Comparing the performance of VarNet, MoDL, and SSDU in fastMRI image reconstruction

Author: Yenju Lu
Supervisor: Zhengguo Tan

1 Introduction

This project evaluates three MRI reconstruction models: VarNet (Variational Network), MoDL (Model-based), and SSDU (Self-supervised learning via data undersampling), to determine their effect using PSNR and SSIM metrics. VarNet leverages a UNET-based variational approach, whereas MoDL employs a CNN for model-based reconstruction that optimizes image consistency from sparse multichannel data, streamlining the process with end-to-end training. SSDU stands out with its self-supervised learning strategy, using a ResNet-based framework to train on undersampled data, overcoming the need for fully sampled datasets. By integrating the physics of MRI, SSDU provides a robust alternative capable of matching the performance of traditional supervised methods, even in data-limited scenarios. Each model aims to improve MRI reconstruction quality and accelerate the imaging process while tackling the challenges of artifact presence and the complexity of biological tissue representation.

The MoDL framework advances model-based image recovery by tackling multichannel, noisy, and sparse data challenges commonly found in MRI and various imaging modalities. This method employs a forward model to fine-tune the consistency between predicted images and actual measurements. Addressing this inherently complex issue, MoDL integrates priors to navigate the recovery process effectively. Deep convolutional neural networks (CNNs) have recently been adapted for these tasks, demonstrating significant potential in both direct inversion and noise reduction. MoDL synergizes model-based reconstruction with deep learning, alternating between leveraging CNNs to capture image redundancies and reinforcing data consistency. This is particularly effective for complicated forward models such as multichannel MRI. Through end-to-end training, MoDL streamlines the CNN's complexity, customizing network parameters precisely for image recovery and consequently improving performance, which is especially beneficial when training data is scarce.

VarNet is centered on enhancing MRI reconstruction by employing deep learning techniques to accelerate the process. It addresses the shortcomings of traditional methods such as Parallel Imaging and Compressed Sensing, which often struggle with artifacts and the complex rep-

resentation of biological tissue imagery. To sidestep the demanding conditions imposed by Compressed Sensing, VarNet adopts learning methods that mirror the interpretative skills of radiologists, who are adept at identifying patterns amidst artifacts. The proposed variational network focuses on learning the complexities of transforming raw data into coherent images, thus advancing image quality and strengthening its relevance in clinical settings.

SSDU confronts MRI's slow data acquisition bottleneck by leveraging self-supervised learning, circumventing the drawbacks of traditional techniques such as Parallel Imaging and Compressed Sensing, which tend to introduce artifacts and noise, especially at higher acceleration rates. This innovative approach stands in contrast to conventional deep learning frameworks that rely heavily on large volumes of fully sampled ground-truth data, often a logistical challenge due to time-sensitive physiological processes and patient comfort concerns. SSDU's algorithm is designed to learn directly from undersampled data, making the most of limited information without depending on exhaustive reference datasets. In practical applications involving knee and brain scans, SSDU has proven capable of achieving a level of image quality with fully supervised methods, establishing a new frontier in the realm of medical imaging. Its success in maintaining high-quality reconstructions while significantly reducing scan time could revolutionize patient experience by shortening the time spent in the scanner and streamlining the workflow for radiologists.

2 Theories and Methods

2.1 VarNet: UNET and Gradient Descent

It is a Parallel Imaging (PI) -based MRI reconstruction. It proposes to learn the parameters of the inverse transform with UNET combined with a gradient descent scheme. UNET's architecture provides a robust framework for medical image reconstruction, offering detailed feature extraction, precise localization, efficient data use, and high-quality reconstruction outputs. These advantages make it an invaluable tool in advancing medical imaging technologies.

The optimization objective, shown in Equation [1], minimizes the difference between the measured undersampled k-space data b and the k-space representation of the image Ax, scaled by a regularization parameter λ. Here, A represents the linear forward sampling operator, embodying the MRI's data acquisition process. The term R(x) introduces a regularization that enforces prior knowledge about the image to stabilize the solution of this ill-posed problem. Equation [2] describes the iterative update rule in the gradient descent scheme for image reconstruction. At each iteration n, the current estimate x_n is updated. The variable z_n represents the out-

put of UNET, it is an intermediate variable that is updated using the gradients derived from the data consistency and regularization terms. This update is modulated by the step size λ, balancing the trade-off between convergence speed and stability.

$$\min_x \frac{\lambda}{2}\|Ax - b\|^2 + R(x) \tag{1}$$

$$x_{n+1} = x_n - z_n - \lambda(A^H A x_n - A^H b) \tag{2}$$

2.2 MoDL: CNN and Conjugate Gradient Descent

MoDL employs an iterative algorithm that alternates between enforcing data consistency and utilizing pre-trained CNN denoisers to exploit image redundancies. The application of CNN denoiser in medical image reconstruction offers a transformative approach to managing and improving the quality of medical images, facilitating better diagnostic accuracy and patient outcomes with the power of deep learning. It leverages the conjugate gradient algorithm informed by normal equations, with shared weights across iterations for efficiency. The network learns the regularization parameter, training to eliminate aliasing and noise at every step.

The optimization hinges on an objective function detailed in Equation [3], where A represents the linear forward sampling operator applied to the sought image x, and b corresponds to the undersampled k-space data. The regularization component R(x) is integral to this process. Equation [4] describes the reconstruction of the image x through the conjugate gradient descent method.

$$\min_x \|Ax - b\|^2 + \lambda\|x - z_n\|^2 \tag{3}$$

$$x_{n+1} = (A^H A + \lambda I)^{-1}(A^H b + \lambda z_n) \tag{4}$$

2.3 SSDU: ResNet and Conjugate Gradient Descent

SSDU employs an iterative algorithm that alternates between enforcing data consistency and leveraging a pre-trained ResNet model, distinctively operating without the need for fully sampled reference data. In medical image reconstruction, ResNet has been adapted to enhance image quality from under-sampled data, reduce artifacts, and improve the sharpness and clarity of reconstructed images. Its ability to learn residual mappings has been particularly effective in iterative reconstruction algorithms, where ResNet layers can be integrated to refine predictions at each iteration, gradually improving the reconstruction quality.

The optimization is driven by an objective function, depicted in Equation [6], which utilizes a linear forward sampling operator, A_Ω, and b_Ω is the given undersampled k-space data. During training, the network harnesses a set of k-space locations defined by

$$\Theta = \Omega \setminus \Lambda \tag{5}$$

within the data consistency units. Conversely, the set denoted by Λ is instrumental in defining the loss function. The reconstructed image x is obtained through the conjugate gradient descent method, as formulated in Equation [7]. The loss function, described in Equation [8], is characterized by a normalized $\ell_1 - \ell_2$ loss, which contributes to the effectiveness of the SSDU model in image reconstruction tasks.

$$\min_x \|A_\Omega x - b_\Omega\|^2 + \lambda \|x - z_n\|^2 \quad (6)$$

$$x_{n+1} = (A_\Omega^H A_\Omega + \lambda I)^{-1}(A_\Omega^H b_\Omega + \lambda z_n) \quad (7)$$

$$loss = \frac{1}{N} \sum_i L\left(y_\Lambda^i, \hat{A}_\Lambda\left(f\left(y_\Theta^i, \hat{A}_\Theta\right)\right)\right) \quad (8)$$

3 Results

The analysis predominantly utilizes a T2-weighted dataset, assembled from 167 multi-coil brain fastMRI files, each rich in k-space data spanning various slices and coil configurations. Integral to this dataset are elements critical for accurate reconstruction: keys for the ground truth (Org), coil sensitivity maps (CSM), a selected coil number of 16 in the dataset, and Poisson's masks. These components were systematically divided into training and validation sets for the VarNet and MoDL models. For the SSDU approach, a tailored selection involving the CSM, train mask, and loss mask was employed. This diligent method was similarly employed in compiling the testing dataset. The comprehensive dataset comprises 2388 slices designated for training, 272 slices for validation, and 368 slices for testing. The evaluation process was standardized using a batch size of 1, complemented by a StepLR scheduler, facilitating a learning rate of 0.01, a step size of 10, and a gamma value of 0.1, ensuring a consistent and structured assessment framework.

3.1 VarNet: structure evaluations

The evaluation of VarNet focuses on two primary factors. Initially, it explores the impact of the unrolled gradient descent blocks, referred to as K. Subsequently, it investigates the configuration of the encoder and decoder layers within the UNET architecture, designated by N. Since the UNET's layers double the output channel size by two, it can dramatically learn the features of the images. The outcomes of this analysis, employing an x8 acceleration mask, are depicted in Figure 1, while Figure 2 provides a visual representation of the VarNet model's structure. Notably, a configuration with K set to 6 and N to 3 yields the highest PSNR value, highlighting the optimal settings for enhancing image reconstruction quality in this context.

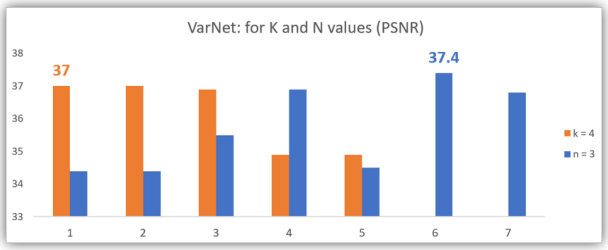


Figure 1: *VarNet: K and N values (PSNR), the model trained with an x8 acceleration mask*

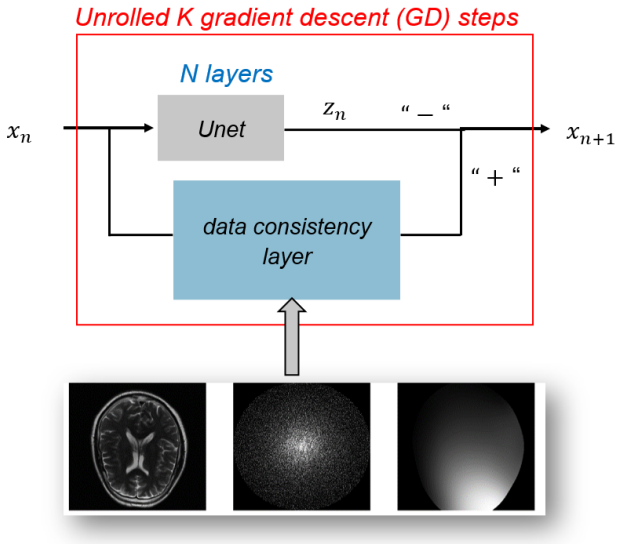


Figure 2: *VarNet*

3.2 MoDL and SSDU: structure evaluations

The assessment of MoDL and SSDU models delves into two crucial elements. Initially, it investigates the influence of the unrolled conjugate gradient descent blocks, labeled as K. Subsequently, it evaluates the convolution layers in the CNN denoiser (MoDL) or the ResNet architecture (SSDU). The findings of this exploration, utilizing an x8 acceleration mask, are presented in Figures 3 and 5, while Figures 4 and 6 offer detailed visualizations of the SSDU and MoDL structures, respectively.

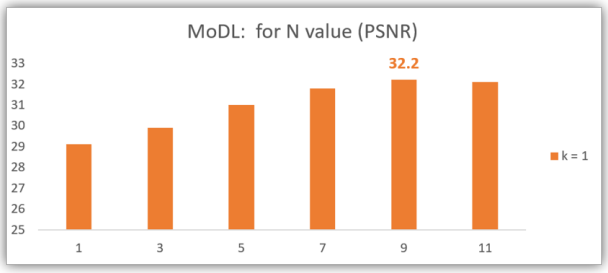


Figure 3: *MoDL: N value (PSNR), the model trained with an x8 acceleration mask*

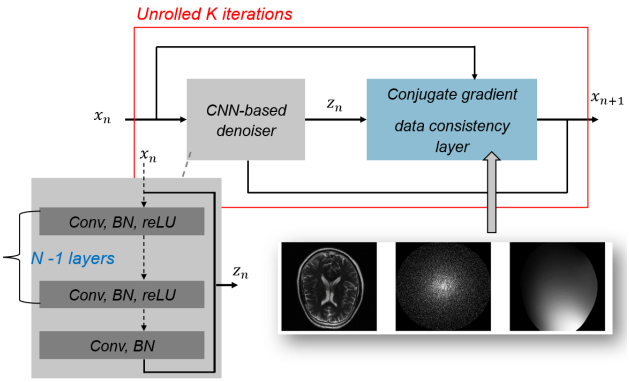


Figure 4: *MoDL*

Remarkably, a configuration with K set to 1 and N to 9 achieves the highest PSNR value for MoDL. Conversely, the ResNet-based SSDU model demonstrates the efficacy of the unrolled blocks K, with a combination of K set to 3 and N to 6 reaching a PSNR value of 34.6 for ResNet(SSDU).

Additionally, the principle of self-supervised learning inherent in SSDU can be adaptively applied to the CNN-based MoDL framework. This adaptation, with K adjusted to 5 and N maintained at 9, culminates in a PSNR performance of 35.3, surpassing the original configuration. This

highlights the flexibility and potential of integrating self-supervised learning techniques to enhance image reconstruction quality further.

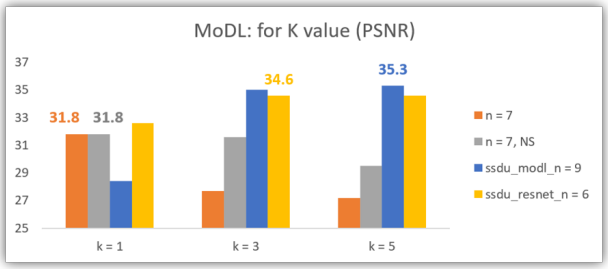


Figure 5: *MoDL,SSDU(MoDL),SSDU(ResNet): K value (PSNR), the model trained with an x8 acceleration mask*

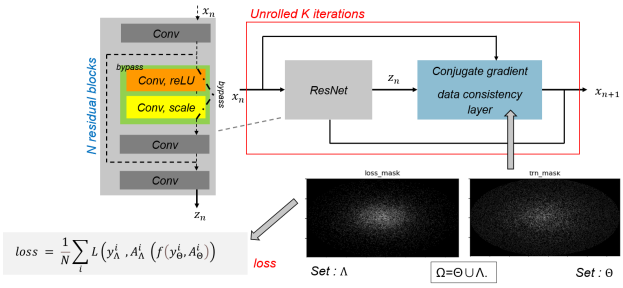


Figure 6: *SSDU(ResNet)*

3.3 Transformer: structure evaluations

The substitution of traditional UNET or CNN structures with a Vision Transformer architecture is evaluated here, as well as the concept of SSDU is also implemented into the Transformer. Figure 7 shows the results of Varnet model with Transformer replacement of the UNET. Figure 8 shows the outcomes of Transformer-based MoDL and the Transformer-based MoDL with SSDU.

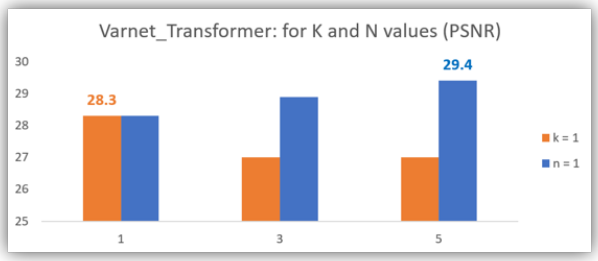


Figure 7: *Varnet Transformer: K and N values (PSNR), the model trained with an x8 acceleration mask*

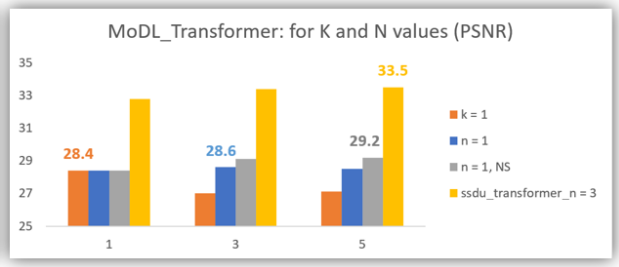


Figure 8: *MoDL Transformer and SSDU(MoDL) Transformer: K and N values (PSNR), the model trained with an x8 acceleration mask*

The proposed Transformer architecture, crafted to revolutionize MRI image reconstruction, showcases a complex assembly of modular components and advanced deep-learning techniques. Central to its design, the Transformer defines essential parameters such as the embedding size at 768 to enrich the model's understanding of each patch, alongside a strategic allocation of 12 attention heads to dissect and analyze image data from multiple perspectives. It manages a staggering 2112 patches illustrating a deeper segmentation of images for finer detail capture. The scale factor same as the patch size, set at 12, mirrors this expansion, facilitating a precise

upscaling process. Embedded within this architecture, the ”VisionTransformer module” leverages these configurations, processing patch embeddings augmented with positional information to preserve spatial relationships. The ”FeatureMapToImage” module stands out for its sophisticated approach to transforming encoded features back into images, utilizing a series of convolutional layers, normalization, and activation functions to refine the output. Integral functions like ”change shape” and ”input embeddings” ensure a seamless transition between image space and embedded patches, optimizing the architecture's efficacy. Through its design, incorporating a notable embedding size of 768, 12 attention heads, and managing an expanded 2112 patches, this architecture enhanced image quality in the application of Transformer models of medical imaging.

3.4 Reconstruction images and tables

This section presents the reconstructed images from the training phase of four different models: MoDL, VarNet, a ResNet-based model with SSDU, and MoDL with SSDU, each under an x8 acceleration mask, as depicted in Figures 9 and 10. These images offer a visual comparison of each model's capability to reconstruct high-fidelity medical scans from undersampled data. Additionally, the variation in image clarity and detail among the models provides insights into their respective performance and potential clinical applicability.

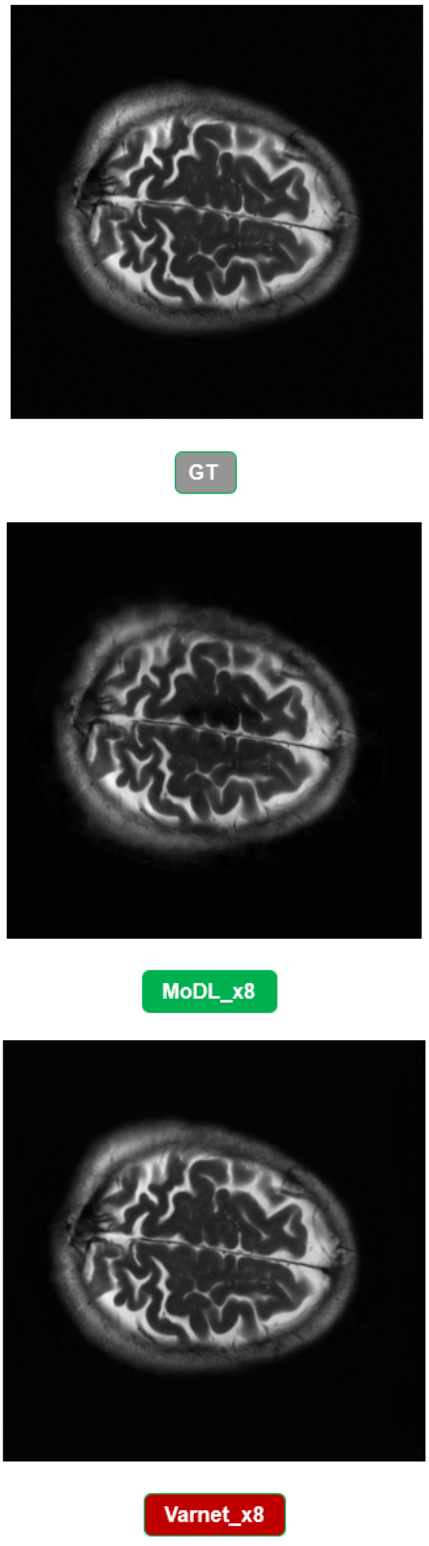
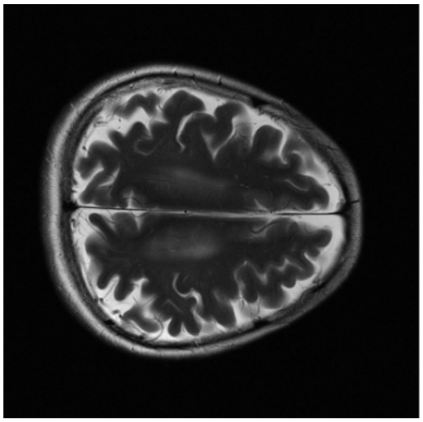
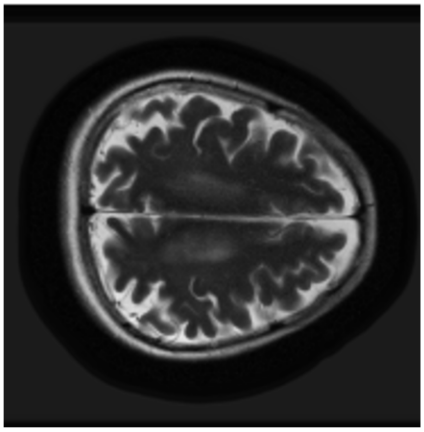


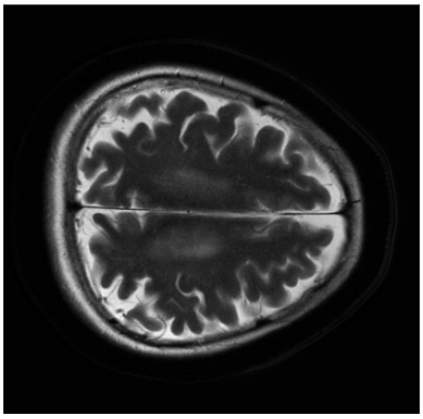
Figure 9: *Training images of MoDL and Varnet, the model trained with an x8 acceleration mask*



GT



SSDU_x8 (MoDL)



SSDU_x8 (ResNet)

Figure 10: *Training images of SSDU(MoDL) and SSDU(ResNet), the model trained with an x8 acceleration mask*

The comparisons of PSNR and SSIM performance between the various models are displayed in Figure 11, which generally shows that the VarNet model possesses the highest reconstruction capability, followed by MoDL with SSDU.

MoDL

Val_x8	PSNR	Test_x8	PSNR	SSIM
k = 1 n = 9	**32.5**		**24.6**	**0.8**

Val_x4	PSNR	Test_x4	PSNR	SSIM
k = 1 n = 9	**33.0**		**24.8**	**0.8**

MoDL with SSDU

Val_x8	PSNR	Test_x8	PSNR	SSIM
k = 6 n = 9	**36.2**		**36.7**	**0.86**

MoDL Transformer-based with SSDU

Val_x8	PSNR
k = 5 n = 3	**33.5**

x4/x8: denotes the model trained with a x4/x8 acceleration mask

Varnet

Val_x8	PSNR	Test	PSNR	SSIM
k =6 n = 3	**38.0**	x4	36.9	0.86
		x8	**37.6**	**0.87**
		x12	36.2	0.84

Val_x4	PSNR	Test	PSNR	SSIM
k =6 n = 3	**39.1**	**x4**	**38.2**	**0.88**
		x8	37.4	0.87
		x12	36	0.85

Varnet Transformer-based

Val_x8	PSNR
k = 5 n = 1	**29.4**

PSNR, SSIM of models

Figure 11: *PSNR and SSIM Performance comparison*

Total trainable parameters

MoDL_T : MoDL with Transformer

	MoDL	VarNet	MoDL_T	SSDU
K, N = 1, 1	2.5 K	100 K	9500 K	110 K
K, N = 5, 1	2.5 K	500 K	9500 K	110 K
K, N = 1, 5	114 K	31000 K	31000 K	410 K

Figure 12: *Model's trainable parameters*

The required trainable parameters for each model are listed in Figure 12, demonstrating that even the model with the simplest structure, where both K and N equal one, has a significantly higher count of trainable parameters compared to other models.

4 Discussion

The supervised learning model VarNet demonstrates superior performance, achieving PSNR values above 38.0 in fastMRI image reconstruction, leveraging the UNET architecture with a manageable number of trainable parameters. Additionally, the study explores the integration of self-supervised learning via data undersampling (SSDU) with the UNET structure. However, simulation results indicate that the SSDU approach may not be compatible with the UNET structure employed in the VarNet model.

Conversely, the SSDU methodology exhibits promising results when paired with CNN-based, ResNet-based, or Transformer-based models. This approach, especially when utilizing weight sharing across unrolled blocks, leads to a significant reduction in the number of trainable parameters and diminishes the risk of overfitting.

As for the Transformer architecture, its implementation in this study did not result in improved image reconstruction quality when replacing the UNET and CNN models. Despite attempts to optimize through adjustments of K and N values, no enhancement in PSNR was observed. This lack of improvement could be due to an inadequate exploration of the Transformer architecture, particularly in terms of patch size and embedding dimensions. Therefore, a deeper investigation into these specific aspects is warranted to fully exploit the Transformer's potential in the domain of medical imaging.

5 Conclusion

In conclusion, the investigation into MRI image reconstruction models—VarNet, MoDL, and SSDU—reveals distinct outcomes regarding their performance and adaptability to different structures such as UNET, CNN, ResNet, and the Transformer. The VarNet model, leveraging a UNET-based approach, demonstrates superior performance, achieving the highest PSNR values, which underscores the effectiveness of its variational network in enhancing image quality with a reasonable number of trainable parameters. The SSDU model, particularly when combined with CNN and

ResNet architectures, showcases the viability of self-supervised learning in scenarios with limited data, achieving commendable image reconstruction quality.

However, the introduction of the Transformer architecture in place of traditional UNET and CNN components did not result in improved image reconstruction quality. Despite variations in K and N values, PSNR did not exhibit any enhancement. This suggests that the Transformer architecture's potential is yet to be fully realized, likely due to an insufficient exploration of its capabilities, especially in terms of patch size and embedding dimensions within the context of this study.

Thus, while VarNet and SSDU (when integrated with CNN and ResNet) present promising avenues for advancing MRI image reconstruction, the application of Transformer architecture necessitates further examination. Future research should focus on refining Transformer configurations, particularly emphasizing not only patch size and embedding dimensions but also the dimension of the feed-forward network, the number of attention heads, and the dropout rate. Optimizing these parameters is key to unleashing the full potential of Transformers in enhancing MRI image quality. Such exploration is crucial for advancing the field of computational imaging, paving the way for more precise and efficient medical imaging technologies.

References

[1] Mani M. P. Jacob M. Aggarwal, H. K. Modl: Model based deep learning architecture for inverse problems. *IEEE Transaction on Medical Imaging*, 38(2):394 – 405, 2018. doi:10.1109/TMI.2018.2865356.

[2] Klatzer T. Kobler E. Recht M. P. Sodickson D. K. Pock T. Knoll F. Hammernik, K. Learning a variational network for reconstruction of accelerated mri data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018. doi:10.1002/mrm.26977.

[3] Hossein Hosseini S. A. Moeller S. Ellermann J. Uğurbil K. Akçakaya M. Yaman, B. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84(6):3172–3191, 2020. doi:10.1002/mrm.28378.

[1] [2] [3]