Accelerating MRI Reconstruction via Deep Learning

Mike Chou
Georgia Institute of Technology (OMSA)
cchou85@gatech.edu

Luis Robaina
Georgia Institute of Technology (OMSCS)
lrobaina3@gatech.edu

Abstract

In this paper, we report a study of different approaches to enhance MRI image reconstruction using the fastMRI dataset. We study the successful U-Net architecture and the integration of attention gates. U-Net serves as the core network to address the challenge of reconstructing MRI images from highly under-sampled k-space data. By training on the fastMRI dataset, which comprises diverse clinical MRI data, the U-Net architecture achieves high-quality reconstructions with reduced artifacts. Furthermore, it has been shown that the incorporation of attention gates refines the reconstructed MRI images by highlighting salient features and suppressing noise, leading to sharper and more accurate results, particularly in complex anatomical structures. We also explore our own version of the Attention UNet which we call the Sparse Attention UNet. ***Source code, example notebooks, hyperparameters, and experiment results are all available in the project Github repository IMPACT-MRI.***

1. Introduction

Since September 1971, magnetic resonance imaging (MRI) scans have been helping doctors all across the world in diagnosing a myriad of diseases and injuries due to their superior capability of delivering accurate medical imaging for soft-tissue contrasts. While MRIs can produce hundreds of images from almost any direction and in any orientation, these procedures can often take up to 90 minutes in a confined space where the patient must remain still in an enclosed machine. Patients that are older or are claustrophobic may have trouble lying completely still and if they move during the scan, it may produce motion artifacts in the resulting images which the radiologists cannot use for an effective diagnosis. Additionally, images have to be acquired sequentially in k-space (an array of numbers representing spatial frequencies in the MR image) and the speed at which it collects the data is limited by hardware constraints [3].

Back in the day, most MR imaging sped up acquisitions via Parallel Imaging which allowed the collection of multiple data points rather than traditional sequential ordering, but it suffered tremendously from a reduction in signal-to-noise ratio (noise amplification) [2]. People then utilized different sub-sampling techniques such as Compressed Sensing to improve faster image acquisition [1]; however, undersampling in the k-space often led to aliased images due to the violation of the Nyquist-Shannon theorem. In recent years with the rise of Deep Learning and open-source datasets such as fastMRI (collection of both raw MR measurements and clinical MR images), subsampled data have been trained using CNN models, particularly the U-Net architecture, which have shown promise in medical image reconstruction tasks.

In this project, our primary objective is to study the task of reconstructing MRI images obtained from the fastMRI dataset (https://fastmri.med.nyu.edu/) using different modifications of the U-Net architecture. The fastMRI dataset is a large-scale collection of clinically acquired MRI data that contains a comprehensive collection of MRI data designed to facilitate the development and evaluation of accelerated MRI reconstruction techniques [10]. It was created to address the limitations of traditional methods used in clinical MRI imaging, which are often slow and can lead to prolonged scanning times for patients. The dataset also comprises images from different anatomical regions, including the knee, brain, and cardiac; however, we will only be utilizing the single-coil knee portion due to time constraints.

One of the central features of the fastMRI dataset is that it includes highly undersampled k-space data which enables accelerated MRI acquisitions, but it also introduces challenges in accurately reconstructing the images [10]. Conventional reconstruction methods applied to this undersampled data often reduce image quality and introduce artifacts, impeding accurate clinical diagnosis. To address this challenge, we explore the different variations of the U-Net deep learning architecture for image reconstruction: Vanilla U-Net, Attention U-Net, and Sparse Attention U-Net. **Our goal is to study the attention mechanism on U-Net and the addition of attention representational biases to emphasize critical image features while suppressing noise and irrelevant information, ultimately improving the**

overall image quality and diagnostic utility of the reconstructed MRI images. Along with the undersampled k-space data, the fastMRI dataset also provides fully sampled reference images for each MRI scan which serve as ground truth for evaluation purposes during the reconstruction process. With the inclusion of both fully-sampled reference images and corresponding undersampled k-space data, it will enable us to train our deep learning model to reconstruct high-quality MRI images from the undersampled data.

The success of our research can significantly impact both medical imaging practitioners and patients. By successfully reducing the duration of MRI scan procedures without compromising the quality of the images, we can significantly increase the efficiency of hospitals and decrease the financial burden on patients, thereby optimizing the overall healthcare system. This, in turn, can lead to better-informed medical decisions, more targeted treatment plans, and ultimately improved patient outcomes. Additionally, the integration of attention gates into the U-Net architecture can further advance the field of deep learning in medical imaging, encouraging further exploration and innovation in accelerated MRI reconstruction.

2. Approach

In this section, we provide details on the deep learning architectures explored in this research project which include: Vanilla U-Net (Section 2.1) Attention U-Net (Section 2.2), and our novel contribution Sparse Attention U-Net (Section 2.3).

2.1. U-Net

The U-Net was first introduced in 2015 by Ronneberger et al [7]. It is a deep learning model that consists of a contracting path, which captures context and compresses the image, and an expansive path, which enables the model to refine reconstructions, ensuring accurate localization and preservation of fine details (see Figure 1.). Traditionally, the contracting path follows the typical architecture of a convolutional network which is composed of multiple blocks, each typically consisting of two convolutional layers followed by a max-pooling layer. These convolutional layers perform feature extraction, while the max-pooling layers reduce the spatial dimensions, effectively down-sampling the feature maps. The number of filters in the convolutional layers typically increases with each block, allowing the network to learn more complex features as it goes deeper into the contracting path. Each block is usually followed by a rectified linear unit (ReLU) activation function, which introduces non-linearity and aids the network's ability to learn complex mappings (for **ConvBlock** implementation details refer to Appendix 8.1). At the bottom of the contracting path, there is a bottleneck layer, which represents the deepest level of feature compression. This layer is often formed by multiple convolutional layers and is crucial for forcing the model to learn a compressed representation containing only the most significant features of the input image before reconstruction. What follows is the expansive path which is responsible for precise localization and reconstructing the learned compression from the bottleneck layer. It mirrors the structure of the contracting path, but instead of downsampling, it performs upsampling operations to restore the spatial dimensions of the feature maps. Each block in the expansive path consists of two convolutional layers followed by an upsampling layer (transposed convolution), which increases the feature map size and allows the network to recover the finer details of the original input image.

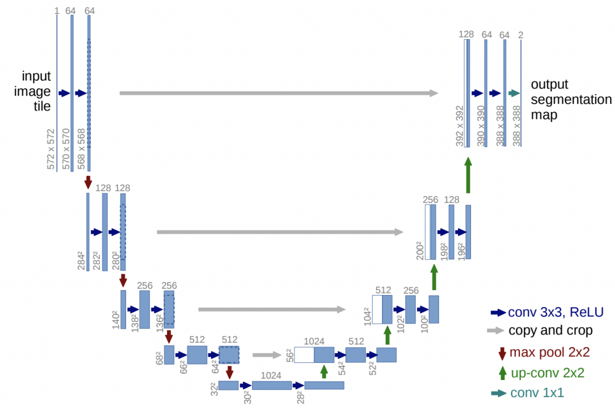

Figure 1. U-Net Architecture [7]

The unique aspect of the U-Net architecture is the incorporation of skip connections between corresponding layers in the contracting and expansive paths. These skip connections provide a shortcut for the information to directly skip several layers and flow directly from the contracting path to the corresponding layer in the expansive path instead without the risk of losing crucial information along the way. This enables the network to merge both local and global features during the upsampling process, aiding in precise localization. **As it will be discussed in Section 2.2 and 2.3, these skip connections are key components to the variations of U-Net we explore in this research.**

Along with the skip connections, the feature maps from the contracting path are concatenated with the feature maps of the corresponding layer in the expansive path. As the feature maps undergo transpose convolution and are restored back to parts of their original image shape in the expansive path, it also gains information from the contracting path via skip connections about what the original image patches are supposed to entail. This is why this concatenation works so effectively and allows the network to combine high-level semantic information with detailed spatial information.

While the U-Net architecture demonstrates exceptional performance in various medical image reconstruction tasks, it does come with certain limitations. One prominent concern is the substantial memory and computational requirements imposed by the symmetric design. Because the U-Net facilitates precise localization and detailed feature reconstruction through skip connections between the symmetric pathways, this results in a large number of parameters, demanding considerable memory resources during training and inference. Moreover, there is a risk of over-fitting due to the extensive parameterization, particularly when dealing with limited training data. This could cause the architecture to struggle to accurately capture intricate anatomical structures and handle ambiguous regions in the images, leading to sub-optimal results in certain scenarios.

2.2. Attention U-Net

The attention gate mechanism introduces a valuable refinement to the U-Net, allowing the network to selectively focus on relevant image features while suppressing noise and irrelevant information. By incorporating attention gates, the network gains the ability to allocate more "attention" to salient regions, effectively improving the quality of reconstructed MRI images. **While the integration of attention gates into the U-Net architecture isn't a novel approach, it has not been widely adopted as a means to expedite MRI image reconstruction.** This attention-based refinement addresses the limitations of traditional U-Net-based networks, resulting in sharper and more accurate reconstructions and thus improving the overall image quality [8].

The structure of the attention gate is shown in more detail in Figure 2. Let x^l represent the feature maps from the contracting path (encoder) before downsampling. Let g represent the feature maps from the expansive path (decoder) after upsampling. Linear transformations are applied on these two feature maps which are computed using channel-wise 1x1x1 convolutions and then concatenated together to be passed through a ReLU activation function. We then apply another linear transformation ψ on the output of the activation function and then pass the result through a Sigmoid function which outputs the attention coefficients α. α is set between 0 to 1 with higher coefficients corresponding to more relevant information and smaller coefficients likely to represent noise. Finally, as proposed by the authors of this architecture, we element-wise multiply these attention coefficients with the original feature map from the encoder and treat it as the input from the skip connection just like the original U-Net [8].

One of the primary advantages of the attention gate lies in its ability to enhance feature representation. Traditional U-Net architectures utilize skip connections to propagate information from the contracting path to the expansive path.

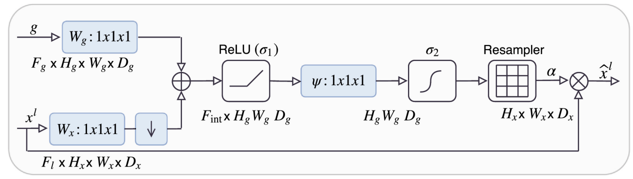

Figure 2. Attention Gates [8]

While skip connections aid in gradient flow and improve information transfer, they may also propagate noise and artifacts from earlier layers to later ones. Attention gates offer a more sophisticated mechanism to selectively amplify or suppress feature maps based on the magnitude of the coefficients as stated earlier, effectively reducing the impact of irrelevant information and enhancing the relevant features. This adaptability allows the network to handle diverse clinical scenarios more effectively, making the attention gate a valuable addition to the U-Net for medical imaging applications.

2.3. Sparse Attention Gates

Inspired by the success of explicit architectural biases in deep learning, we propose to study what we call sparse attention gates, an attention gate similar to the one shown in Figure 2 but with explicit algorithmic biases to encourage sparse attention weights, to further improve the performance of U-Net for image reconstruction. One clear example of successful encoding of domain knowledge in the form of architectural biases is the Convolutional Layer computation which enforces sparse interactions, parameter sharing, and equivariant representations [12]. **Our hypothesis is that this explicit bias will encourage attention gates that filter more noise and only focus on strongly relevant features.**

Figure 3. shows how we implement the Sparse Attention U-Net to encourage more sparse attention weights. We compute the L1 norm of the attention matrix α for each U-Net up-sampling layer on the forward pass. We accumulate the norms and normalize them with the batch size and the number of up-sampling layers. The sum of L^1 norms is added to the overall reconstruction loss to encourage sparser attention weights.

We base our decision to penalize the sum of L^1 norms of the up-sampling layer attention weights because the L1 norm has desirable properties when compared to other norms. For example, in the context of regularization, L^1 regularization results in solutions that are more sparse compared to L^2 [12]. While we use the example of regularization to illustrate the use of L^1 norm, it is important to stress that Sparse Attention Gates are not a regularization technique but rather an explicit modification of the loss landscape to encourage more attention sparsity.

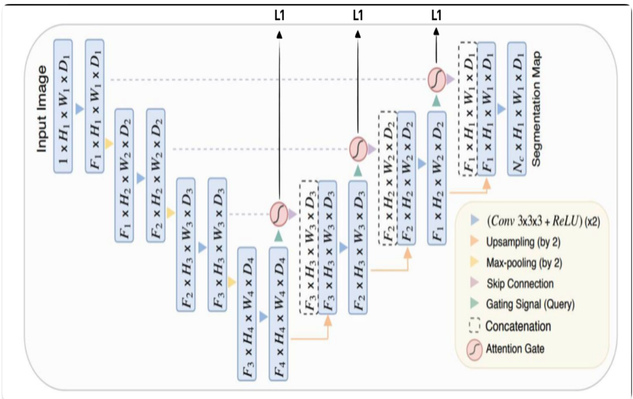

Figure 3. Sparse Attention Computation with L1 Norm

Formally the L^1 norm of the up-sampling layer attention weights α is defined as:

$$||\alpha||_1 = \sum_i |\alpha_i|$$

that is, it is the sum of absolute values of attention weights.

2.4. Methodology and Challenges

During the implementation process, we anticipated challenges related to the complex nature of MRI data and the impact of undersampling on image quality. We expected that the U-Net might struggle to accurately reconstruct images with high acceleration factors due to the loss of critical information during undersampling. On the bright side, the fastMRI repository already contained an existing U-Net architecture source code [10], so all we had to do initially was to train the model and tune the various hyperparameters. However, the real challenge of the project came in trying to implement the attention gate and training the new model architecture with the fastMRI image dataset. We designed a simpler version of the attention gate (https://idiotdeveloper.com/attention-unet-in-pytorch/) than what was implemented in the author's repository [11] which still proved to be equally demanding as we had to incorporate it into the existing U-Net framework.

The very first attempt to integrate attention gates did show some improvements in image quality compared to the standard U-Net reconstruction. However, the initial implementation was far from optimal, and extensive fine-tuning and experimentation were necessary to achieve the desired level of performance. Overfitting to the training data was also a concern, given the large number of parameters in the original U-Net architecture. Despite the difficulties, the iterative nature of our approach allowed us to identify the best configurations for the attention gates, leading to substantial enhancements in the final MRI reconstructions.

3. Experiments

We use the L^1 reconstruction loss function to train the U-Net and Attention U-Net models.

$$L_1(v, \hat{v}) = ||v - \hat{v}||_1$$

For the Sparse Attention U-Net model we used a loss function that penalizes the L^1 of the up-sampling layers attention scores:

$$L_{sparse} = ||v - \hat{v}||_1 + \gamma \frac{1}{m|L|} \sum_{l \in L} (\sum_i |\alpha_i^l|)$$

where m is the batch size L is the set of up-sampling layers of the U-Net and α^l is the attention matrix computed at layer l. γ weights the relative contribution of the attention norm penalty.

The first step of this project was to train the vanilla U-Net model to serve as a baseline. We used the PyTorch U-Net implementation available in the fastMRI repository to the model on the single-coil knee portion of the fastMRI dataset. Next, we implemented attention gates into the existing U-Net code and trained the Attention U-Net model on the single-coil knee data. At the end of the training, the validation set of the single-coil knee data was used to evaluate the effectiveness of the entire process. Finally, we trained our implementation of the Sparse Attention U-Net and collected the relevant metrics (discussed in Section 3.2).

Once we had come up with the best hyperparameters which yielded the lowest L1 loss, we used this new model to compare the results against the vanilla U-Net model. Figures 4 and 5 show an example of a sub-sampled MRI image in the testing set and also a reconstructed MRI image created by the U-Net architecture based on the same sub-sampled test image. As the images show evidently, the Baseline U-Net already does an impeccable job at removing noise and reconstructing sub-sampled images. More information on the reconstruction results is shown below.

Many iterations of training were conducted before the optimal hyperparameters were selected. Some of the more crucial parameters include weight decay, dropout probability, number of pooling layers, and number of top-level U-Net channels. Because the U-Net is prone to overfitting due to its tremendous amount of free parameters caused by skip connections and additional layers in the expansive path, it is extremely important that we tune these 4 hyperparameters well. A sufficient amount of weight decay can penalize large weights while dropout can stop specific neurons in the net from learning too much information. Similarly, the number of pooling layers and top-level channels helps us indicate how deep and complex we want our U-Net architecture. The more pooling layers and channels we have, the

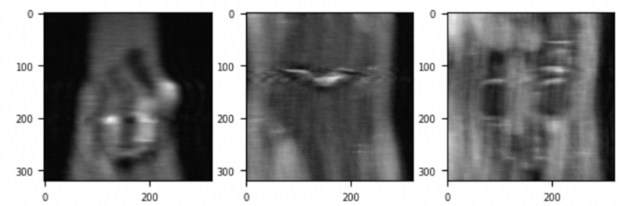

Figure 4. Subsampled Knee MRI Scans

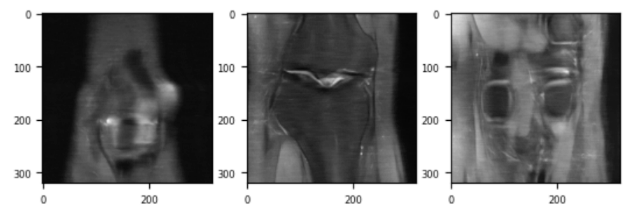

Figure 5. Reconstructed Knee MRI from Vanilla U-Net

deeper the network, the more information is learned, and the easier it is to overfit.

4. Metrics

To further evaluate the effectiveness and authenticity of our MRI reconstructions, we compute a myriad of metrics including Normalized Mean Square Error (NMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity (SSIM), proposed by [4]. We also developed a new metric to study the effect of attention gates and sparse attention gates: Mean Post Attention Norm Difference (Section 3.2).

4.0.1 Normalized Mean Square Error (NMSE)

NMSE measures the average squared difference between the reconstructed image $\hat{v}$ and the reference image v, normalized by the squared Euclidean norm of the reference image.

$$NMSE(v, \hat{v}) = \frac{||\hat{v} - v||_2^2}{||v||_2^2}$$

A lower NMSE value indicates a closer match between the predicted and ground truth, signifying better reconstruction accuracy.

4.0.2 Peak Signal-to-Noise Ratio (PSNR)

PSNR assesses the quality of the reconstructed image by measuring the ratio of the peak signal power (maximum pixel value squared, e.g., 255 for 8-bit images) to the mean squared error between the predicted and ground truth images. Higher PSNR values imply better image quality, indicating that the degree of image information rises above background noise.

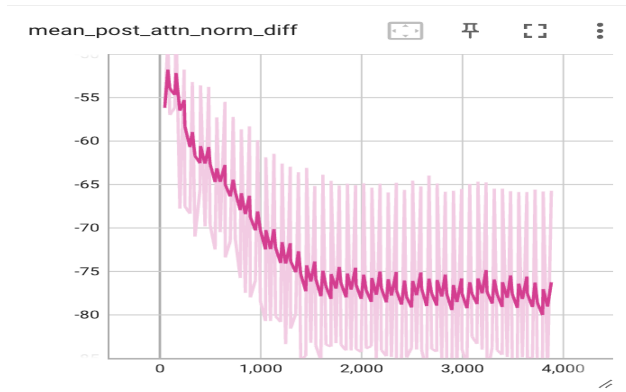

Figure 6. Attention U-Net MPAND

4.0.3 Structural Similarity (SSIM)

SSIM evaluates the structural similarity between the reconstructed and ground truth images. Specifically, SSIM provides insights into the similarity between two images by exploiting the inter-dependencies among nearby pixels. It generates a similarity index ranging from -1 to 1, where 1 represents a perfect match and -1 signifies complete dissimilarity.

4.1. Mean Post Attention Norm Difference

We propose and implement a new metric useful to evaluate prior assumptions about the role of attention in U-Net reconstruction tasks. As illustrated in Figure 12. the Mean Post Attention Norm Difference (MPAND) is the norm difference between the feature representations in the skip connections of each U-Net up-sampling layer after and before attention weights have been applied. Attention has previously been proposed as a mechanism to allow the network to selectively focus on relevant image features while suppressing noise and irrelevant information, our experiments support this claim by demonstrating that over time the MPAND of an Attention U-Net decreases. Figure 6 shows the MPAND metric computed over 50 epochs of Attention U-Net training, we can clearly see that the attention computation is creating more sparse feature representations as the value of MPAND decreases over the first training epochs until it converges to some optimal representation.

5. Results

In this section we highlight experiment results. It is worth noticing that most experiments were repeated with 3 different random seeds and the number of training epochs was kept at 50 due to computational budget constraints, the

Model	NMSE	PSNR	SSIM
Vanilla U-Net	0.034	31.9	0.73
U-Net with attention	0.033	32	0.73
(ours) U-Net with sparse-attention	0.045	30.2	0.71

Table 1. Validation Set Results (using the best hyperparameter configuration for each model)

Figure 7. Validation Set NMSE (top: two random seeds of vanilla U-Net, bottom: Attn. U-Net)

size of the dataset, and long training times. Results show target metrics computed over the validation set are used to compared different methods with a focus of Structural Similarity (SSIM).

5.1. Vanilla U-Net vs Attention U-Net

We compare the performance of vanilla U-Net with Attention U-Net. Given the bounds of our experimental budget, we observe comparable performances among the two architectures as shown in Figure 7 with Attention U-Net getting slightly better performance on validation metrics (more experiment results can be found in Figures 13 and 14 of the Appendix).

5.2. Attention and Sparse Attention

One of the insightful results we generated is a comparison between the performance of the same U-Net architecture trained with the original attention gate and our proposed sparse attention gates. As shown in Figure 8, the validation set SSIM shows better performance for the traditional attention gate when compared to different sparse attention weights (the only different in the value of the γ parameter). As discussed in Section 3, the γ parameter of the Sparse Attention U-Net loss function weights the relative contribution of the attention norm penalty, with larger values encouraging more sparse representations. The results shown in Figure 11. demonstrate that simple attention gates are already a powerful architectural bias. We observe a degradation of performance with the Sparse Attn. U-Net specially as the value of γ increases. **This result suggests that, contrary to our original hypotheses, explicit biases to encourage sparsity in attention weights does not help achieve better reconstructions and attention gates are already a powerful tool to support reconstructions of MRI data.**

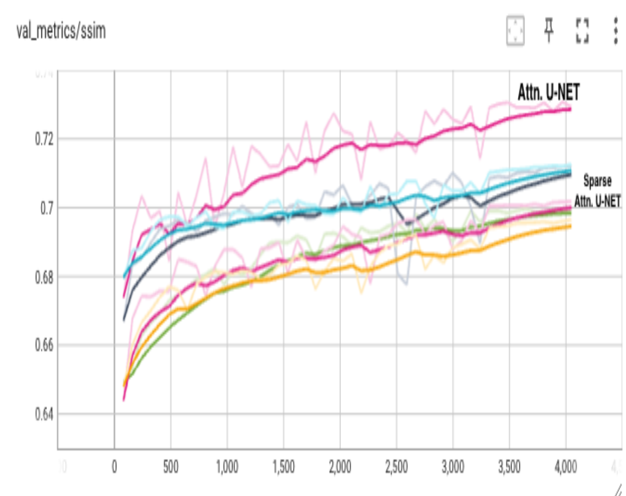

Figure 8. SSIM val Set Comparison between Attn.U-Net and Sparse Attn U-Net.

6. Conclusion

In conclusion, the U-Net's symmetric design, coupled with skip connections, allows for efficient capture of both local and global features, enabling precise localization and reconstruction of intricate anatomical structures in MRI scans. By incorporating attention gates, the model gains the ability to selectively focus on relevant image features while suppressing noise and irrelevant information, leading to improved image quality and accuracy in the reconstruction process. As medical imaging technology continues to evolve, the U-Net with attention gates offers a versatile and robust framework for tackling the challenges posed by undersampled and noisy MRI data, promising enhanced clinical diagnoses and treatment planning. Conducting supervised learning on ground-truth fully sampled data is still extremely expensive as there are no shortcuts to quickly obtain fully-sampled MRI scans. In the era of big data, researchers have made tremendous strides in the field of computer vision and machine learning and are earnestly looking at more efficient methods for reconstructing MR images such as self-supervised or few-shot learning. As researchers and engineers continue to push the boundaries of medical imaging, the U-Net and attention gate combination stands as a beacon of innovation and progress, with the potential to revolutionize MRI reconstruction and ultimately improve patient care in the realm of medical diagnostics.

7. References

[1] "Deep MRI Reconstruction with Radial Subsampling", Yiasemis et al

[2] "Self-Supervised Learning of Physics-Guided Reconstruction Neural Networks without Fully-Sampled Reference Data", Yaman et al

[3] "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction", Schlemper et al

[4] "fastMRI: An Open Dataset and Benchmarks for Accelerated MRI", Zbontar et al

[5] "Deep Learning Tuning Playbook", https://github.com/google-research/tuningplaybook

[6] "What Matters In On-Policy Reinforcement Learning? A Large-Scale Empirical Study" (https://arxiv.org/abs/2006.05990)

[7] "U-Net: Convolutional Networks for Biomedical Image Segmentation", Ronneberger et al, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-428

[8] "Attention U-Net: Learning Where to Look for the Pancreas", Oktay et al

[9] Burhaneddin, Y., "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data", (2020), https://github.com/byaman14/SSDU

[10] Zbontar, J., (2018). "fastMRI: An Open Dataset and Benchmarks for Accelerated MRI" ArXiv e-prints. Retrieved from https://arxiv.org/abs/1811.08839

[11] Oktay, O., https://github.com/ozan-oktay/Attention-Gated-Networks/

[12] Ian Goodfellow, "Deep Learning" Chapter 7 Regularization for Deep Learning

8. Appendix

8.1. Model Components

This section provides more information about the architectures used for the experiments. All architectures are components are implemented in PyTorch and available in the project GitHub repository (IMPACT-MRI).

8.1.1 ConvBlock

```
ConvBlock(
(layers): Sequential(
(0): Conv2d(38, 32)
(1): InstanceNorm2d(32, eps=1e-05)
(2): LeakyReLU(negative_slope=0.2)
(3): Dropout2d(p=0.0, inplace=False)
(4): Conv2d(32, 32, kernel_size=(3, 3)
(5): InstanceNorm2d(32, eps=1e-05)
(6): LeakyReLU(negative_slope=0.2)
(7): Dropout2d(p=0.0, inplace=False)
)
```

Figure 9. ConvBlock Architecture

8.1.2 TrasnsposeConvBlock

```
TransposeConvBlock(
(layers): Sequential(
(0): ConvTranspose2d(128, 64)
(1): InstanceNorm2d(64, eps=1e-05)
(2): LeakyReLU(negative_slope=0.2)
)
```

Figure 10. TransposeConvBlock Architecture

8.1.3 Attention Gate

See Figure 11.

8.2. Architecture Memory Complexity

Figure 15. shows that the extra memory required to host the parameters of the attention gates has a minor effect on the model size, most of the model size comes from convolution blocks and fully connected layers.

8.3. Architecture Time/Space Complexity

In Figure 16. we compare the forward pass running time between comparable U-Net architectures with and without attention gates. Showing that the attention computation has a small effect on forward pass running time.

```
attention_gate(
(Wg): Sequential(
    (0): Conv2d(256, 128)
    (1): InstanceNorm2d(128)
)
(Wx): Sequential(
    (0): Conv2d(256, 128)
    (1): InstanceNorm2d(128
)
(relu): ReLU(inplace=True)
(psi): Sequential(
    (0): Conv2d(128, 1)
    (1): InstanceNorm2d(1)
)
(Sigmoid): Sigmoid()
)
```

Figure 11. Attention Gate Architecture

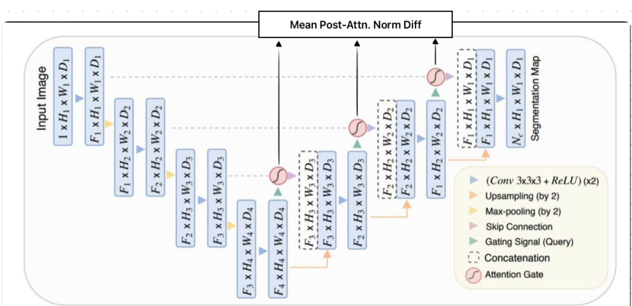

Figure 12. Computation of MPAND from up-sampling U-Net Layers (and attention gates)

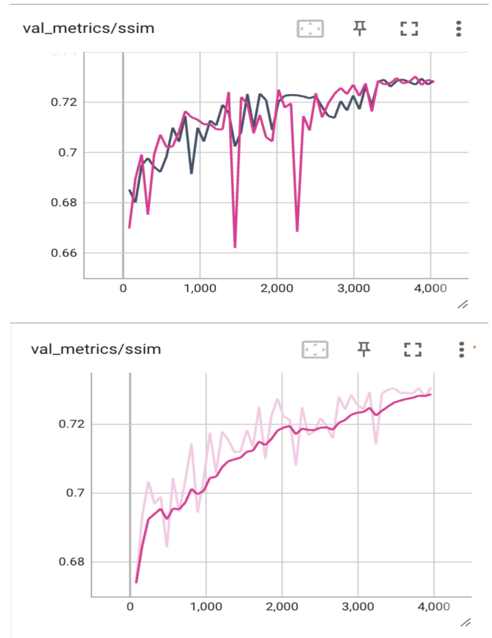

Figure 13. Validation Set SSIM (top: vanilla U-Net, bottom: Attn. U-Net)

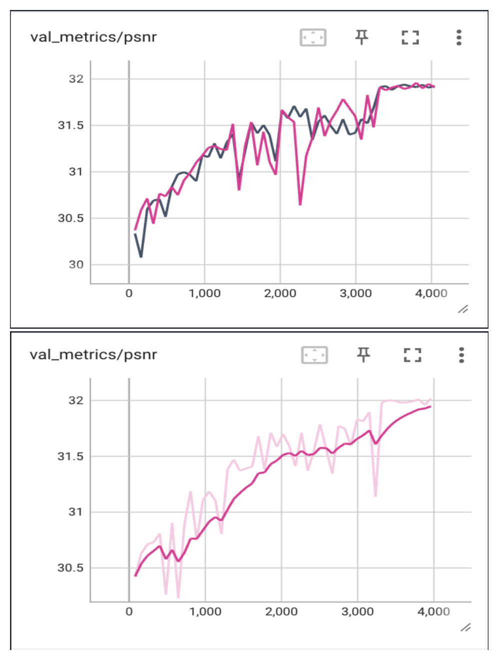

Figure 14. Validation Set PSNR (top: vanilla U-Net, bottom: Attn. U-Net)

8.4. Team Contributions

Table 2. describes the contribution of each team member.

Student Name	Contributed Aspects	Details
Mike Chou	Image Reconstruction and Attention Implementation	Analyzed both fully-sampled and masked MRI images in k-space, reconstructed them using a pre-trained U-Net model, and also implemented the attention gate into the existing U-Net architecture
Luis Robaina	Training Setup and Execution	Proposed and implemented the sparse attention U-Net formulation, performed experiments and generated results plot.

Table 2. Contributions of team members.

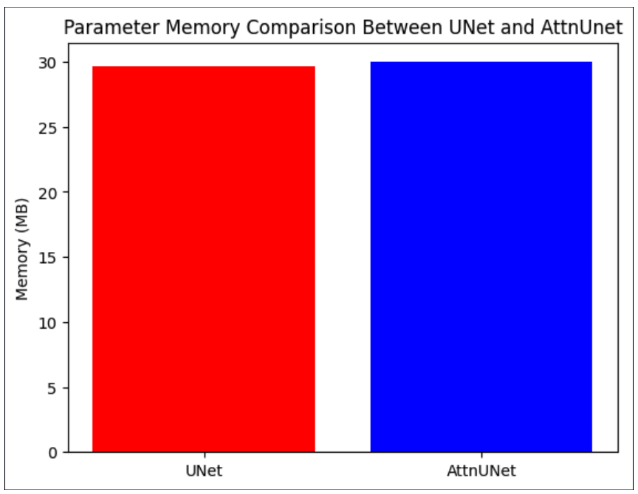

Figure 15. UNet and Attn. UNet Memory Comparison

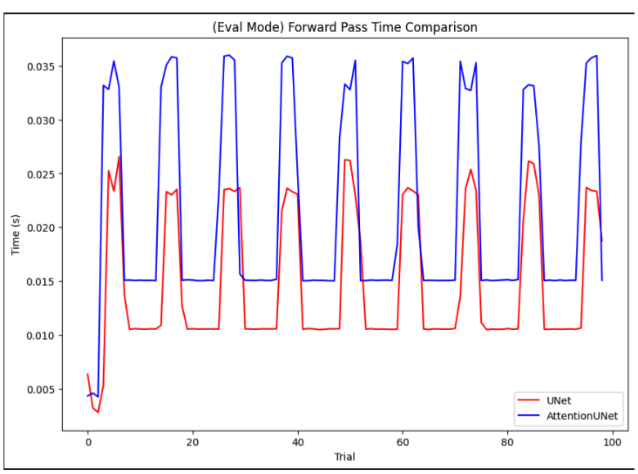

Figure 16. U-Net and Attn.U-Net Forward Pass Running Time Comparison