MoCo-Diff: Adaptive Conditional Prior on Diffusion Network for MRI Motion Correction

Feng Li[1], Zijian Zhou[1], Yu Fang[1], Jiangdong Cai[1], and Qian Wang[1,2(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[2] Shanghai Clinical Research and Trial Center, Shanghai, China
qianwang@shanghaitech.edu.cn

Abstract. Magnetic Resonance Image (MRI) is a powerful medical imaging modality with non-ionizing radiation. However, due to its long scanning time, patient movement is prone to occur during acquisition. Severe motions can significantly degrade the image quality and make the images non-diagnostic. This paper introduces MoCo-Diff, a novel two-stage deep learning framework designed to correct the motion artifacts in 3D MRI volumes. In the first stage, we exploit a novel attention mechanism using shift window-based transformers in both the in-slice and through-slice directions to effectively remove the motion artifacts. In the second stage, the initially-corrected image serves as the prior for realistic MR image restoration. This stage incorporates the pre-trained Stable Diffusion to leverage its robust generative capability and the ControlUNet to fine-tune the diffusion model with the assistance of the prior. Moreover, we introduce an uncertainty predictor to assess the reliability of the motion-corrected images, which not only visually hints the motion correction errors but also enhances motion correction quality by trimming the prior with dynamic weights. Our experiments illustrate MoCo-Diff's superiority over state-of-the-art approaches in removing motion artifacts and retaining anatomical details across different levels of motion severity. The code is available at https://github.com/fengza/MoCo-Diff.

Keywords: Motion correction · Prior-conditioned diffusion model · Dual branch transformer · Magnetic resonance imaging

1 Introduction

Magnetic Resonance Imaging (MRI) is crucial for medical imaging and diagnosis. However, its long acquisition time causes motion-induced artifacts, degrading image quality and diagnostic efficacy. Various solutions have emerged to address the challenge of motion [26]. Among those, retrospective motion correction (MoCo) is being actively investigated because it does not complicate the scanning process and can be elegantly achieved with computational methods [12]. Furthermore, the utilization of deep learning approaches has shown promising results [5,8].

Because of the sequential acquisition of k-space data in MRI, motion artifacts can have strong spatial dependencies in the imaging volume. Recently, physics-based approaches combining deep learning with the MR imaging process were

proposed [3,7]. In such methods, parameters that quantify the subject motion during signal acquisition are estimated and used for the MoCo problem [2,19]. However, most of these studies were performed in 2D MR slices, and diverse scanning protocols and artifact patterns may prevent real usage [18].

Nonetheless, deep learning networks hold great potential in understanding the complex patterns of motion artifacts. Challenges can arise for CNN-based models, which may not effectively extract the through-slice features due to the misaligned neighbouring slices [20]. Transformer-based networks can be suitable for 3D MR MoCo because they can capture features with long-distance dependencies [21]. Inspired by the success of transformer architectures like Restormer [27] and Swin Transformer (SwinIR) [10], early efforts [21] have employed self-attention mechanisms to exploit long-distance spatial dependencies associated with the motion for artifact correction. However, most of these networks require volume registration for 3D MoCo. Moreover, due to prioritizing the minimization of Euclidean error between the corrupted and clean images, these networks risk generating blurred images with suboptimal perceptual quality.

In contrast, the pioneering diffusion-based models show exceptional performance in capturing complex data distributions to yield high-quality images. The well-trained Stable Diffusion [16] shifts the computation into a latent space and shows its efficacy across a range of applications, notably in natural image restoration [11]. A few studies [15,25] exploited its potential for MRI MoCo. However, this field remains largely unexplored. Specifically, unrealistic details may emerge in the restored image if conditional priors are not properly used.

To tackle the aforementioned challenges, we propose a two-stage pipeline, MoCo-Diff, which conditions diffusion model on adaptive prior, to advance the development of 3D MRI MoCo. **First, we endeavour to simultaneously improve the synthesis fidelity and perception** within the MoCo domain. Specifically, we introduce a Dual Branch Transformer (DBT), which integrates a bi-directional through-slice transformer (T-Module) with an in-slice transformer (I-Module) to efficiently learn 3D motion features through a 2D computation framework. **Second, we present an adaptive prior strategy for the Diffusion model (AP-Diff)**, which controls each step of the generation process with the prior derived from the first stage. In this way, we effectively mitigate the inclusion of "fake" details in medical images. We validate the performance of MoCo-Diff in artifact removal and detail preservation using multiple datasets with simulated and real motion artifacts. We also evaluate the impact of the recovered tissue details on downstream segmentation, which helps gauge the quality of the motion-corrected images.

2 Method

Our proposed two-stage MoCo-Diff framework, depicted in Fig. 1, is tailored for robust and superior MRI MoCo performance, applicable in real-world scenarios. The first stage employs a Dual Branch Transformer (DBT) model to generate prior, ensuring restoration fidelity. In the second stage, a Stable Diffusion (SD)

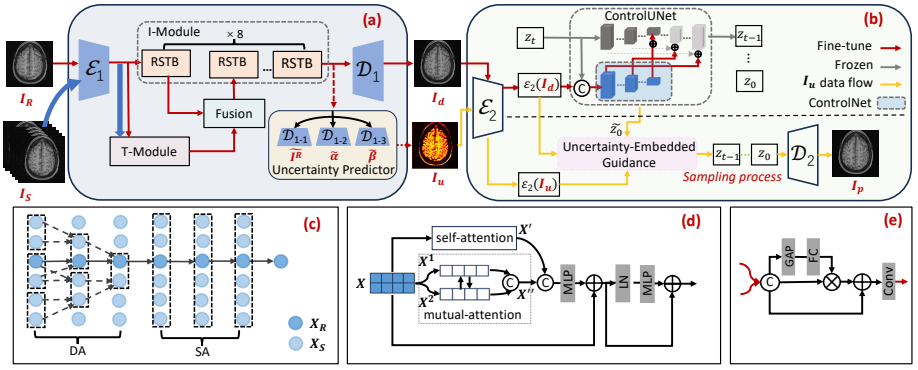

Fig. 1. Overview of our two-stage MoCo-Diff framework: (a) Dual Branch Transformer (DBT) for estimating target distribution prior; (b) Pre-trained Stable Diffusion model fine-tuned on the conditioned prior (AP-Diff); (c) T-Module for through-slice mutual attention; (d) Dual-slice attention operation of the DA block in T-Module; (e) Fusion block integrating features from both transformer branches.

model, guided by an adaptive prior strategy (AP-Diff), produces clean MR images conditioned on the prior.

2.1 Dual Branch Transformer (DBT)

Directly using motion-corrupted images as guidance in diffusion steps can hinder accurate target distribution capture, degrading restoration due to motion artifacts. To mitigate this, we introduce a Dual Branch Transformer (DBT, Fig. 1(a)) to convert motion-corrupted images (I_R) into motion-free ones (I_d), offering a conditional prior for controlling denoising steps in the next stage.

The DBT training involves using a motion-corrupted slice (I_R) and its adjacent slices (I_S) from a 3D motion-corrupted subject, with a motion-free ground-truth (I_{gt}) as the target. The network employs a pixel unshuffle operator [17] for downsampling features into the latent space (encoder $\mathcal{E}_1$). Eight Residual Swin Transformer Blocks (RSTB) from SwinIR [10] serve as an in-slice transformer (I-Module) and integrate with a bi-directional through-slice transformer (T-Module) to exploit dependencies in adjacent MR slices for 3D MoCo. Additionally, a Fusion block (Fig. 1(e)) re-weights features from the two attention modules. The decoder $\mathcal{D}_1$ produces the final prediction (I_d) by mapping features to the original image space. Parameters are optimized using the L_2 pixel loss:

$$\mathcal{L}_{\mathrm{d}} = \left\| I_d - I_{gt} \right\|_2^2. \tag{1}$$

Through-Slice Transformer (T-Module) This module (Fig. 1(c)) adaptively integrates neighbouring slice features X_S with the reference slice features X_R, akin to implicit motion estimation and feature warping. It comprises a

stacked dual-slice attention (DA) block for cross-pair mutual attention calculation and a stacked self-attention (SA) block to seek complementary sharp information from adjacent and reference slice features.

For the mutual attention in the DA block (Fig. 1(d)), it is calculated on every slice pair $X \in \mathbb{R}^{LM^2 \times C}$, where LM^2 is the window size and C is the channel number. $X^1, X^2 \in \mathbb{R}^{\frac{LM^2}{2} \times C}$ are splited features from X. One branch calculates the self-attention of the pair, while another branch calculates the mutual attention, during which X^1 and X^2 are warped. The process can be formulated as follows:

$$\begin{aligned} MA(X^1, X^2) &= Softmax(Q^1 (K^2)^T / \sqrt{D}) V^2, \\ X'' &= Concat(MA(X^1, X^2), MA(X^2, X^1)), \end{aligned} \tag{2}$$

where $Q^1 = X^1 P^Q, K^2 = X^2 P^K, V^2 = X^2 P^V$ by linear projectors. The resulting bi-directional warped feature X'' is concatenated with X' and passed through a multi-layer perceptron (MLP) for dimension reduction.

For multiple adjacent slices, the window shifts slice-wise by $\lfloor \frac{L}{2} \rfloor * (i\%2)$ slices in layer i for cross-pair connections and complexity reduction. The receptive field size increases to 6 slices when stacking only three layers ($L = 2$). To understand through-slice spatial dependencies, the SA block comprises three self-attention operations with a large window size ($L = 6$) [9]. For improved MoCo on the reference slice, we use the reference slice features X_R as query and the neighbouring slice features X_S as key and value in the last self-attention operation.

2.2 Adaptive Prior-Conditioned Diffusion (AP-Diff)

In severe cases of corruption, stage one removes most artifacts but may lead to texture loss and over-smoothness, resulting in poor MR image quality. To tackle this, stage two introduces AP-Diff, conditioned on the estimated target distribution, ensuring realistic and high-quality images amidst extensive 3D motion artifacts, while avoiding fake details. Assessing the model's confidence in the prior is crucial, achieved through an adaptive strategy measured by the difference between estimated and target distributions, aided by the uncertainty predictor. Stage two combines a pre-trained Stable Diffusion (SD) model [16] with a trainable ControlNet [28] integrated into the Unet architecture as an additional encoder branch, termed ControlUNet.

ControlUNet The input to the ControlNet consists of the concatenation of noisy latent z_t and condition latent $\mathcal{E}_2(I_d)$, which is mapped from I_d by the encoder $\mathcal{E}_2$ of a well-trained VAE within SD. The outputs of the ControlNet are then combined with the original Unet decoder, where the prompt condition c is left empty. The detailed structure is shown in Fig. 1. In the diffusion process, Gaussian noise with variance $\beta_t \in (0, 1)$ at time t is added to the encoded latent $z = \mathcal{E}_2(I_{gt})$ for producing the noisy latent. To fine-tune the denoising

ControlUNet ϵ_θ, we adopt the simplified objective [11] as:

$$\begin{aligned} z_t &= \sqrt{\bar{\alpha}_t} z + \sqrt{1-\bar{\alpha}_t}\epsilon, \\ \mathcal{L}_{Diff} &= \mathbb{E}_{z_t,c,t,\epsilon,\mathcal{E}_2(I_d)}\left[\|\epsilon - \epsilon_\theta\left(z_t, c, t, \mathcal{E}_2(I_d)\right)\|_2^2\right], \end{aligned} \tag{3}$$

where ϵ is sampled from a standard Gaussian distribution, $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

Adaptive Prior Strategy In the sampling process, our ControlUnet estimates intermediate variable $\tilde{z}_0$ from the noise z_t under the guidance $\mathcal{E}_2(I_d)$ in the latent space as follows:

$$\tilde{z}_0 = \frac{z_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1-\bar{\alpha}_t}\epsilon_\theta\left(z_t, c, t, \mathcal{E}_2\left(I_d\right)\right)}{\sqrt{\bar{\alpha}_t}}. \tag{4}$$

To control image consistency and reduce the wrong details in the restoration process, the uncertainty-embedded guidance is defined as follows:

$$\begin{aligned} \mathcal{L}_{AP}\left(\tilde{z}_0, \mathcal{E}_2\left(I_d\right), \mathcal{E}_2(I_u)\right) &= \sum_i \frac{1}{C_i \times H_i \times W_i} \|(\mathbf{I} - \mathcal{E}_2(I_u)\left(\tilde{z}_0 - \mathcal{E}_2\left(I_d\right)\right)\|_2^2, \\ z_{t-1} &\in \mathcal{N}\left(\mu_\theta\left(z_t\right) - s\nabla_{\tilde{z}_0}\mathcal{L}_{AP}, \sigma_t^2\right). \end{aligned} \tag{5}$$

The gradient scale s [11] introduces personal preferences to corrected images. However, it's more rational to guide this process by the confidence probability of the conditional prior I_d. To ensure reliable guidance, we introduce the Uncertainty Predictor (Fig. 1(a)) at the end of stage one to quantify I_d's uncertainty. We adopt BayesCap [22] to generate pixel-wise uncertainty maps (I_u) using our trained DBT model. The Uncertainty Predictor mimics the DBT's network structure but replaces the Decoder module with three copies to produce $\tilde{I}_d$, $\tilde{\alpha}$, and $\tilde{\beta}$. Full loss formulation $\mathcal{L}_{\phi^*}$ for the Uncertainty Predictor and inference of I_u are provided in the Supplementary Materials. Ultimately, with this iterative guidance, the final clean MR image I_p is obtained using the VAE decoder $\mathcal{E}_2$ within SD.

3 Experiments

This study employs T1-weighted MR images from the Human Connectome Project (HCP, 314 subjects) [23], augmented with simulated head motions for model development. Extracting 90 axial slices from each volume yields 22,590 training and 5,670 validation slices. External validation includes datasets from UNC/UMN Baby Connectome Project (BCP) [4], MR-ART [14], and our in-house data. In-house T1-weighted images have parameters: repetition/echo time 6.5/2.1 ms, slice thickness/interlayer gap 0.8/0.4 mm, 240 slices, and field of view 256 × 224 mm.

The MoCo-Diff input resolution is 512×512. DBT model training comprises 30 batches for 2.5k iterations [11]. AP-Diff uses Stable Diffusion 2.1-base [16] as

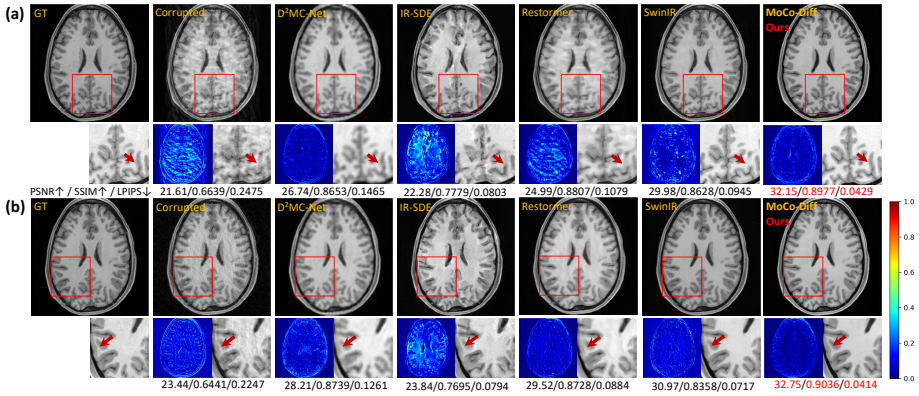

Fig. 2. Comparison of qualitative results obtained by different methods under Gaussian motion trajectory: (a) motion severity of 40%; (b) motion severity of 20%.

the prior, fine-tuning only the ControlNet with 20 batches for 5k iterations. The training involves 1000 diffusion steps, inference uses 50 steps. The evaluation focused on the absolute error map, SSIM, PSNR, and LPIPS metrics [29]. Both models are trained on four NVIDIA A100 GPUs with 80GB memory, using Adam optimizer [6] and initial learning rate 10^{-4} in PyTorch.

Head Motion Simulation We follow the method from [1] to mimic real-world MRI motion artifacts. Artifacts are simulated in k-space via 3D translations and rotations, with parameters chosen randomly from a Gaussian distribution $\mathcal{N}(0, 10)$. Motion severity varies randomly from 0-40%, representing different levels. We use three motion trajectories: piecewise constant/transient and Gaussian, to represent different head motions. Our model's performance is evaluated using 40%, 30%, and 20% subgroups of each trajectory.

Comparisions with State-of-the-Arts Compared to state-of-the-art MoCo models, including D^{2}MC-Net [24], IR-SDE [13], Restormer [27], and SwinIR [10], Fig. 2 and Table 1 show that our proposed MoCo-Diff exhibits superior performances in both image fidelity and perception, even under various levels of motion corruption. The motion-corrupted image without correction is labelled "Corrupted" and the motion-free image is labelled "GT". The SOTA physics-based D^{2}MC-Net, trained on complex images with the simulated phase of the HCP dataset, focuses on reducing pixel-level disparities but tends to produce excessively smoothed images. IR-SDE, a diffusion-based approach, faces challenges in capturing structural distribution due to limited training data and its mean-reverting design. Although Transformer-based Restormer and SwinIR outperform other methods in SSIM, they do not reach a satisfactory level. Our method outperforms others across all evaluation metrics, especially in SSIM and LPIPS,

Table 1. Quantitative comparison of different methods on the HCP dataset under Gaussian motion trajectory, spanning severe to mild motion severities, in terms of PSNR (dB), SSIM, and LPIPS.

Corrupted Phase Lines	40%			30%			20%		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Corrupted	21.81	0.5877	0.2934	22.37	0.6258	0.2456	23.34	0.6741	0.1817
D^2MC-Net	25.23	0.8018	0.1267	26.11	0.8189	0.1651	27.67	0.8584	0.1275
Restormer	24.99	0.8206	0.1686	26.02	0.8468	0.1431	27.37	0.8733	0.1206
IR-SDE	22.76	0.7331	0.1462	23.74	0.7448	0.1296	24.29	0.7607	0.1005
SwinIR	27.61	0.8125	0.1211	28.49	0.8321	0.1173	29.67	0.8548	0.1087
MoCo-Diff	**29.02**	**0.8741**	**0.0947**	**29.74**	**0.8909**	**0.0831**	**30.64**	**0.9089**	**0.0741**

Table 2. Quantitative ablation study of the key components: (a) I-Module; (b) T-Module; (c) motion-corrupted distribution (I_R) as prior; (d) estimated target distribution (I_d) as prior; (e) uncertainty-embedded guidance.

DBT		AP-Diff			PSNR↑	SSIM↑	LPIPS↓
(a)	(b)	(c)	(d)	(e)			
✓					29.37(+6.38)	0.8445(+0.1889)	0.1117(-0.0958)
✓	✓				29.89(+6.90)	0.8945(+0.2389)	0.1161(-0.0914)
		✓			24.51(+1.52)	0.7622(+0.1066)	0.1598(-0.0477)
✓			✓		28.32(+5.33)	0.8488(+0.1932)	0.0982(-0.1093)
✓	✓		✓		29.86(+6.87)	0.8895(+0.2339)	0.0858(-0.1217)
✓	✓		✓	✓	**30.38**(+**7.20**)	**0.9001**(+**0.2445**)	**0.0795**(-**0.1280**)

with improvements of 0.0535 and 0.0264 compared to the best alternative at a motion severity of 40%. Results for other motion trajectories can be found in Supplementary Materials (Table S1).

Ablation study We conduct ablation studies on MoCo-Diff's components, revealing promising findings in Table 2. Incorporating the T-Module into the I-Module significantly enhances SSIM scores, indicating improved anatomical structure restoration through dual-branch attentions focusing on 3D motion features. Integrating the SD model improves perception despite minor declines in objective metrics. A similar study [11] shows promising results using the same SD model framework conditioned solely on a prior that is processed by I-Module. With the addition of uncertainty-embedded guidance (MoCo-Diff), substantial improvements are observed across all metrics, underlining each component's importance and their role in enhancing image quality and subjective perception.

Segmentation Evaluation We apply FMRIB's Automated Segmentation Tool (FAST) [30] to obtain segmentations at 40% motion severity under a Gaussian motion trajectory with GT as reference. Table 3 highlights MoCo-Diff's superior

Table 3. Quantitative comparison of segmentation results on corrected MR images using different methods. The unit of DSC is percentage (%).

DSC	Corrupted	D^{2}MC-Net	IR-SDE	Restormer	SwinIR	**MoCo-Diff**
GM	66.16±4.64	81.95±2.48	62.93 ± 5.28	76.82±3.76	80.84±3.52	**83.56 ± 2.58**
WM	81.76±2.93	90.19±1.32	81.32 ± 3.06	87.72±2.12	90.35±1.58	**91.06 ± 1.36**
Avg	73.96±8.71	86.07±4.57	72.12±10.15	82.27±6.24	85.60±5.48	**87.31 ± 4.28**

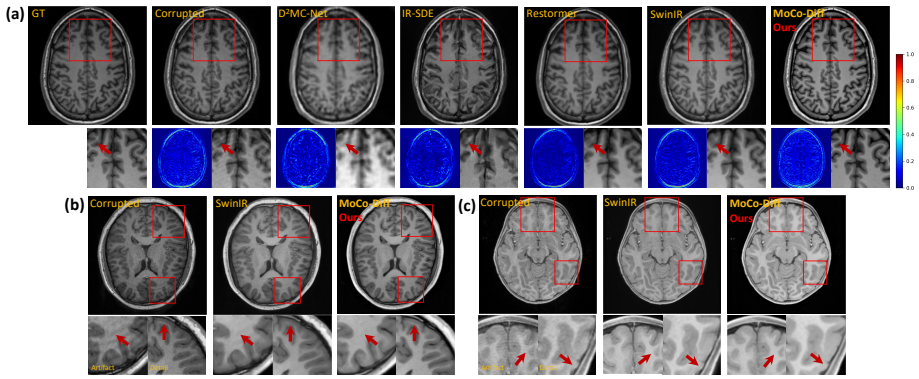

Fig. 3. Qualitative results of different methods under real motion artifacts on three external validation sets: (a) MR-ART; (b) In-house; (c) BCP. We selectively show the top two methods only in (b) and (c) for easy comparison.

segmentation performance compared to other methods, demonstrating its effectiveness in restoring anatomical structures. Visualization results are provided in Fig. S1.

Robustness Effectiveness We validate our algorithm on three external datasets with real motion artifacts. Note that the model here comes from the previous experiment, without new training or fine-tuning. Fig. 3 and Fig. S2 demonstrate its effectiveness in removing these artifacts from MR images, showcasing its clinical potential. In comparison, D^{2}MC-Net struggles with blurred reconstruction, while SwinIR and Restormer tend to lose some details and fail to fully remove artifacts. Our approach excels in robustness and effectiveness for artifact removal and detail preservation. Additional results are available in Fig. S2.

4 Conclusion

In conclusion, our proposed MoCo-Diff can achieve excellent motion artifact correction in 3D MR volume. It can also preserve the anatomic details without introducing fake structures. To our knowledge, MoCo-Diff represents the first model capable of providing pixel-wise uncertainty for the motion-corrected MR

images, ensuring their reliability and can be potentially used for clinical applications. Besides high performance on diverse motion types, our model, like other diffusion-based models, should also address acceleration and lightweight challenges.

Acknowledgments. This work was partially supported by STI 2030-Major Projects (2021ZD0200514) and National Natural Science Foundation of China (62131015).

Disclosure of Interests. The authors declare no competing interests.

References

1. Duffy, B.A., Zhao, L., Sepehrband, F., Min, J., Wang, D.J., Shi, Y., Toga, A.W., Kim, H., Initiative, A.D.N., et al.: Retrospective motion artifact correction of structural mri images using deep learning improves the quality of cortical surface reconstructions. NeuroImage **230**, 117756 (2021)
2. Feinler, M.S., Hahn, B.N.: Retrospective motion correction in gradient echo mri by explicit motion estimation using deep cnns (2023), arXiv preprint arXiv:2303.17239
3. Hossbach, J., Splitthoff, D.N., Cauley, S., Clifford, B., Polak, D., Lo, W.C., Meyer, H., Maier, A.: Deep learning-based motion quantification from k-space for fast model-based magnetic resonance imaging motion correction. Medical Physics **50**(4), 2148–2161 (2023)
4. Howell, B.R., Styner, M.A., Gao, W., Yap, P.T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., et al.: The unc/umn baby connectome project (bcp): An overview of the study design and protocol development. NeuroImage **185**, 891–905 (2019)
5. Johnson, P.M., Drangova, M.: Conditional generative adversarial network for 3d rigid-body motion correction in mri. Magnetic Resonance in Medicine **82**(3), 901–910 (2019)
6. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014), arXiv preprint arXiv:1412.6980
7. Kuzmina, E., Razumov, A., Rogov, O.Y., Adalsteinsson, E., White, J., Dylov, D.V.: Autofocusing+: Noise-resilient motion correction in magnetic resonance imaging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 365–375 (2022)
8. Küstner, T., Armanious, K., Yang, J., Yang, B., Schick, F., Gatidis, S.: Retrospective correction of motion-affected mr images using deep learning frameworks. Magnetic Resonance in Medicine **82**(4), 1527–1540 (2019)
9. Liang, J., Cao, J., Fan, Y., Zhang, K., Ranjan, R., Li, Y., Timofte, R., Van Gool, L.: Vrt: A video restoration transformer (2022), arXiv preprint arXiv:2201.12288
10. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1833–1844 (2021)
11. Lin, X., He, J., Chen, Z., Lyu, Z., Fei, B., Dai, B., Ouyang, W., Qiao, Y., Dong, C.: Diffbir: Towards blind image restoration with generative diffusion prior (2023), arXiv preprint arXiv:2308.15070
12. Loktyushin, A., Nickisch, H., Pohmann, R., Schölkopf, B.: Blind retrospective motion correction of mr images. Magnetic Resonance in Medicine **70**(6), 1608–1618 (2013)

13. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Image restoration with mean-reverting stochastic differential equations (2023), arXiv preprint arXiv:2301.11699
14. Nárai, Á., Hermann, P., Auer, T., Kemenczky, P., Szalma, J., Homolya, I., Somogyi, E., Vakli, P., Weiss, B., Vidnyánszky, Z.: Movement-related artefacts (mr-art) dataset of matched motion-corrupted and clean structural mri brain scans. Scientific Data **9**(1), 630 (2022)
15. Oh, G., Jung, S., Lee, J.E., Ye, J.C.: Annealed score-based diffusion model for mr motion artifact reduction. IEEE Transactions on Computational Imaging (2023)
16. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R.: Sdxl: Improving latent diffusion models for high-resolution image synthesis (2023), arXiv preprint arXiv:2307.01952
17. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1874–1883 (2016)
18. Singh, N.M., Dey, N., Hoffmann, M., Fischl, B., Adalsteinsson, E., Frost, R., Dalca, A.V., Golland, P.: Data consistent deep rigid mri motion correction. In: Medical Imaging with Deep Learning. pp. 368–381 (2024)
19. Singh, N.M., Iglesias, J.E., Adalsteinsson, E., Dalca, A.V., Golland, P.: Joint frequency and image space learning for mri reconstruction and analysis. The Journal of Machine Learning for Biomedical Imaging **2022** (2022)
20. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8934–8943 (2018)
21. Tsai, T.H., Lin, Y.H., Lin, T.H.: Motion artifact correction in mri using gan-based channel attention transformer. In: IEEE Biomedical Circuits and Systems Conference. pp. 1–5 (2023)
22. Upadhyay, U., Karthik, S., Chen, Y., Mancini, M., Akata, Z.: Bayescap: Bayesian identity cap for calibrated uncertainty in frozen neural networks. In: European Conference on Computer Vision. pp. 299–317 (2022)
23. Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., et al.: The human connectome project: a data acquisition perspective. NeuroImage **62**(4), 2222–2231 (2012)
24. Wang, J., Yang, Y., Yang, Y., Sun, J.: Dual domain motion artifacts correction for mr imaging under guidance of k-space uncertainty. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 293–302 (2023)
25. Xie, Y., Li, Q.: Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 655–664 (2022)
26. Zaitsev, M., Maclaren, J., Herbst, M.: Motion artifacts in mri: A complex problem with many partial solutions. Journal of Magnetic Resonance Imaging **42**(4), 887–901 (2015)
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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5728–5739 (2022)

28. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847 (2023)
29. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)
30. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging **20**(1), 45–57 (2001)