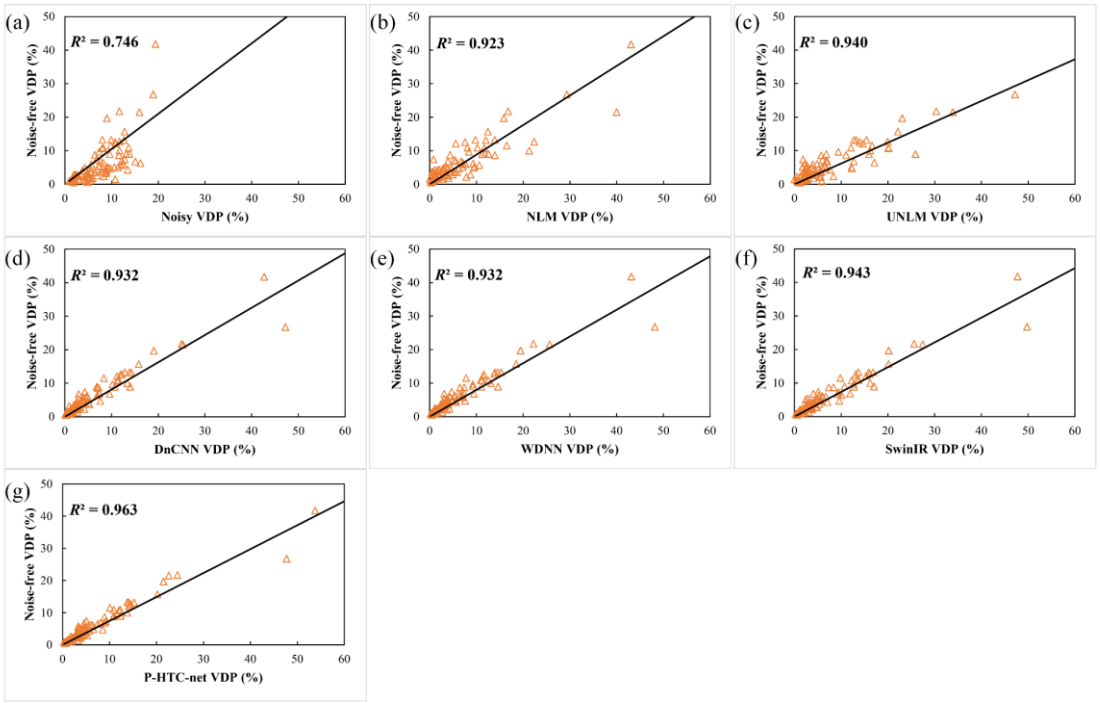

FIGURE S-2 VDP comparisons of noise-free images and different methods (SNR < 8). (a) Scatter plots of VDP calculated from noisy images and noise-free images. (b) Scatter plots of VDP calculated from NLM and noise-free images. (c) Scatter plots of VDP calculated from UNLM and noise-free images. (d) Scatter plots of VDP calculated from DnCNN and noise-free images. (e) Scatter plots of VDP calculated from WDNN and noise-free images. (f) Scatter plots of VDP calculated from SwinIR and noise-free images. (g) Scatter plots of VDP calculated from P-HTC-net and noise-free images. R-squared values are listed in the scatter plots.

2. Results on real noisy MR images

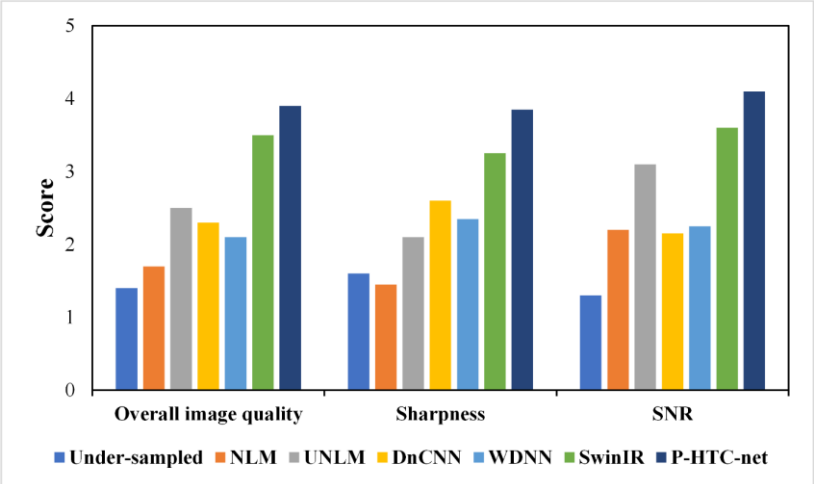

FIGURE S-3 Quantitative comparisons of image quality scores among various methods. Image quality scores are evaluated by two experts in terms of overall image quality, sharpness and SNR.

We present denoised results of different methods on real noisy hyperpolarization (HP) ^{129}Xe images to verify the effectiveness of the proposed method in Figure S-4. Due to lacking

real paired noisy and noise-free MR images, we use the models trained on synthetic pulmonary HP ^{129}Xe MRI dataset with SNRs ranging from 8 to 15 to perform this experiment. As shown in figure S-1, the proposed method produces clearer bronchi and recovers more details than other methods. Additionally, in comparison to other deep learning methods, the proposed method effectively recovers the low signal regions in MR images.

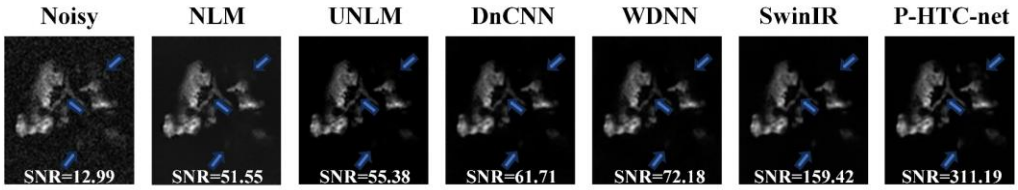

FIGURE S-4 Denoising results on real noisy pulmonary HP ^{129}Xe image.

3. Ablation studies on various network configurations on the IXI dataset

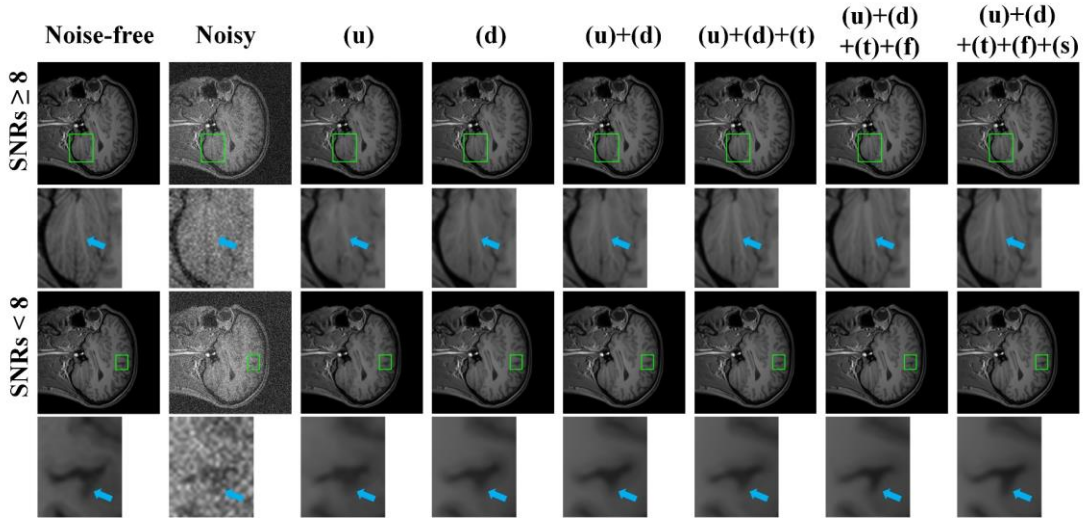

FIGURE S-5 The denoising results of various network configurations on the IXI dataset: (u) denoising with only the CNN branch, (d) denoising with only the Transformer-CNN branch without Transformer encoder, (t) the Transformer encoder, (f) the residual fusion block, (s) the self-supervised pre-training.

4. Ablation studies on pre-training strategy

To better demonstrate the advantages of self-supervised pre-training, we conduct additional experiments, including training the network with half of the paired images, pre-training with all noisy images followed by training with half of the paired images, and training with all paired images. The results are presented in Table S-4.

TABLE S-4 Ablation studies on pre-training strategy on the pulmonary HP ^{129}Xe MRI dataset and the IXI dataset (mean ± standard deviation).

Dataset	Method	SNRs = (8, 15)		SNRs = (5, 8)	
		PSNR	SSIM	PSNR	SSIM
HP ^{129}Xe MRI dataset	50%paired	29.877±2.378**	0.866±0.057**	27.718±1.897**	0.818±0.064**
	50%paired+pretrain	29.973±2.411**	0.868±0.057**	27.825±1.908**	0.821±0.063
	100%paired	30.007±2.429	0.869±0.057	27.804±1.925	0.820±0.065
IXI dataset	50%paired	35.140±1.240**	0.971±0.006**	31.749±1.225**	0.941±0.011**
	50%paired+pretrain	35.369±1.249**	0.972±0.006**	32.118±1.257**	0.946±0.011**
	100%paired	35.392±1.238	0.972±0.006	32.105±1.203	0.947±0.010

* and ** denote p-value < 0.05 and p-value < 0.01, respectively.

After pre-training with 847 noisy HP ^{129}Xe images, we compare the results of supervised training using various numbers of paired images. As shown in Figure S-3, self-supervised pre-training can reduce the requirement for paired images.

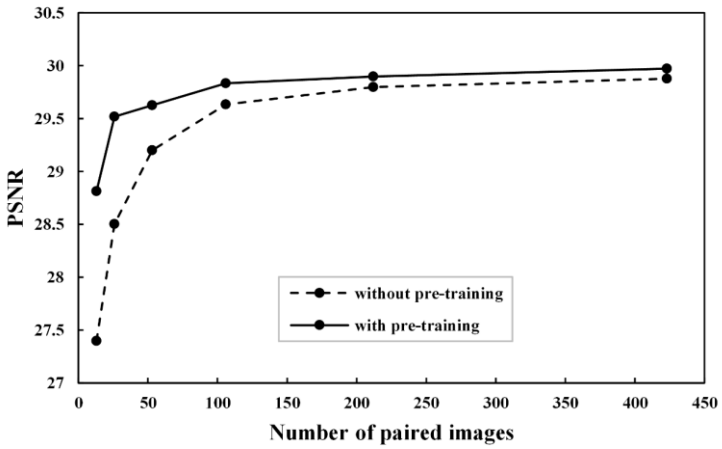

FIGURE S-6 The average PSNR of HTC-net with different number of paired images.

5. Visualization of the feature maps and attention score maps

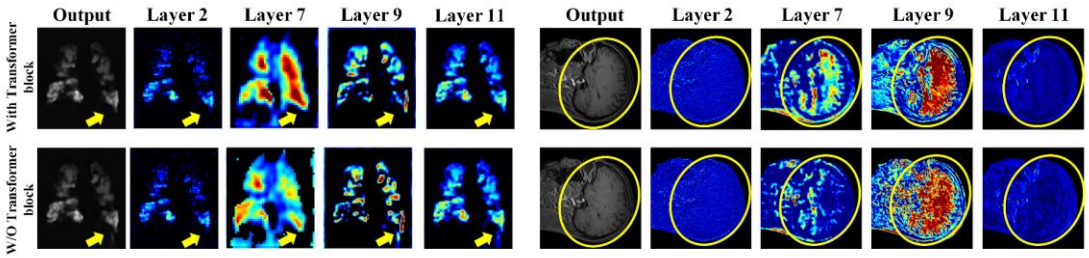

FIGURE S-7 Visualization of the feature maps and output of HTC-net with or without the Transformer blocks.

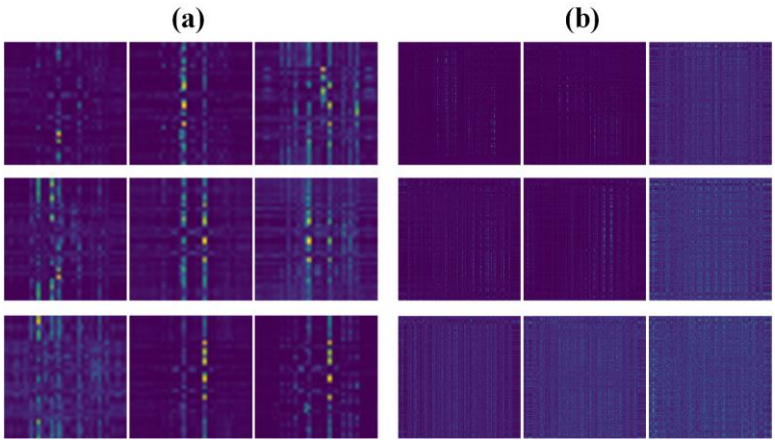

FIGURE S-8 Visualization of attention score maps. (a) Attention maps of P-HTC-net in the pulmonary HP ^{129}Xe MRI dataset. (b) Attention maps of P-HTC-net in the IXI dataset.