UC San Diego

Comparison of E2E-VarNet and MoDL for Accelerated MRI Reconstruction

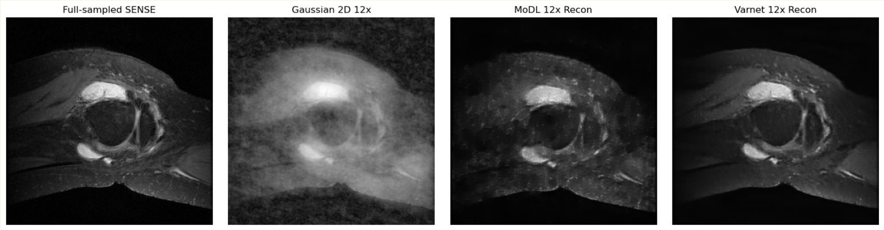
Project 2

BENG 280A Group 11

Ziyuanchang Zhang, Zhaowei Zhang

Project Synopsis

- Background
  - Deep learning has rapidly advanced MRI reconstruction. Deep unfolding networks (DUNs) unroll iterative optimization algorithms into trainable layers, enabling fast, learned reconstruction.
- Objective
  - Evaluate and compare two DUN architectures—MoDL and End-to-End VarNet—for MRI reconstruction performance.
- Method
  - Used ATOMMIC toolbox
  - Tested pretrained MoDL and E2E-VarNet on Stanford Knees 2019 dataset
  - Compared performance using quantitative metrics (PSNR, SSIM, MSE, NMSE) and qualitative image inspection.
- Results
  - E2E-VarNet outperformed MoDL across all metrics, with higher PSNR/SSIM and lower MSE/NMSE.
  - Visual inspection confirmed cleaner images, fewer artifacts, and sharper anatomical details for E2E-VarNet.
- Conclusion
  - E2E-VarNet's superior performance is likely due to its higher-capacity U-Net, k-space update strategy, and learned sensitivity maps, whereas MoDL is constrained by shallow CNNs, weight sharing, image-space update and fixed ESPIRiT maps.
  - E2E-VarNet provides significantly higher-fidelity MRI reconstructions.



Metric	Meaning	MoDL (Baseline)	VarNet (Ours)	Improvement
PSNR (dB)	***Signal Clarity***	**23.73**	**31.31**	**+7.58 dB** ⬆️
SSIM (0-1)	***Structural Perception***	**0.581**	**0.762**	**+31%** ⬆️
NMSE	***Normalized Error***	**0.1856**	**0.0388**	**-79% Error** ⬇️
MSE	***Mean Sq. Error***	**0.0049**	**0.0011**	**-77% Error** ⬇️

UC San Diego

Background - Parallel Imaging

- An MR scanner measures k-space data with receiver coils, and the image is recovered by applying an inverse Fourier transform to those measurements.
- Modern MRI scanners use multiple receiver coils, each capturing k-space data weighted by its spatial sensitivity.

$$\mathbf{k}_i = \mathcal{F}(S_i\mathbf{x}) + \epsilon_i, i = 1, 2, \ldots, N,$$

$$\tilde{\mathbf{k}_i} = M\mathbf{k}_i$$

- MRI can be speed up by undersampling k-space, but directly inverse-Fourier transforming the undersampled data produces aliasing artifacts.
- Parallel imaging accelerates MRI by using k-space samples redundancies, but requires a sufficiently large fully sampled Auto-Calibration Signal(ACS) region to estimate sensitivity maps, which limits the maximum acceleration.

UC San Diego

Background - Compressed Sensing

- Compressed Sensing enables reconstruction of images by using fewer k-space measurements by enforcing suitable priors. Classical compressed sensing methods solve the following optimization problem:

$$\begin{aligned}\hat{\mathbf{x}} &= \operatorname{argmin}_{\mathbf{x}} \frac{1}{2}\sum_i \left\| M\mathcal{F}(S_i\mathbf{x}) - \tilde{\mathbf{k}_i} \right\|^2 + \lambda\Psi(\mathbf{x}) \\ &= \operatorname{argmin}_{\mathbf{x}} \frac{1}{2}\left\| A(\mathbf{x}) - \tilde{\mathbf{k}} \right\|^2 + \lambda\Psi(\mathbf{x}),\end{aligned}$$

- The problem can be solved by iterative gradient descent methods:

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \eta^t \left(A^*(A(\mathbf{x}) - \tilde{\mathbf{k}}) + \lambda\Phi(\mathbf{x}^t) \right)$$

UC San Diego

Background - Deep unfolding networks

- Deep Unfolding Networks (DUNs) are a class of network architectures that combine traditional optimization algorithms with deep learning models.
- Deep Unfolding Networks unroll iterative optimization algorithms (e.g., gradient descent) into trainable neural-network layers, enabling faster, learned versions of traditional optimization procedures.
- A classical example of a deep unfolding architecture is the Variational Network (VarNet):

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \eta^t A^*(A(\mathbf{x}^t) - \tilde{\mathbf{k}}) + \mathrm{CNN}(\mathbf{x}^t),$$

UC San Diego

Model Architecture - End-to-End Variational Network

- End-to-end VarNet is an enhanced version of VarNet that:
  - Replace the shallow CNN with a higher-capacity U-Net to enable richer feature extraction and multi-scale modeling.
  - Use k-space intermediate variables **kt** for reconstruction instead of image-space intermediate variables **xt**.
  - Incorporate a Sensitivity Map Estimation (SME) module based on a U-Net to learn coil sensitivity maps, rather than relying on fixed maps computed by the ESPIRiT algorithm.

$$\mathbf{k}^{t+1} = \mathbf{k}^t - \eta^t M(\mathbf{k}^t - \tilde{\mathbf{k}}) + G(\mathbf{k}^t)$$

where G is the refinement module given by:

$$G(\mathbf{k}^t) = \mathcal{F} \circ \mathcal{E} \circ \mathrm{CNN}(\mathcal{R} \circ \mathcal{F}^{-1}(\mathbf{k}^t)).$$

$$\mathbf{x}^t = \mathcal{R} \circ \mathcal{F}^{-1}(\mathbf{k}^t).$$

UC San Diego

Model Architecture - End-to-End Variational Network

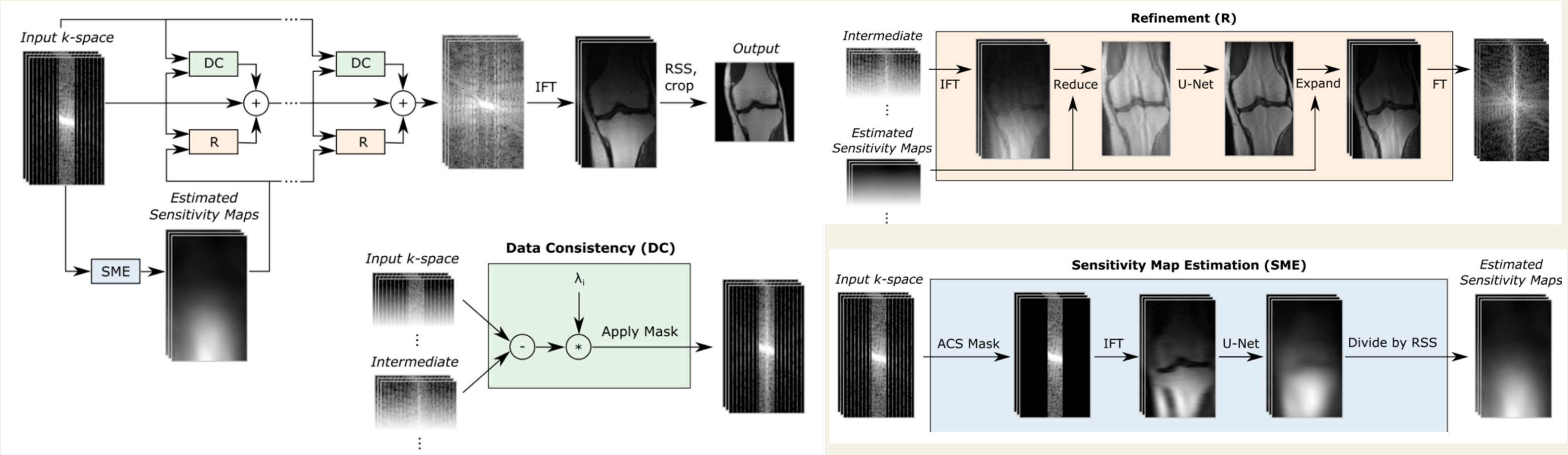


Fig. 1[3]: **Left: Block diagram** of our model which takes under-sampled k-space as input and applies several cascades, followed by an inverse Fourier transform (IFT) and an RSS transform. The **Data Consistency (DC)** module computes a correction map that brings the intermediate k-space closer to the measured k-space values. The **Refinement (R)** module maps multi-coil k-space data into one image, applies a U-Net, and then back to multi-coil k-space data. The **Sensitivity Map Estimation (SME)** module estimates the sensitivity maps used in the Refinement module.

UC San Diego

Model Architecture - Model Based Deep Learning

- Model Based Deep Learning is another deep unfolding architecture model that formulate the reconstruction of the image x as the optimization problem:

$$\mathbf{x}_{\mathrm{rec}} = \arg\min_{\mathbf{x}} \underbrace{\|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|_2^2}_{\text{data consistency}} + \lambda \underbrace{\|\mathcal{N}_{\mathbf{w}}(\mathbf{x})\|^2}_{\text{regularization}} .$$

$$\mathcal{N}_{\mathbf{w}}(\mathbf{x}) = (\mathcal{I} - \mathcal{D}_{\mathrm{w}})(\mathbf{x}) = \mathbf{x} - \mathcal{D}_{\mathbf{w}}(\mathbf{x}).$$

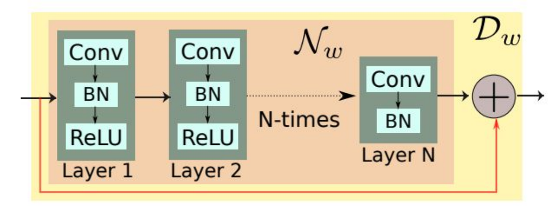
$$\Rightarrow \quad \mathbf{x}_{\mathrm{rec}} = \arg\min_{\mathbf{x}} \|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x} - \mathcal{D}_{\mathbf{w}}(\mathbf{x})\|^2$$

which can be solved by the iterative algorithm:

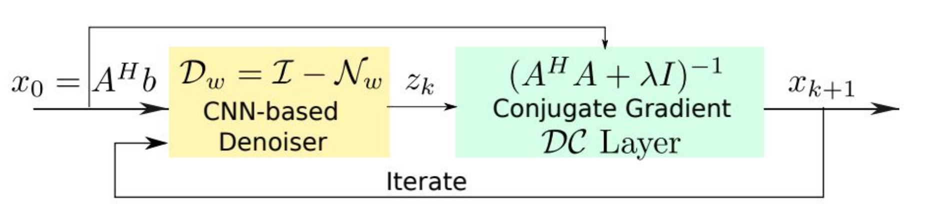
$$\mathbf{x}_{n+1} = \arg\min_{\mathbf{x}} \|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x} - \mathbf{z}_n\|^2,$$
$$\mathbf{z}_n = \mathcal{D}_{\mathbf{w}}(\mathbf{x}_n)$$

$$\mathbf{x}_{n+1} = \underbrace{\left(\mathcal{A}^H \mathcal{A} + \lambda \mathcal{I}\right)^{-1}}_{\mathcal{Q}} \left(\mathcal{A}^H(\mathbf{b}) + \lambda \mathbf{z}_n\right)$$

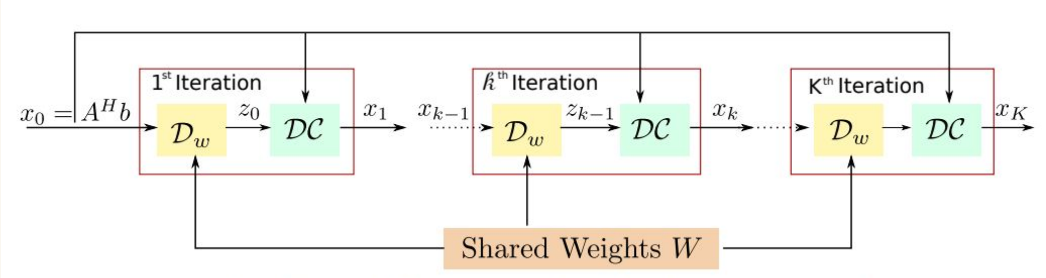
Model Architecture - Model Based Deep Learning



(a) The Residual learning based denoiser



(b) Proposed Model-based Deep Learning (MoDL) architecture



(c) Unrolled architecture as described in Eq. 4 and 5.

Fig. 2[2]: **MoDL: Proposed Model-based Deep Learning framework** for image reconstruction. (a) shows the **CNN based denoising block Dw**. (b) is the recursive MoDL framework that alternates between **denoiser Dw** and the **data-consistency (DC) layer.** (c) is the **unrolled architecture for K iterations**. The denoising blocks **Dw share the weights across all the K iterations.**

UC San Diego

Dataset——Stanford Fully-sampled 3D FSE Knee 2019

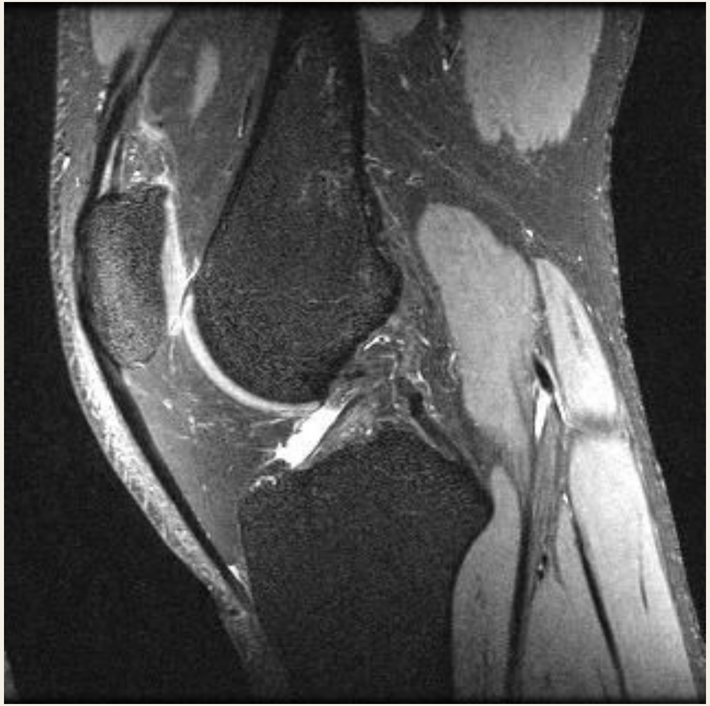


Fig 3: Sample Image from the Stanford 3D Knee MRI Dataset[4]

- The dataset contains 20 fully sampled 3D knee MRI volumes with ISMRMRD raw data stored stored in HDF5 format.
- Provides 8 coil k-space data acquired at 3T with a fast spin echo sequence.
- It is used for accelerated MRI reconstruction, compressed sensing, etc.
- The fully sampled data serve as high-quality ground truth for evaluating accelerated MRI reconstruction.

UC San Diego

Experiments

1. Set up the ATOMMIC[5] environment
2. Load and preprocess the dataset:
   a. Convert ISMRMRD raw data into HDF5 format.
   b. Extract multi-coil k-space, estimate coil sensitivity maps, and generate SENSE and RSS reference images.
   c. Split the dataset into training (70%), validation (15%), and testing (15%).
3. Undersampling Simulation:
   a. Create a 2D Gaussian 12× undersampling mask.
   b. Apply the mask to fully sampled k-space
   c. Reconstruct the masked data using iFFT + RSS
4. Run MoDL and VarNet inference using updated configuration files and pretrained checkpoints.
5. Perform evaluation: visualize reconstructed slices and compare metrics.

UC San Diego

Evaluation Metrics - MSE and NMSE

- Mean Squared Error(MSE) calculates the average of the squared differences between each pixel value of the reconstructed image (yi) and the reference image (xi)
- A smaller MSE value indicates less error between the reconstructed image and the original image, suggesting better reconstruction performance

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2,$$

- Normalized Mean Square Error(NMSE) is the mean square error normalized by the sum of the squares of all pixel values of the original image.
- When image signal intensities vary, NMSE is more meaningful than MSE, as it measures the relative deviation of the reconstructed result from the ground truth.

$$NMSE = \sum_{i=1}^{N}(x_i - y_i)^2 / \sum_{i=1}^{N} x_i^2$$

UC San Diego

Evaluation Metrics - PSNR and SSIM

- Peak Signal-to-Noise Ratio(PSNR) evaluates image quality by calculating the ratio between the maximum possible signal and the noise within the image.
- Higher PSNR values typically indicate smaller pixel-wise differences between the reconstructed image and the original image, suggesting better image quality.

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

- Structural Similarity Index(SSIM) is primarily used to measure the structural similarity between the reconstructed image and the original image.
- Unlike traditional pixel-level metrics, SSIM provides a quality measure that better aligns with human visual perception by considering the image's brightness, contrast, and sharpness.

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

UC San Diego

Results - Qualitative Comparison

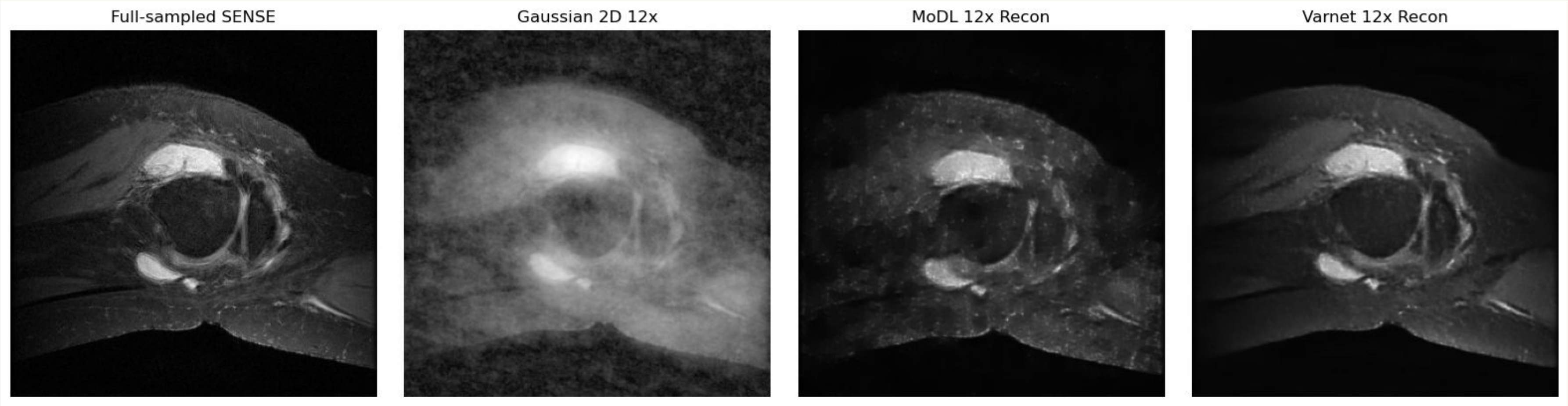


Fig 4: Comparison of End-to-End Variational Network (VarNet) versus Model-Based Deep Learning (MoDL) on a highly accelerated knee MRI scan (R=12x)
(generated using Group11_code.ipynb, lines 308-332.)

UC San Diego

Results - Quantitative Comparison

Metric	Meaning	MoDL (Baseline)	VarNet (Ours)	Improvement
PSNR (dB)	*Signal Clarity*	**23.73**	**31.31**	**+7.58 dB⬆️**
SSIM (0-1)	*Structural Perception*	**0.581**	**0.762**	**+31% ⬆️**
NMSE	*Normalized Error*	**0.1856**	**0.0388**	**-79% Error ⬇️**
MSE	*Mean Sq. Error*	**0.0049**	**0.0011**	**-77% Error ⬇️**

Conclusion

- **VarNet uses a higher-capacity U-Net backbone,**

  allowing it to capture complex anatomy and recover fine details, while MoDL's shallow CNN behaves like a denoiser and oversmooths structures.

- **VarNet performs updates in k-space rather than only in image space,**

  which preserves measured high-frequency information; MoDL's image-space updates cannot guarantee this and therefore blur edges.

- **VarNet learns coil sensitivity maps instead of using fixed maps,**

  improving PI conditioning and reducing artifacts, whereas MoDL relies on fixed ESPIRiT maps and cannot correct map errors

- **Future work may shift from denoising-based priors to learnable optimizers,**

  embedding physical constraints more deeply for better fidelity in highly ill-posed settings.

References

[1] R. Heckel, M. Jacob, A. Chaudhari, O. Perlman, and E. Shimron, "Deep learning for accelerated and robust mri reconstruction," Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 37, no. 3, pp. 335–368, 2024.

[2] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Modl: Model-based deep learning architecture for inverse problems," IEEE transactions on medical imaging, vol. 38, no. 2, pp. 394–405, 2018.

[3] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson, "End-to-end variational networks for accelerated mri reconstruction," in International conference on medical image computing and computer-assisted intervention, pp. 64–73, Springer, 2020.

[4] C. FSE'XL, P. PD, and F. FAT, "Creation of fully sampled mr data repository for compressed sensing of the knee," in SMRT 22nd Annual Meeting, Salt Lake City, Utah, USA, Citeseer, 2013.

[5] D. Karkalousos, I. Iˇsgum, H. Marquering, and M. W. Caan, "Atommic: An advanced toolbox for multitask medical imaging consistency to facilitate artificial intelligence applications from acquisition to analysis in magnetic resonance imaging," 2024.

Thank you!

UC San Diego