MoSE: Mixture-of-Scale-Experts for Medical Image Restoration

Xingyu Li[1], Zhaohu Xing[1], Zekai Liu[1], Shuaibo Li[1], Hongqiu Wang[1], Song Fei[1], and Lei Zhu[1,2] ✉

[1] The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

[2] The Hong Kong University of Science and Technology, Hong Kong, China

leizhu@hkust-gz.edu.cn

Abstract. Medical image restoration (MedIR) is crucial for obtaining high-quality medical images and enabling accurate diagnosis. Existing unified restoration models often suffer from parameter redundancy and high computational cost due to their shared architecture design. While Mixture-of-Experts (MoE) introduces sparse computation, the homogeneous design of experts struggles to handle diverse degradation patterns, resulting in texture distortions and anatomical structure loss. To address these challenges, we propose Mixture-of-Scale-Experts (MoSE) framework, which employs scale-heterogeneous experts operating with different anatomical context ranges, and dynamically selects the most suitable expert for each input. A scale-aware load-balancing loss further promotes balanced expert utilization. Furthermore, we develop an Orientation-Contrast Guided (OCG) filter, combining Haar wavelet–based directional sensing with local contrast enhancement to guide high-frequency reconstruction of fine-grained structures. Finally, we propose a novel Anatomical Texture-Level Similarity (ATLS) metric to better evaluate the structural fidelity of restored anatomical textures. Extensive experiments on a private endoscopic super-resolution dataset and a public all-in-one MedIR benchmark (CT, PET, and MRI) demonstrate that MoSE significantly outperforms state-of-the-art methods in both quantitative metrics and visual quality while maintaining high inference efficiency.

Keywords: Mixture-of-scale-experts · Anatomical texture-level similarity · Medical image restoration · Endoscopy image.

1 Introduction

High-quality medical images are important for accurate diagnosis and clinical decision-making [2,8,30]. However, medical images are often acquired under constrained conditions, which can result in low-quality images with diverse degradation patterns. For example, endoscopic images derived from medical video are affected by camera motion, hardware conditions, and physiological movement [4,15,6], leading to blur, noise, and low resolution. In cross-modality datasets containing CT, MRI, and PET images, degradation sources further vary

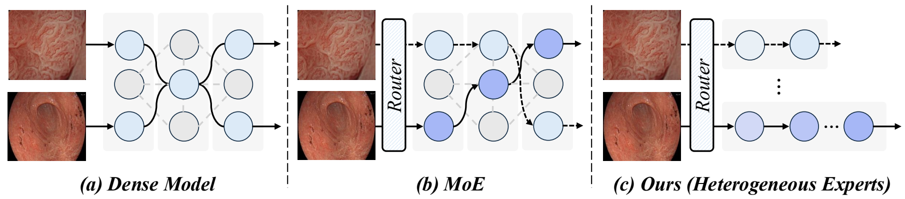


Fig. 1: **Motivation:** (a) Dense restoration models [12,14] share parameters across degradations, causing redundant computation. (b) MoE [23,26] enables sparse computation, but its homogeneous expert design cannot adapt to diverse medical degradation patterns. (c) Our method employs heterogeneous scale experts, achieving accurate structural restoration and better texture preservation.

due to modality-specific acquisition mechanisms [9]. Unlike natural images, medical images exhibit complex anatomical structures [30], which are highly sensitive to subtle structural distortions. Therefore, the main challenge in medical image restoration (MedIR) is restoring high-quality anatomical structures from diverse degradation patterns while preserving fine texture details.

In recent years, unified restoration frameworks [1,3,5] have improved restoration performance across multiple degradation types. As shown in Fig. 1a, due to variations in anatomical contexts, a single model often fails to fully leverage its parameters for medical images with certain degradation types, leading to redundancy and increased computational cost. Fig. 1b illustrates the Mixture-of-Experts (MoE) framework, which dynamically routes inputs through different computational paths [21]. In the medical image restoration domain, AMIR [23] incorporates a feature routing mechanism but still employs a homogeneous-expert MoE architecture. The expert structures are uniform in the Mixture-of-Experts (MoE) [10,11,22], lacking specialization for medical images with diverse anatomical contexts. Consequently, when processing different anatomical structures, the experts fail to achieve functional specialization, which may lead to structural distortions and blurred anatomical textures.

To address these challenges, we propose Mixture-of-Scale-Experts (MoSE), a novel framework for medical image restoration. Our MoSE introduces heterogeneous experts with progressive receptive fields for medical image restoration. A routing module dynamically selects the most suitable expert for each input. Shallow–wide experts specialize in recovering local high-frequency details (edges and fine textures), whereas deep–narrow experts capture long-range dependencies and contextual anatomical priors, enabling structurally consistent restoration. To ensure balanced utilization of experts, we introduce a scale-aware load-balancing loss. Furthermore, an Orientation–Contrast Guided (OCG) filter combines a Haar wavelet–based orientation branch with a local contrast enhancement branch for more accurate detail reconstruction. Additionally, we design an Anatomical Texture–Level Similarity (ATLS) metric to evaluate restoration quality at the patch level. Experimental results on our self-constructed endo-

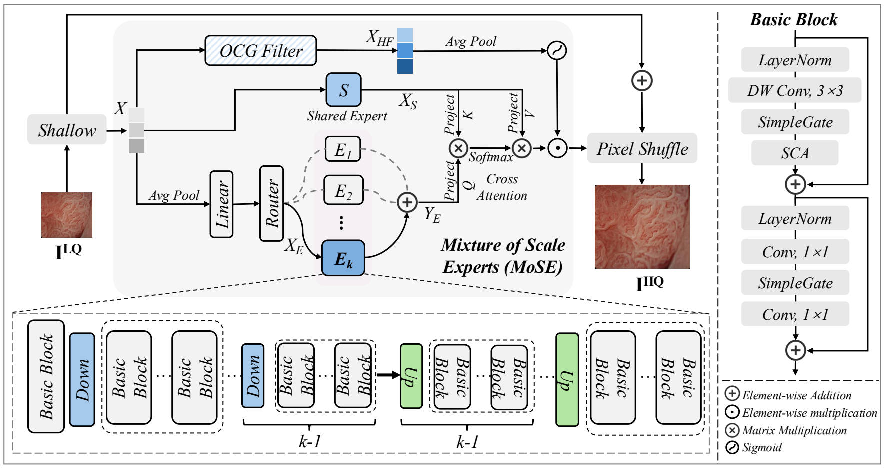


Fig. 2: **Proposed MoSE framework.** Our MoSE employs heterogeneous experts to handle diverse degradation patterns in medical images. Each expert is designed with a different downsampling depth. An Orientation–Contrast Guided (OCG) filter provides high-frequency guidance for more accurate restoration, while a global residual connection preserves low-frequency consistency.

scopic super-resolution dataset and the public all-in-one MedIR dataset show that MoSE outperforms state-of-the-art methods in both quantitative metrics and visual quality.

2 Method

Traditional Mixture-of-Experts (MoE) [17,18] designs lack the architectural flexibility required for diverse medical image degradations. To address this, as illustrated in Fig. 2, we propose the Mixture-of-Scale-Experts (MoSE) framework, which employs heterogeneous expert branches with progressive receptive fields to handle scale-variant features, ranging from local textures to global semantics.

2.1 Mixture-of-Scale-Experts

In our MoSE design, the scale expert uses a U-shaped [19] structure, but with a graduated increase in downsampling layers and network depth to achieve scale-aware functional specialization. In particular, shallow-wide experts are dedicated to capturing local high-frequency textures and boundary refinements, mid-level experts model mesoscale tissue relationships, and deep-narrow experts focus on capturing global anatomical structures and long-range dependencies. However, relying solely on dynamic routing can lead to instability in critical high-frequency details, such as micro-vascular orientations and lesion contours. Thus, we introduce a shared expert $\mathcal{E}_s$ that remains active across all passes to provide a consistent geometric prior. Specifically, the selected scale expert produces features $Y_{\mathrm{E}} = \mathcal{E}_k(x)$, while the shared expert outputs $X_{\mathrm{S}} = \mathcal{E}_s(x)$. To integrate these features efficiently, we employ a channel cross-attention (C-XCA) mechanism [28], where Y_{E} serves as the query Q, and X_{S} serves as key K and value V. The fused feature F_{comb} is then explicitly modulated by the orientation contrast guide X_{HF} to stabilize high-frequency details. The overall process is formulated as:

$$F_{\mathrm{comb}} = Y_{\mathrm{E}} + \gamma_{\mathrm{attn}} \cdot \mathbf{W}_{\mathrm{out}}\mathrm{C\text{-}XCA}(Y_{\mathrm{E}}, X_{\mathrm{S}}), \tag{1}$$

$$\mathcal{F}_\theta(\mathbf{x}) = F_{\mathrm{comb}} \odot (1 + \gamma_{\mathrm{hf}} \cdot \sigma(\mathrm{GAP}(X_{\mathrm{HF}}))), \tag{2}$$

where $\mathbf{W}$ denotes convolution layers, $\odot$ is element-wise multiplication, and GAP is global average pooling. The learnable scaling coefficients γ_{attn} and γ_{hf} are zero-initialized for training stability, and τ is a learnable temperature parameter.

Scale-Aware Load-Balancing Loss. To prevent the model from relying on a single expert and promote balanced expert utilization, we introduce a scale-aware load-balancing auxiliary loss [18]. For each expert e in a batch, we define its importance I_e as the sum of routing probabilities assigned to it, and its load L_e as the number of samples actually routed to it. The auxiliary loss minimizes the squared coefficient of variation (CV^2) of the importance $\mathbf{I}$ and load $\mathbf{L}$ across all k experts:

$$\mathcal{L}_{\mathrm{aux}} = \frac{1}{2}\left(\frac{\mathrm{Var}(\mathbf{I})}{\bar{\mathbf{I}}^2} + \frac{\mathrm{Var}(\mathbf{L})}{\bar{\mathbf{L}}^2}\right). \tag{3}$$

The total training loss is then defined as

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{recon}} + \lambda\,\mathcal{L}_{\mathrm{aux}}, \tag{4}$$

where $\mathcal{L}_{\mathrm{recon}}$ is the L2 (MSE) loss for the restored medical images, and λ is set to 0.1 to weight the auxiliary term.

2.2 Orientation–Contrast Guided (OCG) Filter

Medical images are typically characterized by low contrast and complex tissue interfaces [31,25], where vessels, tumor margins, and organ boundaries often appear as weak yet directionally continuous high-frequency patterns. To improve the accurate reconstruction of anatomical structures, we introduce an Orientation–Contrast Guided (OCG) filter. We use local contrast to highlight structural regions while suppressing homogeneous backgrounds. A Haar wavelet transform [16] is applied to decompose the feature map into high-frequency subbands and introduce an orientation-consistency measure to emphasize regions exhibiting stable directional continuity. This filter produces a direction-sensitive high-frequency representation, which can be expressed as follows:

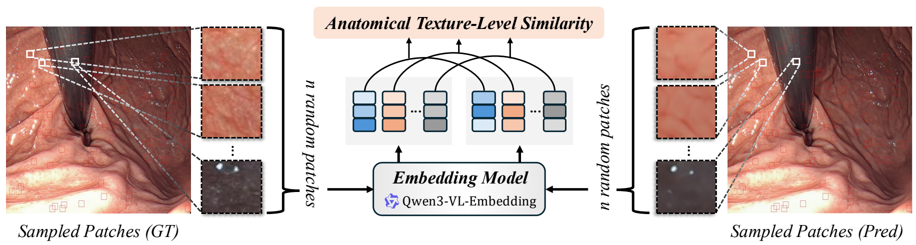


Fig. 3: Pipeline of the proposed ATLS metric. ATLS evaluates local texture-level accuracy using patch-wise similarity in a shared semantic latent space.

$$\mathbf{F} = \lambda_1(\tilde{\mathbf{W}} + \tilde{\mathbf{C}}) + \lambda_2(\tilde{\mathbf{W}} \odot \tilde{\mathbf{C}}), \tag{5}$$

where $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{C}}$ denote the normalized wavelet and local contrast features, $\odot$ is element-wise multiplication, and we set $\lambda_1 = \lambda_2 = 0.5$.

The OCG filter produces a high-frequency guidance map that highlights anatomical structures, helping expert branches restore fine details and maintain structural consistency in medical images, while avoiding noise amplification.

2.3 Anatomical Texture–Level Similarity (ATLS)

Commonly used metrics such as PSNR and SSIM mainly measure average pixel-wise similarity but fail to explicitly evaluate subtle anatomical textures critical for diagnosis. To address this, we propose the Anatomical Texture-Level Similarity (ATLS) metric. As shown in Fig. 3, given a restored image I_p and its corresponding ground truth I_{gt}, we randomly sample N spatially aligned local patches $\{P_i^r\}_{i=1}^N$ and $\{P_i^{gt}\}_{i=1}^N$. Each patch is mapped to a high-level embedding representation using a pre-trained visual encoder $f(\cdot)$. To accurately capture texture-level features, we employ Qwen-3-VL-Embedding [13], the state-of-the-art visual-language embedding model. Its cross-modal pretraining ensures high sensitivity to fine textures, and patch embeddings enable precise comparison of local features for detecting subtle anatomical differences. The ATLS score is defined as the average cosine similarity between corresponding patch embeddings:

$$\text{ATLS} = \frac{1}{N}\sum_{i=1}^{N} \frac{f(P_i^p) \cdot f(P_i^{gt})}{\|f(P_i^p)\| \, \|f(P_i^{gt})\|}. \tag{6}$$

By measuring similarity patch-wise, ATLS captures structural and semantic consistency of anatomical textures, providing a clinically meaningful assessment of whether fine-scale anatomical structures are correctly modeled by the restoration method. This metric complements conventional full-reference measures by explicitly emphasizing texture-level anatomical fidelity.

Table 1: Quantitative comparison on multiple medical datasets. PSNR (dB) ↑, SSIM ↑, LPIPS ↓, and ATLS ↑. The best and second-best results are highlighted in bold and underlined, respectively.

Methods	Endoscopic SR				MRI SR				CT Denoising				PET Synthesis			
	PSNR	SSIM	LPIPS	ATLS	PSNR	SSIM	LPIPS	ATLS	PSNR	SSIM	LPIPS	ATLS	PSNR	SSIM	LPIPS	ATLS
AMIR [23]	28.08	0.633	0.515	0.710	34.54	0.842	0.153	0.876	34.06	0.893	0.185	0.837	39.10	0.945	0.156	0.805
TAT [24]	31.95	0.683	0.545	0.723	36.73	0.901	0.097	**0.905**	36.82	0.936	0.101	0.882	35.99	0.829	0.424	0.760
AirNet [12]	32.67	0.743	0.516	0.786	33.58	0.823	0.190	0.836	33.67	0.848	0.194	0.830	36.87	0.861	0.338	0.791
NAFNet [3]	33.22	0.757	0.402	0.815	36.89	0.921	0.101	0.872	36.46	0.911	0.178	0.880	39.57	0.953	0.137	**0.820**
MambaIR [5]	32.67	0.725	0.477	0.802	36.13	0.901	0.108	0.871	36.36	0.913	0.179	0.857	39.66	0.954	0.139	0.813
SwinIR [14]	31.64	0.642	0.593	0.751	35.46	0.814	0.131	0.870	36.79	0.927	0.133	0.892	37.98	0.943	0.148	0.804
MoCE-IR [27]	33.24	0.761	0.382	0.822	36.01	0.910	0.151	0.876	36.73	0.923	0.106	0.897	36.08	0.887	0.189	0.761
Ours	**33.88**	**0.796**	**0.323**	**0.834**	**37.18**	**0.929**	**0.094**	0.899	**37.05**	**0.953**	**0.099**	**0.905**	**40.39**	**0.961**	**0.109**	0.814

3 Experiments

3.1 Datasets and Implementation Details

Endoscopic Super-Resolution Dataset (ESR). We initially collected 4,000 endoscopic image pairs in collaboration with a clinical institution. After manually processing low-quality images, we selected 3,225 high-quality paired images, each with a high-resolution (HR) image of 2880×1920 and its corresponding low-resolution (LR) image.

All-in-One MedIR Dataset. Following TAT [24], we adopt the all-in-one medical restoration dataset provided by AMIR [23] for evaluation, which includes MRI super-resolution, CT denoising, and PET synthesis. The dataset consists of 11,078 PET pairs, 2,378 CT cases, and 57,728 MRI samples.

Implementation Details. For model training, we employ the AdamW optimizer with a learning rate of 3×10^{-3}. For fair comparison, all methods are trained and evaluated on the same data splits (8:1:1 for training, validation, and testing) and on the same device. For the endoscopic super-resolution task, the super-resolution factor is set to 4×. To evaluate restoration quality, we employ standard metrics including PSNR [7], SSIM [20], and LPIPS [29], alongside our proposed Anatomical Texture-Level Similarity (ATLS) metric, we randomly sample 100 patches to evaluate the fidelity of clinical structures.

3.2 Comparison with SOTA Methods

We compare MoSE with seven state-of-the-art (SOTA) methods, including two approaches specifically designed for medical image all-in-one restoration (AMIR [23] and TAT [24]), and five methods developed for natural image all-in-one restoration (AirNet [12], NAFNet [3], MambaIR [5], SwinIR [14], and MoCE-IR [27]).

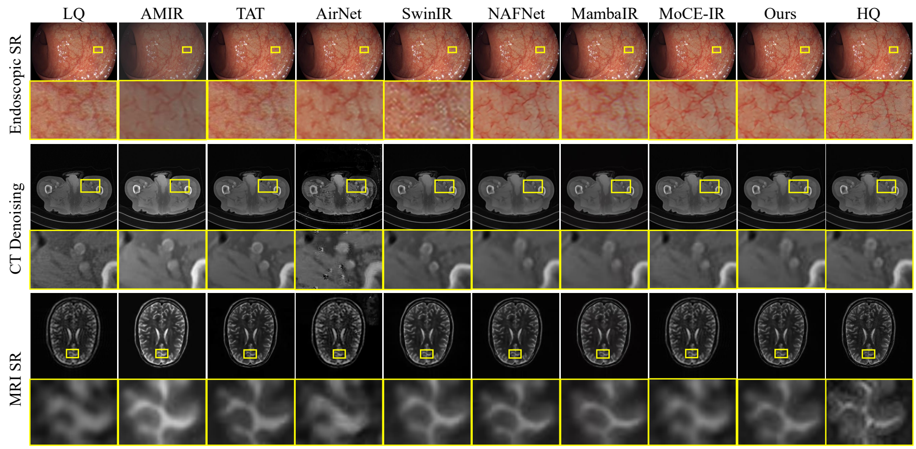
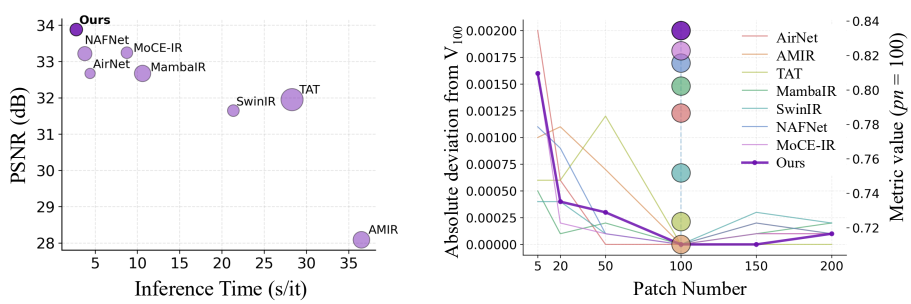


Fig. 4: Visual comparison between the proposed MoSE and other methods.



(a) Inference efficiency comparison. (b) ATLS variation with patch number.

Fig. 5: Inference efficiency and ablation study on patch number.

Quantitative Comparison. The quantitative results are shown in Table 1. Our method achieves the best performance on most benchmarks across four restoration tasks, demonstrating strong cross-task robustness to diverse degradation patterns. In particular, it consistently performs well in terms of PSNR, SSIM, and LPIPS, reflecting improved preservation of anatomical structures and fine details. Although some methods achieve higher ATLS scores, our approach attains second-best performance in those cases while maintaining more balanced and stable results across different modalities. Overall, these findings validate the effectiveness and generalization capability of the proposed MoSE framework.

Visual Comparison. We present visual comparisons in Fig. 4. In the endoscopic super-resolution, compared methods suffer from color deviations, noise

Table 2: Ablation study of our MoSE.

Model	Modules			Endoscopic SR			Medical benchmark avg.		
	MoSE	L_{aux}	OCG	SSIM ↑	PSNR ↑	ATLS ↑	SSIM ↑	PSNR ↑	ATLS ↑
$\mathcal{E}_1$				0.7456	32.5852	0.7848	0.9005	36.7454	0.8277
$\mathcal{E}_2$				0.7370	32.5745	0.7836	0.8961	36.4070	0.8289
$\mathcal{E}_3$				0.7138	32.1464	0.7832	0.8895	36.6681	0.8253
M1	✓			0.7771	33.5642	0.8053	0.9210	37.6682	0.8509
M2	✓	✓		0.7897	33.7541	0.8191	0.9382	37.9161	0.8676
Ours	✓	✓	✓	**0.7964**	**33.8812**	**0.8345**	**0.9480**	**38.2111**	**0.8879**

amplification, or blurred boundaries. Our method reconstructs vessel structures faithfully and maintains consistent textures, outputting the results that more closely match the high-resolution images. For CT and MRI restoration, compared approaches tend to over-smooth details or introduce local distortions, especially in low-contrast regions. In contrast, our model better preserves tissue boundaries and anatomical structures. Combined with the quantitative results, these findings underscore the effectiveness of our method.

Inference Efficiency. As shown in Fig. 5a, we compare the inference time and restoration performance of different methods on the ESR dataset, where the bubble size represents the model parameter scale. Our method achieves the highest PSNR while maintaining one of the lowest inference times. Compared with representative restoration models, our method achieves state-of-the-art restoration accuracy with improved parameter efficiency and reduced computational cost, validating the effectiveness of the proposed MoSE framework.

3.3 Ablation Studies

Effectiveness of Key Components in the MoSE. We conduct ablation studies to evaluate the contributions of the MoSE architecture, L_{aux}, and the OCG module, as summarized in Table 2. Three single-expert baselines ($\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$) achieve comparable results but none consistently achieves the best results. Integrating scale-specific experts within the MoSE framework (M1) achieves the largest performance increase across all metrics, highlighting the effectiveness of MoSE architecture in improving medical image restoration quality. Incorporating L_{aux} (M2) promotes balanced expert usage, resulting in improved performance across all evaluation metrics. Finally, by adding the OCG module, we obtain the final model (Ours), which surpasses all baselines. The ablation results validate the effectiveness of each component within the proposed MoSE framework.

The Effect of the Patch Number in ATLS. To assess the effect of the number of patches, we measure ATLS under various sampling settings. As shown in

Fig. 5b, small numbers of patches lead to noticeable deviations due to insufficient coverage of global anatomical structures. As the number increases, the deviation decreases, and the curves stabilize at around 100 patches. Considering both stability and efficiency, we adopt 100 patches as the default setting.

4 Conclusion

In this work, we propose a novel Mixture-of-Scale-Experts (MoSE) framework for medical image restoration. MoSE employs heterogeneous scale experts with different receptive fields to adaptively handle diverse degradation patterns, with a scale-aware load-balancing loss designed to promote balanced utilization of all experts. We further introduce an Orientation–Contrast Guided (OCG) filter to provide high-frequency guidance for more accurate and detailed reconstruction. Finally, we propose a novel Anatomical Texture–Level Similarity (ATLS) metric for patch-wise evaluation of anatomical details. Experiments on a private ESR dataset and a public all-in-one MedIR benchmark demonstrate that MoSE consistently outperforms state-of-the-art methods while achieving higher inference efficiency, indicating strong potential for clinical applications.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Project No.82572383) and the Program of Guangdong Education Department.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Chen, H., Yang, Z., Hou, H., Zhang, H., Wei, B., Zhou, G., Xu, Y.: All-in-one medical image restoration with latent diffusion-enhanced vector-quantized codebook prior. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 67–77. Springer (2025)
2. 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)
3. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European conference on computer vision. pp. 17–33. Springer (2022)
4. Chow, L.S., Paramesran, R.: Review of medical image quality assessment. Biomedical signal processing and control **27**, 145–154 (2016)
5. 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 (2024)
6. Hayat, M., Gupta, M., Suanpang, P., Nanthaamornphong, A.: Super-resolution methods for endoscopic imaging: A review. In: 2024 12th International Conference on Internet of Everything, Microwave, Embedded, Communication and Networks (IEMECON). pp. 1–6 (2024)

7. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
8. 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)
9. Jiang, J., Zuo, Z., Wu, G., Jiang, K., Liu, X.: A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025)
10. Jiang, S., Zheng, T., Zhang, Y., Jin, Y., Yuan, L., Liu, Z.: Med-moe: Mixture of domain-specific experts for lightweight medical vision-language models. In: Findings of the association for computational linguistics: EMNLP 2024. pp. 3843–3860 (2024)
11. Jiang, Y., Shen, Y.: M4oe: A foundation model for medical multimodal image segmentation with mixture of experts. In: international conference on medical image computing and computer-assisted intervention. pp. 621–631. Springer (2024)
12. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-In-One Image Restoration for Unknown Corruption. In: IEEE Conference on Computer Vision and Pattern Recognition (2022)
13. Li, M., Zhang, Y., Long, D., Chen, K., Song, S., Bai, S., Yang, Z., Xie, P., Yang, A., Liu, D., Zhou, J., Lin, J.: Qwen3-vl-embedding and qwen3-vl-reranker: A unified framework for state-of-the-art multimodal retrieval and ranking. arXiv (2026)
14. 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)
15. Liu, X., Sun, G., Wang, C., Yuan, Y., Konukoglu, E.: Medvsr: Medical video super-resolution with cross state-space propagation. In: ICCV 2025 (2025)
16. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE transactions on pattern analysis and machine intelligence **11**(7), 674–693 (1989)
17. Puigcerver, J., Riquelme, C., Mustafa, B., Houlsby, N.: From sparse to soft mixtures of experts. arXiv preprint arXiv:2308.00951 (2023)
18. Riquelme, C., Puigcerver, J., Mustafa, B., Neumann, M., Jenatton, R., Susano Pinto, A., Keysers, D., Houlsby, N.: Scaling vision with sparse mixture of experts. Advances in Neural Information Processing Systems **34**, 8583–8595 (2021)
19. 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. pp. 234–241. Springer (2015)
20. Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through fsim, ssim, mse and psnr—a comparative study. Journal of Computer and Communications **7**(3), 8–18 (2019)
21. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538 (2017)
22. Wang, G., Ye, J., Cheng, J., Li, T., Chen, Z., Cai, J., He, J., Zhuang, B.: Sam-med3d-moe: Towards a non-forgetting segment anything model via mixture of experts for 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 552–561. Springer (2024)
23. 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)

24. Yang, Z., Zhang, J., Yi, Y., Liang, J., Wei, B., Xu, Y.: Tat: Task-adaptive transformer for all-in-one medical image restoration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 565–575. Springer (2025)
25. You, C., Dai, W., Min, Y., Staib, L., Duncan, J.S.: Implicit anatomical rendering for medical image segmentation with stochastic experts. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 561–571. Springer (2023)
26. Yu, X., Zhou, S., Li, H., Zhu, L.: Multi-expert adaptive selection: Task-balancing for all-in-one image restoration. IEEE Transactions on Circuits and Systems for Video Technology **35**(5), 4619–4634 (2024)
27. Zamfir, E., Wu, Z., Mehta, N., Tan, Y., Paudel, D.P., Zhang, Y., Timofte, R.: Complexity experts are task-discriminative learners for any image restoration. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 12753–12763 (2025)
28. 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)
29. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
30. 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)
31. Zhou, S., Nie, D., Adeli, E., Yin, J., Lian, J., Shen, D.: High-resolution encoder–decoder networks for low-contrast medical image segmentation. IEEE Transactions on Image Processing **29**, 461–475 (2019)