

Tackling Hallucination from Conditional Models for Medical Image Reconstruction with DynamicDPS

Seunghoi Kim^{*1,2}, Henry F. J. Tregidgo^{*1,2}, Matteo Figini^{1,3}, Chen Jin⁵,
Sarang Joshi⁴, Daniel C. Alexander^{1,3}

¹ Centre for Medical Image Computing, University College London,

² Department of Medical Physics and Biomedical Engineering, University College London,

³ Department of Computer Science, University College London,

⁴ The University of Utah, USA,

⁵ Centre for AI, DS&AI, AstraZeneca, UK

Abstract. Hallucinations are spurious structures not present in the ground truth, posing a critical challenge in medical image reconstruction, especially for data-driven conditional models. We hypothesize that combining an unconditional diffusion model with data consistency, trained on a diverse dataset, can reduce these hallucinations. Based on this, we propose DynamicDPS, a diffusion-based framework that integrates conditional and unconditional diffusion models to enhance low-quality medical images while systematically reducing hallucinations. Our approach first generates an initial reconstruction using a conditional model, then refines it with an adaptive diffusion-based inverse problem solver. DynamicDPS skips early stage in the reverse process by selecting an optimal starting time point per sample and applies Wolfe’s line search for adaptive step sizes, improving both efficiency and image fidelity. Using diffusion priors and data consistency, our method effectively reduces hallucinations from any conditional model output. We validate its effectiveness in Image Quality Transfer for low-field MRI enhancement. Extensive evaluations on synthetic and real MR scans, including a downstream task for tissue volume estimation, show that DynamicDPS reduces hallucinations, improving relative volume estimation by over 15% for critical tissues while using only 5% of the sampling steps required by baseline diffusion models. As a model-agnostic and fine-tuning-free approach, DynamicDPS offers a robust solution for hallucination reduction in medical imaging. The code will be made publicly available upon publication.

Keywords: Diffusion models · Image enhancement · Out-of-distribution generalization

^{*} These authors contributed equally to this work

Corresponding author E-mail: seunghoi.kim.17@ucl.ac.uk

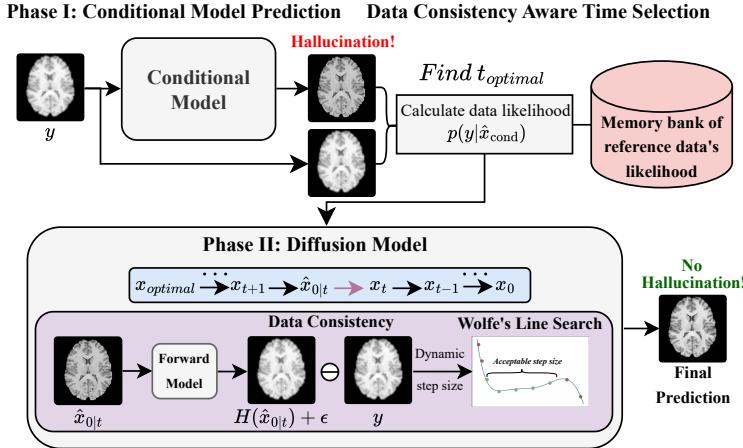


Fig. 1: The overview of our proposed framework. The framework first predicts a high-quality image using a conditional model. Then, it calculates optimal starting time point in the reverse process via Data Consistency Aware Time Selection (DCATS). At last, the prediction is refined using DynamicDPS that adaptively adjusts step size in each time step using Wolfe’s line search.

1 Introduction

A key impediment to the adoption of widespread generative models for the reconstruction and enhancement of Magnetic Resonance Imaging (MRI), is the possibility of *hallucinations*. This particular class of reconstruction error can produce spurious, and yet realistic-looking structures that are not present in the imaging target. Such errors can be detrimental to the accuracy of downstream analysis, preventing clinical use in health critical settings. Hence, we present a novel framework for directly addressing the hallucinations from conditional models, while aiming to minimize additional training and inference requirements. While our method could be applied more broadly, we examine the explicit case on low-field MRI enhancement using Image Quality Transfer (IQT).

IQT [2] is a machine learning framework combining aspects of super-resolution, de-noising and contrast enhancement, enhancing visible tissue properties in low-quality images to the equivalent high-quality counterpart. IQT has been applied using various approaches, ranging from classical regression and random forests to deep learning architectures [16,21], contrast-agnostic models [12,10], attention-based CNNs [15], and 3D conditional diffusion-based methods [14] that can also work robustly under heavily under-sampled MRI. This contrasts to super-resolution work [9,24,30,8,4,23] which focus more directly on the resolution problem only.

Recent work [13,31,1,29] has shown that hallucinations typically arise in image regions that are *out-of-distribution (OOD)* for the conditional model being applied. To date relatively little work has appeared on quantifying hallucinations in medical imaging [3,22], and efforts to minimize hallucinations from conditional generative models have been limited to handling OOD regions separately to *in distribution (IND)* regions [13]. However, we hypothesize that unconditional diffusion models for inverse problems (or score-matching networks) [6,7,19,26,33,5] may provide a framework for hallucination reduction due to the combination of prior domain knowledge and data consistency optimization.

One example, Diffusion Posterior Sampling (DPS) [6], adopts a Bayesian framework, using the score-matching network as strong domain prior and likelihood estimation for data consistency in the reverse process to provide feasible solutions for given measurements. However, this approach typically requires a large number of sampling steps at inference time (e.g. T=1000). Similarly, Lin et al. [17] extended DPS to low-field MRI enhancement, but incorporated test-time model-parameter optimization, increasing inference cost.

In this work, we propose *DynamicDPS*, a novel framework for MRI reconstruction that extends DPS [6] to mitigate hallucinations produced by pre-trained conditional models. Figure 1 illustrates an overview of our method. First, a conditional model generates a high-field-like prediction that serves as a prior. DCATS then determines the optimal intermediate starting point for the reverse process to reduce the number of diffusion steps. Finally, the prediction is iteratively refined using score-matching and data consistency optimization. We dynamically optimize data-consistency step size to reduce hallucinations inconsistent with the low-field measurements and accelerate sampling by skipping unnecessary diffusion steps. Our approach is model-agnostic and seamlessly integrates with existing conditional models. We demonstrate the effectiveness of DynamicDPS for low-field MR enhancement using IQT, showing that it not only reduces hallucinations but also achieves more than 80% faster inference compared to previous diffusion-based approaches. To the best of our knowledge, this is the first work to tackle hallucinations in MRI reconstruction explicitly.

2 Method

We propose DynamicDPS, a novel framework to mitigate hallucinations. This section provides a detailed overview of our approach and how it addresses the limitations. We first formulate the problem of IQT and hallucinations, followed by our hypothesis and theoretical justification for tackling these challenges. Finally, we present a detailed breakdown of each component within our framework.

2.1 Problem Formulation

Image Quality Transfer (IQT) [2] aims to learn a mapping function between low-field (LF) images $\mathbf{y} \in \mathbb{R}^M$ and their high-field (HF) counterparts $\mathbf{x} \in \mathbb{R}^N$.

However, due to the need of paired data training, most networks [2,16,14] simulate LF images using a degradation model and train in a self-supervised manner. Similarly, we approximate the degradation to generate LF counterparts as:

$$\mathbf{y} = \mathbf{Ax} + \mathbf{n}, \quad \text{where } \mathbf{A} = \text{Blur}(DS_k(\Gamma_\gamma(\mathbf{x}))), \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ represents the degradation operator, which is generally non-invertible, Blur , DS_k , Γ_γ denote Gaussian blur kernel, downsampling with a factor of k and gamma transform with the coefficient γ , respectively, and \mathbf{n} denotes measurement noise. To approximate the inverse of \mathbf{A} , a deep neural network f_θ is trained to reconstruct the HF image as $\hat{\mathbf{x}} = f_\theta(\mathbf{y})$.

However, the ill-posed nature of this inverse problem can induce *hallucinations*, spurious structures not present in the ground truth \mathbf{x}_{true} . To assist with formulation of our hypothesis, we categorize hallucinations into either *Intrinsic* or *Extrinsic* [3,27], characterized by:

$$\mathbf{A}\hat{\mathbf{x}} \neq \mathbf{Ax}_{\text{true}}, \quad (\mathbf{I} - \mathbf{A}^+ \mathbf{A})\hat{\mathbf{x}} \neq (\mathbf{I} - \mathbf{A}^+ \mathbf{A})\mathbf{x}_{\text{true}}, \quad (2)$$

The first term represents intrinsic hallucinations, where the reconstructed image $\hat{\mathbf{x}}$ violates data consistency, meaning its projection onto the measurement space differs from $\mathbf{Ax}_{\text{true}}$. The second term captures extrinsic hallucinations, where errors in the null space of \mathbf{A} introduce structures that do not exist in \mathbf{x}_{true} but appear in the reconstruction. These hallucinations may obscure or mimic clinically significant details, potentially leading to misdiagnosis.

2.2 Hypothesis and Justification

We hypothesize that combining a diffusion model, trained on a diverse dataset of HF MR images, with a conditional model can mitigate both intrinsic and extrinsic hallucinations. While conditional models excel at generating HF-like predictions from LF inputs, they are prone to hallucinations, particularly in out-of-distribution scenarios. In contrast, diffusion models provide a robust prior by learning the full distribution of plausible HF MR images rather than explicitly mapping LF to HF. This broader representation, combined with data consistency enforcement, helps to suppress errors introduced by conditional models. Below, we provide a theoretical justification of how this approach reduces hallucinations.

During the reverse diffusion process, DynamicDPS reduces hallucinations from the conditional model by leveraging:

- **Data Prior:** The diffusion model learns from a broad HF MRI distribution, mitigating *extrinsic* hallucinations from incomplete measurements.
- **Data Consistency:** A correction term enforces alignment with y , mitigating *intrinsic* hallucinations.

As in [6], using the trained score function, the gradient of the posterior log-density at each time step is formulated as:

$$\nabla_{x_t} \log p_t(x_t|y) \simeq s_\theta(x_t, t) - \rho_t \nabla_{x_t} \|y - A(\hat{x}_0)\|_2^2, \quad (3)$$

where $s_\theta(x_t, t)$ is the learned score function and $\rho_t > 0$ is the step size.

Hence, extrinsic hallucinations are reduced by the diffusion model's prior, $s_\theta(x_t, t)$, while intrinsic ones are addressed through the data consistency term, $\nabla_{x_t} \|y - A(\hat{x}_0)\|_2^2$, systematically reducing both hallucinations in the conditional model's prediction.

2.3 Phase I: Conditional Model Prediction

In the first phase, any conditional IQT model can be used to predict a HF-like image \hat{x}_{cond} from the LF input y . While this output may still contain hallucinations, it serves as a strong initialization for the next step. Our framework is model-agnostic and can integrate with any conditional model.

DynamicDPS reduces sampling steps by starting the reverse process from an intermediate time point. To balance efficiency and fidelity, stronger hallucinations require to initiate at earlier (noisier) stages [13]. Incorrect time selection may push the conditional prediction outside the model's distribution, degrading performance. To address this, we introduce Data-Consistency-Aware Time Selection (DCATS), which optimizes the starting time point t_{optimal} per sample.

DCATS begins by creating a memory bank that, for each time step t , stores the average data likelihood computed over a reference dataset. Then, during testing, this reference data likelihood is compared to the scaled likelihood of the conditional model's prediction to select t_{optimal} by minimizing the discrepancy between the likelihoods of the two sets:

$$t_{\text{optimal}} = \arg \min_t \left| [p(y | \hat{x}_{\text{cond}})]^\tau - \mathbb{E}_{\text{ref},t} [p(y_{\text{ref}} | \hat{x}_{\text{ref},0|t})] \right|. \quad (4)$$

Here, $[p(y | \hat{x}_{\text{cond}})]^\tau$ represents the likelihood of the test measurement given the predicted image from conditional models, scaled with the temperature hyper-parameter, $\tau \leq 1$, and $p(y_{\text{ref}} | \hat{x}_{\text{ref},0|t})$ denotes the likelihood of a set of reference data at time step t . By selecting t_{optimal} to minimize this likelihood mismatch, the method accelerates sampling while maintaining image quality, achieving a significant improvement over standard DPS.

2.4 Phase II: Diffusion Model

In the second phase, pre-trained diffusion models on HF MR scans handle hallucinations by enforcing data consistency at each time step to reduce intrinsic hallucinations. Meanwhile, the diffusion prior corrects extrinsic hallucinations, ensuring that the reconstructed HF image remains faithful to the measurements.

After computing t_{optimal} , we use \hat{x}_{cond} from Phase I as a warm start for the diffusion process, significantly reducing sampling steps. Since most time steps are skipped, computational resources are focused on optimizing the data consistency term to suppress intrinsic hallucinations effectively.

Data Consistency Loss Function Vanilla DPS uses only an ℓ_2 penalty, $\|\mathbf{y} - A(\hat{\mathbf{x}}_\theta)\|_2^2$, which can be suboptimal for noisy or heavily undersampled inputs. To address this, we add two auxiliary terms, resulting in the total data consistency loss, L_{DC} as:

$$L_{DC} = \|\mathbf{y} - A(\hat{\mathbf{x}}_\theta)\|_2^2 + \lambda_1 \text{Edge}(\mathbf{y}, A(\hat{\mathbf{x}}_\theta)) + \lambda_2 \text{SSIM}(\mathbf{y}, A(\hat{\mathbf{x}}_\theta)), \quad (5)$$

where Edge and SSIM denote ℓ_2 -Sobel Edge and ℓ_2 -SSIM losses, respectively. These preserve anatomical boundaries and enhance local structural similarity.

Dynamic Step-Size Optimization via Wolfe’s Line Search A key limitation of data-consistency-based diffusion models is sensitivity to fixed step sizes, which can be sub-optimal for each test input to reduce intrinsic hallucination effectively under a limited number of iterations. In our reverse diffusion process, we adopt **Wolfe’s line search** [28] to dynamically select the step size α_t for the update of x_t and define $\phi(\alpha)$ as:

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t + \alpha_t \mathbf{p}_t, \quad \phi(\alpha) = f(\mathbf{x}_t + \alpha \mathbf{p}_t), \quad (6)$$

where \mathbf{p}_t is a direction (e.g., negative gradient of an objective) and $f(\cdot)$ includes both the data fidelity term and diffusion prior. Wolfe’s line search optimizes α_t satisfying 1. **Armijo rule**, $\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0)$, ensuring a non-trivial decrease in $f(\cdot)$, and 2. **Curvature condition**, $|\phi'(\alpha)| \leq c_2 |\phi'(0)|$, preventing overshooting. Here, $0 < c_1 < c_2 < 1$, and $\phi'(0) = \nabla f(\mathbf{x}_t)^\top \mathbf{p}_t$. These conditions can be interpreted as finding an upper and lower bounds of step size, which allows faster reduction of intrinsic hallucination without overshooting.

3 Experiments

3.1 Experimental Setup

Datasets We evaluate our method on the *Human Connectome Project (HCP)* dataset [20]. The simulated low-field (LF) test set (600 images) is split into: (1) **In-distribution**: LF images at $2.8mm$ resolution with $\gamma = 0.7$, matching training conditions; (2) **Out-of-distribution**: lower contrast ($\gamma = 0.4$) or lower spatial resolution ($4.2mm$). For real MRI, LF T1-weighted scans were acquired on a 0.36T MagSense 360 scanner with non-isotropic voxels ($1.0 \times 1.0 \times 7.2mm^3$, slice thickness: $6.0mm$, gap: $1.2mm$), and the corresponding HF image ($1.0 \times 1.0 \times 1.0mm^3$ isotropic) was registered with the LF for visual comparison.

Baselines & Metrics We evaluate our approach against four baselines: U-Net [18], ESRGAN [25], and DPS [6]. DPS and DynamicDPS use the same score-matching model pre-trained on the HCP dataset for a fair comparison. Methods requiring test-time parameter optimization are excluded. For quantitative evaluation, we use PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), and LPIPS [32] (Learned Perceptual Image Patch Similarity).

	In-distribution			Out-of-distribution			Inf. time (s)
	PSNR (↑)	SSIM (↑)	LPIPS (↓)	PSNR (↑)	SSIM (↑)	LPIPS (↓)	
U[18]	28±.97	.87±.03	.10±.02	24±1.9	.80±.06	.18±.05	-
G[25]	27±.02	.85±.71	.09±.01	17±5.6	.76±.04	.24±.03	-
DPS[6]	27±.92	.84±.03	.10±.02	26±1.9	.81±.06	.14±.04	196±0.6
Ours	29±1.0	.88±.03	.10±.02	27±1.9	.86±.06	.13±.03	-
Ours+U	29±1.0	.89±.03	.09±.02	26±1.9	.86±.06	.13±.04	37±0.8
Ours+G	29±.90	.89±.03	.09±.02	26±.06	.86±.06	.13±.03	38±0.8

Table 1: Quantitative comparison of image quality across in/out-of-distribution datasets, where an upward arrow indicates that a higher value is better. "Ours" refers to DynamicDPS with diffusion models only ($T=1000$). "U" and "G" denote U-Net [18] and ESRGAN [25], respectively.

3.2 Main Results

Table 1 shows a quantitative comparison against baselines. While conditional models perform similarly to diffusion-based methods on IND data, their performance degrades on OOD, underscoring their sensitivity to data shifts. In contrast, all of our approaches perform robustly on OOD and outperform baselines. Notably, *Ours with ESRGAN* on OOD data boosts PSNR and LPIPS by over 50%, demonstrating the ability to refine the output under a few time steps. Additionally, *Ours* outperforms DPS on both IND and OOD data, validating the effectiveness of our data consistency term.

Efficiency is assessed by comparing inference speed on IND data. Although DynamicDPS is computationally heavy due to the data consistency step, leveraging the conditional model's output as a prior speeds up inference by over 80% compared to DPS [6], using only 50 time steps with U-Net vs. 1000 for DPS.

Figure 2 presents a qualitative comparisons against baseline conditional models. As indicated by the red arrows, conditional models generate both intrinsic and extrinsic hallucinations such as false sulci and contrast artifacts. In contrast, DynamicDPS effectively reduces both types of hallucinations while preserving anatomical structures of tissues. These results demonstrate that our method not only enhances image quality but also visually reduces hallucinations across diverse conditional models.

3.3 Further Analysis

Figure 3a shows visual results tested on real LF and HF scans. While ESRGAN struggles to enhance contrast and fails on the HF scan, introducing extrinsic hallucinations, ours improves tissue contrast without visible hallucinations in both scans. Notably, even with *minimal tuning of the data consistency parameter*, our approach achieves superior visual contrast and image quality than the HF scan (second row), underscoring its robustness across varying degradation levels.

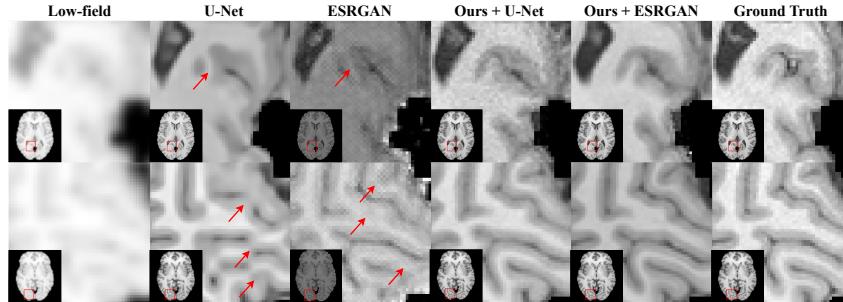
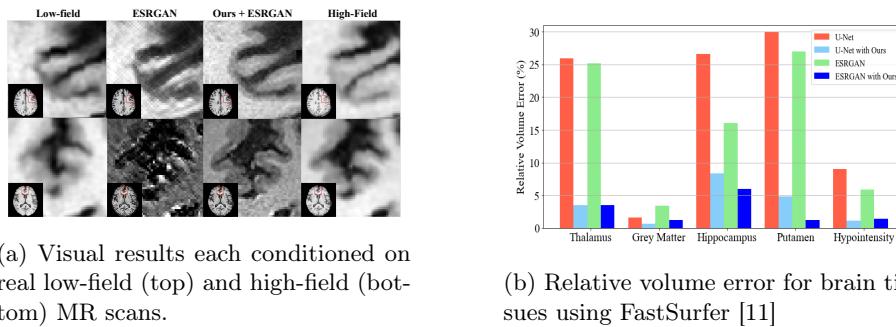


Fig. 2: Visual comparisons on OOD data. Zoomed-in regions are marked with red boxes, and red arrows indicate hallucinated areas.



(a) Visual results each conditioned on real low-field (top) and high-field (bottom) MR scans.

(b) Relative volume error for brain tissues using FastSurfer [11]

Fig. 3: Comparisons against baselines on (a) real low-field and high-field MR scans with zoomed-in regions are marked with red boxes, (b) volume estimation for hallucination evaluation.

Standard image quality metrics may not capture hallucinations effectively. Instead, we estimated brain tissue volumes using FastSurfer [11] and computed relative volume error as in [16]. Figure 3b shows that our method significantly reduces volume error by more than 15% in critical tissues such as the thalamus, hippocampus, putamen, and hypointensity across all conditional models. These structures are prone to hallucination due to their small, variable size and lower SNR compared to the cortex in some sequences, yet crucial for applications such as Alzheimer’s diagnosis. While hallucination metrics [3,22] were discussed earlier, we opted volume estimation to provide a clinically relevant assessment due to the unavailability of implementations. In future work, we plan to incorporate these established metrics into our evaluation framework.

4 Conclusion

We introduce DynamicDPS, a diffusion-based framework that tackles hallucinations by integrating conditional predictions with diffusion models. By selecting

an optimal starting time point via DCATS and initiating the reverse process with conditional outputs using wolfe’s line search-based data consistency, our method dynamically enforces data consistency to reduce hallucinations. Experimental results on both synthetic and real MR scans confirm its superior performance.

However, the approach depends on accurate estimation of data consistency parameters, and the uniform weighting may be sub-optimal when hallucinations occur locally. Future work will focus on automatic data consistency parameter estimation and spatially adaptive strategies. As a fine-tuning-free, model-agnostic solution, DynamicDPS has broad applicability in low-field MRI, advancing the democratization of high-quality medical imaging while enhancing safety.

References

1. Aithal, S.K., Maini, P., Lipton, Z.C., Kolter, J.Z.: Understanding hallucinations in diffusion models through mode interpolation. In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024)
2. Alexander, D.C., Darko, Z., Ghosh, A., Tanno, R., Wottschel, V., et al.: Image quality transfer and applications in diffusion mri. *NeuroImage* **152**, 283–298 (2017)
3. Bhadra, S., Kelkar, V.A., Brooks, F.J., Anastasio, M.A.: On hallucinations in tomographic image reconstruction. *IEEE Transactions on Medical Imaging* **40**(11), 3249–3260 (2021)
4. Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C.: Activating more pixels in image super-resolution transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
5. Chung, H., Kim, J., Kim, S., Ye, J.C.: Parallel diffusion models of operator and image for blind inverse problems. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023)
6. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: The Eleventh International Conference on Learning Representations (2023)
7. Chung, H., Sim, B., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. In: Advances in Neural Information Processing Systems (2022)
8. Delannoy, Q., Pham, C.H., Cazorla, C., Tor-Díez, C., Dollé, G., et al.: Segsrgan: Super-resolution and segmentation using generative adversarial networks — application to neonatal brain mri. *Computers in Biology and Medicine* **120**, 103755 (2020)
9. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: European Conference on Computer Vision (ECCV) (2014)
10. Gopinath, K., Hoopes, A., Alexander, D.C., Arnold, S.E., Balbastre, Y., et al.: Synthetic data in generalizable, learning-based neuroimaging. *Imaging Neuroscience* **2**, 1–22 (2024)
11. Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., et al.: Fastsurfer - a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage* **219**, 117012 (2020)
12. Iglesias, J.E., Billot, B., Balbastre, Y., Magdamo, C., Arnold, S.E., et al.: Synthsrs: A public ai tool to turn heterogeneous clinical brain scans into high-resolution t1-weighted images for 3d morphometry. *Science Advances* **9**(5) (2023)

13. Kim, S., Jin, C., Diethe, T., Figini, M., Tregidgo, H.F.J., et al.: Tackling structural hallucination in image translation with local diffusion. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) European Conference on Computer Vision (ECCV) (2024)
14. Kim, S., Tregidgo, H.F., Eldaly, A.K., Figini, M., Alexander, D.C.: A 3d conditional diffusion model for image quality transfer—an application to low-field mri. arXiv preprint arXiv:2311.06631 (2023)
15. Lau, V., Xiao, L., Zhao, Y., Su, S., Ding, Y., et al.: Pushing the limits of low-cost ultra-low-field mri by dual-acquisition deep learning 3d superresolution. Magnetic Resonance in Medicine **90**(2), 400–416 (2023)
16. Lin, H., Figini, M., D’Arco, F., Obole, G., Tanno, R., et al.: Low-field magnetic resonance image enhancement via stochastic image quality transfer. Medical Image Analysis **87**, 102807 (2023)
17. Lin, X., Du, C., Wu, Q., Tian, X., Yu, J., et al.: Zero-shot Low-field MRI Enhancement via Denoising Diffusion Driven Neural Representation . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 (2024)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI. pp. 234–241 (2015)
19. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., et al.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021)
20. Sotiroopoulos, S.N., Jbabdi, S., Xu, J., Andersson, J.L.R., Moeller, S., et al.: Advances in diffusion mri acquisition and processing in the human connectome project. NeuroImage **80**, 125–143 (2013)
21. Tanno, R., Worrall, D.E., Kaden, E., Alexander, D.C.: Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion mri. NeuroImage **225** (2020)
22. Tivnan, M., Yoon, S., Chen, Z., Li, X., Wu, D., et al.: Hallucination Index: An Image Quality Metric for Generative Reconstruction Models . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 (2024)
23. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018)
24. Wang, J., Chen, Y., Wu, Y., Shi, J., Gee, J.: Enhanced generative adversarial network for 3d brain mri super-resolution. In: IEEE/CVF Winter Conference (WACV) (2020)
25. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., et al.: Esrgan: Enhanced super-resolution generative adversarial networks. In: The European Conference on Computer Vision Workshops (ECCVW) (2018)
26. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. The Eleventh International Conference on Learning Representations (2023)
27. Wang, Y., Wang, Y., Zhao, D., Xie, C., Zheng, Z.: Videohallucer: Evaluating intrinsic and extrinsic hallucinations in large video-language models. arxiv (2024)
28. Wolfe, P.: Convergence conditions for ascent methods. SIAM Review **11**(2), 226–235 (1969)
29. Yu, R., Liu, S., Yang, X., Wang, X.: Distribution shift inversion for out-of-distribution prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)

30. Zhang, K., Hu, H., Philbrick, K.A., Conte, G.M., Sobek, J.D., et al.: Soup-gan: Super-resolution mri using generative adversarial networks. *Tomography* **8**, 905 – 919 (2021)
31. Zhang, M., Zhang, A., McDonagh, S.G.: On the out-of-distribution generalization of probabilistic image modelling. In: *Neural Information Processing Systems* (2021)
32. 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018)
33. Zhu, Y., Zhang, K., Liang, J., Cao, J., Wen, B., et al.: Denoising diffusion models for plug-and-play image restoration. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops (NTIRE)* (2023)