Real-Time Hardware-Free HIFU Interference Suppression via Teacher-Student Diffusion Framework

Dejia Cai[†1], Ali Abdollahi[†1], Xi Wang[1], Kun Yang[2], Zhaohui Guo[3], Xiaowei Zhou[2(✉)], and Hao Chen[1,4,5,6,7(✉)]

[1] Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
jhc@cse.ust.hk
[2] State Key Laboratory of Ultrasound Engineering in Medicine, Chongqing Medical University, Chongqing 400016, China
zhou.xiaowei@cqmu.edu.cn
[3] School of Microelectronics, Tianjin University, Tianjin 300072, China
[4] Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[5] Division of Life Science, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[6] HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, The Hong Kong University of Science and Technology, Futian, Shenzhen, China
[7] State Key Laboratory of Nervous System Disorders, The Hong Kong University of Science and Technology, Hong Kong SAR, China

Abstract. High-Intensity Focused Ultrasound (HIFU) is a non-invasive therapy, yet its safety is often degraded by severe acoustic interference during continuous ultrasound guidance. Existing suppression methods heavily rely on proprietary raw Radio-Frequency (RF) data or hardware synchronization, limiting their clinical utility and preventing real-time implementation. We propose Manifold-Constrained Hyper-Connections Diffusion (mHC-Diff), an image-domain diffusion framework for real-time interference suppression without specialized hardware synchronization, disentangling complex interference from anatomical structures while ensuring high reconstruction fidelity. To achieve clinical real-time application, mHC-Diff first learns an anatomy-aware prior with a high-fidelity multi-step diffusion Teacher, then distills it into a one-step Student. On a clinically representative dataset spanning diverse therapeutic scenarios, mHC-Diff achieves 26.65 dB PSNR and real-time inference (∼20 FPS) on a single NVIDIA RTX 4090, a ∼6.8× speedup over iterative diffusion baselines such as HIFU-Diff. These results suggest a practical route to deployable, hardware-free interference suppression for ultrasound-guided HIFU interventions. Code is available at https://github.com/caidejia/HIFU-mHC-Diff.

Keywords: HIFU · Interference Suppression · Diffusion Models · Manifold Learning · Model Distillation · Real-time Imaging.

[†] Equal contribution.

1 Introduction

High-Intensity Focused Ultrasound (HIFU) has emerged as a widely adopted non-invasive therapeutic modality [1, 6, 11], used to treat various pathologies, including uterine fibroids [8], breast fibroadenomas [12], and adenomyosis [15]. Although HIFU surpasses conventional surgical interventions in patient recovery and tissue preservation, its clinical success depends fundamentally on high-precision, real-time monitoring. Ultrasound imaging provides cost-effective, dynamic intra-operative guidance. However, the intense acoustic waves emitted by the therapeutic HIFU transducer interfere with the imaging pulses, degrading the signal-to-noise ratio of the acquired images. To mitigate this, current protocols rely on a "stop-and-shoot" approach, pausing sonication during imaging. This compromise can prolong procedure times, reduce treatment efficacy, and increase clinical risks associated with unguided sonication [12].

To enable concurrent ultrasound monitoring during continuous sonication, various signal processing and deep learning algorithms have been propose [2, 14, 17, 19–21, 23, 24]. Traditional methods, like FRPCA [24] or encoded excitations [19–21], often lack clinical generalization or impose high computational overhead. Recently, deep learning architectures, ranging from FUS-Net [14] to diffusion-based HIFU-Diff [3], have emerged as powerful alternatives for robust interference suppression. However, these methods share a major bottleneck: they predominantly rely on accessing raw Radio-Frequency (RF) data or demand complex hardware synchronization. In clinical practice, raw RF data is typically proprietary and restricted on commercial scanners. Furthermore, processing raw RF signals is computationally demanding, complicating real-time deployment and limiting clinical versatility.

Consequently, there is an urgent need for a hardware-free, image-domain solution operating on standard B-mode images. We conceptualize the restoration of corrupted images as mapping contaminated observations back into the clean anatomical domain. While diffusion models [7, 9, 10, 26] effectively capture such complex manifolds [16, 18], translating them to real-time HIFU guidance is challenged by slow iterative sampling and the limited capacity of existing U-Nets under extreme interference. To bridge this gap, we propose mHC-Diff, an image-domain diffusion framework that enables high-fidelity prior learning and facilitates knowledge distillation for single-step, real-time acceleration. Specifically, our framework leverages the mHC-UNet architecture, which incorporates manifold-constrained Hyper-Connections (mHC) to expand model capacity via multi-stream residual pathways with minimal computational overhead. By enforcing manifold constraints inspired by [22], this design stably disentangles intense acoustic artifacts from anatomical structures, ensuring high structural fidelity without destructive signal loss. Our primary contributions are:

- We introduce mHC-Diff, an image-domain diffusion framework for HIFU interference suppression operating on standard B-mode inputs, obviating proprietary raw RF access and specialized hardware synchronization.

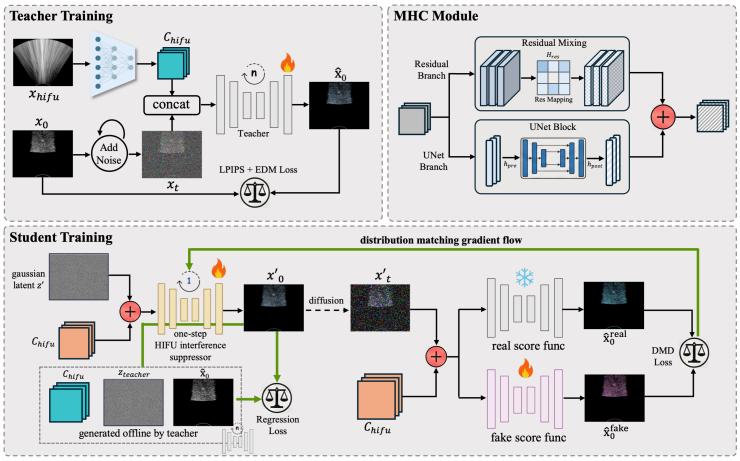

Fig. 1. The mHC-Diff framework. A two-stage pipeline for real-time interference suppression: (i) prior learning via a diffusion-based Teacher, and (ii) knowledge distillation into a single-step Student for real-time inference.

- We propose the mHC-UNet architecture and a two-stage knowledge distillation strategy that disentangles interference from anatomy. This approach achieves superior restoration (26.65 dB PSNR), while accelerating inference to real-time clinical standards (∼20 FPS).
- We constructed a large-scale clinical equivalent HIFU interference suppression dataset comprising 17,464 image pairs from ex vivo and in vivo tissues, covering three typical imaging modes and four HIFU irradiation intensities to support the development of HIFU interference suppression methods.

2 Methods

mHC-Diff is a framework for real-time restoration of ultrasound images corrupted by high-intensity focused ultrasound (HIFU) interference, as show in Fig. 1. It operates in two stages: (i) manifold prior acquisition, where a conditional diffusion Teacher is trained to capture anatomy-consistent clean anatomical distributions; and (ii) where the learned prior is distilled into a single-step Student for real-time inference. Both stages utilize mHC-UNet (Sec. 2.1) to disentangle acoustic interference from anatomical structures.

2.1 Network Architecture: Concat-Conditioned mHC-UNet

We propose a shared denoiser backbone for the Teacher, Student, and critic: an mHC-UNet augmented by manifold constraints inspired by [22].

Conditioning and network input. Given a corrupted frame x_{hifu} and a noisy target x, We extract spatial conditioning features $C_{\text{hifu}} = E_\psi(x_{\text{hifu}})$,

where E_ψ denotes the HIFU-aware conditioning encoder. We concatenate them as $u = [x, C_{\text{hifu}}]$ and feed u into the U-Net stem, where the initial feature map is explicitly broadcast into $n = 4$ parallel latent streams. The noise level σ is injected into all blocks as a global conditional signal, and the n streams are aggregated via feature averaging prior to the final prediction layer.

Multi-stream residual pathways with mHC routing. Following the paradigm in [22], our architecture maintains $n = 4$ parallel streams exclusively along the residual pathways throughout the UNet structure, applying shared-weight projections per stream to improve parameter efficiency.

Cross-stream interaction and integration with standard U-Net blocks are achieved via the adopted mHC module. Because the core U-Net operator $\mathcal{F}(\cdot, \sigma)$ requires a single unified input, the module predicts non-negative read and write gates, $h_{\text{pre}}(p), h_{\text{post}}(p) \in \mathbb{R}^n_+$, alongside a residual stream-mixing matrix $H_{\text{res}}(p) \in \mathbb{R}^{n\times n}$. At spatial location p, the read gates aggregate the n residual streams $x_k(p)$ into a unified representation $r(p)$ for the U-Net block. The write gates then distribute the resulting update $u(p)$ back to the individual streams, which are concurrently mixed by H_{res}:

$$\begin{aligned} r(p) &= \textstyle\sum_{k=1}^{n} h_{\text{pre},k}(p)\, x_k(p), \qquad u(p) = \mathcal{F}(r(p),\, \sigma)\,, \\ x_i^{\text{out}}(p) &= \textstyle\sum_{j=1}^{n} H_{\text{res},ij}(p)\, x_j(p) \;+\; h_{\text{post},i}(p)\, u(p), \qquad \forall i \in \{1, \ldots, n\}. \end{aligned} \tag{1}$$

To ensure stability and identity mapping, we enforce manifold constraints: H_{res} is Sinkhorn-normalized to the Birkhoff polytope to conserve feature magnitude, while sigmoid-activated gates prevent destructive signal cancellation.

2.2 Teacher Prior Learning and Single-Step Distillation

To balance restoration fidelity and real-time efficiency, we propose a two-stage strategy: first, capturing an anatomy-aware prior via an Elucidated Diffusion Model (EDM) [13], and then accelerating inference through Distribution Matching Distillation (DMD) [25].

Stage I: Teacher Optimization via Elucidated Diffusion Model. We first train a Teacher μ_{Teacher} to capture the clean anatomical manifold. Following EDM, we corrupt the clean target x_0 with noise level $\hat{\sigma}$ to obtain $x_t = x_0 + \hat{\sigma}\epsilon$. To preserve fine acoustic textures, the Teacher is optimized to reconstruct x_0 conditioned on frozen C_{hifu} via a perceptual-augmented loss:

$$\mathcal{L}_{\text{Teacher}} = \mathbb{E}_{\hat{\sigma}}\left[\hat{\sigma}^{-2}\, \|\hat{x}_0 - x_0\|_2^2 + \lambda_{\text{perc}}\, \text{LPIPS}(\hat{x}_0, x_0)\right], \tag{2}$$

where $\hat{x}_0 = \mu_{\text{Teacher}}([x_t, C_{\text{hifu}}], \hat{\sigma})$ is the reconstructed image, $\hat{\sigma}$ is the standard deviation of injected noise. Once fully trained, the Teacher is frozen as a reference denoiser μ_{real} for the subsequent phase.

Stage II: One-Step Distillation and Denoising Field Matching. We distill μ_{real} into a one-step Student generator G_ϕ via DMD, involving the Student G_ϕ, the frozen Teacher μ_{real}, and a trainable critic μ_{fake} sharing the mHC-UNet backbone. The Student generates $x'_0 = G_\phi([\sigma_g z', C_{\text{hifu}}], \sigma_g)$ where $z' \sim \mathcal{N}(0, I)$

and σ_g is the generation noise scale and we align the Student's denoising field with the Teacher's prior. For a noisy state $x'_t = x'_0 + \sigma_t \epsilon$ perturbed at a sampled noise level σ_t within the EDM distribution, both denoisers are evaluated:

$$\hat{x}_0^{\text{real}} = \mu_{\text{real}}([x'_t, C_{\text{hifu}}], \sigma_t), \qquad \hat{x}_0^{\text{fake}} = \mu_{\text{fake}}([x'_t, C_{\text{hifu}}], \sigma_t). \tag{3}$$

The update direction g is computed as:

$$w = \text{mean}_{c,p}\big[\,|x'_0 - \hat{x}_0^{\text{real}}|\,\big], \qquad g = \frac{\hat{x}_0^{\text{fake}} - \hat{x}_0^{\text{real}}}{w + \epsilon}. \tag{4}$$

where $\epsilon = 10^{-6}$ is a small constant for numerical stability. To ensure stability, the Student and critic are optimized via decoupled gradient flows:

$$\begin{aligned} \mathcal{L}_{\text{DMD}} &= \frac{1}{2}\,\|x'_0 - \text{sg}(x'_0 - g)\|_2^2\,, \\ \mathcal{L}_{\text{critic}} &= \mathbb{E}_t\left[\left(\frac{1}{\sigma_t^2} + \frac{1}{\sigma_{\text{data}}^2}\right)\|\mu_{\text{fake}}([x'_t, C_{\text{hifu}}], \sigma_t) - x'_0\|_2^2\right], \end{aligned} \tag{5}$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operation. To prevent mode collapse in distillation procedure, we apply a supervised anchor $\mathcal{L}_{\text{reg}} = \text{LPIPS}(\hat{x}_0, x'_0)$ on a reference generation, and the final Student objective is $\mathcal{L}_G = \mathcal{L}_{\text{DMD}} + \lambda_{\text{reg}}\,\mathcal{L}_{\text{reg}}$.

3 Experiments

3.1 Datasets and Implementation

To evaluate the proposed framework, we constructed a large-scale dataset using a clinically equivalent HIFU therapy system (Focused Ultrasound Tumor Therapeutic System, JC200, Chongqing Haifu Medical Technology Co., Ltd.) integrated with a Verasonics Vantage 256 ultrasound research platform (Verasonics Inc., USA). The HIFU waveform was generated by a PC-controlled RIGOL DG4202 signal generator as a continuous 0.97 MHz signal, amplified, and delivered through an electrical matching box to the HIFU transducer. As illustrated

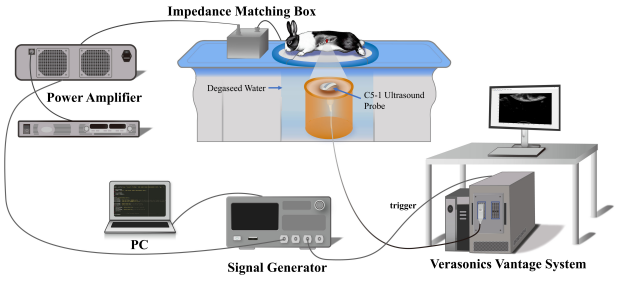

Fig. 2. Schematic illustration of the experimental setup. A clinical-grade HIFU therapy system is integrated with a 256-channel ultrasound research platform.

in Fig. 2, the HIFU transducer (0.97 MHz) was configured coaxially with an imaging probe to ensure aligned visualization and consistent monitoring of the focal region. The therapeutic transducer had a 0.97 MHz central frequency, 220 mm outer diameter, and 80 mm inner diameter, enabling coaxial placement of the imaging transducer.

Data Acquisition and Protocol. A consistent imaging depth of 135–140 mm was maintained across all imaging techniques. We implemented three clinically established beamforming strategies, i.e., Diverging Wave, Wide Beam, and Line Scan imaging, across four power levels (123 W–277 W). Paired corrupted and clean frames were acquired using a pulse-width modulation protocol (200 ms cycles), yielding 50 labeled paired samples per session. The final dataset comprises 17,464 image pairs, acquired at 512×512 and spatially resized to 256×256 for training, from ex vivo bovine and porcine tissues and in vivo rabbit models. To ensure independence, data were split at the subject/session level (7:1:2), yielding 12,225 training, 1,746 validation, and 3,493 test pairs. No session contributed frames to more than one split.

Architecture and Training Details. Both teacher and student models utilize an mHC-UNet backbone (128 channels, $n = 4$, multipliers: $[2, 2, 2]$) coupled with an interference-aware encoder featuring five NAF blocks [5]. We adopt a two-stage training strategy: the encoder is jointly pre-trained alongside the teacher and then frozen during distillation to provide stable feature guidance. EDM parameters are set as $\sigma_{min} = 0.002, \sigma_{max} = 80, \sigma_{data} = 0.5$. The teacher underwent 200 epochs of training (batch size 16, LR 1×10^{-4}), and the student used the same teacher configuration to ensure manifold consistency for DMD. The implementation was executed using PyTorch mixed precision on an NVIDIA H100, requiring $\sim$24 hours for end-to-end training. For fair comparison, all learning-based baselines followed the same preprocessing pipeline and subject/session-level splits, and supervised baselines were trained with identical restoration losses.

3.2 Performance Evaluation and Comparison

We compare mHC-Diff against SOTA baselines, including signal processing methods (FRPCA) operating on raw RF signal, and deep learning architectures (FUS-Net, SwinUnet, and HIFU-Diff) in the image domain. We evaluate restoration quality using PSNR and SSIM (reported as mean with 95% bootstrap confidence intervals).

Quantitative Evaluation. As shown in Table 1, mHC-Diff consistently outperforms all baselines across all beamforming modes. Specifically, in line-scan imaging, it yields a 27.29 dB PSNR, surpassing the strongest baseline (SwinUnet) by 2.05 dB. This superiority persists in more challenging wide-beam and diverging-wave scenarios, with narrow 95% confidence intervals (via 1,000 bootstrap resamples) confirming its statistical stability. While FUS-Net and SwinUnet show moderate suppression, their lower SSIM indicates an inability to fully preserve fine anatomical details. SwinUnet benefits from global attention, but lacks an interference-conditioned prior for lesion fidelity. Notably, standard

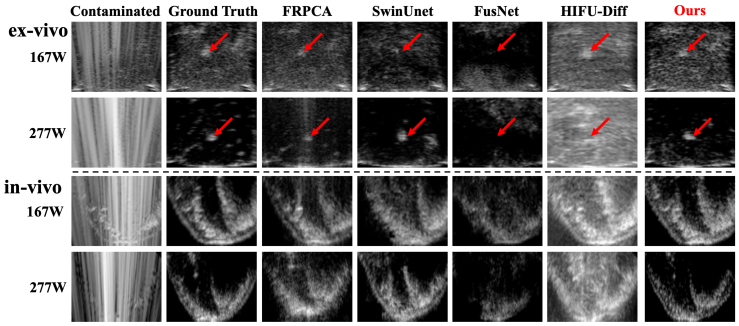

Fig. 3. Qualitative comparison across ex-vivo (top) and in-vivo (bottom) scenarios under varying power levels. **Red arrows** indicate critical HIFU-induced lesions.

HIFU-Diff fails in the image domain as the absence of RF phase information in B-mode data creates nonlinear coupling between artifacts and anatomy, necessitating the robust priors of mHC-Diff for reliable disentanglement.

Qualitative Analysis. As shown in Fig. 3, mHC-Diff achieves robust interference suppression while preserving anatomical structures across both ex vivo and in vivo settings. Conventional methods such as FRPCA and FUS-Net either fail to remove dense vertical streaks or suppress them at the cost of noticeable signal loss and contrast degradation. In contrast, mHC-Diff restores more natural speckle patterns and tissue contrast, even at the maximum power level (277 W). Crucially, HIFU-induced lesions (red arrows) are reconstructed with sharp edges and clear hyperechoic contrast, whereas baseline results are often blurred and partially merge lesion regions with background textures. With real-time inference at ∼20 FPS on a single NVIDIA RTX 4090, our framework provides reliable visual feedback for continuous ultrasound-guided HIFU procedures.

3.3 Inference Efficiency

Fig. 4 compares computational efficiency across all modalities to assess intraoperative suitability. Traditional FRPCA exhibits prohibitive latency (e.g., 123 s/frame in line-scan mode), rendering it clinically incompatible. The standard HIFU-Diff requires 0.355 s/frame due to iterative sampling, creating a severe

Table 1. Quantitative comparison of different methods across three imaging modalities: Line Scan, Wide Beam, and Diverging Wave. Results are presented as Mean (95% CI). Best results are highlighted in bold.

Method	Line Scan		Wide Beam		Diverging Wave	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FUS-Net [14]	23.83 (23.66-24.00)	0.812 (0.809-0.814)	23.82 (23.68-23.96)	0.809 (0.806-0.812)	22.82 (22.63-22.99)	0.805 (0.802-0.808)
HIFU-Diff [3]	8.84 (8.50-9.16)	0.587 (0.576-0.598)	7.31 (7.02-7.58)	0.535 (0.524-0.545)	7.65 (7.35-7.96)	0.537 (0.528-0.548)
SwinUnet [4]	25.24 (25.08-25.41)	0.838 (0.835-0.841)	25.02 (24.88-25.17)	0.838 (0.835-0.841)	25.01 (24.83-25.20)	0.836 (0.833-0.840)
FRPCA [24]	15.59 (15.33-15.81)	0.740 (0.732-0.747)	10.77 (10.56-10.99)	0.522 (0.515-0.529)	13.44 (13.28-13.60)	0.585 (0.581-0.589)
Ours (mHC-Diff)	**27.29 (27.14-27.44)**	**0.858 (0.856-0.861)**	**26.20 (26.05-26.35)**	**0.858 (0.855-0.861)**	**25.49 (25.29-25.66)**	**0.850 (0.846-0.853)**

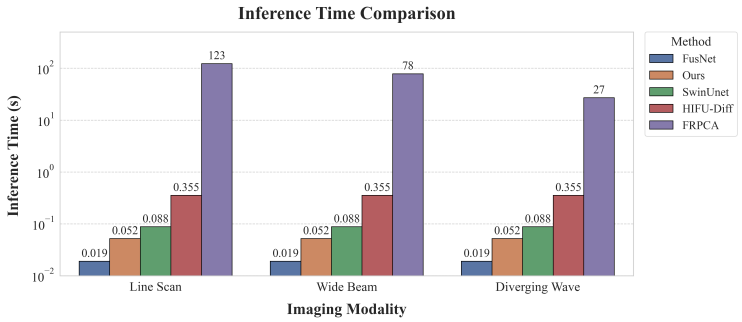

Fig. 4. Inference latency comparison. Our mHC-Diff maintains real-time throughput (≈0.05 s) across modalities, significantly faster than HIFU-Diff and FRPCA baselines.

Table 2. Performance comparison and ablation study. The **Baseline** denotes the iterative EDM model without mHC or distillation, while **mHC-Diff** integrates both components. Best results are in **bold**.

Method	PSNR ↑	SSIM ↑	MSE ↓	MS-SSIM ↑	LPIPS ↓	Time (ms) ↓
FUS-Net [14]	23.50	0.809	0.0054	0.6988	0.2068	**19.0**
HIFU-Diff [3]	7.93	0.553	0.2540	0.3779	0.2761	355.0
SwinUnet [4]	24.59	0.828	0.0040	0.8561	0.1539	88.0
FRPCA [24]	14.33	0.648	0.0474	0.5930	0.1671	7.6×10^4
Baseline	26.52	0.843	0.0022	0.8697	0.1431	499.0
Baseline + mHC	**27.35**	0.851	**0.0018**	**0.8892**	0.1426	592.5
Baseline + Distill	25.39	0.839	0.0029	0.8658	0.1499	38.5
mHC-Diff (Ours)	26.65	**0.856**	0.0022	0.8875	**0.1402**	52.0

temporal bottleneck. Conversely, our one-step Student achieves a consistent 0.052 s/frame (∼20 FPS) across all modalities. Although FUS-Net is the fastest (0.019 s/frame), its failure to preserve therapeutic features (Sec. 3.2) renders it clinically unreliable. SwinUnet (0.088 s/frame) remains significantly slower than our approach. By bypassing iterative sampling, our framework strikes an optimal balance between manifold integrity and real-time throughput. This enables seamless clinical integration for high-fidelity, low-latency visual feedback during active treatment.

3.4 Ablation Study

Table 2 quantifies the individual and synergistic contributions of the manifold-constrained Hyper-Connections (mHC) and Single-step Distillation (SD).

Impact of Single-Step Distillation. The standard iterative baseline (EDM teacher without mHC) achieves strong restoration (26.52 dB) but incurs high latency (499.0 ms). Distilling this teacher into a one-step student (Baseline+Distill) substantially reduces latency to 38.5 ms, at the cost of a PSNR drop to 25.39 dB. Despite this trade-off, the distilled model still outperforms supervised baselines such as SwinUnet (24.59 dB) and FUS-Net (23.50 dB). The teacher therefore serves as a training-time prior, not a deployed module. In contrast, directly ap-

plying HIFU-Diff to B-mode data yields very low PSNR (7.93 dB), suggesting that a dedicated image-domain prior is crucial for stable interference suppression.

Synergy of mHC Routing and Distillation. Integrating mHC to the iterative baseline improves the upper-bound restoration quality, where the mHC-equipped teacher achieves 27.35 dB PSNR and 0.0018 MSE. After distillation, our full model mHC-Diff recovers much of the fidelity typically lost in the one-step setting. Compared to Baseline+Distill (25.39 dB), mHC-Diff reaches 26.65 dB with the best structural and perceptual quality (0.856 SSIM; 0.1402 LPIPS). These results indicate that mHC routing effectively compensates for information loss introduced by distillation, yielding a favorable trade-off between image quality and real-time throughput for ultrasound-guided HIFU procedures.

4 Conclusion

We presented mHC-Diff, a two-stage framework for real-time HIFU interference suppression. By coupling manifold-constrained Hyper-Connections (mHC) with a one-step distillation paradigm, our approach effectively disentangles acoustic artifacts from anatomy, achieving 20 FPS inference. Extensive evaluations on a physically-consistent and clinically-representative corpus demonstrate superior reconstruction accuracy and structural fidelity over state-of-the-art benchmarks. The high-fidelity visualization of lesion boundaries in clinical scenarios underscores its potential for precise intraoperative guidance. Future work will focus on clinical hardware integration and broader cross-domain validation to further establish its robustness and therapeutic efficacy.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grants 12304508 and 62402458, HKUST Research Project (Project No. 24251460C049), Hong Kong Innovation and Technology Commission (Project No. MHP/002/22), and Shenzhen Science and Technology Innovation Committee Fund (Project No. KCXFZ20230731094059008).

Disclosure of Interests

The authors have no competing interests to declare.

References

1. Bond, A.E., Shah, B.B., Huss, D.S., Dallapiazza, R.F., Warren, A., Ma, J., Kassell, N.F., Elias, W.J.: Safety and efficacy of focused ultrasound thalamotomy for patients with medication-refractory, tremor-dominant parkinson disease: a randomized clinical trial. JAMA Neurology **74**(12), 1412–1418 (2017)

2. Cai, D., Yang, K., Liu, X., Xu, J., Ran, Y., Xu, Y., Zhou, X.: A novel diffusion-based deep learning model to suppress acoustic interference for real-time monitoring in ultrasound-guided hifu surgery. In: IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS). pp. 1–4. IEEE (2024). https://doi.org/10.1109/UFFC-JS60046.2024.10793861
3. Cai, D., Yang, K., Liu, X.B., Xu, J., Ran, Y., Xu, Y., Zhou, X.: Suppressing the hifu interference in ultrasound guiding images with a diffusion-based deep learning model. Computer Methods and Programs in Biomedicine **254**, 108304 (2024)
4. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European Conference on Computer Vision (ECCV) Workshops. pp. 205–218. Springer Nature Switzerland, Cham (2022)
5. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 17–33. Springer (2022)
6. Chen, Z., Cheng, L., Zhang, W., He, W.: Ultrasound-guided thermal ablation for hyperparathyroidism: current status and prospects. International Journal of Hyperthermia **39**(1), 466–474 (2022)
7. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
8. Dou, Y., Zhang, L., Liu, Y., He, M., Wang, Y., Wang, Z.: Long-term outcome and risk factors of reintervention after high intensity focused ultrasound ablation for uterine fibroids: a systematic review and meta-analysis. International Journal of Hyperthermia **41**(1), 2299479 (2024)
9. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12873–12883 (2021)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
11. Hynynen, K., Darkazanli, A., Unger, E., Schenck, J.F.: Mri-guided noninvasive ultrasound surgery. Medical Physics **20**(1), 107–115 (1993)
12. Imankulov, S., Tuganbekov, T., Razbadauskas, A., Seidagaliyeva, Z.: Hifu treatment for fibroadenoma-a clinical study at national scientific research centre, astana, kazakhstan. J Pak Med Assoc **68**(9), 1378–80 (2018)
13. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. arXiv preprint arXiv:2206.00364 (2022). https://doi.org/10.48550/arXiv.2206.00364
14. Lee, S.A., Konofagou, E.E.: Fus-net: U-net-based fus interference filtering. IEEE transactions on medical imaging **41**(4), 915–924 (2021)
15. Liu, Z., Liu, Z., Wang, Y., Wan, X., Huang, X.: Machine learning-based predictive analysis of energy efficiency factors necessary for the hifu treatment of adenomyosis. Frontiers in Physiology **16**, 1602866 (2025). https://doi.org/10.3389/fphys.2025.1602866
16. Niu, A., Pham, T.X., Zhang, K., Sun, J., Zhu, Y., Yan, Q., Kweon, I.S., Zhang, Y.: Acdmsr: Accelerated conditional diffusion models for single image super-resolution. IEEE Transactions on Broadcasting **70**(2), 492–504 (2024)
17. Payen, T., Crouzet, S., Guillen, N., Chen, Y., Chapelon, J.Y., Lafon, C., Catheline, S.: Passive elastography for clinical hifu lesion detection. IEEE Transactions on Medical Imaging **43**(4), 1594–1604 (2023)

18. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE transactions on pattern analysis and machine intelligence **45**(4), 4713–4726 (2022)
19. Shen, C.C., Lin, R.C., Wu, N.H.: Golay-encoded ultrasound monitoring of simultaneous high-intensity focused ultrasound treatment: a phantom study. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **69**(4), 1370–1381 (2022)
20. Shen, C.C., Wu, N.H.: Ultrasound monitoring of simultaneous high-intensity focused ultrasound (hifu) therapy using minimum-peak-sidelobe coded excitation. Ultrasonics **138**, 107224 (2024)
21. Song, J.H., Chang, J.H.: Correspondence-an effective pulse sequence for simultaneous hifu insonation and monitoring. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **61**(9), 1580–1587 (2014)
22. Xie, Z., Wei, Y., Cao, H.b., Zhao, C., Deng, C., Li, J., Dai, D., Gao, H., Chang, J., Yu, K., et al.: mhc: Manifold-constrained hyper-connections. arXiv preprint arXiv:2512.24880 (2025)
23. Yang, K., Li, Q., Liu, H., Zeng, Q., Cai, D., Xu, J., Zhou, Y., Tsui, P.H., Zhou, X.: Suppressing hifu interference in ultrasound images using 1d u-net-based neural networks. Physics in Medicine & Biology **69**(7), 075006 (2024)
24. Yang, K., Li, Q., Xu, J., Tang, M.X., Wang, Z., Tsui, P.H., Zhou, X.: Frequency-domain robust pca for real-time monitoring of hifu treatment. IEEE Transactions on Medical Imaging **43**(8), 3001–3012 (2024)
25. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. arXiv preprint arXiv:2311.18828 (2024). https://doi.org/10.48550/arXiv.2311.18828
26. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4791–4800 (2021)