Accelerated MRI Reconstruction with SwinUNet: Enhancing Image Quality through Transformers

Feolu Kolawole [1] Yogesh Seenichamy [1] Kesavan Ramakrishnan [1]

[1]Computer Science, Stanford

Introduction

Magnetic Resonance Imaging (MRI) is a critical medical imaging technique that provides detailed anatomical information without harmful radiation. However, the lengthy acquisition time of fully-sampled MRI scans presents significant challenges in clinical settings. As a result, clinicians often undersample MRI data. This project uses accelerated MRI reconstruction to repair undersampled MRI data, improve image quality, and reveal important anatomical boundaries and small pathological features.

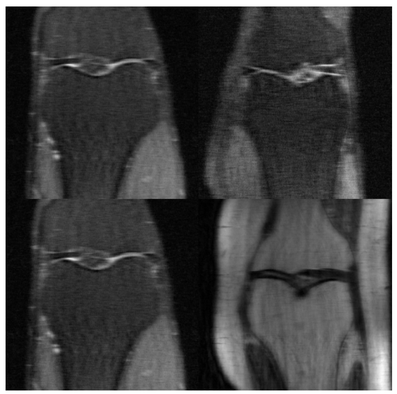

Figure 1. Raw and Undersampled Knee MRI Data

In this project, we address the problem of accelerated MRI reconstruction from undersampled k-space data using deep learning approaches. We propose a hybrid architecture that combines the strengths of U-Net with the Swin Transformer to effectively capture both local features and global dependencies in MRI images.

Background: MRI Reconstruction

Traditional approaches to accelerated MRI reconstruction include:

- **Parallel imaging techniques** like SENSE and GRAPPA
- **Compressed sensing methods** that exploit sparsity
- **Deep learning-based methods** that have emerged as powerful alternatives

While traditional approaches have shown promise, they often suffer from lengthy reconstruction times, increased noise, or dependence on specific sampling patterns. Deep learning methods have demonstrated superior performance in terms of both reconstruction quality and speed.

Dataset: fastMRI

We utilized the fastMRI dataset, a large-scale collection of MRI data specifically designed for machine learning approaches to MR image reconstruction.

- **Focus**: Single-coil knee MRI subset
- **Contrast types**: Proton Density weighted with and without fat suppression (PD and PDFS)
- **Training data**: 973 volumes (middle slices)
- **Validation data**: 199 volumes (middle slices)
- **Acceleration factor**: 4× with 8% central k-space fully sampled

The middle slice was selected as it typically contains substantial anatomical information and provides a consistent reference point across different scans.

Technical Approach

Our approach involves a systematic comparison of multiple architectures:

1. **Baseline U-Net**: A standard convolutional architecture widely used for image-to-image tasks
2. **Transformer at Bottleneck (BT)**: A U-Net with transformer blocks at the bottleneck to capture long-range dependencies
3. **SwinUNet**: A hierarchical vision transformer adapted for MRI reconstruction that uses shifted windows for efficient attention computation

SwinUNet Architecture

The SwinUNet architecture adapts the hierarchical Swin Transformer design to a U-Net-like encoder-decoder structure, offering several advantages:

1. **Hierarchical Feature Representation**: The Swin Transformer's hierarchical design naturally aligns with the multi-scale feature extraction paradigm of U-Net
2. **Shifted Window Attention**: Instead of global self-attention, which is computationally expensive, Swin Transformer uses shifted window-based self-attention
3. **Linear Complexity**: The window-based attention mechanism reduces the computational complexity from quadratic to linear with respect to image size

Our SwinUNet implementation consists of:

- **Patch Embedding**: The input image is divided into non-overlapping patches and projected to a higher-dimensional feature space
- **Encoder**: A series of Swin Transformer blocks with patch merging layers that progressively reduce spatial resolution while increasing feature dimension
- **Bottleneck**: Swin Transformer blocks that process the most abstract features
- **Decoder**: A series of Swin Transformer blocks with patch expanding layers that progressively increase spatial resolution while decreasing feature dimension
- **Skip Connections**: Feature maps from the encoder are concatenated with corresponding decoder features to preserve spatial details

Training Strategy

All models were trained using the following strategy:

- **Loss Function**: Combined L1 loss and SSIM loss to optimize both pixel-wise accuracy and structural similarity
- **Optimizer**: Adam optimizer with learning rates ranging from 1e-5 to 1e-3
- **Learning Rate Scheduling**: Cosine annealing learning rate scheduler
- **Regularization**: Weight decay (1e-4 to 1e-5) and dropout in transformer layers
- **Data Augmentation**: Random flips and rotations
- **Batch Size**: 4 to 16, depending on model complexity

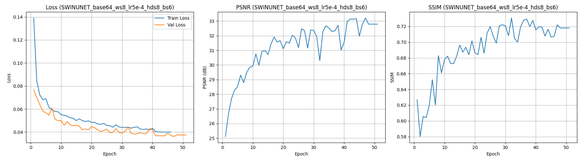

Figure 2. Training progression of the optimized SwinUNet model over 50 epochs, showing: (a) Training and Validation Loss curves (left); (b) Validation PSNR (center); and (c) Validation SSIM (right).

Ablation Studies

We conducted several ablation studies to understand the contribution of different components:

1. **Effect of Window Size**: Increasing the window size from 7 to 8 improved performance by allowing the model to capture slightly larger contextual regions in each attention operation.
2. **Impact of Base Feature Dimension**: Increasing the base feature dimension from 64 to 80 slightly decreased performance while significantly increasing computational requirements, suggesting that 64 provides a good balance between model capacity and efficiency.
3. **Learning Rate Sensitivity**: We found that the SwinUNet was more sensitive to learning rate than the baseline U-Net, with optimal performance achieved at a learning rate of 8e-5.
4. **Data Augmentation**: Removing data augmentation led to faster initial convergence but poorer generalization, confirming the importance of augmentation for preventing overfitting.

Model Efficiency

While the SwinUNet achieves superior reconstruction quality, it comes with increased computational requirements compared to the baseline U-Net.

Model	Parameters (M)	Inference Time (ms)
U-Net Baseline	7.8	18.5
SwinUNet-64	27.3	42.7
SwinUNet-80	42.6	56.3

Table 2. Comparison of model size and inference time.

Despite the increased computational cost, the SwinUNet's inference time remains practical for clinical applications, where reconstruction quality is often prioritized over speed once a certain threshold of efficiency is met.

Results & Future Work

The following shows comparative results between our SwinUNet model and various other proposed models, and qualitative results of our optimized SwinUNet model performed on two undersampled MRI images.

Architecture	Val. Loss	PSNR (dB)	SSIM
U-Net Baseline	0.0496	28.03	0.6935
BT-UNet	0.0412	29.87	0.7102
SwinUNet	0.0352	33.10	0.7274

Table 3. Our proposed SwinUNet compared to Baseline and BT

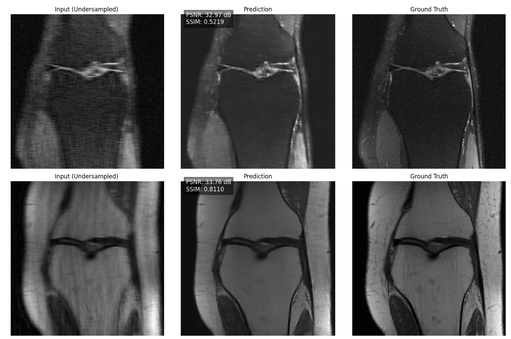

Figure 3. Visual results of the optimized SwinUNet model over 50 epochs, showing: (a) Raw Under-sampled Data(left); (b) SwinUNet reconstruction (center); and (c) Ground Truth Fully-sampled Data (right).