UniH^3: Unifying Hierarchical Homogeneity and Heterogeneity for All-in-One Medical Image Restoration

Zhiwen Yang[1], Jiayin Li[1], Chengyu Liu[1], Hui Zhang[2], Bingzheng Wei[3], and Yan Xu[1†]

[1] School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China
[2] Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China
[3] ByteDance Inc., Beijing 100086, China
[†] Corresponding Author
upyzwup@buaa.edu.cn xuyan04@gmail.com

Abstract. All-in-One medical image restoration (MedIR) aims to address diverse tasks across modalities and degradation types using a single universal model. Existing methods typically prioritize modeling inter-task heterogeneity (e.g., distinct data distributions and degradation types). However, they largely neglect the inherent homogeneity present in medical images, such as widely shared anatomical structures within and across modalities, which can be leveraged to ease model training and improve generalization. To this end, we propose **UniH**^{3}, a novel framework that **Uni**fies **H**ierarchical **H**omogeneity and **H**eterogeneity for all-in-one medical image restoration. Specifically, to comprehensively exploit homogeneity, we introduce a Hierarchical Homogeneity Memory (H^{2}M) module that progressively distills intra- and inter-task homogeneity priors from high-quality images during training, and adaptively retrieves the most relevant priors tailored to the input for guided restoration. These retrieved priors are then injected into the restoration pipeline via an efficient Homogeneity-Guided Attention (HGA) mechanism. Furthermore, to comprehensively address heterogeneity, we design a Hierarchical Heterogeneity Balancer (H^{2}B) that mitigates both inter- and intra-task conflicts during optimization, facilitating balanced and effective multi-task learning. Extensive experiments on two large-scale benchmarks—MedIR-2D-500K and MedIR-3D-3K—demonstrate that UniH^3 achieves state-of-the-art performance on both all-in-one and single-task medical image restoration. We hope this work establishes a strong benchmark and advances the development of general-purpose medical image restoration models. Code is available at https://github.com/Yaziwel/UniH3.

Keywords: Medical Image Restoration · All-in-One · Universal Model

1 Introduction

Medical image restoration (MedIR) aims to recover a high-quality (HQ) image from a degraded low-quality (LQ) acquisition. Since each medical image modality (e.g., PET, CT, MRI) operates under distinct physical principles and is often studied independently, most MedIR research has focused on a single-task setting, in which researchers train specialized models to address the primary degradation introduced by the imaging physics of each modality. Typical MedIR includes PET image denoising [56, 67, 68, 77], CT image denoising [5, 40, 51], MRI image super-resolution [9, 25, 45, 53]. Despite their success in specific scenarios, single-task models have limited practicality for two main reasons. First, in complex scenarios where multiple MedIR tasks coexist (e.g., multimodal PET/CT and PET/MRI), single-task models trained for one task often underperform on others. Moreover, training separate models for each task leads to inefficiencies in both deployment and maintenance. Second, the single-task paradigm hinders progress toward more general intelligence in MedIR. These limitations motivate interest in a universal model that can handle diverse MedIR tasks.

Recent advances in computer vision have fostered the emergence of All-in-One restoration frameworks [8, 11, 32, 42, 48, 63, 66]. Pioneering research in the medical domain [4, 8, 63, 66], particularly the first work on AMIR [63], has established the feasibility of unified modeling for MedIR. To manage diverse tasks within a single model, existing approaches predominantly focus on modeling task heterogeneity—that is, distinguishing between tasks to apply specialized processing. Techniques such as contrastive learning [32], degradation classification [23], visual prompting [42], and mixture-of-experts (MoE) [63,69] are widely employed to distinguish between tasks.

However, we argue that current All-in-One methods suffer from two critical limitations. **First**, they largely overlook the inherent homogeneity of medical images. Compared with natural images, medical images from different modalities and tasks often exhibit more consistent anatomical structures and share stronger biological priors. Neglecting this shared knowledge prevents models from exploiting cross-task synergies, thereby increasing the difficulty of learning as the number of tasks grows. **Second**, regarding heterogeneity, existing methods typically address only inter-task differences (e.g., different degradation types or modalities) while ignoring intra-task variations (e.g., variations caused by different scanners, centers, or patient demographics). This coarse-grained approach fails to resolve optimization conflicts that arise from subtle intra-task distribution shifts. Therefore, it is imperative to simultaneously model both homogeneity and heterogeneity at a hierarchical level (inter- and intra-task) to achieve robust and effective All-in-One MedIR.

To address these challenges, we propose **UniH**^{3}, a novel framework that **Uni**fies **H**ierarchical **H**omogeneity and **H**eterogeneity for All-in-One medical image restoration. On the one hand, to fully exploit shared knowledge, we introduce a Hierarchical Homogeneity Memory (H^{2}M) module. This module progressively distills both inter- and intra-task priors from HQ images into a memory bank during training. During inference, it adaptively retrieves the most relevant structural priors tailored to the input, which are then injected into the network via an efficient Homogeneity-Guided Attention (HGA) mechanism to guide restoration. On the other hand, to manage task conflicts comprehensively, we design a

Hierarchical Heterogeneity Balancer (H^2B). Unlike traditional weighting strategies [27, 58] that only balance loss functions at the task level, H^2B dynamically mitigates optimization conflicts at both inter- and intra-task levels, ensuring balanced convergence across diverse data distributions. Finally, to validate the effectiveness of UniH^3, we construct a benchmark comprising two large-scale datasets: MedIR-2D-500K, containing 509,200 2D image pairs across seven 2D MedIR tasks, and MedIR-3D-3K, containing 3,522 3D volume pairs across three 3D MedIR tasks. Extensive experiments on this benchmark indicate that UniH^3 achieves state-of-the-art (SOTA) performance in both all-in-one and single-task medical image restoration.

In summary, our contributions are as follows:

- We propose UniH^3, a novel framework that simultaneously models hierarchical inter- and intra-task homogeneity and heterogeneity for effective all-in-one medical image restoration.
- We present a Hierarchical Homogeneity Memory (H^2M) module that can adaptively distill and retrieve homogeneity priors to guide the restoration process. Additionally, an efficient Homogeneity-Guided Attention (HGA) mechanism is introduced to fully exploit the retrieved prior for guided restoration.
- We develop a Hierarchical Heterogeneity Balancer (H^2B), which achieves fine-grained task balancing by resolving optimization conflicts arising from both inter-task distinctions and intra-task variations.

2 Related Work

Single-Task Medical Image Restoration. Because different medical imaging modalities are typically studied independently, most MedIR research focuses on single-task problems that address the primary degradations encountered in each modality. Typical MedIR tasks include PET image denoising [56, 67, 68, 77], CT denoising [5, 40, 51], MRI super-resolution [9, 25, 53], X-ray denoising [47], OCT denoising [13], ultrasound denoising [2], and pathology image super-resolution [31]. With the development of deep learning—especially recent advances in network architectures such as convolutional neural networks (CNNs) [5, 30], Transformers [50,64], Mamba [18,40], and RWKV [41,65]—single-task MedIR methods have made substantial progress. However, these single-task models often suffer large performance drops when applied to other MedIR tasks, which limits their practical applicability in broader contexts such as multi-modal imaging scenarios.

All-in-One Medical Image Restoration. All-in-One image restoration [8, 11, 32, 42, 48, 63, 66]. aims to address multiple degradation types and modalities using a single unified model. Early attempts in computer vision, such as TransWeather [48], relied on task-specific encoder–decoder heads to handle distinct weather conditions, inevitably increasing parameter counts as tasks multiplied. To achieve parameter-efficient unified modeling, AirNet [32] introduced a contrastive learning approach to generate task-specific latent representations,

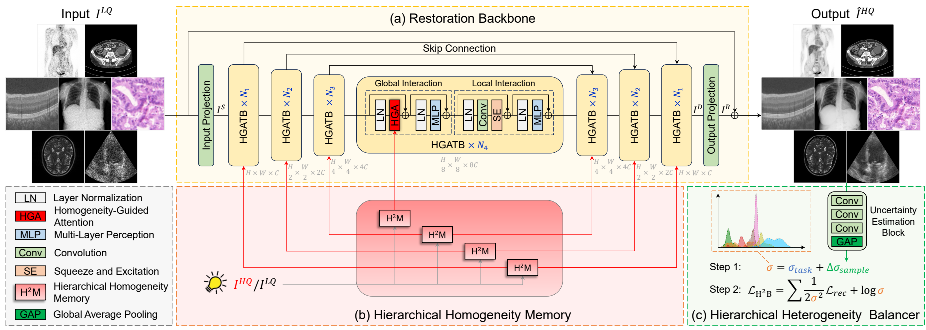

Fig. 1: The framework of UniH^3.

which serve as prompts to guide a shared restoration network. This prompt-based paradigm has become the dominant strategy, with subsequent methods like PromptIR [42], AdaIR [11], and others [8, 66] proposing various mechanisms to learn and inject discriminative prompts for effective task adaptation. In the medical domain, research on All-in-One frameworks is still in its nascent stage [4, 8, 63, 66]. AMIR [63] represents a pioneering effort, utilizing a mixture-of-experts strategy to adapt to three specific medical restoration tasks. While these methods have successfully demonstrated the feasibility of unified restoration, they predominantly focus on modeling inter-task heterogeneity—i.e., distinguishing between different tasks to apply specific processing. Consequently, they largely neglect two critical aspects: the inherent homogeneity of anatomical structures shared across medical modalities, and the fine-grained intra-task heterogeneity arising from variations in scanners and protocols. In contrast, our work formulates a hierarchical learning paradigm that jointly models intra- and inter-task homogeneity and heterogeneity, offering a unified perspective for all-in-one MedIR.

3 Method

Fig. 1 illustrates the UniH^3 pipeline for all-in-one medical image restoration. UniH^3 comprises three main components: a U-shaped restoration backbone (see Fig. 1 (a)) responsible for the basic feature extraction and reconstruction, a Hierarchical Homogeneity Memory (H^{2}M, Fig. 1 (b)) module that employs both intra- and inter-task homogeneity priors to guide the restoration, and a Hierarchical Heterogeneity Balancer (H^{2}B, see Fig. 1 (c)) addresses intra- and inter-task heterogeneity by dynamically balancing task relationships during training. Given a LQ input image $I^{LQ} \in \mathbb{R}^{H\times W\times 1}$, UniH^3 first applies a 3×3 convolutional input projection to produce shallow features $I^S \in \mathbb{R}^{H\times W\times C}$, where $H \times W$ denotes the spatial dimensions and C the number of channels. I^S is then processed by a 4-level asymmetric encoder–decoder and transformed into deep features $I^D \in \mathbb{R}^{H\times W\times C}$. Each encoder–decoder level contains multiple

Homogeneity-Guided Transformer Blocks (HGATBs, see Fig. 1 (a)) that extract features under the guidance of H^{2}M-generated priors. Considering that both global and local information are important for medical image restoration [19,65], each HGATB contains two consecutive transformer-layer variants: the first replaces standard self-attention [14] with a Homogeneity-Guided Attention (HGA) to model global interactions, and the second replaces self-attention with a convolution plus Squeeze-and-Excitation (SE) [24] layer to capture local interactions. Finally, I^D is projected to a residual image $I^R \in \mathbb{R}^{H \times W \times 1}$ by a 3×3 convolution, and the restored HQ output is obtained via the residual connection $\hat{I}^{HQ} = I^{LQ} + I^R$. We next introduce our core innovations: the H^{2}M module (Sec. 3.1), the HGA mechanism (Sec. 3.2), and the H^{2}B strategy (Sec. 3.3).

3.1 Hierarchical Homogeneity Memory

We propose the Hierarchical Homogeneity Memory (H^{2}M) to alleviate the escalating learning difficulty associated with the growing number of restoration tasks and imaging domains. Drawing inspiration from multi-task learning [3], which leverages shared knowledge to reduce learning burdens and accelerate convergence, we observe that HQ medical images exhibit rich homogeneous priors at two hierarchical levels: **intra-task homogeneity** (i.e., consistent anatomical structures among varying patients within the same modality) and **inter-task homogeneity** (i.e., shared structured representations of the human anatomy across different imaging modalities). To explicitly model these hierarchical properties, H^{2}M establishes two structurally symmetric components: a memory bank $M \in \mathbb{R}^{(T+1)L \times C'}$ and a learnable prototype matrix $P \in \mathbb{R}^{(T+1)L \times C'}$. Both are organized into T task-specific slots and one task-shared slot (each length of L), as illustrated in Fig. 2. While M is updated via momentum to store distilled HQ anatomical priors, P is a set of learnable parameters that serves as an addressing mechanism, learning how to optimally store and retrieve information from M. The H^{2}M mechanism operates in two phases: Homogeneity Distillation and Homogeneity Retrieval.

Homogeneity Distillation. To acquire compact homogeneity priors that facilitate all-in-one restoration, we distill clean anatomical structures from HQ medical images and progressively archive them into M using an Exponential Moving Average (EMA) during training. Concretely, we project paired LQ–HQ images I^{LQ}, I^{HQ} to the target resolution via pixel-unshuffle downsampling followed by a 3×3 convolution, obtaining paired features $F^{LQ}, F^{HQ} \in \mathbb{R}^{H'W' \times C'}$. A learnable prototype $P \in \mathbb{R}^{L \times C'}$ with the length of L is then used to query and aggregate crucial HQ priors from F^{HQ} through cross-attention:

$$\text{CrossAttention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\intercal}{\sqrt{C'}}\right)V, \tag{1}$$

$$V^{HQ} = \text{CrossAttention}(P, F^{LQ}, F^{HQ}), \tag{2}$$

where $V^{HQ} \in \mathbb{R}^{(T+1)L \times C'}$. This operation allows P to learn which HQ features are most representative of the clean anatomical structures. Based on the current

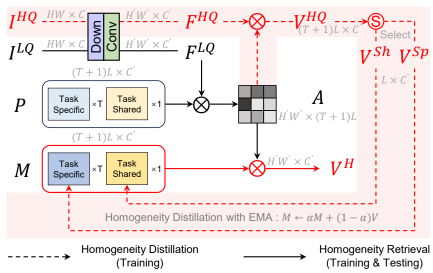

Fig. 2: Hierarchical Homogeneity Memory.

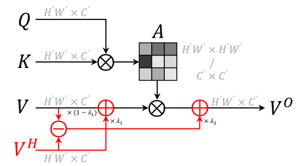

Fig. 3: Homogeneity-Guided Attention.

task index, we select features $V^{Sh} \in \mathbb{R}^{L \times C'}$ and $V^{Sp} \in \mathbb{R}^{L \times C'}$ from V^{HQ}, which are correspondingly stored into the task-shared slot (to store inter-task homogeneity prior) and task-specific slot (to store intra-task homogeneity prior) of M via an EMA strategy:

$$M_{Sh/Sp} \leftarrow \alpha M_{Sh/Sp} + (1-\alpha) V^{Sh/Sp}, \tag{3}$$

where α is the momentum coefficient, and $M_{Sh/Sp}$ denotes the corresponding shared or specific slots in M. Initialized as zero, M gradually accumulates generalized intra- and inter-task homogeneity priors from continuous training batches. Note that this distillation procedure (indicated by dashed red arrows in Fig. 2) is performed only during training and discarded at test time.

Homogeneity Retrieval. Once the hierarchical memory M is updated, we retrieve clean homogeneity priors V^H tailored to the LQ input by using the LQ feature F^{LQ} as a query to retrieve the relevant clean prior from the memory M via cross attention:

$$V^H = \mathrm{CrossAttention}(F^{LQ}, P, M), \tag{4}$$

where $V^H \in \mathbb{R}^{H'W' \times C'}$ is the retrieved homogeneity prior. The red arrows in Fig. 2 illustrate the HQ information flow from I^{HQ} through M into the resulting homogeneity prior V^H. Because the obtained V^H is derived from the distilled HQ memory, it is well-suited to compensate for degraded or missing anatomical information in the LQ features.

To facilitate multi-scale guidance, UniH^3 incorporates four H^{2}M modules (see Fig. 1(b)) at different levels of the U-shaped restoration backbone so that the retrieved homogeneity priors provide effective restoration guidance across multiple resolutions.

3.2 Homogeneity-Guided Attention

To guide the restoration process using homogeneity priors, we propose a novel Homogeneity-Guided Attention (HGA) mechanism. Existing methods typically incorporate restoration guidance via Spatial Feature Transformations (SFT) [55] or cross-attention [11], which treat the LQ features as the basis and the guidance features as supplementary. In contrast, HGA fundamentally shifts the learning paradigm: it anchors the learning starting point on the HQ homogeneity priors rather than the degraded LQ features, thereby substantially reducing the learning difficulty and facilitate model convergence. The design of HGA is detailed below.

HGA is highly flexible and can be implemented on either standard self-attention or transposed self-attention. For clarity of exposition, we formulate it here using standard self-attention. Let the query, key, and value be $Q, K, V \in \mathbb{R}^{H'W' \times C'}$, the conventional self-attention output V_0^O is

$$V_0^O = AV, \qquad A = \mathrm{Softmax}(\frac{QK^\top}{\sqrt{C'}}). \tag{5}$$

To incorporate guidance from the homogeneity prior V^H, a straightforward variant is to complement the LQ value V with clean V^H by direct addition:

$$V_1^O = A(V + V^H) = AV + AV^H. \tag{6}$$

In Eq. 6, the attention mechanism treats V and V^H symmetrically. However, the homogeneity prior V^H contains higher-fidelity information than the degraded observation V, the attention mechanism should preferentially exploit the more reliable V^H. To encourage such a preference, we introduce a second variant that biases attention away from the LQ value V and toward the homogeneity prior V^H by adding identity-based terms to the attention map A:

$$\begin{aligned} V_2^O &= (A - I)V + (A + I)V^H \\ &= A(V + V^H) + V^H - V, \end{aligned} \tag{7}$$

where I denotes the identity matrix. The $\pm I$ terms increase the self-contribution of V^H while reducing that of V. $V + V^H$ denotes the mixed value, and $V^H - V$ acts as a preference bias that reinforces more reliance on the homogeneity prior V^H. To stabilize training and increase model expressivity, the final HGA mechanism is obtained by applying channel-wise learnable weighting parameters $\lambda_1, \lambda_2 \in \mathbb{R}^{C'}$:

$$V^O = A[(1 - \lambda_1)V + \lambda_1 V^H] + \lambda_2(V^H - V). \tag{8}$$

When $\lambda_1 = \lambda_2 = 0$, the HGA reduces to the conventional self-attention. The self-attention–based HGA in Eq. 8 has an analogous form to transposed self-attention (see *supplement*). Our proposed UniH^3 adopts the HGA based on transposed self-attention following Restormer [70]. Fig. 3 illustrates the HGA formulation, which augments attention with simple addition and subtraction operations on the value.

3.3 Hierarchical Heterogeneity Balancer

We propose a Hierarchical Heterogeneity Balancer (H^{2}B) to mitigate inter- and intra-task heterogeneity across diverse MedIR tasks during the optimization process. Heterogeneity among tasks induces gradient conflicts that create an imbalance in optimization: some tasks dominate training while others remain undertrained. Previous work in multi-task learning [27] and all-in-one natural image restoration [58] has shown that uncertainty-based loss balancing is a good way of addressing inter-task heterogeneity by dynamically scaling different task losses for a reasonable optimization route:

$$\mathcal{L}_{\mathrm{UB}} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{2\sigma_t^2}\mathcal{L}_{rec}^{(t)} + \log\sigma_t\right), \tag{9}$$

where T denotes the number of tasks, $\mathcal{L}_{rec}^{(t)}$ denotes the reconstruction loss, and σ_t is a learnable scalar that estimates task-level uncertainty. The factor $\frac{1}{2\sigma_t^2}$ adaptively rescales each task's contribution while the $\log\sigma_t$ term regularizes the scaling. When $\mathcal{L}_{rec}^{(t)}$ increases and tends to dominate the total loss, σ_t increases to attenuate its contribution, and vice versa. However, this uncertainty balancing is too coarse: a single scalar σ_t per-task cannot capture intra-task heterogeneity (e.g., scanner/center/anatomy variations), so hard samples still remain insufficiently handled. We therefore introduce a hierarchical uncertainty model. For task t and sample s we define the total uncertainty $\sigma_{t,s}$ as the sum of a global task term σ_t and a sample-specific correction $\Delta\sigma_s$:

$$\sigma_{t,s} = \sigma_t + \Delta\sigma_s, \tag{10}$$

where t indexes tasks and s indexes samples. σ_t is still a learnable scalar for each task while $\Delta\sigma_s$ is predicted by a lightweight Uncertainty Estimation Block (UEB, see Fig. 1(c)) conditioned on sample-specific signals:

$$\Delta\sigma_s = \mathrm{UEB}\left(\mathrm{Concat}[I_s^{LQ}, \mathrm{sg}(\hat{I}_s^{HQ}), I_s^{HQ}]\right), \tag{11}$$

where I_s^{LQ} is the LQ input for sample s, $\hat{I}_s^{HQ}$ is the model prediction, I_s^{HQ} is the HQ ground truth, and $\mathrm{sg}(\cdot)$ denotes stop-gradient to decouple loss balancing from restoration model optimization. The $\mathrm{H^2B}$ loss then aggregates per-task and per-sample contributions as:

$$\mathcal{L}_{\mathrm{H^2B}} = \frac{1}{TS}\sum_{t=1}^{T}\sum_{s=1}^{S}\left(\frac{1}{2\sigma_{t,s}^2}\mathcal{L}_{rec}^{(t,s)} + \log\sigma_{t,s}\right), \tag{12}$$

where S is the batch size. $\mathrm{H^2B}$ retains the theoretical foundation of standard uncertainty-based balancing [27] while refining it to capture uncertainty at two hierarchical levels: a task-level term σ_t to effectively mitigate inter-task heterogeneity, and a sample-level correction $\Delta\sigma_s$ to mitigate intra-task heterogeneity.

4 Experiments

We conduct experiments under two settings, *All-in-One* and *Single-Task*, for both 2D and 3D MedIR tasks. In the *All-in-One* setting, a single universal model is trained to address multiple MedIR tasks within either the 2D or 3D domain.

Table 1: Overview of the MedIR-2D-500K and MedIR-3D-3K datasets.

Dataset	Dimension	Modality	Training	Testing	Total	Data Source
MedIR-2D-500K	2D	PET	77,000	8,600	85,600	[61], Private
		CT	64,000	8,100	72,100	[37, 38] , Private
		MRI	75,000	8,400	83,400	[36, 49]
		X-ray	108,000	11,600	119,600	[10, 43, 52, 54]
		OCT	32,000	3,600	35,600	[16, 17, 33]
		Ultrasound	56,000	5,800	61,800	[20, 22, 29, 39, 62]
		Pathology	46,000	5,100	51,100	[1, 12, 15, 28, 44, 46]
		Total	**458,000**	**51,200**	**509,200**	-
MedIR-3D-3K	3D	PET	1,388	156	1,544	[61], Private
		CT	258	30	288	[37, 38], Private
		MRI	1,520	170	1,690	[36, 49]
		Total	**3,166**	**356**	**3,522**	-

In the *Single-Task* setting, separate models are trained for each MedIR task. We first describe the experimental setup, including datasets, implementation details, and evaluation. We then present comparative results in Sec. 4.1 and Sec. 4.2, and ablation studies in Sec. 4.3.

Datasets. Most existing MedIR datasets are limited in size and narrowly tailored to specific tasks and modalities. To promote the development of general-purpose MedIR methods, we organize publicly available datasets together with private collections into two datasets, as summarized in Tab. 1: **(i) MedIR-2D-500K** comprises 509,200 2D LQ-HQ image pairs across seven distinct 2D MedIR tasks: PET image denoising, CT image denoising, MRI image super-resolution, X-ray image denoising, OCT image denoising, ultrasound image denoising, and pathological image super-resolution. **(ii) MedIR-3D-3K** includes 3,522 3D LQ-HQ volume pairs covering three 3D MedIR tasks: PET image denoising, CT image denoising, and MRI image super-resolution. We expect these two datasets to serve as a useful benchmark for advancing general-purpose MedIR research. More detailed descriptions are shown in the *supplement*.

Implementation. For the UniH^3 architecture, the number of HGATBs are $N_1 = 2$, $N_2 = N_3 = 3$, and $N_4 = 4$. The input channel dimension is $C = 48$. For the H^{2}M module, we set the the number of tasks $T = 7$, memory length $L = 128$ and EMA coefficient $\alpha = 0.99$. During training we use patches of size 128×128 with a batch size of 14. The reconstruction loss $\mathcal{L}_{rec}$ is defined as the L1 loss. The model is optimized using Muon optimizer [35] for 6×10^5 iterations, with an initial learning rate 3×10^{-4} and annealed to 1×10^{-7} using a cosine schedule. To support 3D MedIR, we introduce a 3D variant, UniH^3-3D, obtained by replacing each module in UniH^3 with its 3D counterpart. For UniH^3-3D, the HGATB numbers are $N_1 = N_2 = 1$ and $N_3 = N_4 = 5$, the input channel number is $C = 16$, and the init learning rate is set as 5×10^{-5}. The patch size is $64 \times 64 \times 64$ and the batch size is 6. All other settings are identical to the original UniH^3.

Table 2: All-in-one MedIR comparison results on the MedIR-2D-500K dataset.

Method	#Params (M)	FLOPs (G)	PET		CT		MRI		X-ray		OCT		Ultrasound		Pathology		Average	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
SwinIR [34]	11.50	187.93	44.24	0.9866	43.17	0.9338	38.44	0.9453	35.87	0.9275	35.70	0.8892	27.52	0.8089	28.35	0.7941	36.18	0.8979
Uformer [57]	50.47	21.42	44.51	0.9874	43.45	0.9356	39.01	0.9507	36.59	0.9353	35.84	0.8908	27.61	0.8135	28.55	0.8011	36.51	0.9020
Restormer [70]	26.12	35.21	44.47	0.9873	43.44	0.9353	39.05	0.9515	36.46	0.9335	35.84	0.8909	27.66	0.8138	28.53	0.8006	36.49	0.9018
NAFNet [6]	67.89	15.74	44.40	0.9871	43.32	0.9346	38.90	0.9502	36.33	0.9326	35.82	0.8905	27.59	0.8131	28.49	0.7995	36.41	0.9011
Restore-RWKV [65]	27.91	37.46	44.45	0.9872	43.46	0.9352	39.05	0.9514	36.46	0.9333	35.84	0.8909	27.68	0.8149	28.55	0.8011	36.50	0.9020
MambaIR [19]	31.50	34.35	44.50	0.9873	43.47	0.9355	39.07	0.9517	36.57	0.9341	35.83	0.8904	27.69	0.8146	28.53	0.8004	36.52	0.9020
TransWeather [48]	38.05	1.55	43.73	0.9846	41.12	0.9217	37.62	0.9333	35.28	0.9221	35.12	0.8699	27.23	0.8011	28.17	0.7874	35.47	0.8886
AirNet [32]	7.61	230.48	44.32	0.9868	43.34	0.9348	38.81	0.9489	36.38	0.9322	35.77	0.8900	27.62	0.8127	28.46	0.7981	36.39	0.9005
DRMC [68]	0.62	9.92	43.62	0.9841	42.48	0.9281	37.24	0.9301	34.06	0.9085	35.46	0.8862	27.09	0.7993	28.02	0.7830	35.42	0.8885
AMIR [63]	23.54	31.76	44.49	0.9873	43.47	0.9356	39.09	0.9519	36.47	0.9333	35.89	0.8914	27.69	0.8150	28.57	0.8019	36.52	0.9023
PromptIR [42]	35.59	39.49	44.52	0.9874	43.48	0.9355	39.13	0.9524	36.57	0.9341	35.84	0.8909	27.69	0.8152	28.54	0.8006	36.54	0.9023
AdaIR [11]	28.76	36.74	44.55	0.9875	43.49	0.9356	39.17	0.9527	36.60	0.9344	35.86	0.8915	27.69	0.8150	28.56	0.8015	36.56	0.9026
UniH^3 (Ours)	28.96	26.33	**44.89**	**0.9883**	**43.65**	**0.9368**	**39.55**	**0.9564**	**36.88**	**0.9368**	**35.96**	**0.8921**	**27.80**	**0.8179**	**28.63**	**0.8035**	**36.77**	**0.9045**

Table 3: Single-task MedIR comparison results on the MedIR-2D-500K dataset.

Method	#Params (M)	FLOPs (G)	PET		CT		MRI		X-ray		OCT		Ultrasound		Pathology		Average	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
SwinIR [34]	11.50	187.93	44.08	0.9860	43.53	0.9358	39.19	0.9529	36.30	0.9308	35.80	0.8904	27.68	0.8143	28.50	0.7993	36.44	0.9014
Uformer [57]	50.47	21.42	44.61	0.9876	43.54	0.9362	39.21	0.9526	36.61	0.9363	35.91	0.8915	27.64	0.8151	28.58	0.8022	36.59	0.9031
Restormer [70]	26.12	35.21	44.90	0.9883	43.64	0.9369	39.44	0.9554	36.71	0.9359	35.97	0.8921	27.76	0.8164	28.63	0.8038	36.72	0.9041
NAFNet [6]	67.89	15.74	44.74	0.9880	43.55	0.9362	39.29	0.9540	36.48	0.9343	35.93	0.8919	27.67	0.8139	28.58	0.8024	36.61	0.9030
Restore-RWKV [65]	27.91	37.46	44.93	0.9884	43.64	0.9369	39.57	0.9565	36.77	0.9359	35.97	0.8923	27.76	0.8168	28.62	0.8036	36.75	0.9043
MambaIR [19]	31.50	34.35	44.93	0.9883	43.55	0.9363	39.59	0.9568	36.77	0.9363	35.99	0.8923	27.77	0.8178	28.64	0.8041	36.75	0.9046
UniH^3 (Ours)	28.96	26.33	**45.11**	**0.9889**	**43.77**	**0.9376**	**39.78**	**0.9584**	**37.05**	**0.9384**	**36.04**	**0.8929**	**27.86**	**0.8192**	**28.69**	**0.8050**	**36.90**	**0.9058**

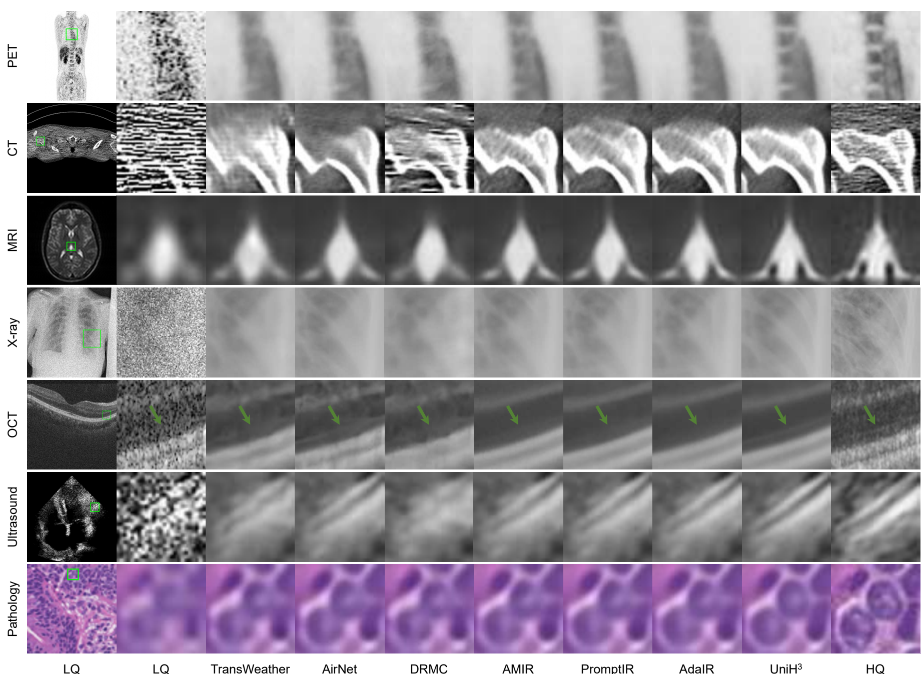

Fig. 4: Visual comparison of methods for all-in-one medical image restoration on the MedIR-2D-500K dataset.

Evaluation. To quantitatively assess image quality, we employ the widely used PSNR and SSIM metrics. In the reported tables, the highest and second-highest scores are indicated in red and blue, respectively.

Table 4: 3D all-in-one MedIR results on the MedIR-3D-3K dataset.

Method	PET		CT		MRI		Average	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
3D-cGAN [56]	48.22	0.9937	40.82	0.9320	38.57	0.9489	42.54	0.9582
MRDG [53]	48.68	0.9943	42.86	0.9388	38.93	0.9519	43.49	0.9617
DRMC [68]	48.58	0.9933	43.33	0.9391	38.81	0.9514	43.57	0.9613
Spach Transformer [26]	48.91	0.9949	42.21	0.9380	39.04	0.9542	43.39	0.9624
Restore-RWKV-3D [65]	49.08	0.9951	43.35	0.9402	39.43	0.9578	43.95	0.9644
UniH^3-3D (Ours)	**49.43**	**0.9955**	**44.05**	**0.9429**	**40.03**	**0.9637**	**44.50**	**0.9674**

Table 5: 3D single-task MedIR results on the MedIR-3D-3K dataset.

Method	PET		CT		MRI		Average	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
3D-cGAN [56]	48.47	0.9944	40.86	0.9309	38.76	0.9510	42.70	0.9588
MRDG [53]	48.40	0.9942	43.14	0.9395	39.34	0.9570	43.63	0.9636
DRMC [68]	48.76	0.9947	43.60	0.9404	38.98	0.9540	43.78	0.9630
Spach Transformer [26]	48.71	0.9946	43.57	0.9409	39.16	0.9546	43.81	0.9634
Restore-RWKV-3D [65]	48.67	0.9945	42.48	0.9375	39.15	0.9551	43.43	0.9624
UniH^3-3D (Ours)	**49.77**	**0.9958**	**45.37**	**0.9448**	**40.73**	**0.9689**	**45.29**	**0.9698**

Table 6: Performance of H^{2}P and H^{2}B on different backbones on the MedIR-2D-500K dataset. † denotes applying H^{2}P and H^{2}B to the corresponding backbone.

Method	#Params (M)	FLOPs (G)	PET		CT		MRI		X-ray		OCT		Ultrasound		Pathology		Average	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Uformer	50.47	21.42	44.51	0.9874	43.45	0.9356	39.01	0.9507	36.59	0.9353	35.84	0.8908	27.61	0.8135	28.55	0.8011	36.51	0.9021
Uformer †	53.15	22.10	**44.80**	**0.9880**	**43.59**	**0.9363**	**39.39**	**0.9549**	**36.74**	**0.9354**	**35.94**	**0.8918**	**27.72**	**0.8155**	**28.57**	**0.8022**	**36.68**	**0.9034**
Restormer	26.12	35.21	44.47	0.9873	43.44	0.9353	39.05	0.9515	36.46	0.9335	35.84	0.8909	27.66	0.8138	28.53	0.8006	36.49	0.9018
Restormer †	27.56	35.83	**44.78**	**0.9880**	**43.58**	**0.9362**	**39.37**	**0.9548**	**36.71**	**0.9352**	**35.93**	**0.8919**	**27.73**	**0.8159**	**28.60**	**0.8027**	**36.67**	**0.9035**
PromptIR	35.59	39.49	44.52	0.9874	43.48	0.9355	39.13	0.9524	36.57	0.9341	35.84	0.8909	27.69	0.8152	28.54	0.8006	36.54	0.9023
PromptIR †	36.15	40.68	**44.85**	**0.9882**	**43.62**	**0.9366**	**39.40**	**0.9551**	**36.76**	**0.9365**	**35.94**	**0.8919**	**27.75**	**0.8165**	**28.59**	**0.8024**	**36.70**	**0.9039**
AdaIR	28.76	36.74	44.55	0.9875	43.49	0.9356	39.17	0.9527	36.60	0.9344	35.86	0.8915	27.69	0.8150	28.56	0.8015	36.56	0.9026
AdaIR †	29.77	37.37	**44.93**	**0.9884**	**43.64**	**0.9367**	**39.53**	**0.9563**	**36.82**	**0.9363**	**35.95**	**0.8920**	**27.76**	**0.8165**	**28.61**	**0.8027**	**36.75**	**0.9041**
Baseline	27.17	25.50	44.52	0.9874	43.47	0.9355	39.10	0.9521	36.39	0.9330	35.88	0.8914	27.68	0.8149	28.57	0.8018	36.52	0.9023
UniH^3 (Ours)	28.96	26.33	**44.89**	**0.9883**	**43.65**	**0.9368**	**39.55**	**0.9564**	**36.88**	**0.9368**	**35.96**	**0.8921**	**27.80**	**0.8179**	**28.63**	**0.8035**	**36.77**	**0.9045**

Table 7: Component analysis.

H^{2}M	H^{2}B	#Params (M)	FLOPs (G)	PSNR↑	SSIM↑
		27.17	25.50	36.52	0.9023
✓		28.96	26.33	36.66	0.9034
	✓	27.17	25.50	36.64	0.9033
✓	✓	28.96	26.33	**36.77**	**0.9045**

Table 8: Ablation studies on HGA.

Method	#Params (M)	FLOPs (G)	PSNR↑	SSIM↑
w/o HGA	27.17	25.50	36.52	0.9023
SFT [55]	37.46	36.99	36.74	0.9041
Cross Attention [11]	29.60	27.43	36.67	0.9035
HGA (Ours)	28.96	26.33	**36.77**	**0.9045**

4.1 All-in-One MedIR Results

2D All-in-One MedIR. We evaluate the 2D *All-in-One* setting on the MedIR-2D-500K dataset. We compare it to several general image-restoration methods (SwinIR [34], Uformer [57], Restormer [70], NAFNet [6], Restore-RWKV [65], and MambaIR [19]) and to SOTA all-in-one approaches (TransWeather [48], AirNet [32], DRMC [68], AMIR [63], PromptIR [42], and AdaIR [11]). Tab. 2 shows that UniH^3 achieves both high efficiency and strong effectiveness. It significantly outperforms all comparison methods across all seven tasks. On average, UniH^3 surpasses the second-best AdaIR by 0.21 dB in PSNR, which is an appreciable improvement given that each of the seven task contains thousands of testing images. Visual comparison in Fig. 4 demonstrates that UniH^3 best restores different types of medical images with higher structural fidelity and finer detail than competing methods.

3D All-in-One MedIR. We assess the UniH^3-3D in the 3D *All-in-One* setting on the MedIR-3D-3K dataset. We compare it with several SOTA 3D restoration methods, including 3D-cGAN [56], MRDG [53], DRMC [68], Spach Transformer [26], and Restore-RWKV-3D [65]. Tab. 4 shows that UniH^3 significantly outperforms all comparison methods across the three 3D MedIR tasks. In particular, UniH^3-3D exceeds the second-best Restore-RWKV-3D by an average margin of 0.56 dB in PSNR. Visual comparisons are shown in the *supplement*.

Table 9: Ablation studies on H^2M.

Inter-task Homogeneity	Intra-task Homogeneity	#Params (M)	FLOPs (G)	PSNR↑	SSIM↑
		28.96	26.33	36.52	0.9023
✓		28.96	26.33	36.61	0.9031
	✓	28.96	26.33	36.72	0.9038
✓	✓	28.96	26.33	**36.77**	**0.9045**

Table 10: Ablation studies on H^2B.

Inter-task Heterogeneity	Intra-task Heterogeneity	#Params (M)	FLOPs (G)	PSNR↑	SSIM↑
		28.96	26.33	36.66	0.9034
✓		28.96	26.33	36.70	0.9036
	✓	28.96	26.33	36.73	0.9041
✓	✓	28.96	26.33	**36.77**	**0.9045**

4.2 Single-Task MedIR Results

2D Single-Task MedIR. We evaluate UniH^3 for 2D single-task MedIR on the MedIR-2D-500K dataset, comparing it to five general image restoration methods. As shown in Tab. 3, UniH^3 significantly outperforms all comparison methods. On average across seven tasks, UniH^3 improves PSNR by 0.15 dB over the second-best MambaIR.

3D Single-Task MedIR. We evaluate 3D single-task MedIR on the MedIR-3D-3K dataset and compare UniH^3-3D to five SOTA 3D restoration methods. As shown in Tab. 5, UniH^3-3D beats all comparison methods. In particular, it improves average PSNR by 1.48 dB over the second-best Spach Transformer [26] across the three tasks.

4.3 Ablation Studies

To evaluate the effectiveness of individual components, we conduct ablation experiments on the 2D *All-in-One* MedIR task with the MedIR-2D-500K dataset.

Component Analysis of H^2M and H^2B. We first perform a component analysis of the H^2M module and the H^2B strategy by selectively disabling each component. To disable H^2M, we remove the H^2M module and replace the HGA mechanism with a transposed self-attention layer [70]. The H^2B is disabled by substituting the loss term $\mathcal{L}_{H^2B}$ with an L_1 loss. Table 7 shows that both components improve model performance with minimal increase in computational cost. This finding is further supported by the visual comparison in Fig. 5, where both components contribute to better preservation of image details. Moreover, we apply the H^2M module and H^2B strategy to other Transformer-based U-shaped backbones, including Uformer, Restormer, PromptIR, and AdaIR. Results in the Tab. 6 demonstrate significant improvements across all these backbones.

Ablation Studies on H^2M. We investigate the effectiveness of both inter- and intra-task homogeneity priors in H^2M by selectively disabling the task-specific and task-shared slots in M. By replacing these slots with naive learnable parameters—thus isolating them from the HQ Homogeneity Distillation procedure—we observe a drop in performance in Tab. 9. Results indicate that both priors independently benefit the restoration process, and their combination achieves the best performance. To further understand the internal mechanics of H^2M, Fig. 6 visualizes the retrieval attention maps for PET and CT tokens. The distributions confirm that tokens primarily query the task-shared slot and their specific modality slot. Furthermore, attention maps reflect clear semantic correlations: anatomically similar tokens within the same modality share highly similar query

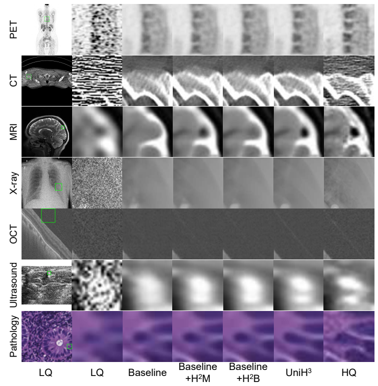

Fig. 5: Visual comparison for component analysis.

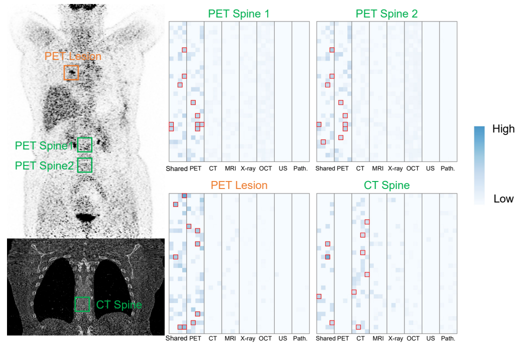

Fig. 6: Retrieval attention map in H^{2}M. The Top-10 scores are marked by red rectangles.

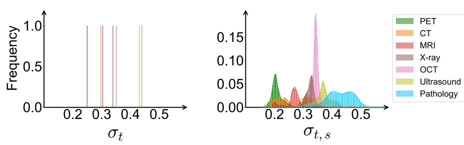

Fig. 7: Estimated uncertainty distribution.

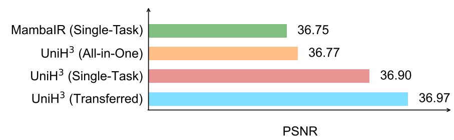

Fig. 8: All-in-One vs. Single-Task.

patterns (8/10 top-score overlap for two PET spine tokens), and cross-modality similarities are also captured (2/10 overlap for PET and CT spine tokens). In contrast, dissimilar anatomies exhibit distinct query patterns (0/10 overlap for PET spine and lesion tokens). This provides strong visual evidence that H^{2}M successfully maps, stores, and retrieves hierarchical anatomical priors.

Ablation Studies on HGA. We validate the impact of HGA by replacing it with alternative guiding mechanisms, including Spatial Feature Transform (SFT) [55] and cross attention [11]. Tab. 8 shows that the proposed HGA performs the best with minimal computation and parameter increase.

Ablation Studies on H^{2}B. We validate the roles of intra- and inter-task heterogeneity in H^{2}B in Tab. 10. The results indicate that mitigating both levels of heterogeneity improves overall restoration performance, with their joint optimization achieving the best results. In Fig. 7, we visualize the estimated uncertainty distributions. While the conventional $\mathcal{L}_{\mathrm{UB}}$ [27] estimates an uncertainty σ_t per task and therefore handles only inter-task heterogeneity, our proposed $\mathcal{L}_{\mathrm{H^2B}}$ estimates uncertainty $\sigma_{t,s}$ at both the task and the sample level: each sample receives its own uncertainty while each task exhibits a distinct uncertainty distribution. This enables our $\mathcal{L}_{\mathrm{H^2B}}$ to capture finer-grained relationships, i.e., both inter-task and intra-task heterogeneity, facilitating convergence toward a more optimal solution for diverse MedIR tasks.

5 Discussion and Limitation

All-in-one medical image restoration is an emerging research field. The practical value of all-in-one models remains under active discussion. In this paper, experiments on a large-scale dataset show that a single all-in-one model, UniH^3, already achieves comparable performance to SOTA single-task MambaIR models across seven MedIR tasks (see Fig. 8). This result strongly supports the practicality of developing all-in-one models instead of separate single-task models for MedIR tasks. Additionally, by fine-tuning the well-trained all-in-one UniH^3 model for each task, we are able to transfer its learned knowledge to individual tasks and achieve further improvements (also shown in Fig. 8), indicating that the all-in-one model can serve as a transferable pretrained backbone. Our study has limitations: we focus only on the primary restoration task within each modality and therefore do not cover other tasks or degradation types that may occur in the same modality. Future work should address these gaps and pursue more universal MedIR models to benefit clinical diagnosis and more downstream tasks [7, 21, 59, 60, 71–76].

6 Conclusion

In this paper, we have presented UniH^3, a novel and unified framework for all-in-one MedIR. By moving beyond the conventional focus on inter-task heterogeneity, UniH^3 effectively models the hierarchical homogeneity and heterogeneity across diverse MedIR tasks. The integration of the Hierarchical Homogeneity Memory (H^{2}M) module and the Homogeneity-Guided Attention (HGA) mechanism allows the model to distill and utilize inter- and intra-task homogeneity priors to guide the restoration process. Simultaneously, the Hierarchical Heterogeneity Balancer (H^{2}B) ensures a stable and balanced multi-task learning process by mitigating both inter- and intra-task heterogeneity. Extensive evaluations on the large-scale MedIR-2D-500K and MedIR-3D-3K benchmarks demonstrate that UniH^3 significantly outperforms existing methods, achieving state-of-the-art performance in both all-in-one and single-task scenarios. In future work, we plan to expand the task spectrum by incorporating additional modalities and degradation types, moving toward more general MedIR models.

Acknowledgements

This work is supported by the National Natural Science Foundation in China under Grant U23B2063 and 62371016, the Bejing Natural Science Foundation Haidian District Joint Fund in China under Grant L2602042, the Prevention and Control of Emerging and Major Infectious Diseases-National Science and Technology Major Project under Grant 2026ZD01910500, the Fundamental Research Funds for the Central University of China from the State Key Laboratory of Software Development Environment in Beihang University in China, the 111 Proiect in China under Grant B13003, the Academic Excellence Foundation of BUAA for PhD Students.

References

1. Aksac, A., Demetrick, D.J., Ozyer, T., Alhajj, R.: Brecahad: a dataset for breast cancer histopathological annotation and diagnosis. BMC research notes **12**(1), 82 (2019)
2. Asgariandehkordi, H., Goudarzi, S., Basarab, A., Rivaz, H.: Deep ultrasound denoising using diffusion probabilistic models. In: 2023 IEEE International Ultrasonics Symposium (IUS). pp. 1–4. IEEE (2023)
3. Caruana, R.: Multitask learning. Machine learning **28**(1), 41–75 (1997)
4. 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)
5. 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)
6. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European conference on computer vision. pp. 17–33. Springer (2022)
7. Chen, L.: Beyond external constraints: The missing dimension of ai governance. Available at SSRN 6449738 (2026)
8. Chen, T., Ma, X., Bai, L., Wang, W., Sun, Y., Zhou, L.: Endoir: Degradation-agnostic all-in-one endoscopic image restoration via noise-aware routing diffusion. arXiv preprint arXiv:2511.05873 (2025)
9. Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D.: Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network. In: International conference on medical image computing and computer-assisted intervention. pp. 91–99. Springer (2018)
10. Chowdhury, M.E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al Emadi, N., et al.: Can ai help in screening viral and covid-19 pneumonia? Ieee Access **8**, 132665–132676 (2020)
11. Cui, Y., Zamir, S.W., Khan, S., Knoll, A., Shah, M., Khan, F.S.: Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In: 13th International Conference on Learning Representations, ICLR 2025. pp. 57335–57356. International Conference on Learning Representations, ICLR (2025)
12. Da, Q., Huang, X., Li, Z., Zuo, Y., Zhang, C., Liu, J., Chen, W., Li, J., Xu, D., Hu, Z., et al.: Digestpath: A benchmark dataset with challenge review for the pathological detection and segmentation of digestive-system. Medical image analysis **80**, 102485 (2022)
13. Dong, Z., Liu, G., Ni, G., Jerwick, J., Duan, L., Zhou, C.: Optical coherence tomography image denoising using a generative adversarial network with speckle modulation. Journal of biophotonics **13**(4), e201960135 (2020)
14. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
15. Drifka, C.R., Loeffler, A.G., Mathewson, K., Keikhosravi, A., Eickhoff, J.C., Liu, Y., Weber, S.M., Kao, W.J., Eliceiri, K.W.: Highly aligned stromal collagen is a negative prognostic factor following pancreatic ductal adenocarcinoma resection. Oncotarget **7**(46), 76197 (2016)

16. Fang, L., Li, S., Nie, Q., Izatt, J.A., Toth, C.A., Farsiu, S.: Sparsity based denoising of spectral domain optical coherence tomography images. Biomedical optics express **3**(5), 927–942 (2012)
17. Geng, M., Meng, X., Zhu, L., Jiang, Z., Gao, M., Huang, Z., Qiu, B., Hu, Y., Zhang, Y., Ren, Q., et al.: Triplet cross-fusion learning for unpaired image denoising in optical coherence tomography. IEEE Transactions on Medical Imaging **41**(11), 3357–3372 (2022)
18. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
19. 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)
20. Guo, Y., Zhou, S., Shi, J., Wang, Y.: Ultrasound image enhancement challenge 2023 (Apr 2023), `https://doi.org/10.5281/zenodo.7841250`, accessed: 2024-02-13
21. Hao, R., Jing, B., Yu, H., Nie, Z.: Styledrive: Towards driving-style aware benchmarking of end-to-end autonomous driving. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 4627–4635 (2026)
22. van den Heuvel, T.L., de Bruijn, D., de Korte, C.L., Ginneken, B.v.: Automated measurement of fetal head circumference using 2d ultrasound images. PloS one **13**(8), e0200412 (2018)
23. Hu, J., Jin, L., Yao, Z., Lu, Y.: Universal image restoration pre-training via degradation classification. arXiv preprint arXiv:2501.15510 (2025)
24. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
25. 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)
26. Jang, S.I., Pan, T., Li, Y., Heidari, P., Chen, J., Li, Q., Gong, K.: Spach transformer: Spatial and channel-wise transformer based on local and global self-attentions for pet image denoising. IEEE Transactions on Medical Imaging (2023)
27. 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)
28. Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O.F., Tsougenis, E., Chen, H., Heng, P.A., Li, J., Hu, Z., et al.: A multi-organ nucleus segmentation challenge. IEEE transactions on medical imaging **39**(5), 1380–1391 (2019)
29. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE transactions on medical imaging **38**(9), 2198–2210 (2019)
30. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation **1**(4), 541–551 (1989)
31. Li, B., Keikhosravi, A., Loeffler, A.G., Eliceiri, K.W.: Single image super-resolution for whole slide image using convolutional neural networks and self-supervised color normalization. Medical Image Analysis **68**, 101938 (2021)
32. 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 Conference on Computer Vision and Pattern Recognition. pp. 17452–17462 (2022)
33. Li, M., Huang, K., Xu, Q., Yang, J., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., Chen, Q.: Octa-500: a retinal dataset for optical coherence tomography angiography study. Medical image analysis **93**, 103092 (2024)

34. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1833–1844 (2021)
35. Liu, J., Su, J., Yao, X., Jiang, Z., Lai, G., Du, Y., Qin, Y., Xu, W., Lu, E., Yan, J., et al.: Muon is scalable for llm training. arXiv preprint arXiv:2502.16982 (2025)
36. LLC, M.: Ixi dataset, `https://brain-development.org/ixi-dataset/`, accessed: 2024-01-15
37. McCollough, C.H., Bartley, A.C., Carter, R.E., Chen, B., Drees, T.A., Edwards, P., Holmes III, D.R., Huang, A.E., Khan, F., Leng, S., et al.: Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. Medical physics **44**(10), e339–e352 (2017)
38. Moen, T.R., Chen, B., Holmes III, D.R., Duan, X., Yu, Z., Yu, L., Leng, S., Fletcher, J.G., McCollough, C.H.: Low-dose ct image and projection dataset. Medical physics **48**(2), 902–911 (2021)
39. Montoya, A., Hasnin, kaggle446, shirzad, Cukierski, W., yffud: Ultrasound nerve segmentation (2016), kaggle
40. Öztürk, Ş., Duran, O.C., Çukur, T.: Denomamba: A fused state-space model for low-dose ct denoising. arXiv preprint arXiv:2409.13094 (2024)
41. Peng, B., Alcaide, E., Anthony, Q., Albalak, A., Arcadinho, S., Cao, H., Cheng, X., Chung, M., Grella, M., GV, K.K., et al.: Rwkv: Reinventing rnns for the transformer era. arXiv preprint arXiv:2305.13048 (2023)
42. Potlapalli, V., Zamir, S.W., Khan, S.H., Shahbaz Khan, F.: Promptir: Prompting for all-in-one image restoration. Advances in Neural Information Processing Systems **36**, 71275–71293 (2023)
43. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Al Maadeed, S., Zughaier, S.M., Khan, M.S., et al.: Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. Computers in biology and medicine **132**, 104319 (2021)
44. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al.: Gland segmentation in colon histology images: The glas challenge contest. Medical image analysis **35**, 489–502 (2017)
45. Song, Z., Qi, Z., Wang, X., Zhao, X., Shen, Z., Wang, S., Fei, M., Wang, Z., Zang, D., Chen, D., et al.: Uni-coal: A unified framework for cross-modality synthesis and super-resolution of mr images. Expert Systems with Applications **270**, 126241 (2025)
46. Tekin, E., Yazıcı, Ç., Kusetogullari, H., Tokat, F., Yavariabdi, A., Iheme, L.O., Çayır, S., Bozaba, E., Solmaz, G., Darbaz, B., et al.: Tubule-u-net: a novel dataset and deep learning-based tubule segmentation framework in whole slide images of breast cancer. Scientific Reports **13**(1), 128 (2023)
47. Thanh, D., Surya, P., et al.: A review on ct and x-ray images denoising methods. Informatica **43**(2) (2019)
48. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2353–2363 (2022)
49. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The wu-minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)

50. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
51. 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)
52. Wang, D., Wang, X., Wang, L., Li, M., Da, Q., Liu, X., Gao, X., Shen, J., He, J., Shen, T., et al.: A real-world dataset and benchmark for foundation model adaptation in medical image classification. Scientific Data **10**(1), 574 (2023)
53. Wang, J., Chen, Y., Wu, Y., Shi, J., Gee, J.: Enhanced generative adversarial network for 3d brain mri super-resolution. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3627–3636 (2020)
54. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017)
55. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 606–615 (2018)
56. Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., Zhou, L.: 3d conditional generative adversarial networks for high-quality pet image estimation at low dose. Neuroimage **174**, 550–562 (2018)
57. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17683–17693 (2022)
58. Wu, G., Jiang, J., Wang, Y., Jiang, K., Liu, X.: Debiased all-in-one image restoration with task uncertainty regularization. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 8386–8394 (2025)
59. Wu, Y., Zhou, Y., Saiyin, J., Wei, B., Lai, M., Shou, J., Xu, Y.: Attriprompter: Auto-prompting with attribute semantics for zero-shot nuclei detection via visual-language pre-trained models. IEEE Transactions on Medical Imaging **44**(2), 982–993 (2024)
60. Wu, Y., Zhou, Y., Saiyin, J., Wei, B., Xu, Y.: Visual textualization for image prompted object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 20900–20910 (2025)
61. Xue, S., Guo, R., Bohn, K.P., Matzke, J., Viscione, M., Alberts, I., Meng, H., Sun, C., Zhang, M., Zhang, M., et al.: A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose pet. European journal of nuclear medicine and molecular imaging **49**(6), 1843–1856 (2022)
62. Yang, J., Ding, X., Zheng, Z., Xu, X., Li, X.: Graphecho: Graph-driven unsupervised domain adaptation for echocardiogram video segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 11878–11887 (2023)
63. 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)
64. 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)
65. Yang, Z., Li, J., Zhang, H., Zhao, D., Wei, B., Xu, Y.: Restore-rwkv: Efficient and effective medical image restoration with rwkv. IEEE Journal of Biomedical and Health Informatics (2025)
66. 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)
67. Yang, Z., Zhou, Y., Chen, H., Zhang, H., Zhao, D., Wei, B., Xu, Y.: Unipet: a universal network for high-quality pet image denoising across varied dose reduction factors. Medical Image Analysis p. 104059 (2026)
68. 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)
69. 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)
70. 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 Conference on Computer Vision and Pattern Recognition. pp. 5728–5739 (2022)
71. Zhou, K., Cai, R., Ma, Y., Tan, Q., Wang, X., Li, J., Shum, H.P., Li, F.W., Jin, S., Liang, X.: A video-based augmented reality system for human-in-the-loop muscle strength assessment of juvenile dermatomyositis. IEEE Transactions on Visualization and Computer Graphics **29**(5), 2456–2466 (2023)
72. Zhou, K., Cai, R., Wang, L., Shum, H.P.H., Liang, X.: A comprehensive survey of action quality assessment: Method and benchmark. Pattern Recognition **179**, 113933 (2026)
73. Zhou, K., Hao, Z., Wang, L., Liang, X.: Adaptive score alignment learning for continual perceptual quality assessment of 360-degree videos in virtual reality. IEEE Transactions on Visualization and Computer Graphics **31**(5), 2880–2890 (2025)
74. Zhou, K., Shum, H.P.H., Li, F.W.B., Zhang, X., Liang, X.: Phi: Bridging domain shift in long-term action quality assessment via progressive hierarchical instruction. IEEE Transactions on Image Processing **34**, 3718–3732 (2025)
75. Zhou, Y., Wu, Y., Saiyin, J., Wei, B., Lai, M., Chang, E., Xu, Y.: Sdpt: synchronous dual prompt tuning for fusion-based visual-language pre-trained models. In: European Conference on Computer Vision. pp. 340–356. Springer Nature Switzerland Cham (2024)
76. Zhou, Y., Wu, Y., Saiyin, J., Wei, B., Xu, Y.: Sdpt: Synchronous dual prompt tuning for visual-language pre-trained models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2026)
77. 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)