Compared Reconstruction of MRI Data based on Model-Based Deep Learning (MoDL), Variational Network (VarNet) and Self-supervised Learning via Data Undersampling (SSDU) Architecture

Author: Chen Jinying
Supervisor: Dr. Zhengguo Tan

1. Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique widely used in clinical diagnosis. However, acquiring high-quality MRI images usually involves lengthy scanning and complex data processing, limiting its widespread application in clinical practice. To tackle this issue, deep learning-based image reconstruction methods have gained significant attention and research. This paper focuses on comparing and analyzing three main deep learning models: Model-Based Deep Learning (MoDL) [1], Variational Network (VarNet) [2], and Self-supervised learning via data undersampling (SSDU) [3], to compare their difference in MRI image reconstruction.

MoDL is a model-based approach for image reconstruction that uses a regularization prior based on convolution neural networks (CNNs) [1]. This proposed framework integrates deep learning with the capabilities of model-based reconstruction techniques. We utilize a variational framework incorporating a trained CNN and a data-consistency term to efficiently capture image redundancy. Additionally, an alternating recursive technique is employed, leading to the construction of a deep network upon unfolding. This network comprises data consistency blocks, which promote alignment with measurements, and interleaved CNN blocks, facilitating the retrieval of relevant information from the image dataset. For simpler cases like single-channel MRI recovery, analytical solutions for the quadratic sub-problem addressed by the data consistency block are available [4]. In more complex scenarios such as multichannel MRI, we recommend employing conjugate gradient (CG) optimization to solve the quadratic sub-problem [1].

Variational network (VarNet) combines the mathematical structure of variational models with deep learning [2]. We want to recreate clinical accelerated multi-coil MR data quickly and with excellent quality using VarNet. In this VarNet model, an unrolled gradient descent system contains a generalized compressed sensing reconstruction that is formulated as a variational model. Through an offline training process, all the parameters of this formulation—including the previous model specified by filter kernels and activation functions as well as the data term weights—are learned. After learning it, the model can be used online with data that has never been seen before.

We propose a self-supervised method, which we call self-supervised learning via data undersampling (SSDU) [3]. It divides the obtained k-space indices into two disjoint sets. The DC unit for the network uses one of these, and the loss function in k-space is defined by the other set. As a result, all assumptions regarding image output or characteristics are left out when training and evaluating the network end-to-end using only the measurements that have been obtained [3].

In conclusion, this paper reproduces the codes of MoDL [1], VarNet [2] and SSDU [3] in MRI reconstructed images, describes their respective methods as well as compares and analyzes their reconstructed images of brain MRI.

2. Methods

2.1. Model-Based Deep Learning (MoDL)

We express the image reconstruction of $x \in C^n$ as the optimization problem:

$$x_{rec} = arg\min_x \underbrace{\|A(x) - b\|_2^2}_{data\ consistency} + \lambda \underbrace{\|N_w(x)\|^2}_{regularization}. \quad (1)$$

In this context, N_w represents a CNN estimator of noise and alias patterns, which is dependent on the learned parameters w. We denote $N_w(x)$ as:

$$N_w(x) = (I - D_w)(x) = x - D_w(x). \quad (2)$$

Here, $D_w(x)$ represents the "denoised" version of x, obtained after eliminating alias artifacts and noise. Utilizing the CNN-based prior $\|N_w(x)\|^2$, which yields high values when x is affected by noise and alias patterns, leads to solutions that maintain data consistency and are minimally influenced by noise and alias patterns. Here, λ is a trainable regularization parameter. By substituting equation (2) into equation (1), we derive:

$$x_{rec} = arg\min_x \|A(x) - b\|_2^2 + \lambda\|x - D_w(x)\|^2. \quad (3)$$

As these approaches depend on forward models, the receptive field of the networks does not necessarily need to cover the entire image size. Moreover, given that the network's objective is to capture redundancies in the images, a network with substantially fewer parameters is adequate to achieve satisfactory results.

We observe that the non-linear mapping $D_w(x_n + \Delta x)$ can be estimated by employing a Taylor series expansion centered around the n-th iteration:

$$D_w(x_n + \Delta x) \approx \underbrace{D_w(x_n)}_{z_n} + J_n^T \nabla x, \quad (4)$$

where J_n is the Jacobian matrix. Setting $x_n + \Delta x = x$, the penalty term can be approximated as

$$\|x - D_w(x_n + \nabla x)\|^2 \approx \|x - z_n\|^2 + \|J_n \Delta x\|^2 \quad (5)$$

It is observed that the second term tends towards zero for small perturbations (i.e., for small values of $\|\nabla x\|$). Given that the above approximation is only valid in the vicinity of x_n, we derive the alternating algorithm that approximates equation (3):

$$x_{n+1} = arg\min_x \|A(x) - b\|_2^2 + \lambda\|x - z_n\|^2, \quad (6a)$$

$$z_n = D_w(x_n) \quad (6b)$$

The sub-problem (6a) can be addressed by solving the normal equations:

$$x_{n+1} = \underbrace{(A^H A + \lambda I)^{-1}}_{Q} (A^H(b) + \lambda z_n) \quad (7)$$

The procedure starts with the initialization of $z_0 = 0$. The overview of the iterative framework is illustrated in Fig 1B.

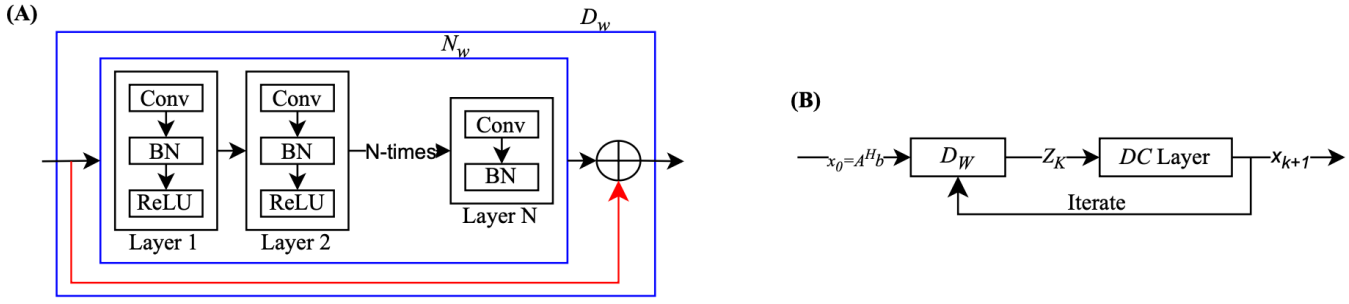

Fig 1. Model-Based Deep Learning (MoDL) framework for image reconstruction. (A) is the CNN based denoising block D_w. (B) shows the recursive MoDL framework, where the denoising block D_w and data-consistency (*DC*) layer alternate.

It illustrates the CNN architecture employed in this study (Fig 1A). For the implementation of the N_w block, we utilized an N-layer model with 64 filters at each layer. Each layer consists of a rectified linear unit (ReLU) as the non-linear activation function ($f(x) = \max(0, x)$), followed by batch normalization (BN) and convolution (Conv). Notably, the N-th layer excludes ReLU to prevent truncation of the negative part of the learned noise patterns. The reconstructed image, serving as the output of the D_w block, is derived by applying the learned noise from the N_w block to its input, following the residual learning technique. Subsequently, this output is fed into the data consistency (*DC*) layer (Fig 1B).

The recursive model proposed is shown in Fig 1B. Specifically, we configured the number of iterations k to be 1, 2, or 10, and the number of layers N to be 5. The data consistency (*DC*) layer operates explicitly with complex inputs and yields a complex output since MR images are complex. The CNN component manages complex data by concatenating

the real and imaginary parts as channels, thus transitioning from $\mathbb{C}^{m\times n}$ space to $\mathbb{R}^{m\times n\times 2}$ space.

In MoDL model, we used mean-squared error (MSE) as our chosen loss function:

$$\mathrm{MSE}(x,y)=\frac{1}{N}\sum_{i=1}^{N}(x_i-y_i)^2, \quad (8)$$

where N is the number of samples, x_i is the i-th element in the reconstructed image, and y_i is the i-th element in the ground truth image.

2.2. Variational Network (VarNet)

The VarNet defined by equation (9) with these parameters: filter kernels K_i^t, activation functions $\Phi_i^{t\prime}$, and data term weights λ^t.

$$x^{t+1}=x^t+\sum_{i=1}^{N_k}(K_i^t)^T\Phi_i^{t\prime}(K_i^t x^t)-\lambda^t A^*(Ax^t-f), \quad 0\le t\le T-1. \quad (9)$$

Here, x is a reconstructed image, f is the given undersampled k-space data with noisy, where missing data are padded by zeros.

The VarNet comprises T gradient descent iterations. To generate a reconstruction, the undersampled k-space data, coil sensitivity maps, and the zero-filling solution are inputted into the VarNet. Given the complex-valued nature of the images, separate filters K_i^t are learned for the real and imaginary planes. The non-linear activation function $\Phi_i^{t\prime}$ combines the filter responses from both feature planes. Throughout the training process, the filter kernels, activation functions, and data term weights λ^t are iteratively learned.

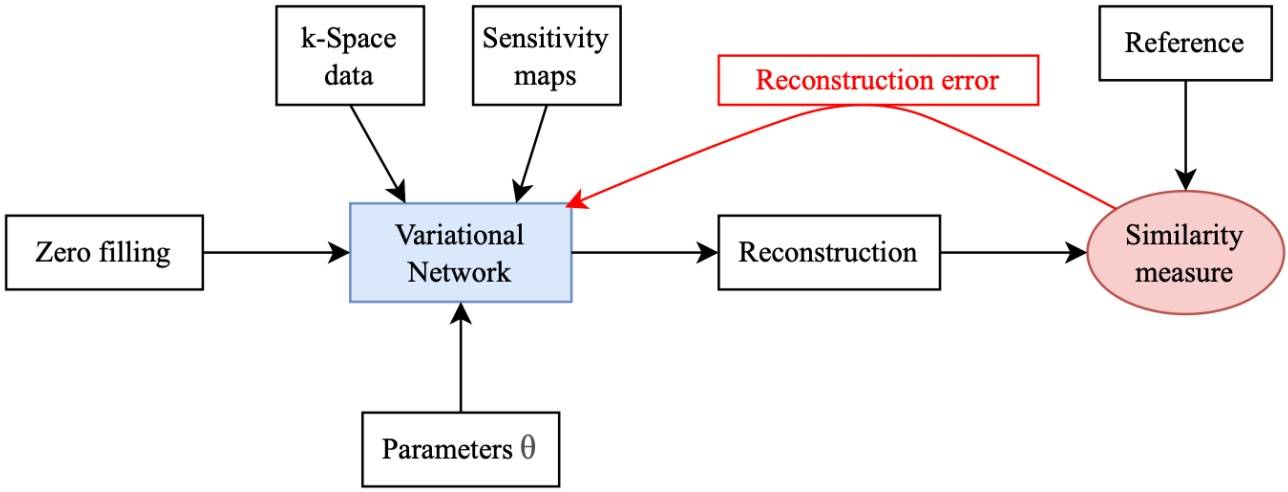

Fig 2. The training process for variational networks (VarNet). During an offline training process, the goal is to learn a set of parameters of the VarNet. To achieve this, we use a similarity measure to compare the current reconstruction image to an artifact-free reference. In order to compute an update of parameters, we propagate the reconstruction error back to the VarNet using this data.

During the offline training process (Fig 2), the objective is to determine an optimal parameter set $\theta=\{\theta^0,\dots,\theta^{T-1}\}$, $\theta^t=\{w_{ij}^t,k_i^t,\lambda^t\}$ for our proposed VarNet in Equation (9). To establish the training procedure, we aim to minimize a loss function over a set of images N with respect to the parameters θ. The loss function quantifies the similarity between the reconstructed image x and a clean, artifact-free reference image y. In this context, we utilize the mean-squared error (MSE) as our chosen loss function:

$$\mathrm{L}(\theta)=\min_{\theta}\frac{1}{N}\sum_{i=1}^{N}(x_i(\theta)-y_i)^2, \quad (10)$$

where N is the number of samples, x_i is the i-th element in the predicted values, and y_i is the i-th element in the true values.

2.3. Self-supervised learning via data undersampling (SSDU)

In this paper, the SSDU model operates on a principle similar to that of the MoDL model. However, a notable distinction arises as we substitute the representation of the 'denoised' version of x in the network structure, typically denoted as D_w (Fig 1B), with ResNet (Fig 3A).

The neural network utilized in this study was constructed using a convolutional neural network (CNN) architecture based on ResNet. This CNN (Fig 3B) comprised an input layer, an output layer, and 15 residual blocks interconnected by skip connections, which facilitate the flow of information during network training. Each residual block contained two convolutional layers, with the first layer being followed by a rectified linear unit (ReLU), and the second layer being followed by a constant multiplication layer (xScale) with a factor of C = 0.1 (Fig 3C). All layers had a kernel size of 3 × 3 and consisted of 64 channels.

In SSDU model, a normalized $\ell_1 - \ell_2$ loss was used:

$$\mathcal{L}(u, v) = \frac{\|u-v\|_2}{\|u\|_2} + \frac{\|u-v\|_1}{\|u\|_1}. \quad (11)$$

In this supervised configuration, the fully sampled k-space and the network output k-space are denoted by u and v, respectively. A fully sampled encoding operator is employed to convert network output images to k-space, resulting in this network output k-space. On the other hand, in the suggested self-supervised training approach, u and v represent the acquired k-space measurements at the locations indicated by the loss mask, as well as the k-space corresponding to the network output image at those same places.

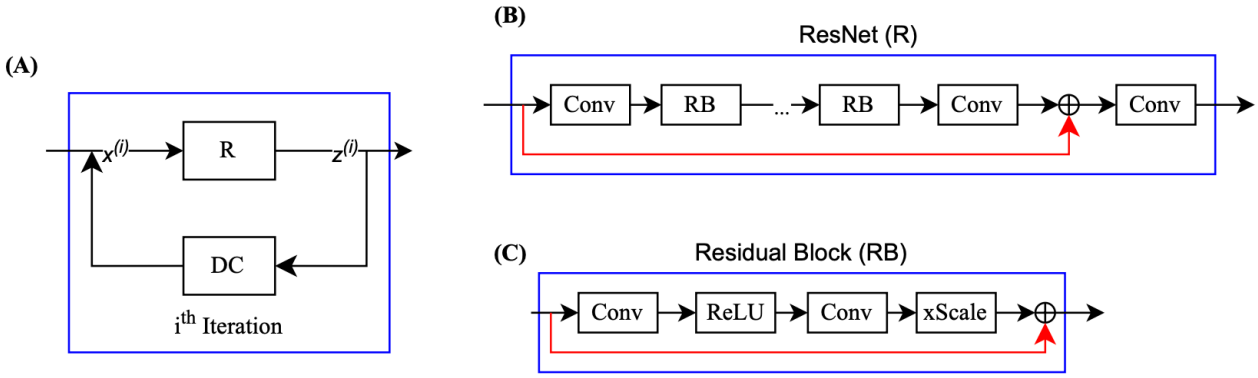

Fig 3. Self-supervised Learning via Data Undersampling (SSDU) framework for image reconstruction. (A) shows the description of a traditional iterative optimization approach for regularized inverse reconstruction issues. Data consistency (DC) and regularization (R) are the two techniques that alternate in these algorithms. (B) is the framework of the ResNet (R). With 3 convolution layers overall and 15 residual blocks (RBs), the ResNet (R) architecture employed as the regularizer in this work. (C) shows the structure of the residual block (RB). It contains two convolution layers, the first of which is followed by a rectified linear unit (ReLU) and the second of which is followed by a constant multiplication layer.

2.4. Data processing

2.4.1. FastMRI Dataset

Data used in the preparation of this article were obtained from the NYU fastMRI Initiative database (fastmri.med.nyu.edu) [5]. This raw dataset includes axial T1 weighted, T2 weighted and FLAIR images.

2.4.2. Choice of Mask

In MoDL model, we generated two kinds of mask: poisson mask (Fig 4A) and cartesian mask (Fig 4B). In the study, the poisson mask was employed to sample each image slice using a Poisson distribution with a parameter set to 8. Similarly, the cartesian mask was utilized to sample each image slice at a 4x acceleration rate.

In VarNet model, we chose poisson mask with the parameter set to 8 (Fig 4A).

In SSDU model, we need to use two masks (Fig 4C). The trn mask (Fig 4C left) and the loss mask (Fig 4C right) are two distinct sets generated using the proposed SSDU technique from the acquired subsampled data. The trn mask serves

as the initial set of indices utilized within the data consistency unit of the unrolled network, while the loss mask is chosen from the acquired k-space points to define the loss function. A uniformly random selection from the elements of the acquired k-space locations was employed to establish the distribution of the loss mask. During training, the output of the network is converted to k-space, and the resulting reconstructed k-space values are compared with the subset of available measurements at the loss mask. Subsequently, the network parameters are adjusted based on this training loss.

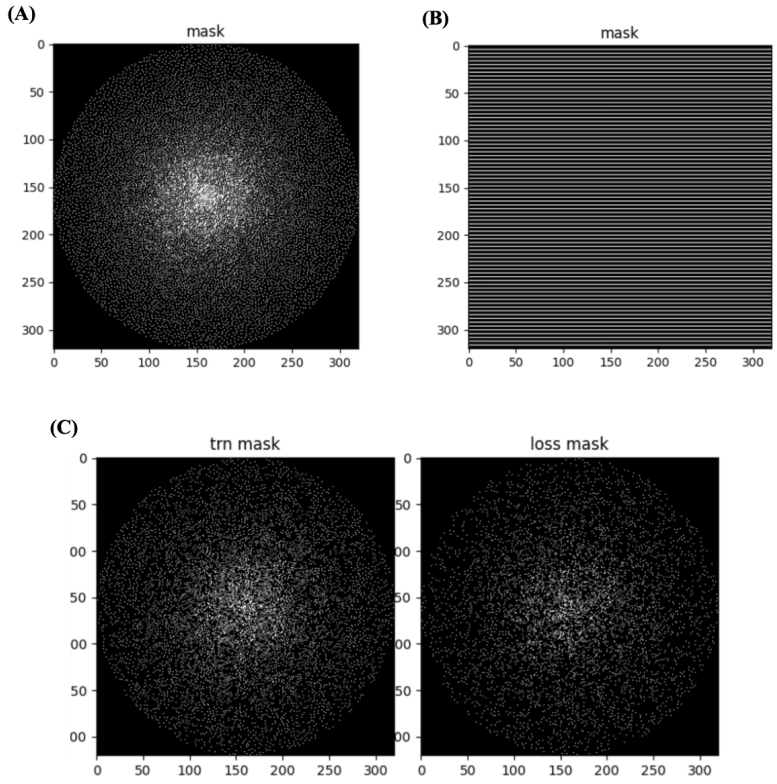

Fig 4. Mask Demonstration. (A) is poisson mask and it is used in MoDL and VarNet model. (B) is cartesian mask and it is used in MoDL model. (C) The trn mask and loss mask are generated by poisson mask and they are used in SSDU model. The trn mask is utilized in the data consistency unit of the SSDU network, while the loss mask defines the training loss function. To compare the accessible subset of measurements at loss mask with the corresponding reconstructed k-space values, the output of the network is converted to k-space during training. The network parameters are then changed in response to this training loss.

2.5. Metrics

2.5.1. Peak Signal-to-Noise Ratio (PSNR)

The Peak Signal-to-Noise Ratio (PSNR) quantifies the relationship between the maximum possible image intensity across a volume and the power of distortion noise and other errors:

$$PSNR(x,y) = 10\log_{10}\frac{\max(y)^2}{MSE(x,y)}. \quad (12)$$

In the provided equation, x represents the reconstructed volume, y denotes the target volume, $\max(y)$ signifies the largest entry in the target volume y, $MSE(x,y)$ represents the mean square error between x and y, defined as $\frac{1}{n}\|x-y\|_2^2$, where n is the number of entries in the target volume y. Higher values of PSNR indicate superior reconstruction quality.

2.5.2. Structural Similarity (SSIM)

The Structural Similarity (SSIM) index evaluates the similarity between two images by exploiting the interdependencies among neighboring pixels. SSIM inherently assesses the structural characteristics of objects within an

image and is computed across various image locations using a sliding window. The resultant similarity between two image patches x (reconstructed volume) and y (target volume) is defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}, \quad (13)$$

where μ_x and μ_y are the means of x and y respectively, σ_x^2 and σ_y^2 are the variances of x and y respectively, σ_{xy} is the covariance between x and y, C_1 and C_2 are constants added for stability, typically set to $(k_1L)^2$ and $(k_2L)^2$ respectively, where k_1 and k_2 are parameters controlling the influence of contrast and structure similarity, and L is the dynamic range of pixel values. In this paper, we set $k_1 = 0.01$ and $k_2 = 0.03$, $L = \max(y) - \min(y)$.

3. Results

All model used Adam as optimizer and the learning rate was set to 0.001. There were 240 training data and 48 test data in T2 dataset, 234 training data and 42 test data in T1 dataset, and also 240 training data and 48 test data in FLAIR dataset.

We compare the effect of using different masks on the reconstruction of T2, T1, and FLAIR images in MoDL model (Fig 5). When the number of iterations k is 1 and the number of epochs is 50, by comparing PSNR and SSIM, as well as reconstructing the image, the results are better using the poisson mask, where T1 has a better score (PSNR=30.5713, SSIM=0.7939), whereas the number of iterations k is 2, FLAIR reconstructed image has a better result (PSNR=25.9402, SSIM=0.7422). In a word, when we use poisson mask in the MoDL model, the reconstructed effect is better. It is possible that we need to perform more epochs to get better results when using the cartesian mask.

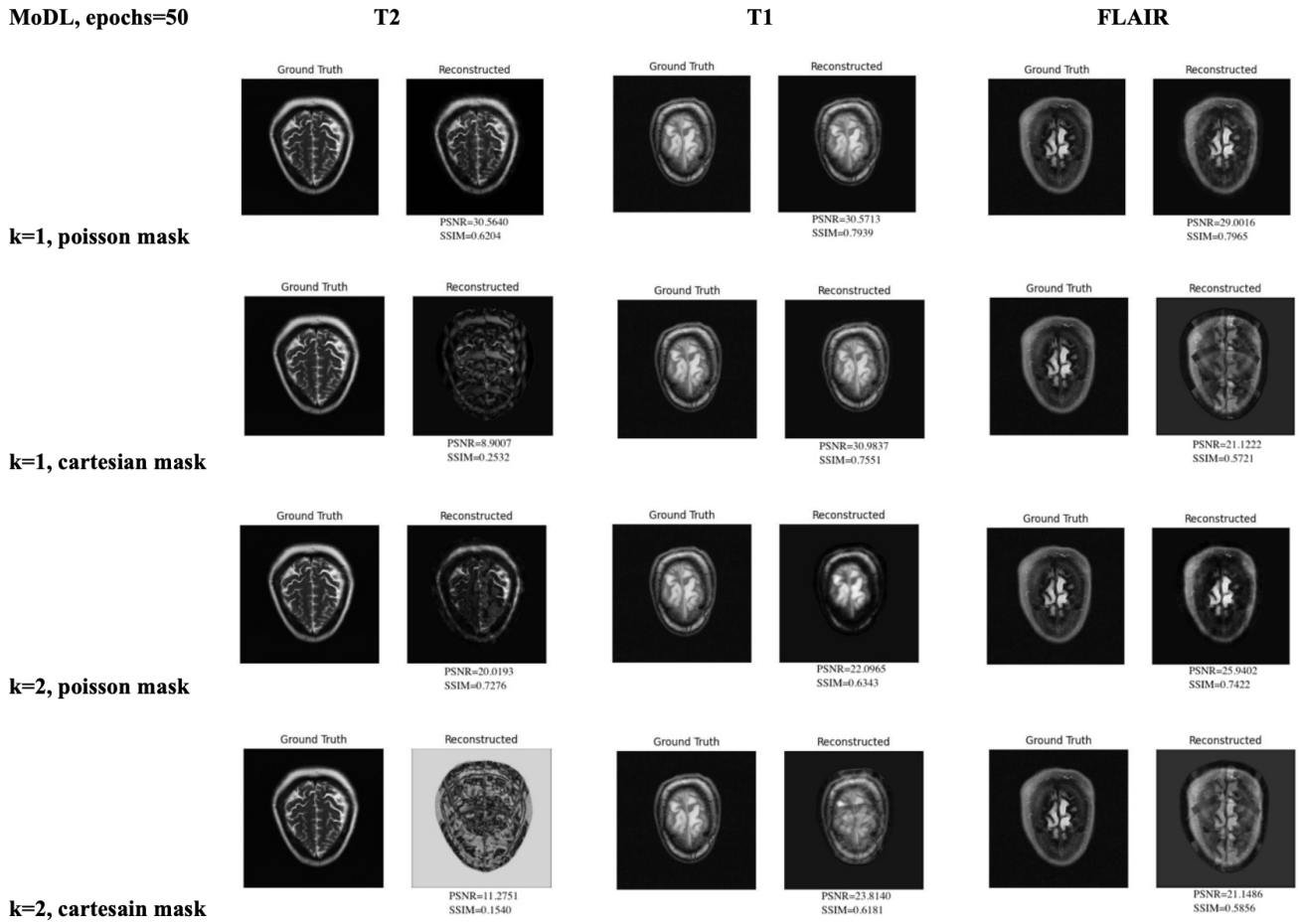

Fig 5. Reconstructed MRI images based on MoDL model with different mask (poisson mask and cartesian mask). It shows the reconstructed image of T2, T1 and FLAIR based on MoDL model with PSNR and SSIM, when the number of epochs is 50, the number of iteration k is 1 and 2 respectively, and different mask (poisson and cartesian) are used.

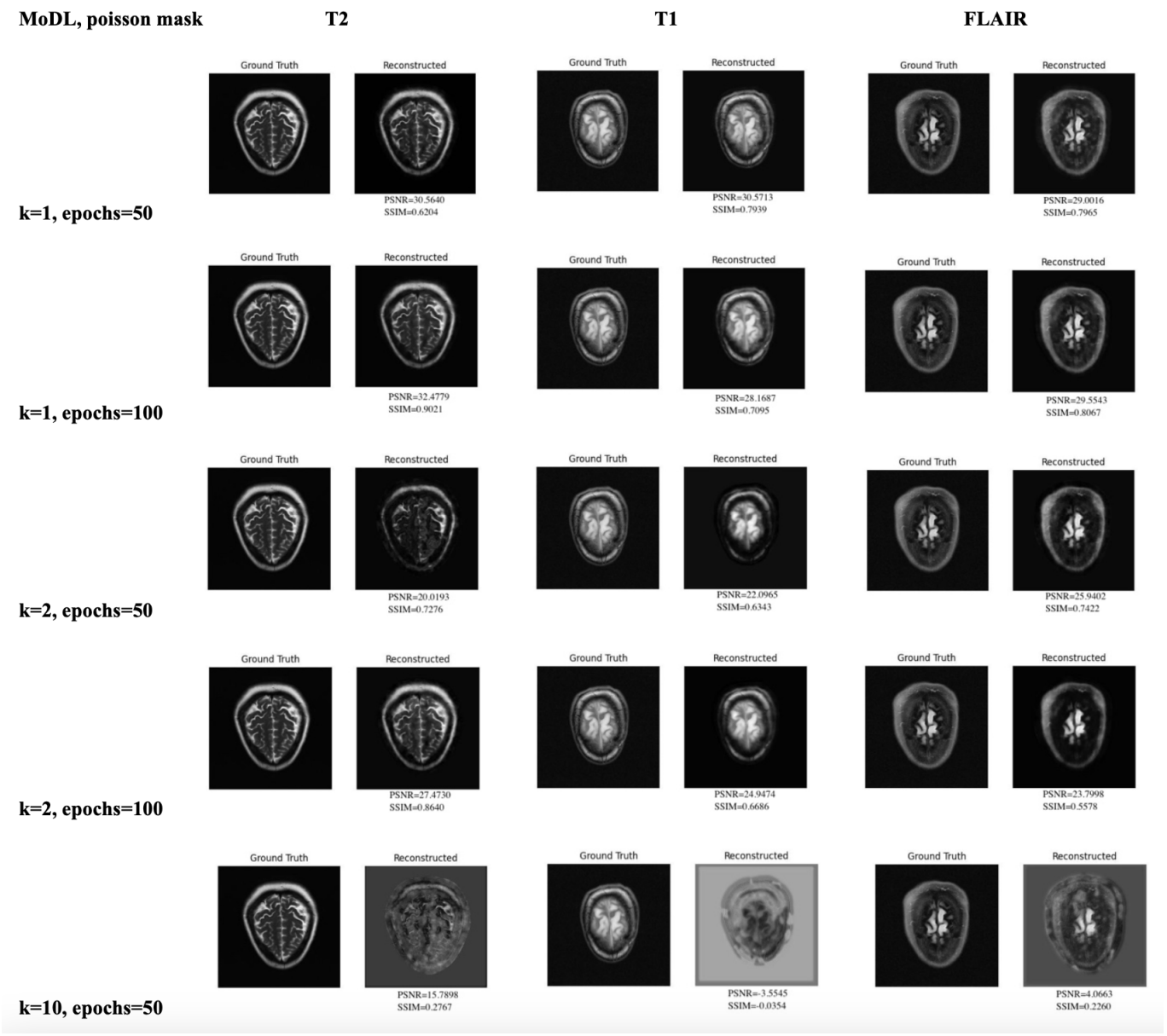

Fig 6. Reconstructed MRI images based on MoDL model with poisson mask. It shows the reconstructed image of T2, T1 and FLAIR based on MoDL model with PSNR and SSIM, when the number of epochs is 50 and 100 respectively, the number of iteration k is 1, 2 and 10 correspondingly are used.

Since better results are obtained with the poisson mask, we use the poisson mask for the remainder of the training process.

We compared the reconstruction results for different number of iterations k (k=1, 2, 10) and different number of epochs (epochs=50, 100) based on MoDL model (Fig 6). For T2 image, it has a better score when the number of iterations k is 1 and the number of epochs is 100 (PSNR=32.4779, SSIM=0.9021). For T1 image, we obtain a better result when the number of iterations k is 1 and the number of epochs is 50 (PSNR=30.5713, SSIM=0.7939). Whereas, for FLAIR image, the reconstruction is better when the number of iterations k is 1 and the number of epochs is 100 (PSNR=29.5543, SSIM=0.8067).

Moreover, we additionally evaluated the reconstruction outcomes across various numbers of iterations (k = 1, 2, 10) and different numbers of epochs (epochs = 50, 100) using the VarNet model (Fig 7). For T2 dataset, they perform well when the number of epochs is 100 and the number of iterations is 1 (PSNR=34.9859, SSIM=0.9156) and 2 (PSNR=34.9828, SSIM=0.9156), respectively. For T1 dataset, it has better reconstruction when the number of iterations

k is 10 and the number of epochs is 50 (PSNR=31.9726, SSIM=0.7906). For FLAIR dataset, we obtain a better score when the number of iteration is 2 and the number of epochs is 100 (PSNR=32.4424, SSIM=0.8351).

In general, satisfactory outcomes were achieved for all kinds of images reconstructed using VarNet, with T2 exhibiting the most favorable performance (PSNR>33.5, SSIM>0.9).

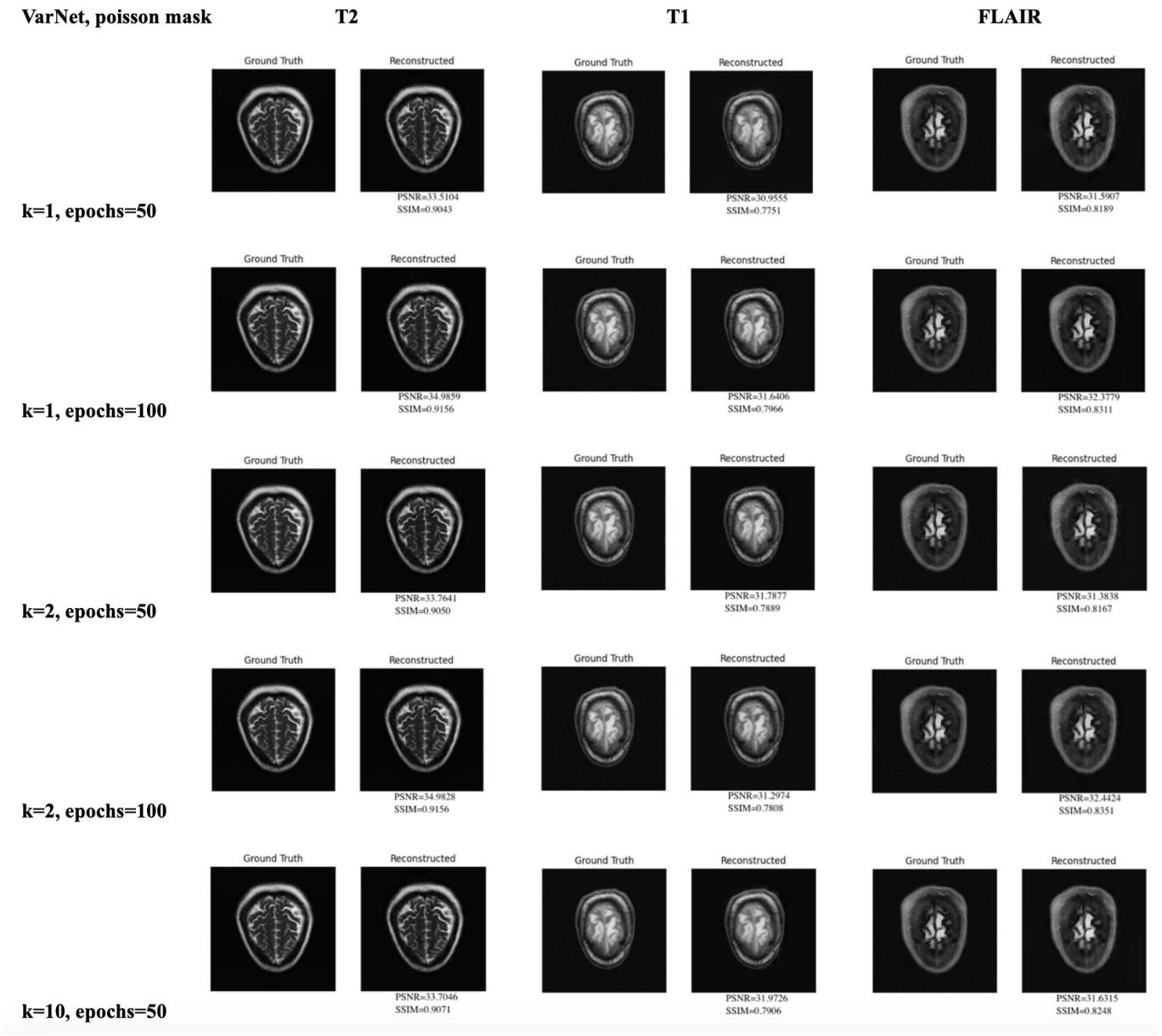

Fig 7. Reconstructed MRI images based on VarNet model with poisson mask. It shows the reconstructed image of T2, T1 and FLAIR based on VarNet model with PSNR and SSIM, when the number of epochs is 50 and 100 respectively, the number of iteration k is 1, 2 and 10 correspondingly are used.

Furthermore, we conducted additional evaluations of the reconstruction results utilizing the SSDU model across the same numbers of epochs (epochs = 50) and varying numbers of iterations (k = 1, 2, 10) (Fig 8). When the number of iterations is 2, we obtain a better reconstructed image based on T2 dataset (PSNR=18.0507, SSIM=0.4135). In the T1 dataset, it has a better PSNR score when the number of iterations is 1 (PSNR=18.4057), whereas it has a better SSIM score when the number of iterations k is 2 (SSIM=0.4591). For FLAIR image, it performs well when the number of iterations k is 10 (PSNR=20.7633, SSIM=0.6222).

Overall, training with FLAIR data in the SSDU model produces superior reconstructed images when the number of iterations is 10 and the number of epochs is 50(PSNR=20.7633, SSIM=0.6222).

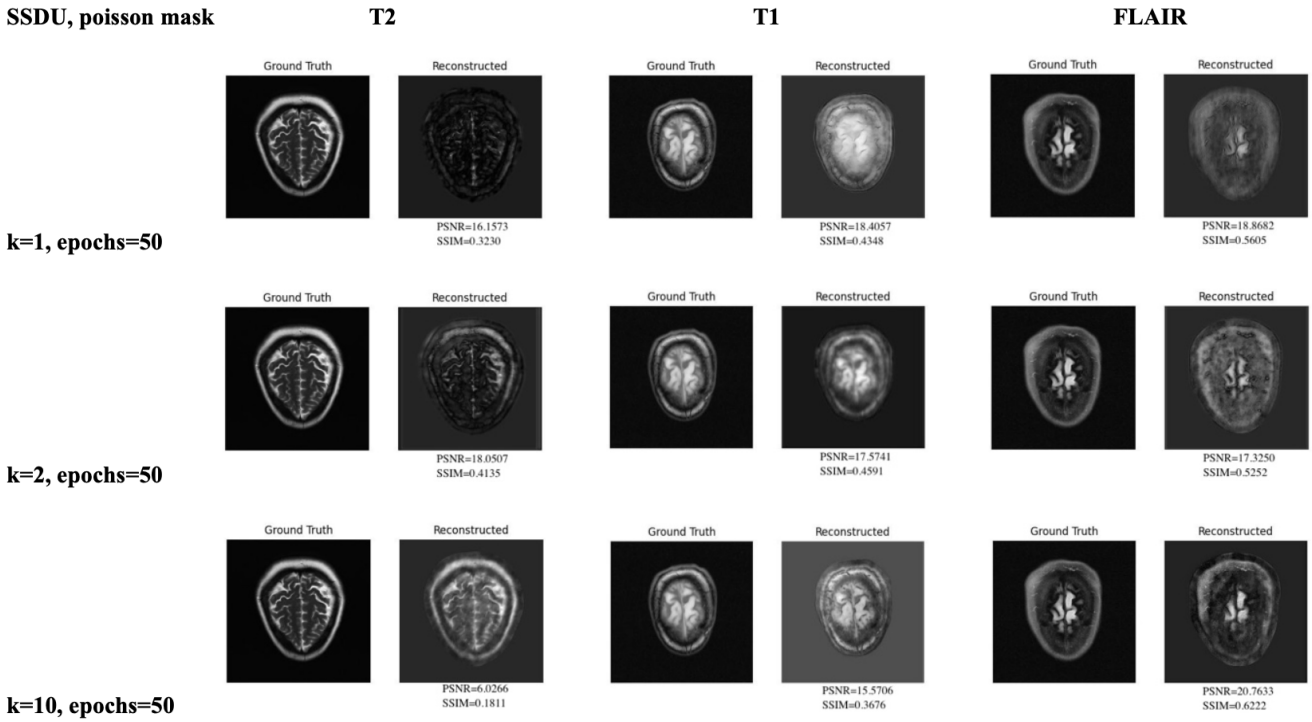

Fig 8. Reconstructed MRI images based on SSDU model with poisson mask. It shows the reconstructed image of T2, T1 and FLAIR based on SSDU model with PSNR and SSIM, when the number of epochs is 50, the number of iteration k is 1, 2 and 10 respectively are used.

4. Discussion

In order to reconstruct MRI images, we investigated the efficacy of several reconstruction models in this work, including MoDL, VarNet, and SSDU. According to our findings, the quality of reconstructed images is highly influenced by the reconstruction model and hyperparameter selection.

First of all, we noticed that the performance of MoDL model was significantly influenced by the type of mask selected, specifically the poisson mask. The poisson mask demonstrated a consistent ability to capture image redundancy and minimize reconstruction artifacts when compared to the cartesian mask in a variety of iterations and epochs. For the purpose of improving reconstruction results, we subsequently trained using the poisson mask.

Additionally, we obtained good PSNR and SSIM scores with a limited number of epochs and iterations in our VarNet model studies, particularly for the T2 dataset. As a consequence, high-fidelity reconstructions are produced by VarNet efficiently utilizing the structural information found in the MRI data. The necessity for meticulous optimization was highlighted by inconsistent performance of VarNet across datasets and hyperparameter settings.

Moreover, MRI image reconstruction using the SSDU model demonstrated promise, especially when trained using FLAIR data. Across several iterations and datasets, its performance showed some variation. It could be possible to improve the performance of SSDU by looking into hyperparameter optimization and modifying the architecture.

5. Conclusion

In conclusion, we show that different deep learning models work well for reconstructing MRI images. The MoDL model consistently produced high-quality reconstructions across several datasets when trained with a poisson mask. Particularly for the T2 dataset, VarNet produced good outcomes, suggesting that it has potential for practical applications. Through training using FLAIR data, SSDU demonstrated competitive performance in the meantime. Overall, our results show the significance that model selection and hyperparameter adjusting are to the best MRI image reconstruction algorithms when designing them for clinical applications. Clinical outcomes and diagnostic accuracy in medical imaging may be enhanced by more study in this field.

6. References

[1] H. K. Aggarwal, M. P. Mani and M. Jacob, "MoDL: Model-Based Deep Learning Architecture for Inverse Problems," in IEEE Transactions on Medical Imaging, vol. 38, no. 2, pp. 394-405, Feb. 2019, doi: 10.1109/TMI.2018.2865356.

[2] Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, Knoll F. Learning a variational network for reconstruction of accelerated MRI data. Magn Reson Med. 2018 Jun;79(6):3055-3071. doi: 10.1002/mrm.26977. Epub 2017 Nov 8. PMID: 29115689; PMCID: PMC5902683.

[3] Yaman B, Hosseini SAH, Moeller S, Ellermann J, Uğurbil K, Akçakaya M. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magn Reson Med. 2020; 84: 3172–3191. https://doi.org/10.1002/mrm.28378

[4] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for MR image reconstruction,” in Proc. Inf. Process. Med. Imag., 2017, pp. 647–658.

[5] Knoll et al Radiol Artif Intell. 2020 Jan 29;2(1):e190007. doi:10.1148/ryai.2020190007. (https://pubs.rsna.org/doi/10.1148/ryai.2020190007), and the arXiv paper: https://arxiv.org/abs/1811.08839