BME^2: A Plug-and-Play Bridge-Based Module for Misalignment Estimation and Elimination in Multi-Scan Image Restoration

Wenxuan Chen[1]*, Caiwen Jiang[2]*, Xiaolei Song[1](✉), Dinggang Shen[2,3,4](✉)

[1] School of Biomedical Engineering, Tsinghua University
songxl@tsinghua.edu.cn
[2] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University
[3] Shanghai United Imaging Intelligence Co., Ltd.
[4] Shanghai Clinical Research and Trial Center
dgshen@shanghaitech.edu.cn

Abstract. The multi-scan imaging procedure, involving both multi-modal and multi-timepoint scans, captures temporal changes and complementary cross-modality information, playing a key role in clinical diagnosis. Multi-scan image restoration (IR), which leverages high-quality reference scans to aid in restoring degraded current scans, holds significant potential for reducing the cost of the multi-scan procedure. However, misalignment between scans, arising from patient physiological or posture changes, impacts the ability of networks to exploit cross-scan correlations and leads to declined restoration performance. To this end, we propose a plug-and-play **B**ridge-**B**ased **M**odule for **M**isalignment **E**stimation and **E**limination (BME^2), which adopts a coarse-to-fine strategy to estimate cross-scan misalignment. Specifically, a lightweight misalignment estimation (ME) network first predicts the initial deformation fields, which are then iteratively refined via a latent Schrödinger bridge-based model to obtain the final estimation. Notably, BME^2 can be added to arbitrary backbones and only introduces mild computational costs. Validated on brain MRI and abdominal CT datasets, BME^2 universally enhances four baselines, achieving average PSNR gains of 0.54 and 0.65 dB on brain and abdominal data, respectively. The codes are available at: https://github.com/ChenWenxuan2021/BME2.

Keywords: Multi-Scan Image Restoration · Misalignment Estimation · Plug-and-Play · Schrödinger Bridge.

1 Introduction

The multi-scan imaging procedure involves acquiring multi-modal and multi-timepoint scans from the same patient, which can capture temporal changes

* Wenxuan Chen and Caiwen Jiang contribute equally to this work.

in disease progression [26] and provide complementary cross-modality information [1], offering significant value for disease monitoring, treatment assessment, and intervention planning. However, traditional multi-scan procedures treat each scan as an independent process, overlooking the substantial overlapping information between scans, which leads to information redundancy and inefficient data utilization. As a solution, multi-scan image restoration (IR) uses high-quality reference (Ref) scans from the same patient to help restore degraded (Deg) current scans [24, 23, 28, 22], thereby improving data efficiency and reducing imaging costs in multi-scan procedures.

In recent years, several studies have focused on multi-scan IR, especially in the multi-contrast MRI field. For instance, Lyu et al. introduced a dual-branch progressive super-resolution network [19]. Feng et al. proposed networks with multi-level integration [6] and separable attention mechanisms [7]. Lei et al. developed a deep variational network to model correlations between multiple scans [13]. However, these methods did not account for spatial misalignment between Ref and Deg scans, which may lead to limited performance. Inspired by reference-based super-resolution [25, 18], some studies attempted to address this issue by performing global-search-based feature matching [14, 5], but they are typically computationally expensive. Moreover, these methods may introduce undesired artifacts when matching and transferring irrelevant textures from Ref scans. Recent studies proposed deformable attention [2] and difference projection discriminator [8] mechanisms to alleviate cross-scan misalignment, but their performance in complex clinical settings, e.g., significant non-rigid misalignment between multi-scan abdominal images, requires further improvement.

Notably, advancements in diffusion models (DMs) [10, 15] provide new insights into overcoming the limitations of current multi-scan IR methods. Leveraging the powerful representation capabilities of diffusion models, some studies have successfully applied them to predict misalignment fields for registration tasks [11]. However, directly applying the diffusion model to predicted cross-scan misalignment still presents some issues. First, conventional DMs performed in the image space are computationally expensive, especially for high-resolution medical images. Secondly, as deep generative models, DMs tend to produce hallucinated deformation fields, particularly when predicting large misalignment, which can be detrimental to detail-sensitive medical images.

Recent studies have explored integrating latent diffusion models (LDMs) into regression-based backbones for computation efficiency [21, 3]. Additionally, DM-related bridge-based methods [17] have been proposed to enhance the utilization of conditional inputs. Inspired by these studies, we consider leveraging deep generative models to refine the misalignment estimation, such as that in [4, 12], whereas remaining regression-based image-restoration backbones to ensure reconstruction fidelity. To this end, this paper proposes a **B**ridge-**B**ased **M**odule for **M**isalignment **E**stimation and **E**limination (BME^2), providing a plug-and-play solution for predicting deformation fields between multi-scan medical images. BME^2 employs a coarse-to-fine two-stage strategy, where a misalignment estimation (ME) network first predicts the initial deformation fields. Subsequently,

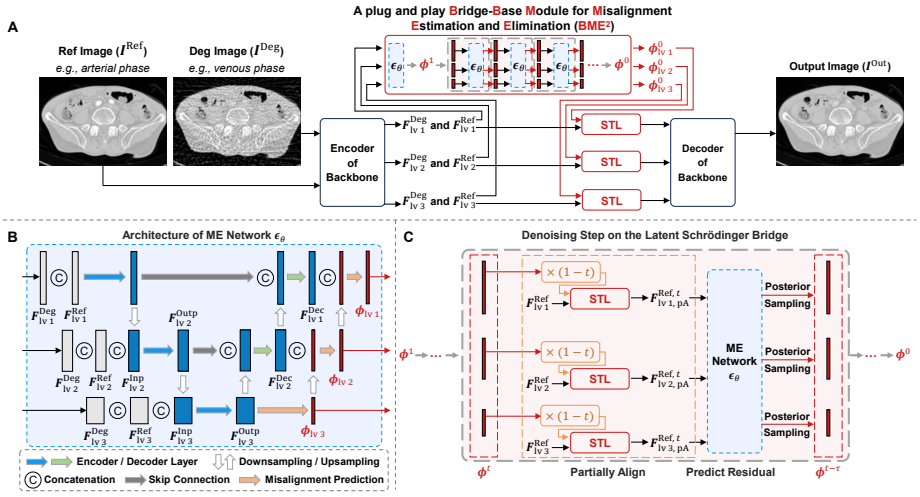

Fig. 1. Overview of our proposed method. (A) Embedding of BME^2 into backbones, e.g., a hierarchical network with 3-scale features. (B) Architecture of the misalignment estimation (ME) network. (C) A typical denoising step on the latent Schrödinger bridge.

the deformation fields are iteratively refined through recursive posterior sampling using a Schrödinger bridge-based model, thereby achieving good alignment between multi-scan features and improving output quality. BME^2 can be added to arbitrary backbones and only introduces mild computation costs. In conclusion, the main contributions of our work include: (1) We introduce the Schrödinger bridge to multi-scan IR for addressing cross-scan misalignment. (2) We design a plug-and-play module for misalignment estimation and elimination that can be embedded into arbitrary multi-scan IR backbones. (3) We conduct extensive experiments on two datasets to validate the effectiveness of our method.

2 Methodology

Fig. 1(A) illustrates embedding our proposed plug-and-play BME^2 into a multi-scan IR backbone. Note that, BME^2 allows the backbone with arbitrary levels, i.e., both single-scale and multi-scale backbones are applicable, and herein we mainly introduce the more complicated case with multi-scale features. Specifically, given the multi-scale Deg and Ref features denoted as $\{\boldsymbol{F}^{\text{Deg}}_{\text{lv}n}, \boldsymbol{F}^{\text{Ref}}_{\text{lv}n}\}$ across N scales ($n \in \{1, 2, \ldots, N\}$), BME^2 obtains correspondingly multi-scale deformation fields $\boldsymbol{\phi}_{\text{lv}n}$. In the following sections, we will introduce its two main components, i.e., the ME network and latent Schrödinger bridge.

2.1 Misalignment Estimation Network

The ME network $\boldsymbol{\epsilon}_\theta$ adopts a lightweight U-shaped architecture, which accepts multi-scale features as inputs. Specifically, the encoder of ME network can be

expressed as

$$\begin{aligned} \boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Outp}} =& Enc_{\mathrm{lv}n}\left(\mathrm{Concat}(\boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Deg}},\ \boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Ref}},\ \boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Inp}})\right), \\ \boldsymbol{F}_{\mathrm{lv}n+1}^{\mathrm{Inp}} =& Ds_{\mathrm{lv}n}(\boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Outp}}), \end{aligned} \tag{1}$$

where $Enc(\cdot)$ and $Ds(\cdot)$ indicate encoder and downsampling layers, and $\boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Inp}}$ denotes the intermediate features obtained by the previous downsampling layer. There is no $\boldsymbol{F}_{\mathrm{lv1}}^{\mathrm{Inp}}$ and $DS_{\mathrm{lv}N}$ at the first and last level, respectively. In the decoder, we first predict the deepest-level misalignment at the bottleneck layer:

$$\boldsymbol{\phi}_{\mathrm{lv}N} = Pred_{\mathrm{lv}N}(\boldsymbol{F}_{\mathrm{lv}N}^{\mathrm{Outp}}), \tag{2}$$

where $Pred(\cdot)$ is implemented by a convolutional layer with two-channel outputs to obtain the deformation in the Y and X directions. The prediction of deformation fields at other levels ($n < N$) can be expressed as follows:

$$\begin{aligned} \boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Dec}} &= Dec_{\mathrm{lv}n}\left(\mathrm{Concat}(Us_{\mathrm{lv}n}(\boldsymbol{F}_{\mathrm{lv}n+1}^{\mathrm{Dec}}),\ \boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Outp}})\right) \\ \boldsymbol{\phi}_{\mathrm{lv}n} &= Pred_{\mathrm{lv}n}\left(\mathrm{Concat}(\boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Dec}},\ \boldsymbol{\phi}_{\mathrm{lv}n+1}\uparrow)\right), \end{aligned} \tag{3}$$

where $Dec(\cdot)$ and $Us(\cdot)$ represent the decoder and upsampling layers, $\boldsymbol{F}_{\mathrm{lv}n}^{\mathrm{Outp}}$ is a skip connection, and $\boldsymbol{\phi}_{\mathrm{lv}n+1}\uparrow$ denotes the bilinear interpolation of the previous-level deformation fields. The outputs of ME network are $\{\boldsymbol{\phi}_{\mathrm{lv1}}, \boldsymbol{\phi}_{\mathrm{lv2}}, \ldots, \boldsymbol{\phi}_{\mathrm{lv}N}\}$.

2.2 Bridge-Based Refinement of Misalignment

Using a single ME network to estimate the deformation fields precisely is difficult when cross-scan misalignment has a large amplitude or complex patterns. To address this, we leverage the powerful representation capability of Schrödinger bridge-based model to refine the initially estimated deformation fields. As illustrated in Fig. 1(C), the bridge adopts an iterative process of "partially aligning and residual prediction". Specifically, given a time step $t \in (0, 1]$, i.e., $t = 1$ and $t = 0$ denote the starting and end of the bridge, the Ref features are warped with a portion of deformation fields. Subsequently, the ME network $\boldsymbol{\epsilon}_\theta$, which shares parameters with that in Sec. 2.1, is used to estimate the residual components of misalignment, expressed as

$$\begin{aligned} \boldsymbol{F}_{\mathrm{pA}}^{\mathrm{Ref},t} &= \mathrm{STL}\left(\boldsymbol{F}^{\mathrm{Ref}}, (1-t)\times\boldsymbol{\phi}^t\right), \\ \boldsymbol{\phi}_{\mathrm{Res}}^t &= \boldsymbol{\epsilon}_\theta\left(\boldsymbol{F}^{\mathrm{Deg}},\ \boldsymbol{F}_{\mathrm{pA}}^{\mathrm{Ref},t}\right), \end{aligned} \tag{4}$$

where the subscript "pA" of $\boldsymbol{F}_{\mathrm{pA}}^{\mathrm{Ref},t}$ indicates the partial alignment of Ref features, and $\mathrm{STL}(\cdot)$ denotes a spatial transform layer. Note that, all equations in Sec. 2.2 omit the subscript "lv$n$" for simplicity because the ME network simultaneously

generates the deformation fields at all scales. Then, we calculate the pseudo-refined deformation fields at the current step, denoted as $\hat{\boldsymbol{\phi}}^{0,t}$, via the addition of current prediction and residual components:

$$\hat{\boldsymbol{\phi}}^{0,t} = (1-t) \times \boldsymbol{\phi}^{t} + \boldsymbol{\phi}^{t}_{\mathrm{Res}}. \tag{5}$$

After that, we recursively update the next-step $\boldsymbol{\phi}$ by posterior sampling:

$$\boldsymbol{\phi}^{t-\tau} = \frac{\sigma_t^2 - \sigma_{t-\tau}^2}{\sigma_t^2}\hat{\boldsymbol{\phi}}^{0,t} + \frac{\sigma_{t-\tau}^2}{\sigma_t^2}\boldsymbol{\phi}^{t}, \tag{6}$$

where σ_t^2 is a scheduling hyper-parameter in the Schrödinger bridge, and τ controls the number of sampling steps ($1/\tau$). Please refer to [17] for more details of the posterior sampling process.

As shown in Fig. 1(A), the input of the Schrödinger bridge is the coarsely-estimated deformation fields ϕ^1. With the iterative sampling progressing, the decreasing amplitude of residual misalignment facilitates the increasing prediction accuracy of the ME network. The bridge concludes at $t = 0$, where we obtain $\boldsymbol{\phi}^0$ as the ultimately-refined outputs, which are then used to align Ref features and eliminate the cross-scan misalignment. The well-aligned Deg and Ref features are input into the backbone to obtain high-quality restored images.

2.3 Loss Functions

Unlike conventional DMs that randomly sample a time step [10, 20], the plug-and-play BME^2 are jointly trained with regression-based backbones in an end-to-end manner by running full sampling steps. There are three loss terms, i.e., reconstruction loss, correlation loss, and smoothness loss, used to optimize the network. Reconstruction loss uses the L1 loss between outputs and ground truth (GT) images. Correlation loss computes the cross-correlation between the aligned Ref and GT images, similar to that in unsupervised image registration models [11, 12]. Smoothness loss is a penalty term on the gradient of the deformation fields. The loss functions can be expressed as

$$\begin{gathered} \mathcal{L}_{\mathrm{Rec}} = \|\boldsymbol{I}^{\mathrm{Out}} - \boldsymbol{I}^{\mathrm{GT}}\|_1, \\ \mathcal{L}_{\mathrm{Cor}} = -\boldsymbol{I}_{\mathrm{A}}^{\mathrm{Ref}} \otimes \boldsymbol{I}^{\mathrm{GT}}, \quad \text{where } \boldsymbol{I}_{\mathrm{A}}^{\mathrm{Ref}} = \mathrm{STL}(\boldsymbol{I}^{\mathrm{Ref}}, \boldsymbol{\phi}^0_{\mathrm{lv1}}), \\ \mathcal{L}_{\mathrm{Smo}} = \textstyle\sum_{n=1}^{N} \left(\sum \|\nabla \boldsymbol{\phi}^0_{\mathrm{lv}n}\|\right), \end{gathered} \tag{7}$$

where $\boldsymbol{\phi}^0_{\mathrm{lv1}}$ is the refined first-level deformation fields with the same size as Ref images. The total loss is obtained by

$$\mathcal{L}_{\mathrm{Total}} = \mathcal{L}_{\mathrm{Rec}} + 0.1 \times \mathcal{L}_{\mathrm{Cor}} + 0.1 \times \mathcal{L}_{\mathrm{Smo}}. \tag{8}$$

3 Experiments

3.1 Dataset

Experiments are conducted on brain MRI and abdominal CT datasets. (1) The **IXI dataset** [9], an open-access dataset for brain MRI, contains 498 paired

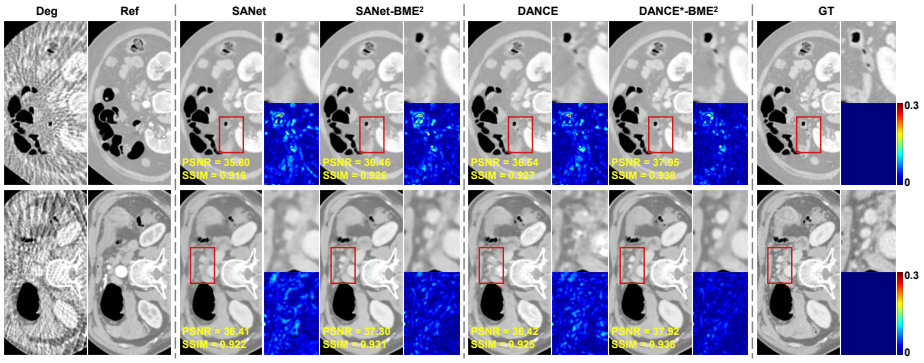

Fig. 2. Qualitative comparisons of two cases from the abdominal CT dataset. Models enhanced by BME^2 provide better output quality than those without BME^2.

T1-weighted (T1w) and T2-weighted (T2w) volumes, split into training (350 volumes), validation (48 volumes), and testing (100 volumes) sets. (2) An **Abdominal CT dataset** includes 84 real clinical cases from *anonymous* hospitals, with each sample containing scans from two points, i.e., arterial phase (AP) and venous phase (VP). We split the abdominal dataset into (60, 8, 16) volumes for training, validation, and testing, respectively.

Data Preparation. For MR images, T1w images are used as high-quality Ref scans, and T2w images are downsampled into $1/4\times$ lower-resolution counterparts by K-space truncating to simulate Deg scans. We manually introduce random rigid misalignment between Ref and Deg scans, with a maximum amplitude of $[-6, 6]$ rotation and $[-8, 8]$ pixel of translation, similar to those operations in [2]. For abdominal CT images, VP images are used as Deg scans, with degradation simulated by uniform undersampling sinograms in the Radon transformation domain. Inherent non-rigid misalignment exists between AP and VP images due to inevitable soft tissue deformation.

3.2 Implementation

Backbones. Four backbones, i.e., MINet (MICCAI 2021) [6], McMRSR (CVPR 2022) [14], SANet (TNNLS 2024) [7], DANCE (MedIA 2025) [2], are used in the experiments. Among them, MINet and SANet are based on the spatial alignment assumption, while MINet-BME^2 and SANet-BME^2 introduce BME^2 to align Ref features with Deg features before feature fusion. McMRSR and DANCE employ global correlation search and deformable attention mechanisms for feature matching; therefore, we replace their feature-matching modules with BME^2, creating McMRSR*-BME^2 and DANCE*-BME^2, where the asterisk indicates removal of their original feature-matching modules. Besides, two single-image IR methods, SwinIR [16] and Restormer [27], are also used as comparative baselines.

Implementation Details of BME^2. The ME network $\boldsymbol{\epsilon}_\theta$ is implemented by a three-layer U-Net, with the number of channels being $(64, 128, 256)$ from the 1$^{\text{st}}$

Table 1. Comparative results of models with or without the proposed BME^2. The unit of PSNR is dB. Higher PSNR and SSIM metrics indicate better results.

		Brain (rigid)		Abdominal (non-rigid)	
		PSNR	SSIM	PSNR	SSIM
Single Image	SwinIR	30.45	0.931	35.18	0.907
	Restormer	30.99	0.935	35.62	0.916
Multi Scan	MINet	30.52	0.933	35.91	0.920
	MINet-BME^2	31.31	0.940	36.59	0.925
	McMRSR	31.34	0.942	36.38	0.921
	McMRSR*-BME^2	31.60	0.944	36.95	0.929
	SANet	30.86	0.939	36.82	0.930
	SANet-BME^2	31.71	0.946	37.70	0.938
	DANCE	31.74	0.950	37.36	0.936
	DANCE*-BME^2	32.01	0.953	37.81	0.939

Table 2. Ablation study on the abdominal dataset. DANCE (abbreviated as D.) is adopted as the backbone for ablation study.

	PSNR	SSIM
D.	37.36	0.936
D.*-w/o bri	37.29	0.933
D.*-w/o csl	37.66	0.938
D.*-BME^2	37.81	0.939

Table 3. Efficiency Study on the abdominal dataset. FLOPs are based on the image size of 512×512. S. and D. are abbreviations of SANet and DANCE.

	FLOPs	PSNR	SSIM
S.	1.658×10^{12}	36.82	0.930
S.-BME^2	1.975×10^{12}	37.70	0.938
D.	2.873×10^{12}	37.36	0.936
D.*-BME^2	2.728×10^{12}	37.81	0.939

to the 3$^{\text{rd}}$ levels. Each level of encoders and decoders of the ME network employs two convolutional layers with 3×3 kernels. The scheduling hyper-parameter σ_t^2 is identical to that in [17]. The latent Schrödinger bridge has 8 sampling steps, i.e., $\tau = 1/8$. All models are trained with an initial learning rate of 1×10^{-4}, attenuated by 0.99 at every 1000 iterations until convergence.

3.3 Comparative Results

Quantitative results on the two datasets are provided in Table 1. Overall, multi-scan IR methods outperform single-image methods by leveraging the guidance from high-quality Ref scans. For all involved backbones, BME^2 consistently improves the model performance in the presence of both rigid and non-rigid misalignment, achieving average gains of 0.54 and 0.65 dB in PSNR on brain and abdominal data, respectively. Fig. 2 provides qualitative comparisons for several cases from the abdominal dataset. The outputs of SANet-BME^2 and DANCE*-BME^2 show improved reconstruction fidelity, with fewer artifacts and blurs compared to their original versions, proving promising potential in eliminating cross-scan misalignment and accelerating multi-scan imaging.

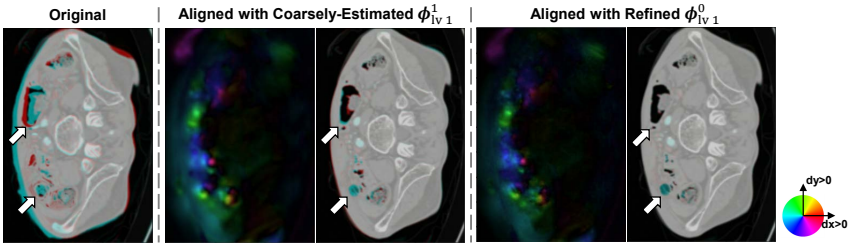

Fig. 3. Visualization of deformation fields estimated by BME^2. Using the alpha blending strategy, it can be observed that the initially estimated ϕ^1 mitigates the cross-scan misalignment, while the refined deformation fields ϕ^0 further eliminate it.

3.4 Ablation and Efficiency Study

Ablation Study. We conduct an ablation study to explore the effectiveness of the key components in BME^2. In Table 2, "w/o bri' indicates removal of the latent Schrödinger bridge and directly using the coarsely-estimated ϕ^1 for misalignment elimination, and "w/o csl" does not utilize the correlation and smoothness loss terms ($\mathcal{L}_{\text{Cor}}$ and $\mathcal{L}_{\text{Smo}}$) for model training. We choose DANCE as the backbone. The full DANCE*-BME^2 outperforms two ablated models. Interestingly, the performance of DANCE*-w/o bri is inferior to the original DANCE with the deformable attention mechanism, suggesting that the coarsely-estimated ϕ^1 from a lightweight ME network may not be accurate, demonstrating the importance of the Schrödinger bridge-based refinement. Besides, Fig. 3 visualizes the deformation fields via the alpha blending strategy. Compared with the coarsely-estimated deformation fields ϕ^1, the refined ϕ^0 possesses more abundant textures and aligns the $\boldsymbol{I}_{\text{A}}^{\text{Ref}}$ to $\boldsymbol{I}^{\text{GT}}$ with better accuracy, suggesting the effectiveness of latent Schrödinger bridge to refine the estimated misalignment.

Efficiency Study. Table 3 compares the efficiency for models with and without the proposed BME^2, providing the floating-point operations (FLOPs) on each 512×512 CT image. Thanks to the lightweight ME network and the relatively small number of iteration sampling steps on the Schrödinger bridge, SANet-BME^2 only introduces 19.1% additional computational costs compared to the SANet backbone. Moreover, DANCE*-BME^2 has fewer FLOPs than the original DANCE with deformable attention mechanisms while delivering better performance, suggesting the success of our proposed BME^2 to prove an effective balance between performance and efficiency.

4 Conclusion

In this paper, we propose a play-and-plug latent bridge-based module misalignment estimation and elimination (BME^2), which first coarsely estimates the cross-scan misalignment by a lightweight ME network, and then refines it by

recursively performing "prediction and posterior sampling". We conduct experiments on brain MRI and abdominal CT datasets with rigid and non-rigid misalignment to validate our method. Experimental results show that BME^2 consistently improves the performance of four involved backbones while only leading to mild computation costs, proving good potential for clinical settings.
Limitations. Our experiments have not covered the long-term temporal variations, such as tumor position and size changes across scans at different time points. Such applications will be explored in our future work.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers 82441023, U23A20295, 62131015, 82394432), the China Ministry of Science and Technology (S20240085, STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), Shanghai Municipal Central Guided Local Science and Technology Development Fund (No. YDZX20233100001001), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), and HPC Platform of ShanghaiTech University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chang, A.E., Matory, Y.L., Dwyer, A.J., Hill, S.C., Girton, M.E., Steinberg, S.M., Knop, R.H., Frank, J.A., Hyams, D., Doppman, J.L.: Magnetic resonance imaging versus computed tomography in the evaluation of soft tissue tumors of the extremities. Annals of surgery **205**(4), 340 (1987)
2. Chen, W., Wu, S., Wang, S., Li, Z., Yang, J., Yao, H., Tian, Q., Song, X.: Multi-contrast image super-resolution with deformable attention and neighborhood-based feature aggregation (DANCE): applications in anatomic and metabolic MRI. Medical Image Analysis **99**, 103359 (2025)
3. Chen, Z., Zhang, Y., Liu, D., Xia, B., Gu, J., Kong, L., Yuan, X.: Hierarchical integration diffusion model for realistic image deblurring. In: NeurIPS (2023)
4. Fan, J., Cao, X., Wang, Q., Yap, P.T., Shen, D.: Adversarial learning for mono-or multi-modal registration. Medical Image Analysis **58**, 101545 (2019)
5. Feng, C.M., Yan, Y.L., Chen, G., Xu, Y., Hu, Y., Shao, L., Fu, H.Z.: Multimodal transformer for accelerated MR imaging. IEEE Transactions on Medical Imaging **42**(10), 2804–2816 (2023)
6. Feng, C.M., Fu, H., Yuan, S., Xu, Y.: Multi-contrast MRI super-resolution via a multi-stage integration network. In: MICCAI (2021)
7. Feng, C., Yan, Y., Yu, K., Xu, Y., Fu, H., Yang, J., Shao, L.: Exploring separable attention for multi-contrast MR image super-resolution. IEEE Transactions on Neural Networks and Learning System **35**(9), 12251–12262 (2024)
8. Feng, Y., Deng, S., Lyuand, J., Cai, J., Wei, M., Qin, J.: Bridging MRI cross-modality synthesis and multi-contrast super-resolution by fine-grained difference learning. IEEE Transactions on Medical Imaging **44**, 373–383 (2025)

9. Gousias, I.S., Edwards, A.D., Rutherford, M.A., Counsell, S.J., Hajnal, J.V., Rueckert, D., Hammers, A.: Magnetic resonance imaging of the newborn brain: manual segmentation of labeled atlases in term-born and preterm infants. Neuroimage **62**(3), 1499–1509 (2012)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)
11. Kim, B., Han, I., Ye, J.C.: Diffusemorph: Unsupervised deformable image registration using diffusion model. In: ECCV. vol. 13691, pp. 347–364 (2022)
12. Kim, B., Zhuang, Y., Mathai, T.S., Summers, R.M.: Otmorph: Unsupervised multi-domain abdominal medical image registration using neural optimal transport. IEEE Transactions on Medical Imaging **44**(1), 165–179 (2025)
13. Lei, P., Fang, F., Zhang, G., Zeng, T.: Decomposition-based variational network for multi-contrast MRI super-resolution and reconstruction. In: ICCV (2023)
14. Li, G., Lv, J., Tian, Y., Dou, Q., Wang, C., Xu, C., Qin, J.: Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast MRI super-resolution. In: CVPR (2022)
15. Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., Chen, Y.: SRDiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing **479**, 47–59 (2022)
16. Liang, J., Cao, J., Sun, G., Zhang, K., Gool, L.V., Timofte, R.: SwinIR: Image restoration using swin transformer. In: ICCV (2021)
17. Liu, G., Vahdat, A., Huang, D., Theodorou, E.A., Nie, W., Anandkumar, A.: I^2SB: Image-to-image Schrödinger bridge. In: ICML. pp. 22042–22062 (2023)
18. Lu, L., Li, W., Tao, X., Lu, J., Jia, J.: MASA-SR: matching acceleration and spatial adaptation for reference-based image super-resolution. In: CVPR (2021)
19. Lyu, Q., Shan, H., Steber, C., Helis, C., Whitlow, C., Chan, M.D., Wang, G.: Multi-contrast super-resolution MRI through a progressive network. IEEE Transactions on Medical Imaging **39**(9), 2738–2749 (2020)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR. pp. 10674–10685 (2022)
21. Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Gool, L.V.: DiffIR: efficient diffusion model for image restoration. In: ICCV. pp. 13049–13059 (2023)
22. Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., Shen, D.: Deep-learning-based multi-modal fusion for fast MR reconstruction. IEEE Transactions on Biomedical Engineering **66**(7), 2105–2114 (2018)
23. Xu, W., Ha, S., Mueller, K.: Database-assisted low-dose CT image restoration. Medical Physics **40**(3), 031109 (2013)
24. Xu, W., Mueller, K.: Efficient low-dose CT artifact mitigation using an artifact-matched prior scan. Medical Physics **39**(8), 4748–4760 (2012)
25. Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning texture transformer network for image super-resolution. In: CVPR (2020)
26. Yip, E., Yun, J., Wachowicz, K., Gabos, Z., Rathee, S., Fallone, B.: Sliding window prior data assisted compressed sensing for MRI tracking of lung tumors. Medical Physics **44**(1), 84–98 (2017)
27. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)
28. Zheng, H., Qu, X., Bai, Z., Liu, Y., Guo, D., Dong, J., Peng, X., Chen, Z.: Multi-contrast brain magnetic resonance image super-resolution using the local weight similarity. BMC medical imaging **17**, 1–13 (2017)