Doctoral oral defense

Candidate	Guanxiong Luo
Title	Development of Advanced Generative Priors for MRI Reconstruction
First Supervisor	Prof. Dr. Martin Uecker
Second Supervisor	Prof. Dr. Markus Haltmeier
Third Supervisor	Prof. Dr. Philipp Wieder
Date	Nov 10, 2023

Outline

- **Background and motivation**
- Bayesian MRI reconstruction using diffusion priors
- Phase augmentation for training priors
- Speed up MR scans with generative prior
- Summary and outlook

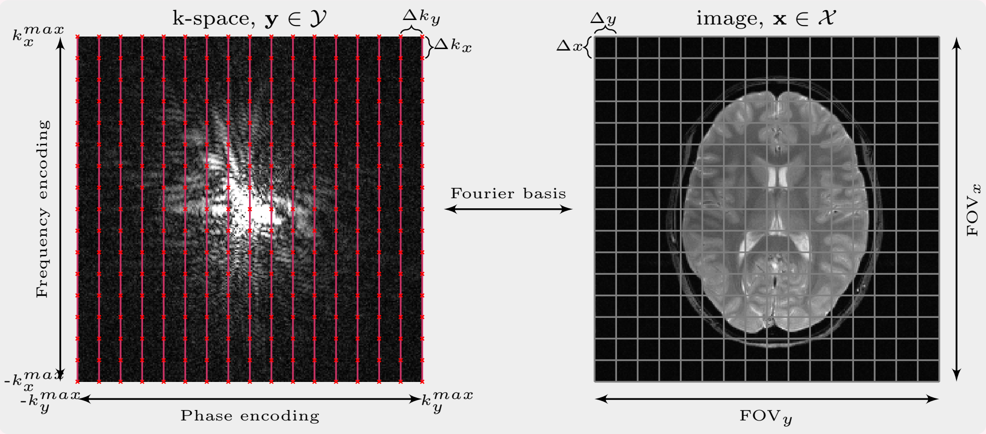
MR scans have revolutionized clinical practice in numerous ways

- non-invasive, radioactive-free
- organs, tissues and skeletal system
- multi-contrast, high-resolution
- functional imaging
- ...

Created by DALL·E·3

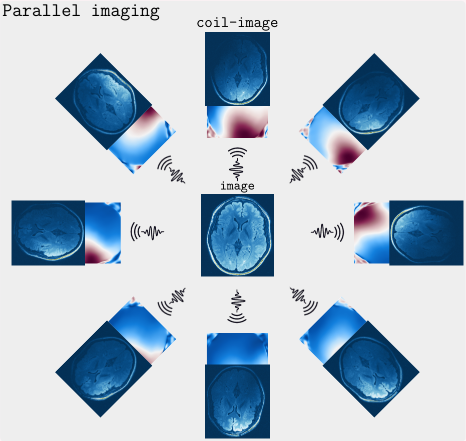
Form the image in k-space

Nyquist Theorem at least requires: $\mathrm{FOV}_x = 1/\Delta k_x,\ \Delta x = 1/(2 * k_x^{max}),\ \mathrm{FOV}_x = \mathrm{N} * \Delta x$.



Accelerate MR scans

- High-performance hardware and well-designed sequence
- Parallel imaging[1] and compressed sensing MRI[2]
- Sampling pattern
- Coil sensitivities estimation[3,4]



[1] Blaimer et al. MRM 2004. [2] Lustig et al. MRM 2007. [3] McKenzie et al. MRM 2002. [4] Uecker et al. MRM 2014

Accelerate MR scans

The reconstruction is formulated as an inverse problem

$$F(\mathbf{x}, \mathbf{c}) := (\mathcal{F}_S(\mathbf{x} \odot c_1), \cdots, \mathcal{F}_S(\mathbf{x} \odot c_N)) = \mathbf{y}$$

The least-squared error estimation is

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{y} - F_{\mathbf{c}}\mathbf{x}\|_2^2$$

- $\mathcal{F}_S$ is an undersampled Fourier transform operator
- Regularization: L1 in wavelet domain, total variation
- Joint estimation via nonlinear inversion

The application of machine learning in fast MRI[1,2,3]

Many unrolling neural networks are trained in a supervised way to solve this problem

$$\arg\min_{\mathbf{x}} \|F(\mathbf{x}, \mathbf{c}) - \mathbf{y}\|_2^2 + R_\theta(\mathbf{x}) \ . \tag{1}$$

The output of a network $F_\theta^\dagger$ parametrized by θ is trained to predict the reconstruction by maximizing the likelihood

$$\begin{aligned} \hat{\theta} &= \arg\max_\theta \mathcal{L}(X = \mathbf{x}, \hat{X} = F_\theta^\dagger(\mathbf{y})) \\ &= \arg\max_\theta \mathbb{E}_D\left[\|F_\theta^\dagger(\mathbf{y}) - \mathbf{x}\|_2^2\right] \ . \end{aligned} \tag{2}$$

A dataset $D = \{(\mathbf{x}_i, \mathbf{y}_i) | i = 1, \cdots, N\}$, consisting of paired the undersampled k-space data $\mathbf{y}$ and the reference images $\mathbf{x}$, and the pre-definition of the forward operator F are required for training.

[1]Yang et al. NIPS 2016. [2]Hammernik et al. MRM 2018. [3]Aggarwal et al. TMI(2019)

Challenges 1: Generalizability

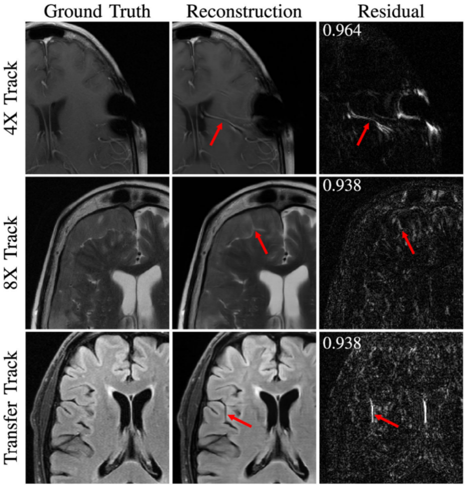
Studies[1,2] reported that tiny perturbations, small structural changes, and sampling pattern variations can fail most inverse networks $F_\theta^\dagger$.
The patterns learned in this way pertain to the pre-defined forward operator, which leads to poor generalizability to the k-space sampled with another forward operator.

[1] Antun et al. PNAS 2020. [2] Knoll et al. MRM 2019.

Challenges 2: Hallucination

Studies[1,2] reported hallucinations generated by a deep neural network $F_\theta^\dagger$.

- (top) a false-generated vessel
- (middle) a linear bright signal mimicking a cleft of cerebrospinal fluid, and blurring of the boundaries of the extra-axial mass.
- (bottom) a false-generated sulcus or prominent vessel.



[1] Muckley et al. IEEE TMI 2021. [2] Bhadra et al. IEEE TMI 2021.

Challenges 3: Data availability

- fastMRI[1], full-sampled k-space data, four types of contrast
- but to collect more is expensive and takes time

[1] Zbontar et al. Radiology 2020

Outline

- Background and motivation
- **Bayesian MRI reconstruction using diffusion priors**
- Phase augmentation for training priors
- Speed up MR scans with generative prior
- Summary and outlook

Reconstruction from Bayesian Perspective

- The image $\mathbf{x}$ conditioned on the acquired k-space data $\mathbf{y}$ is

$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{posterior}} = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \propto \underbrace{p(\mathbf{y} \mid \mathbf{x})}_{\text{likelihood}} \cdot \underbrace{p(\mathbf{x})}_{\text{prior}}, \tag{3}$$

Reconstruction from Bayesian Perspective

- The image $\mathbf{x}$ conditioned on the acquired k-space data $\mathbf{y}$ is

$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{posterior}} = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \propto \underbrace{p(\mathbf{y} \mid \mathbf{x})}_{\text{likelihood}} \cdot \underbrace{p(\mathbf{x})}_{\text{prior}}, \tag{3}$$

- $p(\mathbf{y} \mid \mathbf{x}) \rightarrow$ the probability of the acquired k-space data $\mathbf{y}$ for a given image $\mathbf{x}$
- $p(\mathbf{x}) \rightarrow$ the prior model of image $\mathbf{x}$
- The maximum a posterior estimation (MAP) gives an estimate $\hat{\mathbf{x}}$,

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \left(\log p(\mathbf{y} \mid \mathbf{x}) + \log p(\mathbf{x})\right) \tag{4}$$

The likelihood for k-space

- The likelihood $p(\mathbf{y}|\mathbf{x})$ for observing the $\mathbf{y}$ determined by $\mathbf{y} = \mathcal{A}\mathbf{x} + \boldsymbol{\eta}$
- The noise $\boldsymbol{\eta}$ is zero mean and normal distributed with covariance matrix $\sigma_\eta \mathbf{I}$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathcal{A}\mathbf{x}, \sigma_\eta^2 \mathbf{I}). \tag{5}$$

The likelihood for k-space

- The likelihood $p(\mathbf{y}|\mathbf{x})$ for observing the $\mathbf{y}$ determined by $\mathbf{y} = \mathcal{A}\mathbf{x} + \boldsymbol{\eta}$
- The noise $\boldsymbol{\eta}$ is zero mean and normal distributed with covariance matrix $\sigma_\eta \mathbf{I}$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathcal{A}\mathbf{x}, \sigma_\eta^2 \mathbf{I}). \tag{5}$$

then, we have

$$p(\mathbf{x}|\mathbf{y}) \propto \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp(-\frac{1}{2\sigma_\eta^2}\|\mathbf{y} - \mathcal{A}\mathbf{x}\|_2^2) \cdot p(\mathbf{x}) \tag{6}$$

$$\log p(\mathbf{x}|\mathbf{y}) \propto -\frac{1}{2\sigma_\eta^2}\|\mathbf{y} - \mathcal{A}\mathbf{x}\|_2^2 + \boxed{\log p(\mathbf{x})} \tag{7}$$

Non-learned and learned priors

- Laplace distribution, $p(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$, can promote sparsity in a certain domain.
- We can learn an empirical prior $p_{\theta}^{data}(\mathbf{x})$ from the i.i.d. samples $D_n = \{\mathbf{x}^1, ..., \mathbf{x}^{(n)}\}$, using generative modeling.

Non-learned and learned priors

- Laplace distribution, $p(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$, can promote sparsity in a certain domain.
- We can learn an empirical prior $p_\theta^{data}(\mathbf{x})$ from the i.i.d. samples $D_n = \{\mathbf{x}^1, ..., \mathbf{x}^{(n)}\}$, using generative modeling.

The reconstruction can be achieved by

- Maximize posterior (**MAP**)

$$\hat{\mathbf{x}}_{\text{MAP}}(\mathbf{y}) = \arg\max_{\mathbf{x}} \ \log p(\mathbf{x}|\mathbf{y}) : \arg\max_{\mathbf{x}} \left(\log p(\mathbf{y}|\mathbf{x}) + \log p_\theta^{data}(\mathbf{x})\right)$$

- Minimum mean square error (**MMSE**)

$$\hat{\mathbf{x}} = \arg\min_{\tilde{\mathbf{x}}} \int \|\tilde{\mathbf{x}} - \mathbf{x}\|^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \mathbb{E}[x|y]$$

Simulate samples from high dimensional probability

Take the image as an example

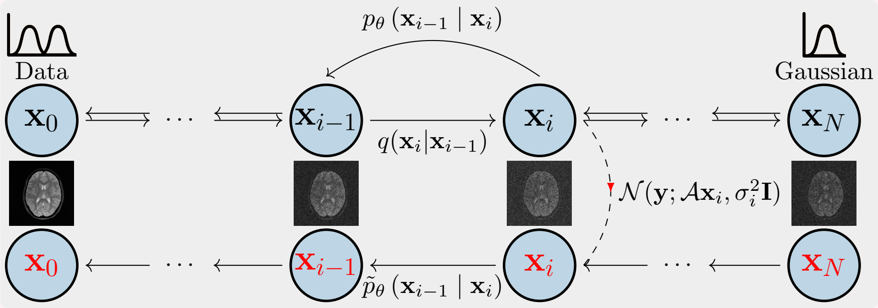
- We don't know probability density function $q(\mathbf{x})$ for them and only have images.
- Images are sparsely distributed in high dimensional space.
- For instance, a 8-bit gray scale image of size 16x16 means that there are $256^{16\times16}$ possible outcomes, but only very tiny amount of them makes sense.

Annealed Langevin Dynamics

For i in $\{\sigma_N, ..., \sigma_1\}$

$$\mathbf{x}_i^k = \mathbf{x}_i^{k-1} - \frac{\gamma}{2} \boxed{\underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{x}^{k-1})}_{\text{grad of logp w.r.t x}}} + \sqrt{\gamma\epsilon}\mathbf{z}_i \tag{8}$$

Sample the posterior



- Transform a data distribution into a known distribution gradually. (Forward)
- Learn to reverse known distribution back to data distribution. (Reversal)
- Incorporate the measurement of k-space into the learned reverse process.

Transform data distribution with forward process

Images $\mathbf{x}_0$ from data distribution $q(\mathbf{x}_0)$ is perturbed with a sequence of noise scales $\{\sigma_i\}_{i=1}^N$.

- Consider the Markov chain

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \sqrt{\sigma_i^2 - \sigma_{i-1}^2}\mathbf{z}_{i-1}, \tag{9}$$

  where $\mathbf{z}_{i-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- $q(\mathbf{x}_i|\mathbf{x}_{i-1}) = \mathcal{N}(\mathbf{x}_i; \mathbf{x}_{i-1}, (\sigma_i^2 - \sigma_{i-1}^2)\mathbf{I})$.

Learn reverse process[1,2,3]

According to Kolmogorov's backward equation, the reverse process has the same form as the forward process.

$$p_{\boldsymbol{\theta}}\left(\mathbf{x}_{i-1} \mid \mathbf{x}_i\right)=\mathcal{N}\left(\mathbf{x}_{i-1} ; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\mathbf{x}_i, i\right), \tau_i^2 \mathbf{I}\right), . \tag{10}$$

The reverse process can be learned by minimizing the KL divergence between the forward and backward.

$$\ell=\sum_{i=2}^N \mathbb{E}_{q\left(\mathbf{x}_0\right)} \mathbb{E}_{q\left(\mathbf{x}_i \mid \mathbf{x}_0\right)} D_{\mathrm{KL}}\left(\boxed{q\left(\mathbf{x}_{i-1} \mid \mathbf{x}_i, \mathbf{x}_0\right)} \| \boxed{p_{\boldsymbol{\theta}}\left(\mathbf{x}_{i-1} \mid \mathbf{x}_i\right)}\right)+C \tag{11}$$

$$=\sum_{i=2}^N \mathbb{E}_{\mathbf{x}_0, \mathbf{z}}\left[\frac{1}{\tau_i^2}\left\|\frac{\sigma_{i-1}^2}{\sigma_i^2} \mathbf{z}+\mathbf{x}_0-\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\mathbf{x}_i, i\right)\right\|_2^2\right]+C . \tag{12}$$

[1] Sohl-Dickstein et al. ICLR 2016. [2] Song et al. ICLR 2019. [3] Ho et al. ICLR 2020

Learn reverse process[1, 2]

Let:

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\mathbf{x}_i, i\right) - \mathbf{x}_0 = \sigma_{i-1}^2 \mathbf{s}_{\boldsymbol{\theta}}\left(\mathbf{x}_i, i\right), \tag{13}$$

then we have

$$\boxed{\mathbb{E}_{\mathbf{x}_0,\mathbf{z}}\left[\left\|\frac{\mathbf{x}_0 - \mathbf{x}_i}{\sigma_i^2} - \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_i, i)\right\|_2^2\right]} = \mathbb{E}_{q(\mathbf{x}_0)}\mathbb{E}_{q(\mathbf{x}_i|\mathbf{x}_0)}\left[\left\|\nabla_{\mathbf{x}_i} \log q(\mathbf{x}_i \mid \mathbf{x}_0) - \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_i, i)\right\|_2^2\right] \tag{14}$$

where $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_i, i)$ denotes the score network.

[1] Lyu Siwei. UAI 2009. [2] Song Yang et al. ICLR 2020

Compute the posterior $p(\mathbf{x}|\mathbf{y})$

- a new sequence of intermediate distributions $\tilde{p}\left(\mathbf{x}_i\right) \propto p\left(\mathbf{x}_i\right) p(\mathbf{y}|\mathbf{x}_i)$
- the modified reverse process is $\tilde{p}\left(\mathbf{x}_i \mid \mathbf{x}_{i+1}\right) \propto p\left(\mathbf{x}_i \mid \mathbf{x}_{i+1}\right) p(\mathbf{y}|\mathbf{x}_i)$
- For i in $\{\sigma_N, \ldots, \sigma_1\}$, $\mathbf{x}_i^{k+1} \leftarrow \mathbf{x}_i^k + \frac{\gamma}{2} \nabla_{\mathbf{x}_i} \log \tilde{p}(\mathbf{x}_i^k \mid \mathbf{x}_{i+1}) + \sqrt{\gamma}\mathbf{z}$

Algorithm

Algorithm 1 Sampling the posterior via Markov chain Monte Carlo

```
Give the acquired k-space y_0
Construct the forward operator 𝒜 with sampling pattern 𝒫 and coil sensitivities 𝒮
Set the Langevin steps K, Langevin step coefficient β, for the tuning factor λ, the start noise level index N.
Initialize x_N^0 with samples from a uniform distribution 𝒰_[−1,1]
for i in {N − 1, ··· , 0} do
    Draw samples from p̃(x_i | x_{i+1}) by running K Langevin steps.
end for
```

Give the acquired k-space $\mathbf{y}_0$

Construct the forward operator $\mathcal{A}$ with sampling pattern $\mathcal{P}$ and coil sensitivities $\mathcal{S}$

Set the Langevin steps K, Langevin step coefficient β, for the tuning factor λ, the start noise level index N.

Initialize $\mathbf{x}_\mathrm{N}^0$ with samples from a uniform distribution $\mathcal{U}_{[-1,1]}$

for i in $\{\mathrm{N}-1, \cdots, 0\}$ **do**

  Draw samples from $\tilde{p}(\mathbf{x}_i \mid \mathbf{x}_{i+1})$ by running K Langevin steps.

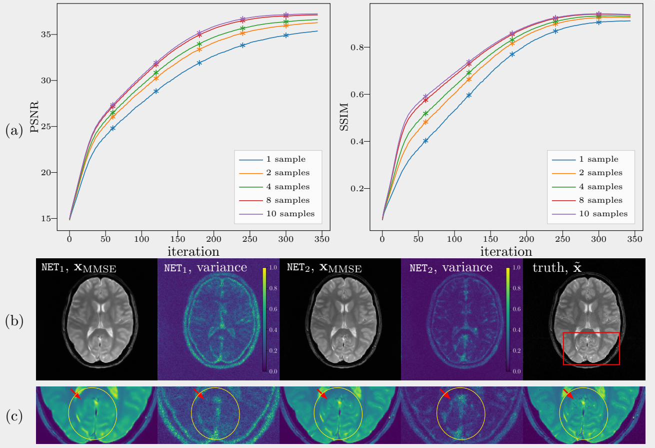
end for

Network, Data, Training

- Refine-Net. 1. conditional instance normalization layer 2. Gaussian position embedding.
- Self-attention modules are added to model high resolution images (320x320).
- Small dataset (1300 images) is used for development. FastMRI dataset is used for performance evaluation.
- $\mathtt{NET}_1$ and $\mathtt{NET}_2$ are trained with small dataset. $\mathtt{NET}_3$ is trained with fastMRI dataset
- $\mathtt{NET}_1$ uses conditional instance normalization layer (10). $\mathtt{NET}_2$ and $\mathtt{NET}_3$ use Gaussian position embedding.
- MCMC sampling algorithm is implemented with TensorFlow and Numpy.

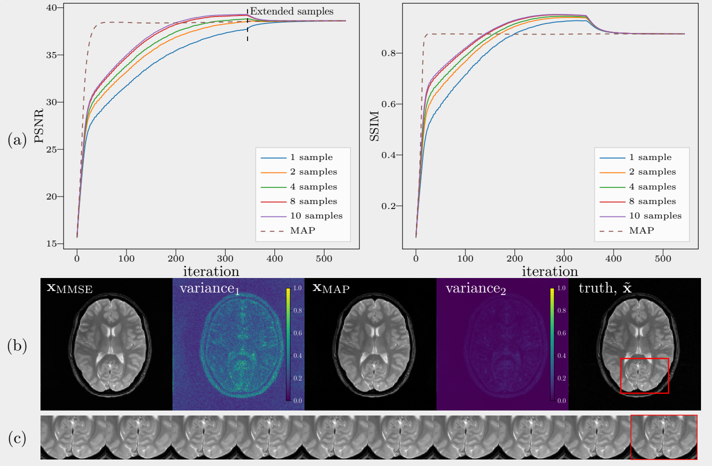
MAP VS MMSE

- variable density poisson disc, the central 20×20 region is fully acquired, and The acquisition mask covers 11.8
- 10 images were drawn. $\mathbf{x}_{\mathrm{MMSE}}$ was computed using different number of samples.
- NET_2 was used to construct transition kernels
- The parameters in Algorithm 1 are $K = 5, \beta = 1, N = 70, \lambda = 25$



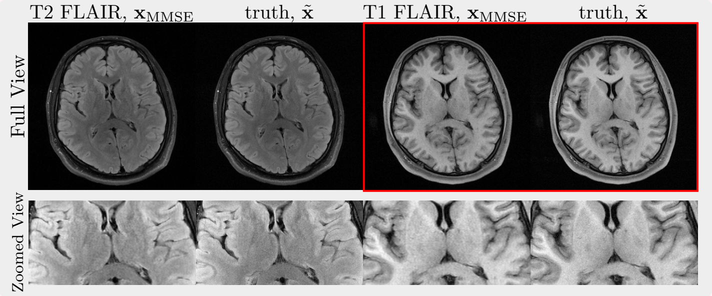
MAP VS MMSE

- Disable the disturbance of noises after random inference in last distribution $\tilde{p}(\mathbf{x}_0 \mid \mathbf{x}_1)$, run 100 iterations more to get extended samples.
- $\mathtt{NET}_2$ was used to construct transition kernels
- The parameters in Algorithm 1 are $\mathrm{K} = 5, \beta = 1, \mathrm{N} = 70, \lambda = 25$.



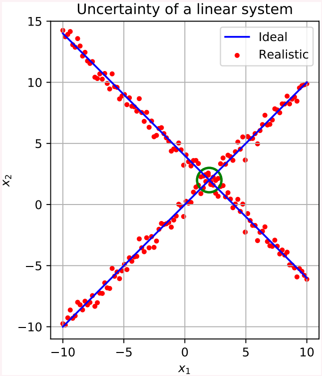
Transferability of learned information

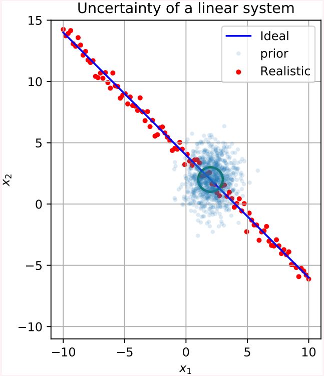
- Reconstruction of T2 and T1 FLAIR images with a prior trained on T2 FLAIR.
- Use a Poisson-disk pattern of 8x undersampling in k-space, 320x320.
- The T2 FLAIR weighted k-space is from FastMRI dataset and the T1 FLAIR weighted k-space is obtained in Göttingen.



Uncertainty

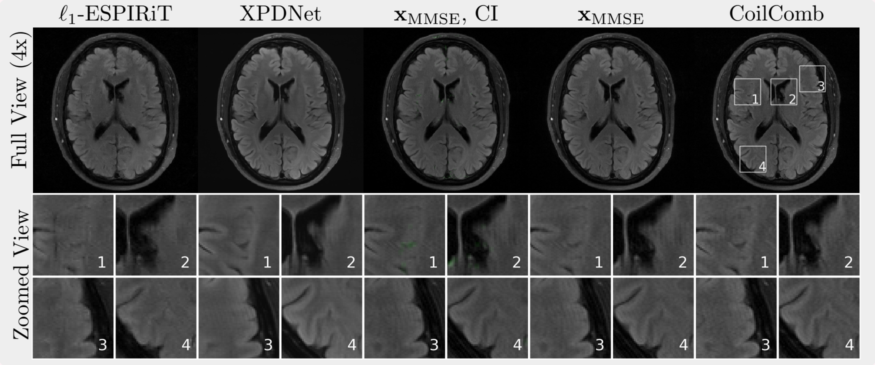
- In our context, the uncertainty means our measurement is not sufficient to exactly determine the image to be reconstructed with our assumed model.
- $x_1 + x_2 = 4 + \eta;\ x_1 - x_2 = \eta; \eta \sim \mathcal{N}(0, 0.5)$





Uncertainty and hallucination

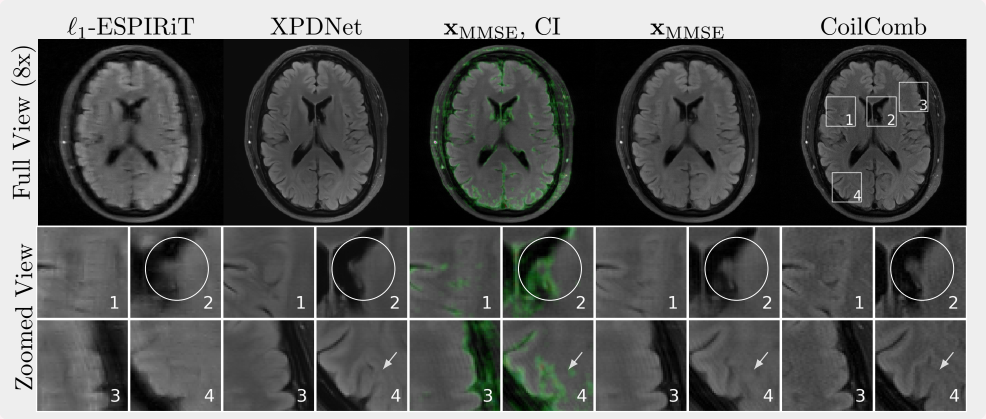
- Reconstructions are ℓ_1-wavelet, XPDNet[1], $\mathbf{x}_{\mathrm{MMSE}}$ highlighted with confidence interval (CI), $\mathbf{x}_{\mathrm{MMSE}}$ and a fully-sampled coil-combined image (Reference). 4x acceleration along phase direction.



[1] Ramzi et al. arXiv (2020)

Uncertainty and hallucination

- Hallucinations appear when using 8.2-fold acceleration and are highlighted with the CI after thresholding. Selected regions of interests are presented in a zoomed view.



Conclusion

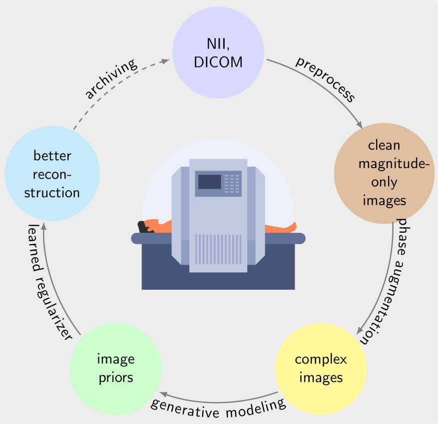
- We combine concepts from machine learning, Bayesian inference and image reconstruction.
- The image reconstruction is realized by drawing samples from the posterior term $p(\mathbf{x}|\mathbf{y})$ using a learned prior.
- This method provides a minimum mean square reconstruction and uncertainty estimation.
- This method shows good performance and transferability to different contrasts and sampling patterns.

Outline

- Background and motivation
- Bayesian MRI reconstruction using diffusion priors
- **Phase augmentation for training priors**
- Speed up MR scans with generative prior
- Summary and outlook

Construct dataset from magnitude-only images

- easy to get magnitude-only images
- especially pathological data



Phase augmentation using a diffusion model

Given the likelihood term of the magnitude $p(\mathbf{m}|\mathbf{x})$ and a prior for complex-valued images $p(\mathbf{x})$, the posterior of the complex image is proportional to

$$p(\mathbf{x}|\mathbf{m}) \propto p(\mathbf{x}) \cdot p(\mathbf{m}|\mathbf{x})$$

Phase augmentation using a diffusion model

Given the likelihood term of the magnitude $p(\mathbf{m}|\mathbf{x})$ and a prior for complex-valued images $p(\mathbf{x})$, the posterior of the complex image is proportional to

$$p(\mathbf{x}|\mathbf{m}) \propto p(\mathbf{x}) \cdot p(\mathbf{m}|\mathbf{x})$$

where

$$\mathbf{m} = \sqrt{\mathbf{x}_r^2 + \mathbf{x}_i^2}\ ,$$

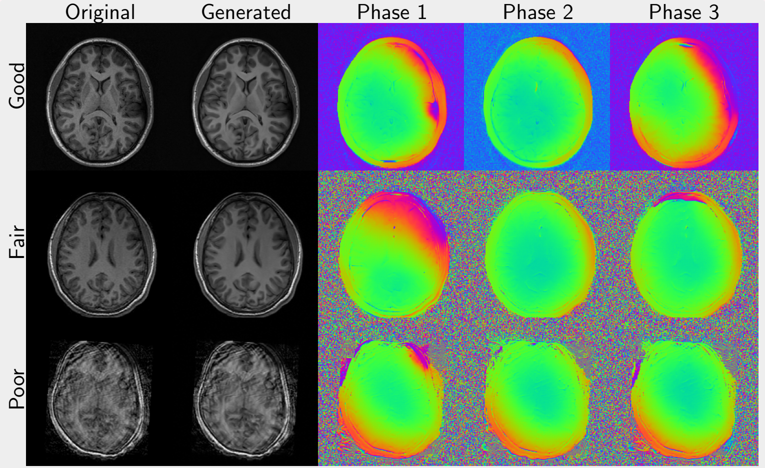
$$p(\mathbf{m}|\mathbf{x}) \propto \exp\left(-\epsilon\left\|\mathbf{m} - \sqrt{\mathbf{x}_r^2 + \mathbf{x}_i^2}\right\|_2^2\right).$$

We initialize samples with random complex Gaussian noise, then transfer them gradually to the distribution of complex images with learned transition kernels $p_\theta(\mathbf{x}_n|\mathbf{x}_{n+1})$, and run unadjusted Langevin iterations sequentially at each intermediate distribution

$$\mathbf{x}_n^{k+1} \leftarrow \mathbf{x}_n^k + \frac{\gamma}{2}\nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}_n^k \mid \mathbf{x}_{n+1}^{\mathrm{K}}) + \frac{\gamma}{2}\nabla_{\mathbf{x}} \log p(\mathbf{m}|\mathbf{x}_n^k) + \sqrt{\gamma}\mathbf{z}\ .$$

Examples of generated phase maps

- Three magnitude images of different quality from the ABIDE[1] dataset
- The corresponding magnitude and phase maps of complex-valued images generated using diffusion prior NET_1



[1] Di Martino et al. Molecular Psychiatry 2014

MRI Image Priors

Table: Datasets and computational resources used to train the six different priors used in this work.

Prior		Model	Phase	Nr. of Images	MR Contrasts	GPUs	Parameters	Time × epochs
P_{SC}	(small, complex)	PixelCNN	preserved	1000	T_1, T_2, T_2-FLAIR, T_2^*	4×A100, 80G	~22M	~40s × 500
P_{SM}	(small, magnitude)	PixelCNN	not available	1000	T_1, T_2, T_2-FLAIR, T_2^*	4×V100, 32G	~22M	~144s × 500
P_{LM}	(large, magnitude)	PixelCNN	not available	23078	MPRAGE	4×A100, 80G	~22M	~748s × 100
P_{LC}	(large, complex)	PixelCNN	generated	23078	MPRAGE	3×A100, 80G	~22M	~1058s × 100
D_{SC}	(SMLD, complex)	Diffusion	generated	79058	MPRAGE	4×A100, 80G	~8M	~2330s × 50
D_{PC}	(DDPM, complex)	Diffusion	generated	79058	MPRAGE	8×V100, 32G	~8M	~1430s × 200

Linear and non-linear reconstruction

Parallel MRI is formulated as an inverse problem

$$F(\mathbf{x},\ \mathbf{c}) := (\mathcal{F}_S(\mathbf{x} \cdot c_1), \cdots, \mathcal{F}_S(\mathbf{x} \cdot c_N)) = \mathbf{y}.$$

- $\mathcal{F}_S$, subsampling operator. $\mathbf{y} = (y_1, \cdots, y_N)$, measured k-space data
- $\mathbf{x}$, image. N, number of coils. $\mathbf{c} = (c_1, \cdots, c_N)$, coil sensitivities

Linear and non-linear reconstruction

Parallel MRI is formulated as an inverse problem

$$F(\mathbf{x},\ \mathbf{c}) := (\mathcal{F}_S(\mathbf{x}\cdot c_1), \cdots, \mathcal{F}_S(\mathbf{x}\cdot c_N)) = \mathbf{y}.$$

- $\mathcal{F}_S$, subsampling operator. $\mathbf{y} = (y_1, \cdots, y_N)$, measured k-space data
- $\mathbf{x}$, image. N, number of coils. $\mathbf{c} = (c_1, \cdots, c_N)$, coil sensitivities

Minimizing the sub-linearized problem at each Gauss Newton step k

$$\min_{\delta\mathbf{x}} \frac{1}{2}\|F'(\mathbf{x}^k)\delta\mathbf{x} + F(\mathbf{x}^k) - \mathbf{y}\|^2 + \beta_k \mathcal{W}(\mathbf{c}^k + \delta\mathbf{c}) + \alpha_k R(\mathbf{x}^k + \delta\mathbf{x}) .$$

or pre-compute sensitivities and minimizing

$$\min_{\mathbf{x}} \frac{1}{2}\|F_{\mathbf{c}}(\mathbf{x}) - \mathbf{y}\|^2 + \alpha R(\mathbf{x})$$

[1]Uecker et al. MRM 2008.

Generative prior as regularization

The proximal operator for the log-prior $R(\mathbf{x}) = \log p(\mathbf{x})$ is defined as

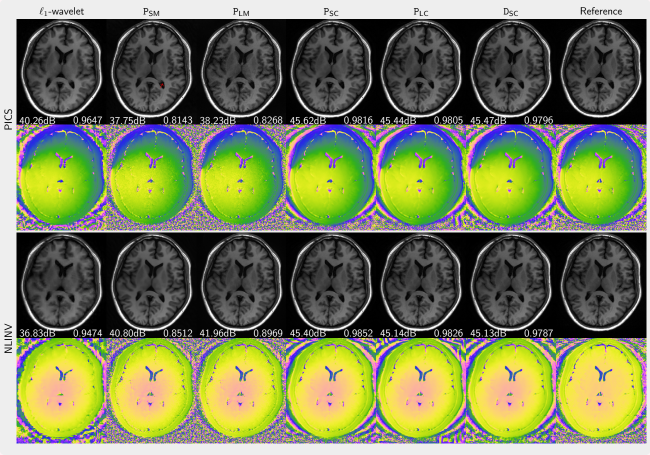
$$\operatorname{prox}_t(\mathbf{z}) \;=\; \arg\min_{\mathbf{x}} \frac{1}{2t}\|\mathbf{x}-\mathbf{z}\|^2 + \log p(\mathbf{x}) \; . \tag{15}$$

then,

$$\operatorname{prox}_t(\mathbf{z}) \approx \mathbf{z} - t\nabla_{\mathbf{x}} \log p(\mathbf{x}) \; ,$$

Influence of phase maps

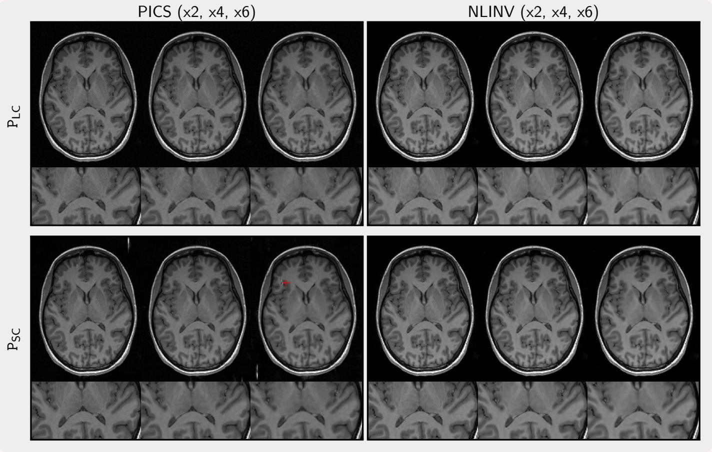
- 8.2x undersampling in k-space with Poisson disc
- Linear and nonlinear reconstruction with BART[1] command PICS and NLINV



[1] Uecker et al. ISMRM 2015

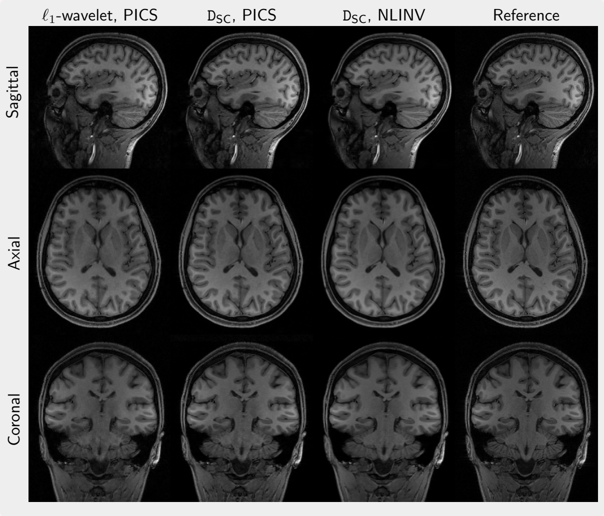
Small complex prior vs Large complex prior

- MPRAGE (Magnetization Prepared-RApid Gradient Echo) sequence
- 2-fold acceleration with grappa and 4/5 partial parallel Fourier imaging
- dimensions (256, 256, 224) and voxel sizes (1mm, 1mm, 1mm).



3D reconstruction

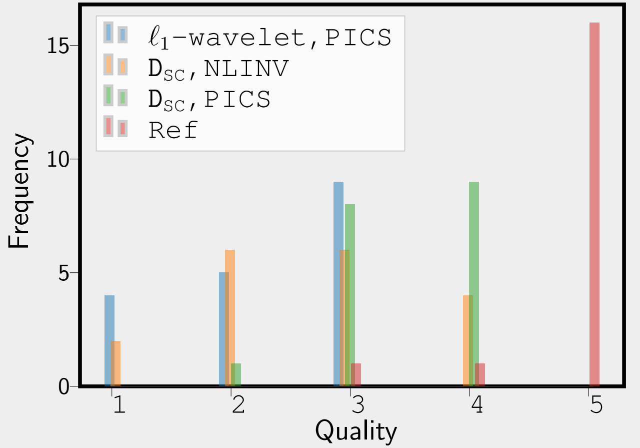
- k-space data were acquired from 6 healthy volunteers using TurboFLASH-3D sequence (TE=3.3ms, TR=2250ms, TI=900ms, flip angle $\alpha = 9^\circ$)
- dimensions (256, 256, 176) and voxel sizes (1mm, 1mm, 1mm).
- retrospective 8.2x-undersampling



Evaluation by radiologists

Three radiologists rated the images reconstructed with different methods

Scale	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad



Conclusion

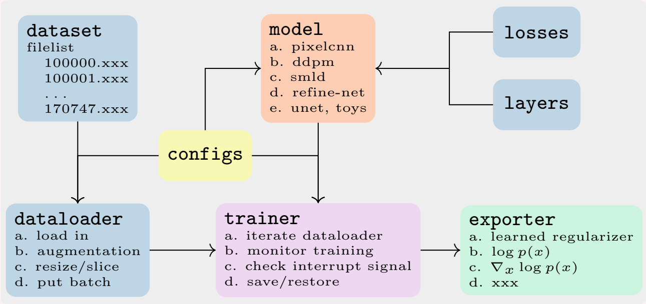
- We demonstrate that priors trained on complex images are superior to priors trained on magnitude-only images
- We leverage a diffusion model trained on a small dataset of complex images to augment a much larger dataset
- We show that we can train more robust priors with a larger training dataset
- We integrate priors at regularization term into linear and non-linear reconstruction technique

Outline

- Background and motivation
- Bayesian MRI reconstruction using diffusion priors
- Phase augmentation for training priors
- **Speed up MR scans with generative prior**
- Summary and outlook

Overview

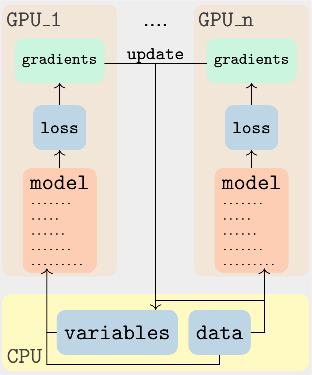
- The loss and layer functions are used to create models.
- The trainer is fed by the dataloader and trains the model on multiple GPUs.
- The exporter is used to customize trained models for deployment.

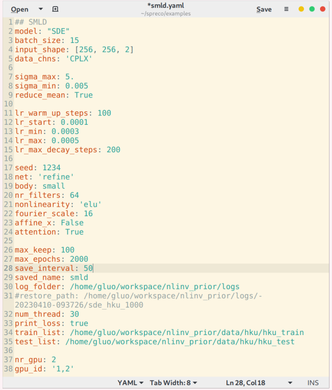


It is implemented with : 1) tensorflow; 2) numpy; 3) pyyaml; 4)matplotlib; 5)scikit-image; 6) tqdm.

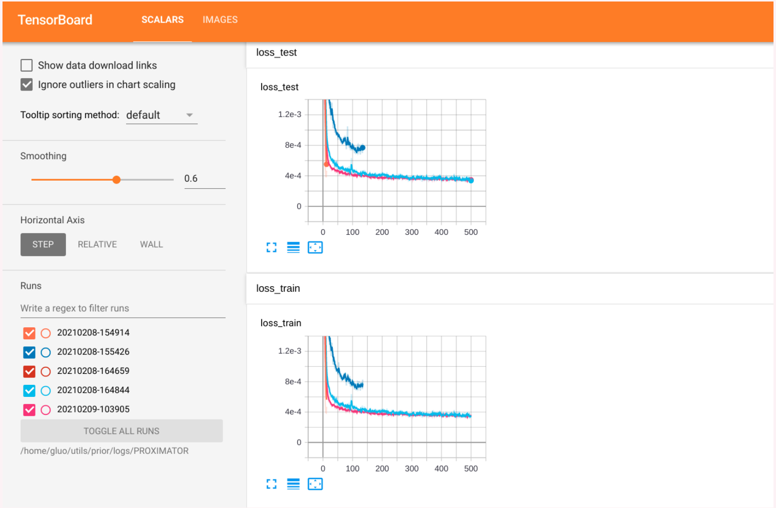
Spreco

- The trainable variables are shared across GPUs
- Dataloader feeds data to every GPU worker.
- Config file for training models





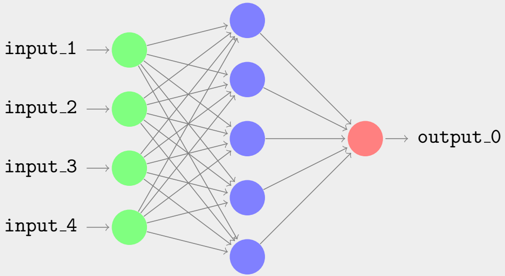
Spreco


TensorBoard
SCALARS
IMAGES
Show data download links
Ignore outliers in chart scaling
Tooltip sorting method: default
Smoothing
0.6
Horizontal Axis
STEP
RELATIVE
WALL
Runs
Write a regex to filter runs
20210208-154914
20210208-155426
20210208-164659
20210208-164844
20210209-103905
TOGGLE ALL RUNS
/home/gluo/utils/prior/logs/PROXIMATOR
loss_test
loss_test
1.2e-3
8e-4
4e-4
0
0
100
200
300
400
500
loss_train
loss_train
1.2e-3
8e-4
4e-4
0
0
100
200
300
400
500

BART with tensorflow computation graph

- Export the trained model with customized labels for inputs $\mathbf{x}$ and outputs $\mathbf{y} = \mathtt{Net}_\theta(\mathbf{x})$;
- Initialize an exported graph, the restoration of a saved model using C API;
- Wrap the exported computation graph into BART's non-linear operator (nlop)

(a) Export the trained model as computation graph



(b) Use the graph as regularization in BART

```
$ bart pics -R TF:<graph_path>:λ <kspace> <coils> <reco>
```

Outline

- Background and motivation
- Bayesian MRI reconstruction using diffusion priors
- Phase augmentation for training priors
- Speed up MR scans with generative prior
- **Summary and outlook**

Summary

- What are the advantages of a prior? Decoupled from the forward operator, which leads to better generalizability to many acquisition scenarios.
- How to use a prior for reconstruction? Sample the posterior or use it as regularization term, which permits uncertainty estimation and flexibility.
- How to get more data for training? Phase augmentation using diffusion priors trained on small complex-valued images.
- How to deploy a prior? BART + TF C API.

Outlook

- How to determine the performance bound of a prior?
- How does it relate to the uncertainties of the reconstructed image?
- How many samples need to be acquired to keep the uncertainty within an acceptable limit?
- How effectively can the prior information transfer to a broader range of scenarios?

Q&A

Generate samples

For a multi-modal distribution, use annealed Langevin dynamics. For σ_i in $\{\sigma_N, ..., \sigma_1\}$,

$$\mathbf{x}_i^k = \mathbf{x}_i^{k-1} - \frac{\lambda}{2}\nabla_{\mathbf{x}} \log p(\mathbf{x}_i^{k-1}) + \omega_i^k$$

A mixture of bivariate Gaussian as prior distribution

- $p(\mathbf{x}) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$
- ϕ_i is the mixture indicator, $\phi_i = \{0.7, 0.2, 0.1\}$
- $\boldsymbol{\mu_i} = \{[5,5], [-5,-5], [5,-5]\}$, $\boldsymbol{\Sigma_i} = \mathbf{I}$.