MRI Super-Resolution Using Fine-Tuned Diffusion Model

Jonas Vinson

Abstract—Magnetic Resonance Imaging (MRI) super-resolution (SR) aims to reconstruct high-resolution anatomically correct brain scans from low-resolution acquisitions, reducing scan time while persevering diagnostic quality. Recently, diffusion-based generative models have achieved strong results in natural image generation, prompting works on MRI specific diffusion models which have had equal success. In this work I investigate the feasibility of fine-tuning a large state-of-the-art diffusion based natural image generator (Flux) for MRI SR. I integrate a ControlNet to guide the reverse diffusion process using low resolution MRI scans and apply Low-Rank adaption (LoRA) to enable fine-tuning of both the ControlNet and the transformer backbone. Training is performed on the IXI brain MRI dataset using 2D 64×64 patches. The loss function is a hybrid loss combining diffusion, SSIM and gradient based terms. Despite stable optimization during training both quantitative and qualitative evaluation shows the model failed to produce accurate SR outputs. PSNR and SSIM remain far below established MRI SR baselines. The generated patches demonstrate coarse structural tendencies but lack anatomical detail. These findings suggest that fine-tuning large natural-image diffusion models is insufficient for medical SR, and that domain-specific diffusion models or fully supervised SR networks remain necessary for high-fidelity MRI reconstruction.

Index Terms—Super-resolution, Diffusion models, Latent diffusion, Magnetic resonance imaging (MRI), Deep learning, ControlNet, Low-rank adaptation (LoRA).

1 Introduction

MAGNETIC Resonance Imaging (MRI) is the primary tool for the diagnosis and treatment of various brain disorders. However, clinicians often acquire low-resolution scans due to the long acquisition times required for high-resolution imaging [1].

Super-resolution (SR) techniques aim to reconstruct high-quality MRIs from low-resolution inputs, revealing anatomical details that may be obscured in standard clinical scans [2]. Recent advances in diffusion-based models have shown promising results for MRI super-resolution [3] [4] [5].

However, these methods are typically trained from scratch on 3D volumetric MRIs, which incurs a large computational cost. In contrast, state-of-the-art large diffusion models such as Flux are trained on billions of parameters and massive natural image datasets where fine-tuning requires much less computational power.

The motivation behind this project is to explore whether these pretrained models can be adapted to the medical imaging domain, specifically to determine whether a model like Flux can effectively perform MRI SR through targeted fine-tuning. To do so, I fine-tune Flux using Low-Rank Adaptation (LoRA) and incorporate a ControlNet conditioning mechanism that injects structural information extracted from a low-resolution MRI directly into the diffusion backbone which guides the reverse diffusion process.

A limitation of this approach is that Flux is inherently a 2D model, which means volumetric (3D) consistency is not enforced; furthermore, the natural-image pretraining distribution differs substantially from MRI, which can impact reconstruction fidelity.

2 Background

Diffusion models are generative models that are used for image generation. Diffusion-based neural networks are trained to progressively "diffuse" samples with random noise, then reverse that diffusion process to generate high-quality images. The training process of a diffusion model is called forward diffusion, whereas the inference portion is called reverse diffusion [6].

2.1 Forward Diffusion

The purpose of the forward diffusion process is to transform clean data from the training set into pure noise. This step by step process is formulated as a Markov chain which adds gaussian noise of varying degree over a sequence of timesteps. Starting from a clean image x_0, each step produces a slightly noisier sample x_t according to a fixed variance schedule β_t [6]. At each step, the transition is defined as

$$x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\, \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I). \tag{1}$$

A key property of diffusion models is that this accumulation of noise can be expressed in closed form [6]. Rather than iteratively adding noise at every step, we can define a noisy sample for an arbitrary time step t using

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \tag{2}$$

where

$$\bar{\alpha}_t = \prod_{s=1}^{t} (1-\beta_s) \tag{3}$$

This closed form makes training efficient: the model is presented with a randomly sampled timestep t, a noisy sample x_t and learns to predict the amount of noise added ϵ [6]. This forward process defines a path from structured data to pure noise, which is what the reverse diffusion process

must learn to invert. By learning this inversion, diffusion models can learn to transform noise back into images [6].

2.2 Reverse diffusion process

The reverse diffusion process learns to invert the additive noise defined by the forward process, transforming pure noise back into a clean image. At each timestep t, the model receives a noisy sample x_t and predicts the current noise in the image $\epsilon_\theta(x_t, t)$. Using this prediction, the model constructs a slightly cleaner sample x_{t-1} [6]. For a standard DDPM scheduler, the transition from x_t to x_{t-1} is described by

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z, \qquad z \sim \mathcal{N}(0, I) \tag{4}$$

where the term involving $\epsilon_\theta(x_t, t)$ removes the predicted noise, and the additional stochastic term $\sigma_t z$ ensures that the reverse process follows the correct probability distribution [6].

Sampling begins with pure Gaussian noise $x_T \sim \mathcal{N}(0, I)$. By iteratively applying the reverse update for $t = T, \ldots, 1$, the model restores coarse structure, mid-level details and eventually fine textures as the noise level decreases [6]. As such, the reverse diffusion process traces a path from true noise back to structured data only guided by the model's learned noise predictions.

The method proposed in this work, along with all approaches discussed in the following sections are based on a specific kind of diffusion model: latent diffusion models (LDM). LDMs extend diffusion models by performing the diffusion process in a compressed latent space rather than directly in the high-dimensional image space, substantially reducing computational requirements during training [7]. A schematic overview of a typical LDM architecture is shown in Figure 2.

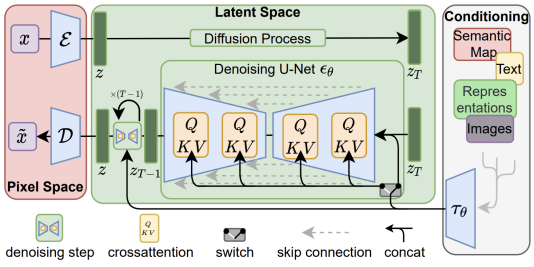

Fig. 1. General latent diffusion model architecture illustrating the encoder–diffusion–decoder pipeline [7].

3 Related Work

3.1 Diffusion based MRI Super resolution

Super-resolution (SR) is a significant area of research in brain MRI reconstruction, and numerous approaches have been proposed to address this problem. Recently, diffusion models have demonstrated state-of-the-art performance in generating realistic natural images, as shown by models such as Flux [6] [7] [8]. Motivated by these advances, several works have applied diffusion-based frameworks to MRI SR [3] [4] [5].

One such example is *DisC-Diff: Disentangled Conditional Diffusion Model for Multi-contrast MRI Super-Resolution* [3]. During training, this method concatenates high-resolution (HR), low-resolution (LR), and an auxiliary contrast image into a single latent representation, which is then processed by a diffusion model to produce a multi-contrast SR reconstruction. The overall structure is similar to that of the present work in that the diffusion process is performed in the latent space; however, because my approach leverages a pretrained foundational model, there will be differences in the data structure and model training.

Another closely related study, *Latent Diffusion Model-based MRI Super-Resolution Enhances Mild Cognitive Impairment Prognostication and Alzheimer's Disease Classification* [4], follows a comparable framework. Their model employs conditional latent diffusion, where Gaussian noise is added to HR images, and the network learns to predict this noise distribution. During the reverse diffusion process, the latent representation of LR MRI serves as conditioning to guide the denoising trajectory, preserving the anatomical structure. This mechanism is similar to the ControlNet component in my model, which similarly guides the reverse diffusion process to ensure anatomically accurate SR outputs.

3.2 Slice-Wise latent diffusion Approaches

Both aforementioned works operate on 3D MRI volumes, whereas Flux is inherently a 2D latent diffusion model. Consequently, my implementation will generate 2D slice-wise reconstructions, akin to the approach proposed in *Temporal and Spatial Super-Resolution with Latent Diffusion Model in Medical MRI Images* [9]. This work combines a VQGAN-based encoder–decoder with a latent diffusion model trained on 2D cardiac MRI slices.

Although many studies adopt similar latent diffusion architectures for MRI SR, no prior work (from what I have found) has explored adapting a large, pretrained natural-image diffusion model (such as Flux) for medical image super-resolution. This project therefore represents a novel direction, aiming to bridge the gap between large-scale generative models and domain-specific medical imaging tasks.

4 Proposed Method

4.1 Dataset and Preprocessing

The dataset used in this project is the IXI Dataset, which contains approximately 600 brain MRI scans. This dataset has been widely used in previous super-resolution (SR) studies [5] [10]. As a preprocessing step, all brain volumes were skull-stripped using the HD-bet brain extraction tool [11]. This process removes non-brain tissues such as the skull, neck, and eyes, ensuring that the model focuses only on relevant anatomical structures during training. The extracted brain volumes were then processed into 2D 8-bit RGB slices by saving each volume layer individually and propagating the grayscale information across all three channels to ensure compatibility with the Flux architecture. Slices below a background intensity threshold were discarded, and the remaining slices were center-cropped to

minimize background regions. The resulting images were used as the high-resolution (HR) component of the LR–HR training pairs. The corresponding low-resolution (LR) slices were generated by downsampling and then upsampling the HR slices by a factor of two, ensuring a sufficient loss of spatial information for the SR task.

4.2 Model Architecture

4.2.1 Flux Backbone and LoRA Adaption

Flux is traditionally a text-to-image diffusion model with billions of parameters, which makes full retraining computationally infeasible. To adapt such a large network to the medical imaging domain, I employ Low-Rank Adaption (LoRA). LoRA is a parameter efficient fine tuning method which injects low rank decomposition matrices into select layers of a pretrained model [12]. The original model weights are then frozen and LoRA learns this pair of injected matrices to approximate the parameter updates in a lower-dimensional subspace [12].

Formally, for a given weight matrix $W_0 \in \mathbb{R}^{d\times k}$, LoRA introduces two low-rank matrices $A \in \mathbb{R}^{d\times r}$ and $B \in \mathbb{R}^{r\times k}$, where $r \ll \min(d,k)$, such that:

$$W = W_0 + \alpha(AB) \tag{5}$$

Here, W_0 represents the pretrained frozen weights, while A and B are trainable and capture task-specific adaptations. The scaling factor α controls the magnitude of the adaptation. This low rank update works to perturb the network activations toward the target distribution while preserving the expressive capacity of the pretrained model [12]. These adapters are generally applied to the self attention and feed-forward layers of the model, where they learn fine grained domain specific corrections without overwriting the overall priors learned from the pretrained model. After training these weights are then merged back into the original weight matrices for inference granting the model new domain specific knowledge [12].

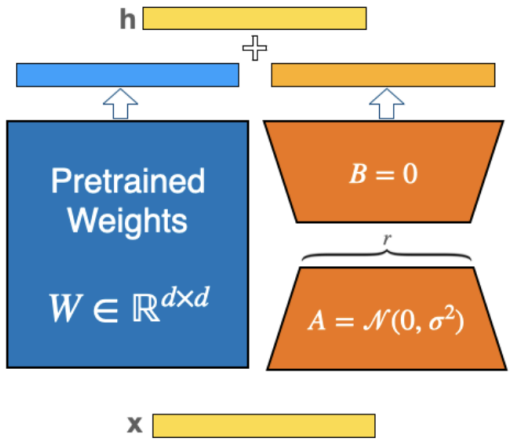

Fig. 2. Low-Rank Adaptation (LoRA) injects a small trainable update BA into a frozen pretrained weight matrix W, enabling efficient fine-tuning with dramatically fewer parameters and reduced memory usage. [12]

4.2.2 ControlNet Integration

Since Flux was originally built for text-to-image generation, an additional conditioning mechanism is needed to enable image-to-image diffusion. To achieve this, a ControlNet module is added to the flux architecture. ControlNet was introduced in [13] as a means of adding condition control to diffusion models. It functions by creating a trainable copy of the denoising backbone that processes conditioning information in parallel with the original backbone [13].

In convolutional diffusion methods, the ControlNet operates by producing feature maps from a conditioning image and injecting this information into the corresponding layers of the Unet backbone [13]. Since flux has replaced the Unet with a transformer based denoising backbone (FluxTransformer2DModel), this method has been adapted to operate on hidden states rather than feature maps.

During the training process, the Controlnet receives the same noisy latent representation x_t and timestep t as the transformer ensuring it's features are aligned with the current timestep of the denoising process. In addition it also receives a conditioning image, in our case the low res MRI. These inputs are then combined into residual embeddings $\Delta F_{c,l}$ at multiple layers of the transformer [13]. The embeddings are then injected into the main transformer through residual addition:

$$\tilde{h}l = h_l + \Delta Fc, l \tag{6}$$

where h_l represents the hidden state of the main transformer at layer l, and $\Delta F_{c,l}$ encodes the structural features extracted from the conditioning image [13].

Formally, the main transformer predicts noise as $\epsilon_\theta(x_t, t)$, while the ControlNet augments this prediction through the residual conditioning process.

$$\tilde{\epsilon}\theta(x_t, t, c) = \epsilon\theta(x_t, t) + \Delta F_c \tag{7}$$

Adapting this ControlNet conditioning mechanism to the transformer-based model of FLux allows the model to incorporate the low resolution MRI directly into the diffusion process. This helps provide anatomical consistency to the model while maintaining its expressiveness.

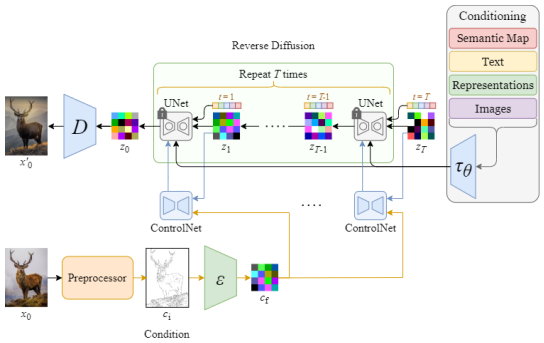

Fig. 3. General architecture of a ControlNet. Note that the model used in this work employs a Transformer-based backbone (Flux) rather than a UNet, but the ControlNet mechanism remains analogous.

4.2.3 Loss function

To ensure the perceptual quality and the structural integrity of the image, I will use a composite loss with three terms. The base of the loss function will be the standard diffusion loss. This is the MSE between the predicted noise of the model and the actual gaussian noise. This is the traditional loss used in latent diffusion models [7]. This will ensure accurate reconstruction of a clean image through the reverse diffusion process.

$$L_{\text{diff}} = \mathbb{E}_{x_t,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(x_t, t, c)\|^2\right] \tag{8}$$

In addition to traditional diffusion loss, two image space regularizers will be added to preserve the general structure and anatomical edges of the image. The first of these will be a Structural Similarity Index (SSIM) loss. This will promote the alignment of local structures between the denoised output and ground truth [14].

$$L_{\text{SSIM}} = 1 - \text{SSIM}(x_{\text{pred}}, x_{\text{true}}) \tag{9}$$

where

$$\text{SSIM}(x_{\text{pred}}, x_{\text{true}}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{10}$$

[14]

The final piece of the loss function will be a gradient loss. This will promote sharper reconstructions and clear edges by aligning gradient information between the denoised and original image.

$$L_{\text{grad}} = \|\nabla x_{\text{pred}} - \nabla x_{\text{true}}\|_1 \tag{11}$$

Combining these three loss terms leads to a composite loss of

$$L_{\text{total}} = \lambda_{\text{diff}} L_{\text{diff}} + \lambda_{\text{ssim}} L_{\text{SSIM}} + \lambda_{\text{grad}} L_{\text{grad}} + \tag{12}$$

The image based losses are computed on reconstructions obtained by decoding the models noise prediction. Although prior diffusion-based MRI SR studies have relied primarily on diffusion noise-prediction loss and perceptual metrics [15] [16], these objectives may not fully preserve fine anatomical detail when considering the domain change presented by this work. To address this, I incorporate a Structural Similarity (SSIM) term to encourage local structural fidelity and a gradient-consistency loss to explicitly enhance edge sharpness.

5 Experimental Results

5.0.1 Experimental Setup

All experiments were performed using the IXI brain MRI dataset. It was preprocessed as outlined in section 4.1. The model architecture follows the fine-tuned Flux ControlNet configuration as outlined in section 4.2. LoRA was applied to both the transformer backbone as well as the ControlNet attention layers.

After preprocessing approximately 40,000 LR HR pairs were obtained, the data was then split into training, testing and validation sections with a split of 0.8, 0.1, 0.1 respectively. Each section was split by patient and not by slice to ensure that during both validation and training, the model was seeing completely new uncorrelated data. The LoRA configuration for both the ControlNet and the transformer backbone had a rank $r = 16$ and a scaling factor $\alpha = 32$. The LoRA adapters were inserted into the multi headed attention layers and feed forward layers of the models, specifically the projection matrices `to_q`, `to_k`, `to_v`, `to_out.0`, `ff.net.0.proj`, and `ff.net.2`. Training was conducted using random 64×64 patches for eight epochs with a batch size of 8, using the AdamW optimizer with a learning rate = 1×10^{-4}.

5.0.2 Quantitative results

Figure 4 shows the training losses over 10 epochs. All losses exhibit a steady downward trend, indicating stable optimization during fine tuning. Both the SSIM and gradient losses begin to plateau around the 5th epoch, suggesting both the structural and edge based improvements begin to saturate early. However, the diffusion loss continues it's downward trend throughout all epochs contributing the strongest to the overall model loss.

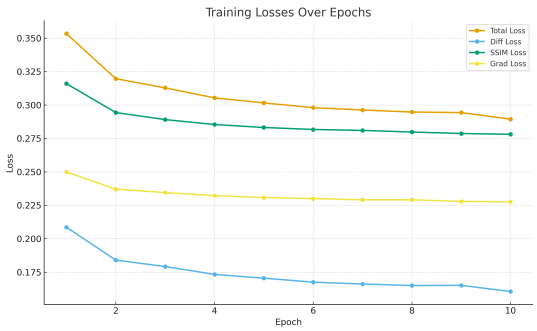

Fig. 4. Training losses over 10 epochs. All loss terms decrease steadily, with diffusion loss showing the strongest improvement. SSIM and gradient losses plateau after about 5 epochs, indicating early saturation of structural and edge-based refinement.

In figure 5 we can see the validation loss over the same epochs. These losses initially mirror the training losses, but they begin to diverge around the 8th epoch. Both the SSIM and the gradient losses increase at this epoch, suggesting that the model is overfitting to the training data and is failing to generalize structural details reliably. The diffusion loss continues to decrease up to the 9th epoch, causing the total loss to decrease as well. This indicates that the weighting of the individual loss components may place too much emphasis on the diffusion loss relative to the structural components.

To evaluate generalization, the final model was tested on 5 slices from 20 subjects never seen before (100 total). Unlike the training phase, which optimizes noise prediction, the test phase measures image quality after the full reverse diffusion process. I tested using Standard image quality metrics.

- PSNR to quantify pixel-level reconstruction fidelity
- SSIM to capture structural similarity and anatomical coherence

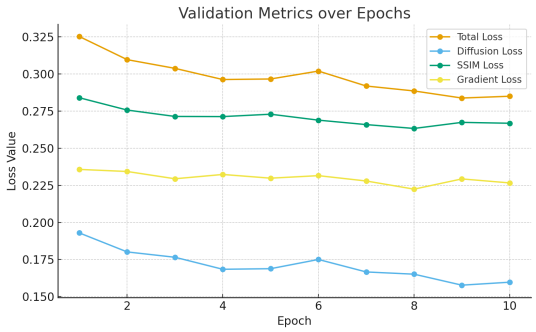

Fig. 5. Validation losses over 10 epochs. Validation losses follow training trends initially but begin to diverge around epoch 7, where SSIM and gradient losses rise—suggesting overfitting—while diffusion loss continues to decline slightly.

For each set of 100 slices, the model was evaluated under varying numbers of diffusion steps and different levels of Controlnet guidance. Table 1 reports the mean PSNR and SSIM across all test slices for each such configuration.

Guidance Scale	Diffusion Steps	PSNR	SSIM
0.5	15	11.22	0.183
0.5	30	11.26	0.146
0.5	50	11.10	0.123
1.0	15	11.07	0.168
1.0	30	11.34	0.161
1.0	50	11.77	0.146
2.0	15	**12.14**	**0.191**
2.0	30	11.83	0.165
2.0	50	11.57	0.147

TABLE 1
Mean PSNR and SSIM over 100 test patches for different ControlNet guidance scales and numbers of diffusion steps. Higher values indicate better reconstruction quality.

The quantitative evaluation shows that both the PSNR and SSIM are relatively low across all guidance scale diffusion step configurations. The PSNR values remain below $13.2db$ and the SSIM values below 0.2. The best performing configuration comes from a guidance scale of 2.0 and 15 diffusion steps, suggesting that stronger conditioning improves structural alignment and coherence in the generated output. This indicates that the ControlNet provides some meaningful guidance during the reverse diffusion process. However, overall performance remains well below the established MRI SR baselines. Unfortunately, it seems as though the model did not successfully adapt the pretrained Flux priors to the medical imaging domain under the given training regime.

The table above references the $2\times$ super-resolution results reported in DisC-Diff [3], which evaluates several MRI SR methods based on the IXI dataset. Although the method proposed in the work differs from DisC-Diff which uses multi contrast conditioning, it provides a relevant baseline using the same dataset, down scaling factor and evaluation metrics. Therefore, we compare our results against these to contextualize the effectiveness of our Flux-based approach.

5.0.3 Qualitative results

The following are the strongest qualitative results produced by the model. Each example shows a 64×64 SR patch and its corresponding HR ground truth patch. Although some very coarse structural patterns begin to emerge, the reconstructions are dominated by noise and lack meaningful anatomical detail. When viewed from further away, the noisy structure starts to loosely align with the underlying brain structure. However, the overall appearance is extremely speckled and the intensity distribution does not match true MRI contrast, further suggesting that the model failed to adapt Flux's latent space to this domain.

These observations are consistent with the low PSNR and SSIM reported earlier: the model only captures the very coarse structure, with anatomical boundaries being fully lost along with a substantial amount of added residual noise. For context, the qualitative slice-wise reconstructions from previous MRI SR work shown in Fig 7 further illustrate the performance gap between this approach and established super-resolution methods [4].

Method	PSNR	SSIM
Bicubic	32.84	0.9622
EDSR [17]	36.59	0.9865
SwinIR [18]	37.21	0.9856
Guided Diffusion [19]	36.32	0.9815
MINet [20]	36.56	0.9806
DisC-Diff [3]	37.64	0.9873
Fine-tuned Flux	12.14	0.191

TABLE 2
Reported $2\times$ downsampled MRI super-resolution performance on the IXI dataset from DisC-Diff [3], including several widely used SR baselines. Values shown are PSNR and SSIM as reported in the original paper. The performance of the fine-tuned Flux model from this work is included for comparison.

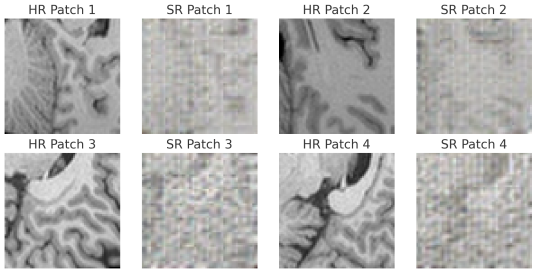

Fig. 6. Best-performing 64×64 super-resolved patches (SR) compared with their corresponding high-resolution (HR) ground-truth patches. Although the SR outputs capture coarse structural patterns, they remain dominated by noise and lack fine anatomical detail.

6 Conclusion

6.0.1 Limitations

While the proposed approach demonstrates that FLux can be fine-tuned to produce coarse noisy MRI-like structures, there are several limitations which constrain its overall effectiveness and usability. First, the domain gap between natural images, which Flux and the ControlNet are trained on, and brain MRIs is substantial. As a result, the model struggled to learn meaningful anatomical priors from the training data. Second, because LoRA adapters only present a lightweight fine-tuning, the models ability to adapt deeper

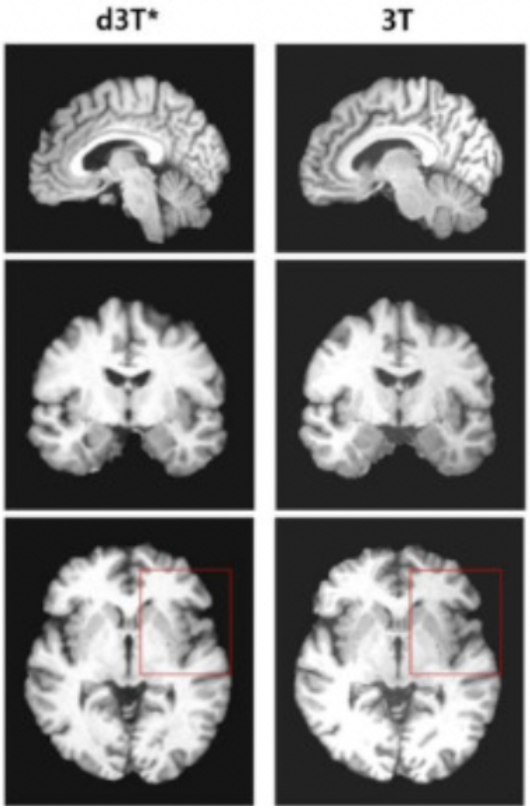

Fig. 7. Slice-wise MRI super-resolution reconstructions from the LDM-based method in [4], with prediction on the left and ground truth on the right, illustrating the level of anatomical fidelity achieved by a domain-specialized model.

representations was hindered. This likely prevented it from capturing fine-grained boundaries and tissue contrast. Third, due to computational costs and model architecture, training was performed using small 2D patches rather than full slices or volumes, limiting structural context and potentially causing reconstructions to exhibit noise artifacts at patch and slice boundaries. Finally, as stated above, the ControlNet was pretrained on natural images as opposed to MRIs which reduced its ability to guide the denoising process in a meaningful way. Collectively, these limitations lead the model to under perform compared to its SR counterparts.

6.0.2 Future work

Several extensions could be pursued to potentially improve model performance. The most impactful of these changes would likely be to train a ControlNet from scratch as opposed to relying on a pretrained version intended for natural images. A domain specific ControlNet could learn appropriate structural features and provide more meaningful conditioning signals during the reverse diffusion process. Additionally, a deeper LoRA adaption could be explored by increasing the rank r, using a larger scaling factor α or inserting adapters into a broader set of transformer blocks. This would give the model a greater ability to shift it's internal representation and priors toward the MRI domain. Another possible improvement is to train on full 2D slices as opposed to random 64×64 patches, which could improve global structure and avoid problems when stitching the outputs. Finally, incorporating MRI-specific conditioning signals, such as segmentation maps, tissue probability maps, or learned edge detectors could be incorporated as alternative ControlNet inputs or as additional channels concatenated to the LR input. They could also be incorporated into the loss function to check the fidelity of the models output. These signals would encode anatomical structure, potentially leading to shaper and more anatomically accurate reconstructions.

6.0.3 Final Remarks

The experiments in this work which adapt the Flux diffusion model and potentially other large pretrained models to MRI SR through LoRA fine-tuning are not a viable approach under the constraints of this project. Despite some structural alignment, the model fails to produce anatomically accurate reconstructions and performs far below established MRI SR baselines. These results suggest that the domain gap between natural images and MRI data is too large for a shallow adaption alone, especially in a domain where accuracy and reliability are critical. In practice training a domain specific diffusion model or a supervised SR model from scratch is far more likely to be effective. While this work demonstrates why cross domain adaption is challenging, it clarifies that future progress will require architecture and priors designed explicit from MRI data.

References

[1] F. Shi, J. Cheng, L. Wang, P.-T. Yap, and D. Shen, "Lrtv: Mr image super-resolution with low-rank and total variation regularizations," *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2459–2466, 2015.

[2] J. Wang, Y. Chen, Y. Wu, J. Shi, and J. Gee, "Enhanced generative adversarial network for 3d brain mri super-resolution," 2019. [Online]. Available: https://arxiv.org/abs/1907.04835

[3] Y. Mao, L. Jiang, X. Chen, and C. Li, "Disc-diff: Disentangled conditional diffusion model for multi-contrast mri super-resolution," 2023. [Online]. Available: https://arxiv.org/abs/2303.13933

[4] D. Yoon, Y. Myong, Y. G. Kim, Y. Sim, M. Cho, B.-M. Oh, and S. Kim, "Latent diffusion model-based mri superresolution enhances mild cognitive impairment prognostication and alzheimer's disease classification," *NeuroImage*, vol. 296, p. 120663, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811924001587

[5] J. Wang, J. Levman, W. H. L. Pinaya, P.-D. Tudosiu, M. J. Cardoso, and R. Marinescu, "Inversesr: 3d brain mri super-resolution using a latent diffusion model," 2023. [Online]. Available: https://arxiv.org/abs/2308.12465

[6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020. [Online]. Available: https://arxiv.org/abs/2006.11239

[7] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," 2022. [Online]. Available: https://arxiv.org/abs/2112.10752

[8] A. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," 2021. [Online]. Available: https://arxiv.org/abs/2102.09672

[9] V. Dubey, "Temporal and spatial super resolution with latent diffusion model in medical mri images," 2024. [Online]. Available: https://arxiv.org/abs/2410.23898

[10] S. Chatterjee, A. Sciarra, M. Dünnwald, A. B. T. Ashoka, M. G. C. Vasudeva, S. Saravanan, V. T. Sambandham, P. Tummala, S. Oeltze-Jafra, O. Speck, and A. Nürnberger, "Beyond nyquist: A comparative analysis of 3d deep learning models enhancing mri resolution," *Journal of Imaging*, vol. 10, no. 9, 2024. [Online]. Available: https://www.mdpi.com/2313-433X/10/9/207

[11] F. Isensee, M. Schell, I. Pflueger, G. Brugnara, D. Bonekamp, U. Neuberger, A. Wick, H. Schlemmer, S. Heiland, W. Wick, M. Bendszus, K. H. Maier-Hein, and P. Kickingereder, "Automated brain extraction of multisequence mri using artificial neural networks," *Human Brain Mapping*, vol. 40, no. 17, p. 4952–4964, Aug. 2019. [Online]. Available: http://dx.doi.org/10.1002/hbm.24750

[12] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "Qlora: Efficient finetuning of quantized llms," 2023. [Online]. Available: https://arxiv.org/abs/2305.14314

[13] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," 2023. [Online]. Available: https://arxiv.org/abs/2302.05543

[14] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[15] M. Safari, S. Wang, Z. Eidex, Q. Li, E. H. Middlebrooks, D. S. Yu, and X. Yang, "Mri super-resolution reconstruction using efficient diffusion probabilistic model with residual shifting," 2025. [Online]. Available: https://arxiv.org/abs/2503.01576

[16] Z. Wu, X. Chen, S. Xie, J. Shen, and Y. Zeng, "Super-resolution of brain mri images based on denoising diffusion probabilistic model," *Biomedical Signal Processing and Control*, vol. 85, p. 104901, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809423003348

[17] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," 2017. [Online]. Available: https://arxiv.org/abs/1707.02921

[18] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," 2021. [Online]. Available: https://arxiv.org/abs/2108.10257

[19] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," 2021. [Online]. Available: https://arxiv.org/abs/2105.05233

[20] C.-M. Feng, H. Fu, S. Yuan, and Y. Xu, "Multi-contrast mri super-resolution via a multi-stage integration network," 2021. [Online]. Available: https://arxiv.org/abs/2105.08949