LOW-FIELD MAGNETIC RESONANCE IMAGE QUALITY ENHANCEMENT USING A CONDITIONAL FLOW MATCHING MODEL

Huu Tien Nguyen[1] *and Ahmed Karam Eldaly*[1][2]

[1]Department of Computer Science, University of Exeter, Exeter, United Kingdom
[2]UCL Hawkes Institute, University College London, London, United Kingdom

ABSTRACT

This paper introduces a novel framework for image quality transfer based on conditional flow matching (CFM). Unlike conventional generative models that rely on iterative sampling or adversarial objectives, CFM learns a continuous flow between a noise distribution and target data distributions through the direct regression of an optimal velocity field. We evaluate this approach in the context of low-field magnetic resonance imaging (LF-MRI), a rapidly emerging modality that offers affordable and portable scanning but suffers from inherently low signal-to-noise ratio and reduced diagnostic quality. Our framework is designed to reconstruct high-field–like MR images from their corresponding low-field inputs, thereby bridging the quality gap without requiring expensive infrastructure. Experiments demonstrate that CFM not only achieves state-of-the-art performance, but also generalises robustly to both in-distribution and out-of-distribution data. Importantly, it does so while utilising significantly fewer parameters than competing deep learning methods. These results underline the potential of CFM as a powerful and scalable tool for MRI reconstruction, particularly in resource-limited clinical environments.

Index Terms— Low-field MRI, super-resolution, image quality transfer, deep learning, conditional flow matching

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a cornerstone of radiology, but image quality is highly dependent on field strength: high-field systems (> 1T) provide superior resolution and contrast, whereas low-field scanners (< 1T) suffer from reduced signal-to-noise ratio and artifacts that limit diagnostic accuracy. There is a pressing need for cost-effective methods that deliver high-field-like quality [1]. To address this challenge, various image enhancement methods have been proposed. Super-resolution approaches attempt to improve spatial resolution by upsampling low-resolution scans, while modality transfer methods consider the problem as a cross-domain translation task, learning mappings across imaging modalities (e.g., CT-to-MR synthesis) to recover richer information [2]. Another line of work, image quality

transfer (IQT), leverages high-quality data from advanced imaging systems to improve routinely acquired scans. Early IQT approaches employed patch-based random forests [3], which captured nonlinear relationships between low- and high-quality image patches. More recently, Lin et al. [4] introduced stochastic simulation combined with a directional U-Net for low-field MRI enhancement, showing further improvements.

Generative modeling has driven additional progress. Diffusion models have emerged as powerful frameworks for image enhancement [5]. In parallel, unsupervised and hybrid dictionary learning approaches have demonstrated improved robustness to out-of-distribution data, an essential property given the variability of clinical imaging [6]. These studies highlight the potential of generative and representation learning methods, yet challenges remain. Existing techniques often rely on large model sizes, iterative sampling procedures, or struggle to generalise reliably across diverse acquisition settings. Thus, in this work, we introduce conditional flow matching (CFM) — a recent generative paradigm that learns continuous flows between noise and data distributions through velocity field regression [7]. Unlike diffusion models, CFM achieves this without iterative denoising, offering a more efficient and compact alternative. To our knowledge, this is the first application of CFM to IQT for low-field MRI enhancement. The main contributions of this work are as follows. (1) We propose a novel CFM-based framework for reconstructing high-field-like images from low-field MRI. (2) We provide a comprehensive benchmarking against state-of-the-art IQT approaches, demonstrating that CFM achieves superior quality while using significantly fewer parameters, with robust performance in both in-distribution and out-of-distribution datasets.
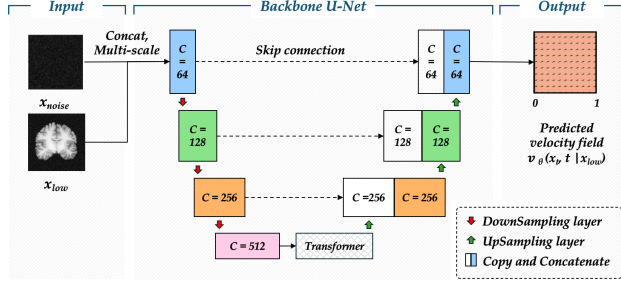
2. METHODS

2.1. Conditional Flow Matching process

CFM assumes a continuous transformation from a source distribution $p_0 \sim x_{noise}$ to a target distribution $p_1 \sim x_{high}$ over a time interval $t \in [0, 1]$, as shown in Fig. 1. This transformation can be modeled as an ordinary differential equation (1)
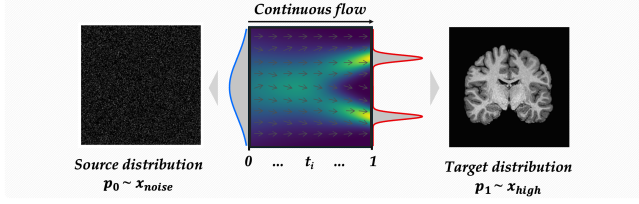
Fig. 1. CFM learns velocity fields between noise and data distributions, then generates samples by following these fields.

Fig. 2. Integration of CFM with U-Net backbone: multi-scale inputs, residual blocks with channel attention, transformer-enhanced bottleneck, and advanced sampling methods.

where x_t is the intermediate image at time t. as follows

$$\frac{dx_t}{dt} = v_\theta(x_t, t \mid x_{\text{low}}). \tag{1}$$

After learning the velocity field $f_\theta(x_t, t \mid x_{low})$, we can sample by solving the differential equation (1). A common method to solve this equation is the Euler method as in Eq. (2), and the final result is an output sample.

$$x_{noise}(t + \Delta t) \approx x_{noise}(t) + f_\theta(x_{noise}, t | x_{low})\Delta t. \tag{2}$$

2.2. Network architecture

Figure 2 shows network architecture, which is built upon the foundation of U-Net [8]. A distinctive feature of U-Net is the symmetric skip connections between encoder and decoder layers, enabling direct concatenation of high-resolution feature maps. This mechanism combines abstract semantic information from deep layers with precise spatial information from shallow layers, crucial for tasks requiring high positional accuracy and image reconstruction. Based on this foundation, we propose a customised U-Net architecture for IQT using CFM, enhanced with several components as follows.
Multi-scale input layer: The network input is a tensor concatenated from the intermediate image (x_t) and the conditional image (x_{low}). To effectively extract initial features, the first convolutional layer of traditional U-Net has been replaced with a multi-scale input layer, where we process it in
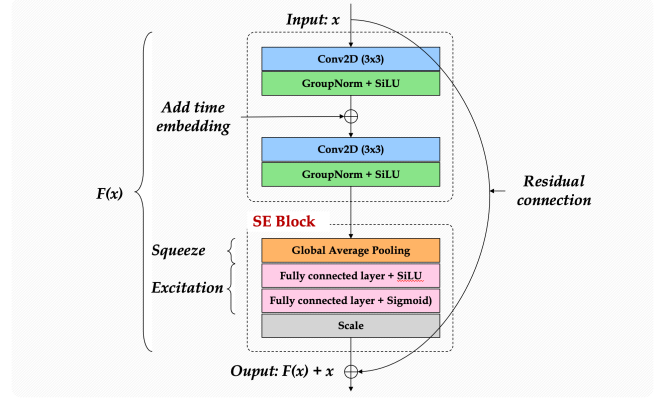
Fig. 3. ResnetBlock serves as the core feature processing component in both encoder and decoder.

parallel through multiple convolutional branches with different kernel sizes (1×1, 3×3, 7×7, and 15×15). The output from each branch is then concatenated along the channel dimension to create a rich input feature map for the first encoder layer. This approach allows the model to simultaneously capture both fine details through small kernels and larger structures through large kernels right from the first layer rather than having to wait until deeper layers in the network.

Residual block with Squeeze-and-Excitation (SE) module: As shown in Fig.3, each Resnet block consists of two 3×3 convolutional layers (Conv2d), with each layer followed by a group normalisation layer (GroupNorm) and a SiLU activation function. To condition the processing according to the time variable t of the CFM method, a time embedding vector is incorporated into the block. This vector is first processed by a small multi-layer perceptron network and then added to the feature map after the first convolutional layer. Additionally, to enhance representational capacity, each block is also integrated with a SE module. This module implements a channel attention mechanism, allowing the model to automatically learn how to adjust the importance of each feature channel [9].

Pixel unshuffle for down-sampling and pixel shuffle for up-sampling: To change spatial resolution, our model employs symmetric pixel unshuffle and pixel shuffle operations instead of conventional max-pooling or ConvTranspose2d layers. During downsampling, pixel unshuffle (space-to-depth) rearranges 2×2 pixel blocks from spatial dimensions into channels, quadrupling channel count while halving spatial dimensions, followed by 1×1 convolution to adjust channels for the next layer. Conversely, during up-sampling, we use pixel shuffle using a 1×1 convolution quadruples channels, then Pixel Shuffle (Depth-to-Space) spreads these channels across spatial dimensions, doubling height and width while restoring original channel count.

Bottleneck transformer block The bottleneck contains a transformer block allowing multi-head self-attention mechanisms

to be applied in this deeper feature map. This allows the model to more effectively identify and enhance relationships between distant image regions before the reconstruction process begins in the decoder part.

3. EXPERIMENTS AND RESULTS

3.1. Datasets

We use high-resolution T1-weighted (T1w) MR images from the Human Connectome Project (HCP) [10]. The original brain images were acquired on a 3 Tesla Siemens Connectome scanner, with parameters of 0.7-mm isotropic voxel, repetition time (TR) of 2400 ms, echo time (TE) of 2.14 ms, and inversion time (TI) of 1000 ms for T1w images. To generate training pairs from T1w MRI images, we apply a stochastic low-field image simulator [4] that transforms high-quality images into low-field counterparts with realistic noise and contrast variations. The InD datasets used for training and testing are synthesised with parameters constrained by a Mahalanobis distance < 1, ensuring that the SNR in white matter (WM) is higher than in gray matter (GM) to maintain tissue contrast compatible with T1-weighted images. On the other hand, the OOD test dataset is synthesised by sampling SNR parameters from a distribution reflecting ultra-low field T1-weighted MRI images, introducing significant deviations from the Gaussian distribution used for the InD data for training. For training, 404 image pairs were randomly split: 80% (323 pairs) for training and 20% (81 pairs) for validation. For testing, we use two evaluation sets: an InD test set with 300 pairs sharing training distribution characteristics, and an OOD test set with 20 pairs having different distribution properties to evaluate generalisation capability and robustness in real-world applications where new data differs from training data.

3.2. Implementation details

All training and testing procedures were implemented in Python 3.9.18 and executed on an NVIDIA H100 NVL GPU system. The IQT-CFM model was trained using the Adam optimizer with an initial learning rate of 10^{-4} and L2 weight decay of 10^{-4} to prevent overfitting. Training was conducted with a batch size of 4 over 200 epochs to ensure loss function convergence. Additionally, we employed a Cosine Annealing learning rate scheduler that gradually decreases the learning rate following a sinusoidal cycle from the initial value of 10^{-4} to a minimum of 10^{-6}. The trained model weights obtained from training process were subsequently used during the inference phase to evaluate performance on both InD and OOD test datasets.

3.3. Comparison with other work and evaluation criteria

The proposed algorithm is compared to the baseline interpolation method which uses a stochastic low-field image simulator [4] to generate low-field versions from high-quality im-

Table 1. Quantitative results using InD and OOD datasets. Best results indicated in bold.

Metric	Interpolation	IQT-DDL	IQT-CFM
InD Dataset			
PSNR↑	28.92 ± 0.443	36.07 ± 0.903	**37.07±1.024**
SSIM↑	0.88 ± 0.013	0.95 ± 0.009	**0.96±0.007**
LPIPS↓	0.12 ± 0.011	0.05 ± 0.010	**0.03±0.008**
Params↓	N/A	9,337,203	**5,253,249**
OOD Dataset			
PSNR↑	24.69 ± 0.602	26.25 ± 0.823	**26.33±0.821**
SSIM↑	0.76 ± 0.050	0.76 ± 0.061	**0.77±0.050**
LPIPS↓	0.16 ± 0.061	0.14 ± 0.065	**0.13±0.062**
Params↓	N/A	9,337,203	**5,253,249**

ages, and the state-of-the-art deep dictionary learning IQT method, (IQT-DDL) [6]. The performance of the proposed approach and the existing methods is assessed using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS).

3.4. Quantitative results

Table 1 presents the quantitative evaluation results on both in-distribution (InD) and out-of-distribution (OOD) datasets. On the InD dataset, the proposed IQT-CFM approach consistently achieved superior performance across all evaluation metrics. It delivered a noticeable improvement in reconstruction fidelity compared with both the state-of-the-art IQT-DDL model and the interpolation baseline, leading to enhanced structural preservation and perceptual quality. Furthermore, these gains were achieved with substantially fewer learnable parameters, highlighting the efficiency of the proposed framework. A similar trend was observed on the OOD dataset, where IQT-CFM continued to outperform the competing methods in terms of reconstruction quality and perceptual similarity, despite the expected drop in absolute performance across all approaches. The method maintained a modest yet consistent advantage, indicating stronger generalisation beyond the training distribution. Overall, these results demonstrate that IQT-CFM not only achieves state-of-the-art image reconstruction quality but also does so with significantly improved parameter efficiency.

3.5. Qualitative results

Figures 4 and 5 show reconstructed brain images from the HCP dataset using the InD and OOD test sets, along with the corresponding error maps that illustrate the absolute differences between the ground-truth high-resolution images and their reconstructions. The qualitative results are consistent with the quantitative findings in Table 1. As expected, interpolation produces noticeable blurring and a loss of structural details, whereas IQT-CFM achieves best reconstructions

with substantially lower errors compared to both interpolation and IQT-DDL. Furthermore, error maps derived from the OOD dataset reveal higher residuals than those from the InD dataset. Such performance reflects the common generalisation challenge of domain shifted data, but simultaneously emphasises the importance of evaluating model generalisation.
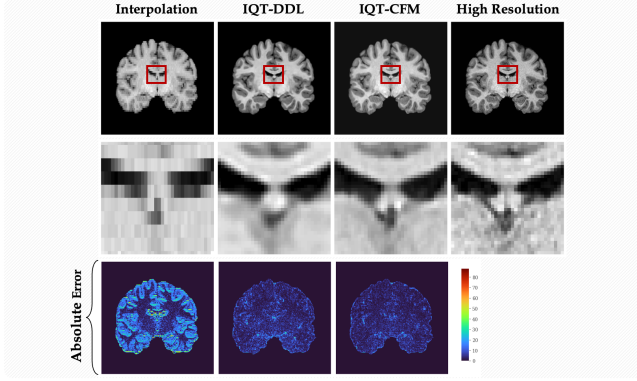
Fig. 4. Example of enhanced brain image using InD data and the corresponding error maps.
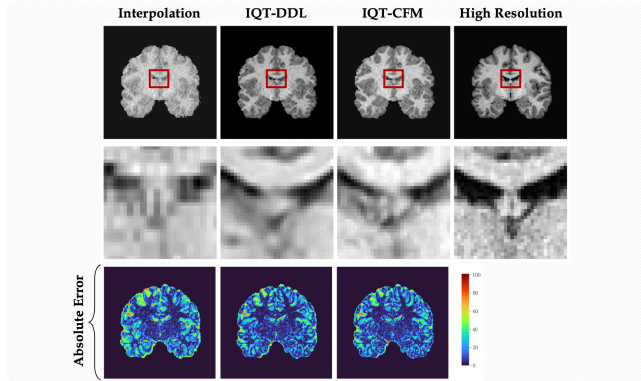
Fig. 5. Example of enhanced brain image using OOD data and the corresponding error maps.

4. CONCLUSIONS

In this work, we introduced a novel conditional flow matching (CFM) framework for image quality transfer (IQT) for low-field MRI quality enhancement. By learning continuous flows between noise and data distributions, CFM provides an efficient alternative to conventional generative approaches. Comprehensive evaluations using both in-distribution and out-of-distribution datasets demonstrated that IQT-CFM achieves state-of-the-art reconstruction quality with significantly fewer parameters than existing IQT methods. This efficiency, coupled with robust generalisation, makes CFM particularly well-suited for clinical deployment in resource-limited settings, where access to high-field scanners is restricted. Future work will explore integrating uncertainty estimation for improved clinical reliability.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using a publicly available brain MRI scan dataset [10]. Ethical approval was not required.

6. REFERENCES

[1] Sanjana Murali, Hao Ding, et al., "Bringing mri to low- and middle-income countries: directions, challenges and potential solutions," *NMR in Biomedicine*, vol. 37, no. 7, pp. e4992, 2024.

[2] Xianfan Gu et al., "Cross-modality image translation: Ct image synthesis of mr brain images using multi generative network with perceptual supervision," *Computer Methods and Programs in Biomedicine*, vol. 237, pp. 107571, 2023.

[3] D. Alexander, D. Zikic, et al., "Image quality transfer via random forest regression: applications in diffusion mri," in *MICCAI*. Springer, 2014, pp. 225–232.

[4] Hongxiang Lin, Matteo Figini, et al., "Low-field magnetic resonance image enhancement via stochastic image quality transfer," *Medical Image Analysis*, vol. 87, pp. 102807, 2023.

[5] S. Kim, H. Tregidgo, et al., "A 3d conditional diffusion model for image quality transfer–an application to low-field mri," *arXiv preprint arXiv:2311.06631*, 2023.

[6] Ahmed K. Eldaly, Matteo Figini, and Daniel C Alexander, "Alternative learning paradigms for image quality transfer," *arXiv preprint arXiv:2411.05885*, 2024.

[7] Yaron Lipman, Ricky Chen, et al., "Flow matching for generative modeling," *arXiv preprint arXiv:2210.02747*, 2022.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on MICCAI*. Springer, 2015, pp. 234–241.

[9] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on CVPR*, 2018, pp. 7132–7141.

[10] S. Sotiropoulos, S. Jbabdi, et al., "Advances in diffusion mri acquisition and processing in the human connectome project," *Neuroimage*, vol. 80, pp. 125–143, 2013.