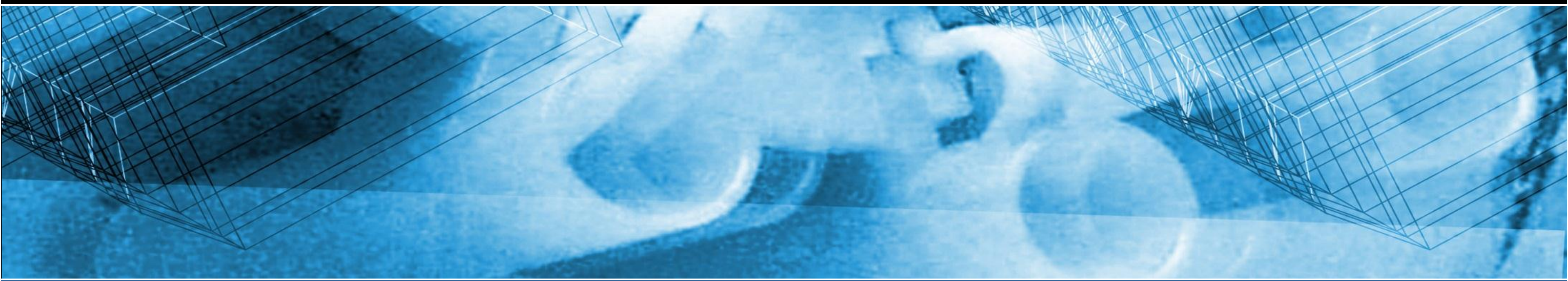


Joint Bayesian Compressed Sensing with Prior Estimate

Berkin Bilgic[1], Tobias Kober[2,3], Gunnar Krueger[2,3], Elfar Adalsteinsson[1,4]

[1] Electrical Engineering and Computer Science, MIT
[2] Laboratory for Functional and Metabolic Imaging, EPFL
[3] Siemens Schweiz AG
[4] Harvard-MIT Division of Health Sciences and Technology

Declaration of Relevant Financial Interests or Relationships

Speaker Name: Berkin Bilgic

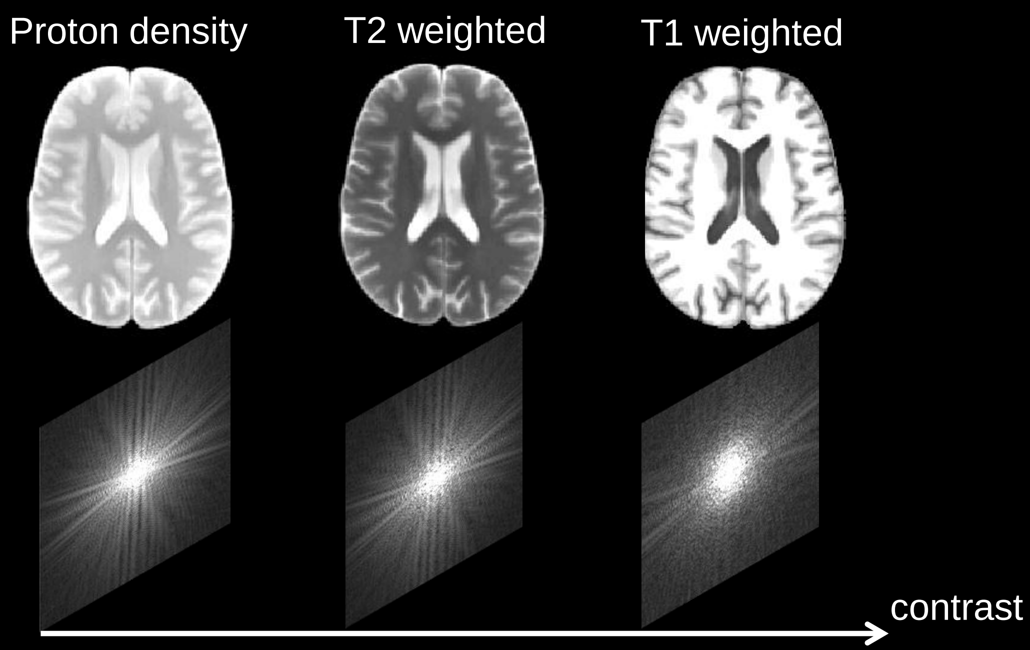
I have no relevant financial interest or relationship to disclose with regard to the subject matter of this presentation.

Multi-contrast Acquisition

- Clinical MRI: acquiring multiple contrast preparations increases the diagnostic power, but also the total scan time

Multi-contrast Acquisition

- Clinical MRI: acquiring multiple contrast preparations increases the diagnostic power, but also the total scan time

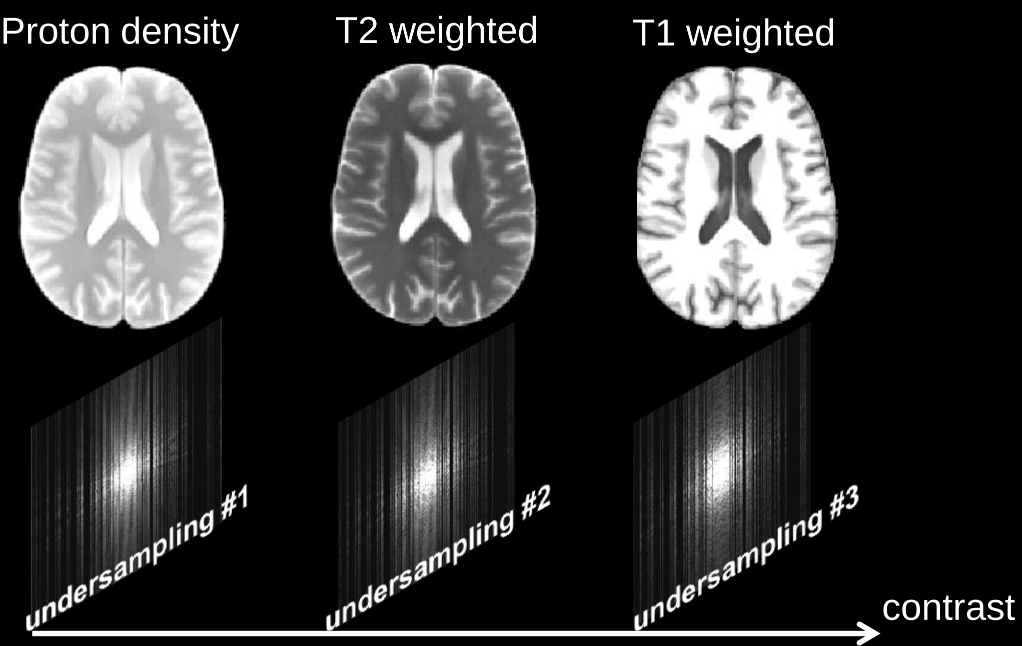


SRI24 atlas[1]

[1] Rohlfing et al. Hum Brain Map, 2010

Multi-contrast Acquisition

- Clinical MRI: acquiring multiple contrast preparations increases the diagnostic power, but also the total scan time

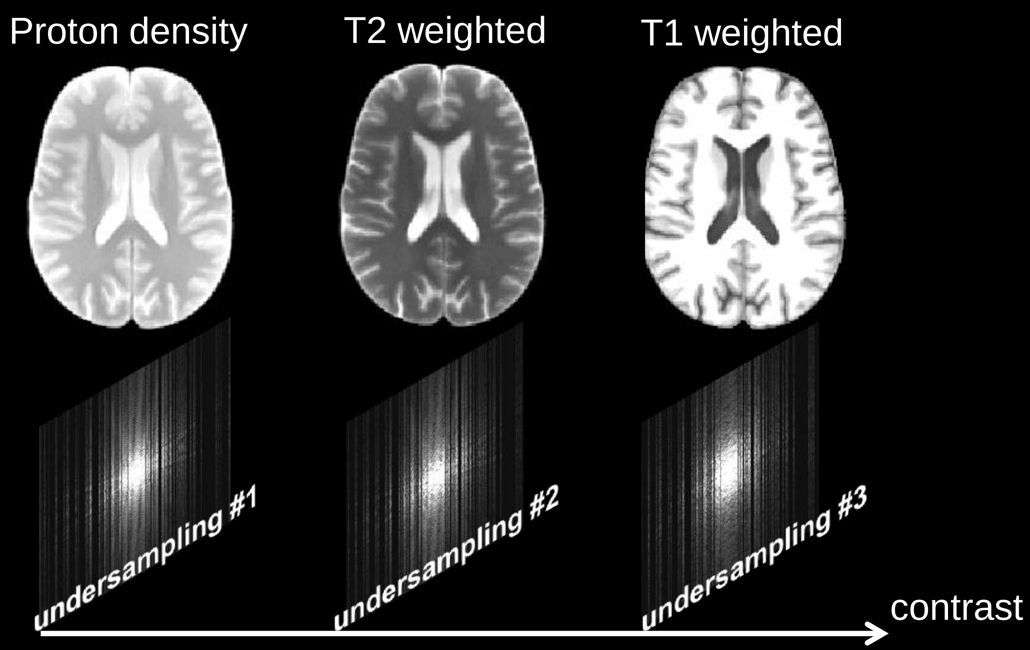


- Joint reconstruction from undersampled acquisitions substantially improves reconstruction quality[1]

[1] Bilgic et al. MRM, 2011

Multi-contrast Acquisition

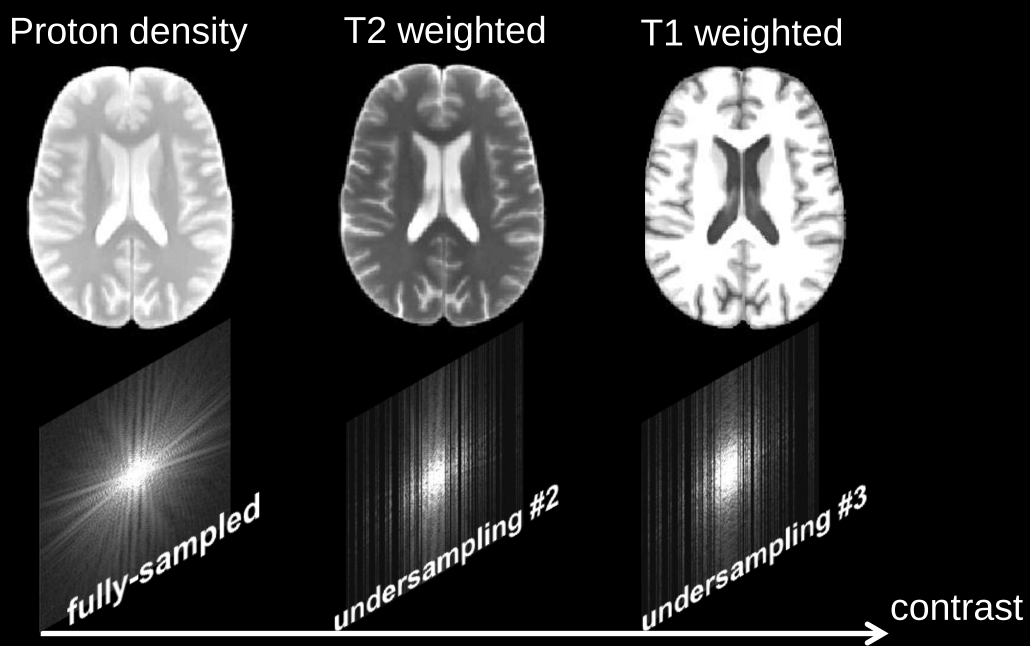
- Clinical MRI: acquiring multiple contrast preparations increases the diagnostic power, but also the total scan time



- Suppose that one of the contrasts can be acquired much faster than the others (e.g. AutoAlign)

Multi-contrast Acquisition

- Clinical MRI: acquiring multiple contrast preparations increases the diagnostic power, but also the total scan time



- Suppose that one of the contrasts can be acquired much faster than the others (e.g. AutoAlign)
- If we fully-sample the fast contrast, can we use it to help reconstruct the others?

Observation model

$$\mathbf{F}\ \boldsymbol{x} = \boldsymbol{y}$$

$\mathbf{F}$: partial Fourier transform

$\boldsymbol{x}$: image to be estimated

$\boldsymbol{y}$: undersampled k-space data

Observation model – sparse representation

$$\mathbf{V}\,\mathbf{F}\;x = \mathbf{V}\,y$$

$$\mathbf{V} = \left(1 - e^{-2\pi j\omega/n}\right)$$

Observation model – sparse representation

$$\mathbf{F}\,\boldsymbol{\delta} = \tilde{\boldsymbol{y}}$$

$\boldsymbol{\delta}$: image gradient to be estimated

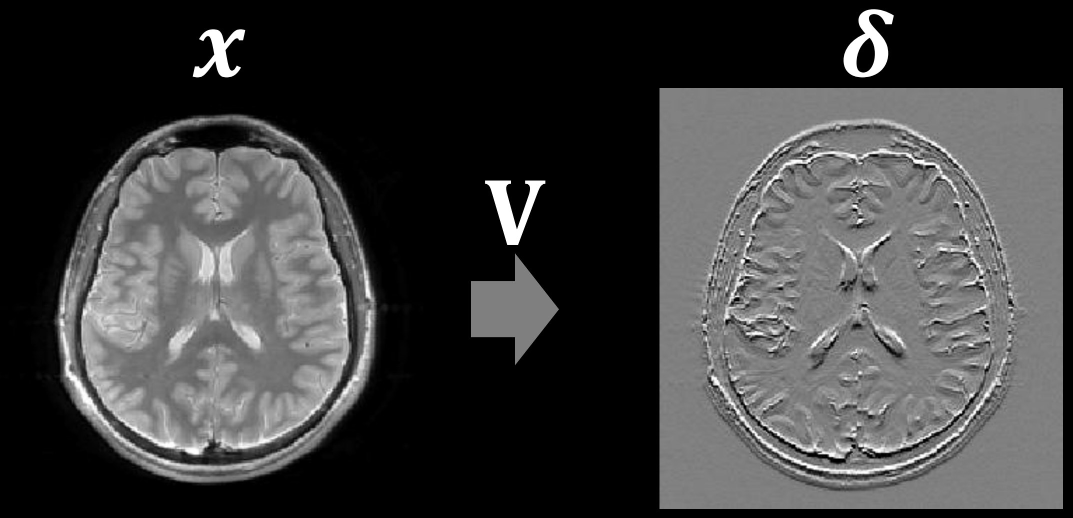
$\tilde{\boldsymbol{y}}$: modified k-space data

Observation model – sparse representation

$$\mathbf{F}\,\boldsymbol{\delta} = \tilde{\boldsymbol{y}}$$

$\boldsymbol{\delta}$: image gradient to be estimated

$\tilde{\boldsymbol{y}}$: modified k-space data



Data likelihood

- Assuming that the k-space data is corrupted by complex-valued Gaussian noise with σ^2 variance,

$$\underbrace{\mathrm{p}(\tilde{\boldsymbol{y}} \mid \boldsymbol{\delta}, \sigma^2)}_{\text{Gaussian likelihood}} \sim \mathcal{N}(\mathbf{F}\boldsymbol{\delta} - \tilde{\boldsymbol{y}}, \sigma^2)$$

Prior distribution on gradient coefficients

- Bayesian CS places hyperparameters γ on each pixel,

$$\underbrace{\mathrm{p}(\delta_i \mid \gamma_i)}_{\text{Gaussian prior}} \sim \mathcal{N}(0, \gamma_i)$$

- So that ith pixel is a zero-mean Gaussian with variance γ_i

Prior distribution on gradient coefficients

- Bayesian CS places hyperparameters γ on each pixel,

$$\underbrace{\mathrm{p}(\delta_i \mid \gamma_i)}_{\text{Gaussian prior}} \sim \mathcal{N}(0, \gamma_i)$$

- So that i^{th} pixel is a zero-mean Gaussian with variance γ_i

- Multiplicative combination of all pixels give the full prior distribution,

$$\mathrm{p}(\boldsymbol{\delta} \mid \boldsymbol{\gamma}) \sim \prod_i \mathcal{N}(0, \gamma_i)$$

Posterior distribution for gradient coefficients

- Using the likelihood and the prior, we invoke Bayes' Rule to arrive at the posterior,

$$\mathrm{p}(\boldsymbol{\delta} \mid \tilde{\boldsymbol{y}}, \boldsymbol{\gamma}) \propto \mathrm{p}(\boldsymbol{\delta} \mid \boldsymbol{\gamma}) \cdot \mathrm{p}(\tilde{\boldsymbol{y}} \mid \boldsymbol{\delta})$$

Posterior distribution for gradient coefficients

- Using the likelihood and the prior, we invoke Bayes' Rule to arrive at the posterior,

$$\underbrace{\mathrm{p}(\boldsymbol{\delta} \mid \tilde{\boldsymbol{y}}, \boldsymbol{\gamma})}_{\text{Gaussian posterior}} \propto \underbrace{\mathrm{p}(\boldsymbol{\delta} \mid \boldsymbol{\gamma})}_{\text{Gaussian prior}} \cdot \underbrace{\mathrm{p}(\tilde{\boldsymbol{y}} \mid \boldsymbol{\delta})}_{\text{Gaussian likelihood}}$$

Posterior distribution for gradient coefficients

- Using the likelihood and the prior, we invoke Bayes' Rule to arrive at the posterior,

$$\mathrm{p}(\boldsymbol{\delta} \mid \widetilde{\boldsymbol{y}}, \boldsymbol{\gamma}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu} = \boldsymbol{\Gamma}\mathbf{F}^{H}\mathbf{A}^{-1}\widetilde{\boldsymbol{y}}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{F}^{H}\mathbf{A}^{-1}\mathbf{F}\boldsymbol{\Gamma}$$

Posterior distribution for gradient coefficients

- Using the likelihood and the prior, we invoke Bayes' Rule to arrive at the posterior,

$$\mathrm{p}(\boldsymbol{\delta} \,|\, \widetilde{\boldsymbol{y}}, \boldsymbol{\gamma}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu} = \boldsymbol{\Gamma}\mathbf{F}^H \mathbf{A}^{-1} \widetilde{\boldsymbol{y}}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{F}^H \mathbf{A}^{-1}\mathbf{F}\boldsymbol{\Gamma}$$

$$\boldsymbol{\Gamma} = diag(\boldsymbol{\gamma})$$

$$\mathbf{A}^{-1} = (\sigma^2 \mathbf{I} + \mathbf{F}\boldsymbol{\Gamma}\mathbf{F}^H)^{-1}$$

Inversion using Lanczos algorithm[1]

[1] Seeger et al. MRM, 2010

EM algorithm for optimization

- Expectation-maximization algorithm[1] is used to estimate the hyperparameters and the posterior iteratively,

Expectation step:

$$\boldsymbol{\mu} = \boldsymbol{\Gamma}\mathbf{F}^H\mathbf{A}^{-1}\widetilde{\boldsymbol{y}}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{F}^H\mathbf{A}^{-1}\mathbf{F}\boldsymbol{\Gamma}$$

Maximization step:

$$\gamma_i = |\mu_i|^2/(1 - \Sigma_{ii}/\gamma_i)$$

[1] Wipf et al. IEEE Trans Signal Process, 2007

EM algorithm for optimization

- Expectation-maximization algorithm[1] is used to estimate the hyperparameters and the posterior iteratively,

Expectation step:

$$\boldsymbol{\mu} = \boxed{\boldsymbol{\Gamma}}\mathbf{F}^H\mathbf{A}^{-1}\tilde{\boldsymbol{y}}$$

$$\boldsymbol{\Sigma} = \boxed{\boldsymbol{\Gamma}} - \boxed{\boldsymbol{\Gamma}}\mathbf{F}^H\mathbf{A}^{-1}\mathbf{F}\boxed{\boldsymbol{\Gamma}}$$

Maximization step:

$$\boxed{\gamma_i} = |\mu_i|^2/(1 - \Sigma_{ii}/\gamma_i)$$

[1] Wipf et al. IEEE Trans Signal Process, 2007

EM algorithm for optimization

- Expectation-maximization algorithm[1] is used to estimate the hyperparameters and the posterior iteratively,

Expectation step:

$$\boxed{\boldsymbol{\mu}} = \boldsymbol{\Gamma}\mathbf{F}^{H}\mathbf{A}^{-1}\tilde{\boldsymbol{y}}$$

$$\boxed{\boldsymbol{\Sigma}} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{F}^{H}\mathbf{A}^{-1}\mathbf{F}\boldsymbol{\Gamma}$$

Maximization step:

$$\gamma_i = |\boxed{\mu_i}|^2/(1 - \boxed{\Sigma_{ii}}/\gamma_i)$$

[1] Wipf et al. IEEE Trans Signal Process, 2007

Using fully-sampled prior image

- If we run EM iterations on the fully sampled image $\boldsymbol{\delta}_{prior}$

Expectation step:

$$\boldsymbol{\mu}_{prior} = \boldsymbol{\delta}_{prior}$$

$$\boldsymbol{\Sigma}_{prior} = \mathbf{0}$$

Maximization step:

$$\gamma_{prior} = |\mu_{prior}|^2$$

Using fully-sampled prior image

- If we run EM iterations on the fully sampled image $\boldsymbol{\delta}_{prior}$

Expectation step:

$$\boldsymbol{\mu}_{prior} = \boldsymbol{\delta}_{prior}$$

$$\boldsymbol{\Sigma}_{prior} = \mathbf{0}$$

Iterations do not alter fully-sampled prior image

Maximization step:

$$\gamma_{prior} = |\mu_{prior}|^2$$

Using fully-sampled prior image

- If we run EM iterations on the fully sampled image $\boldsymbol{\delta}_{prior}$

<u>Expectation step:</u>

$$\boldsymbol{\mu}_{prior} = \boldsymbol{\delta}_{prior}$$

$$\boldsymbol{\Sigma}_{prior} = \mathbf{0}$$

Iterations do not alter fully-sampled prior image

<u>Maximization step:</u>

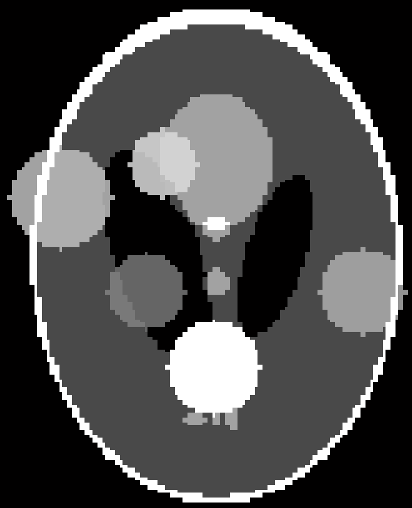
$$\gamma_{prior} = |\mu_{prior}|^2$$

Use γ_{prior} to initialize the EM iterations for undersampled images

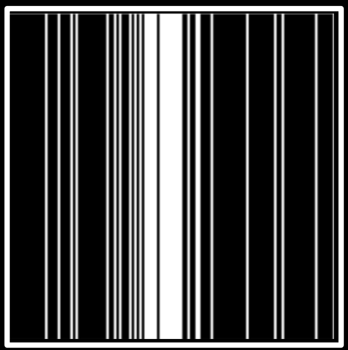
Extended Shepp-Logan Phantoms

fully-sampled prior

undersampled



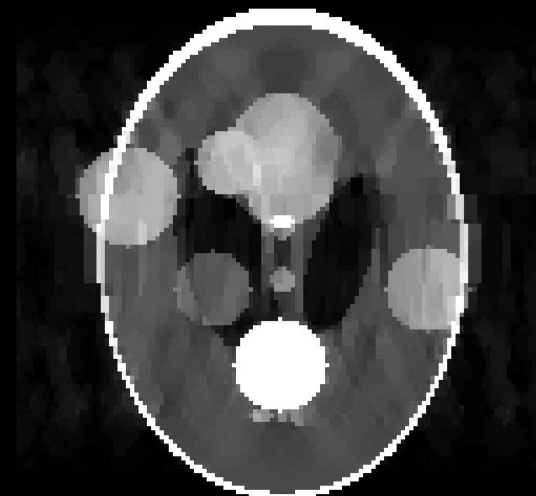
R = 4



sampling pattern

Extended Shepp-Logan Phantoms

sparseMRI[1]: Total Variation



20.1% RMSE

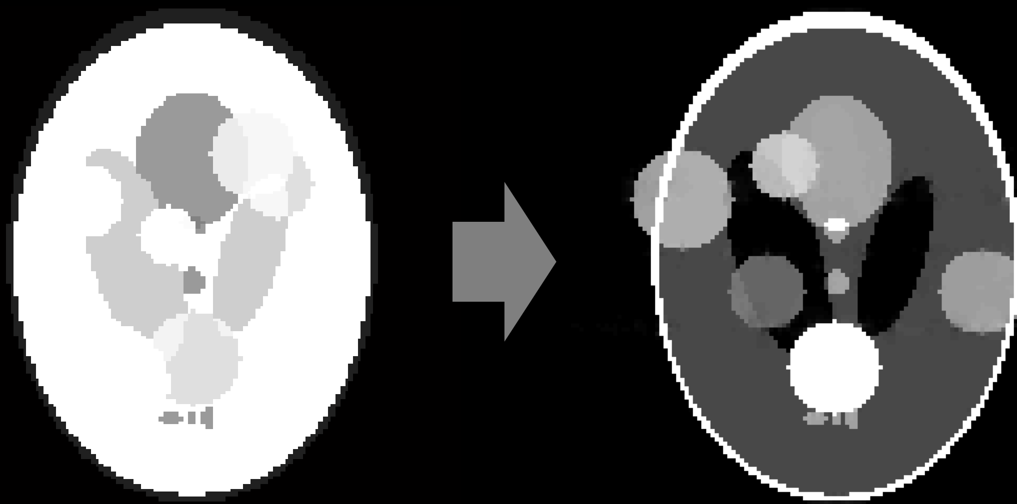
error: scaled 10×

sparseMRI:  20.1% RMSE

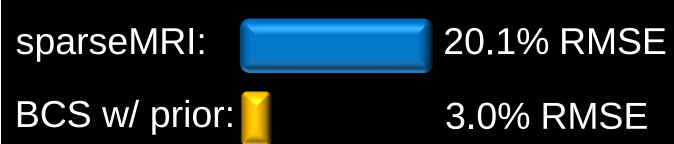
[1] Lustig et al. MRM, 2007

Extended Shepp-Logan Phantoms

fully-sampled prior | Bayesian CS w/ prior



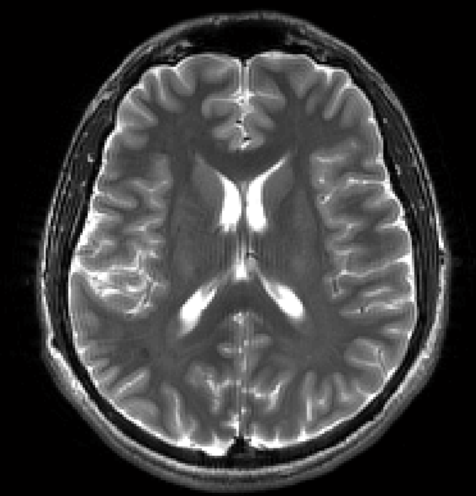
3.0% RMSE



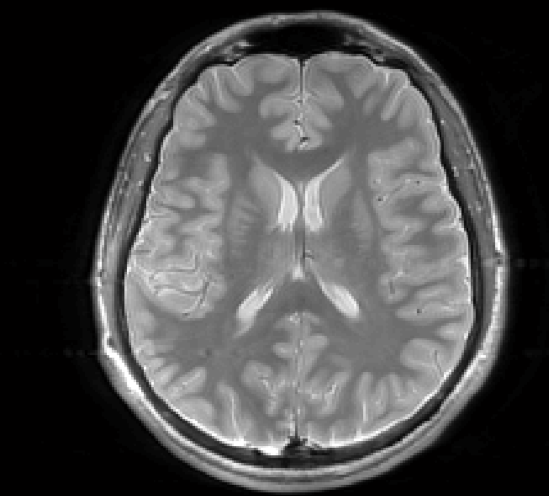
error: scaled 10×

Turbo Spin Echo

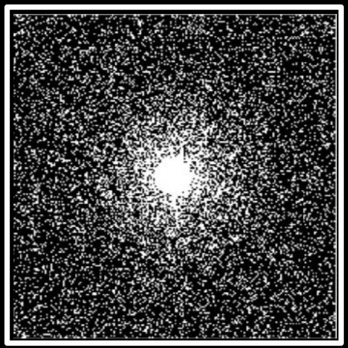
Late Echo
fully-sampled prior



Early Echo
undersampled



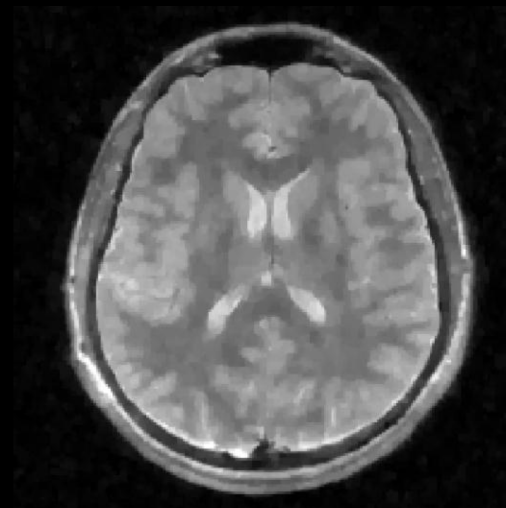
R = 4



sampling pattern

Turbo Spin Echo

sparseMRI[1]: Total Variation



9.3% RMSE

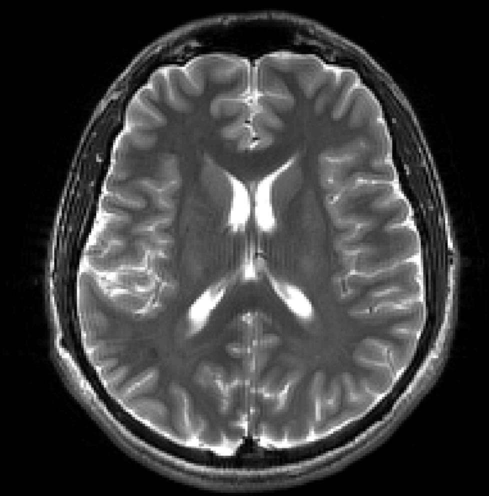
error: scaled 10×

sparseMRI:  9.3% RMSE

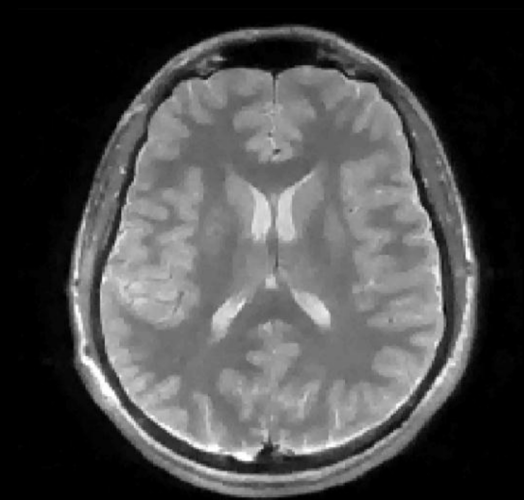
[1] Lustig et al. MRM, 2007

Turbo Spin Echo

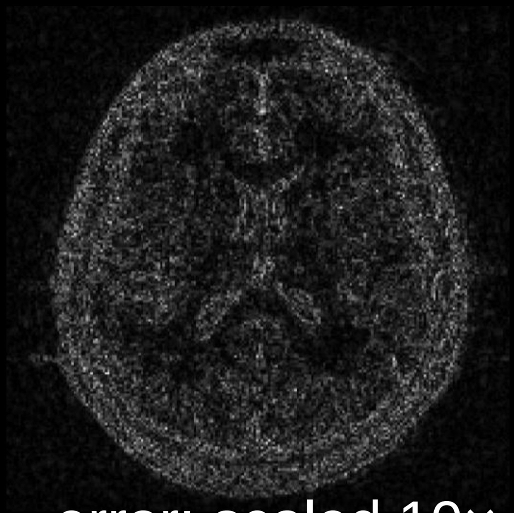
Late Echo
fully-sampled prior



Bayesian CS w/ prior

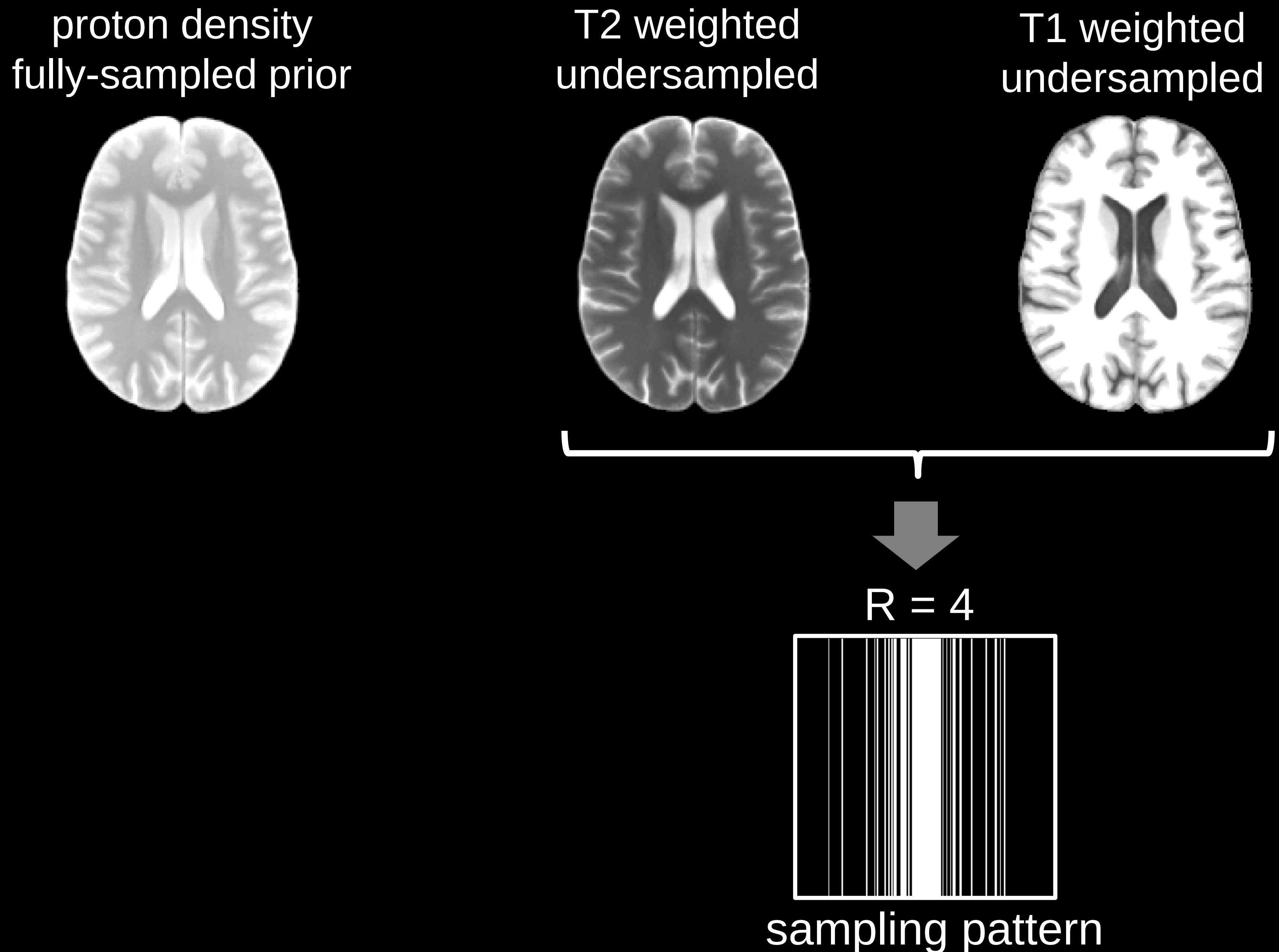


5.8% RMSE



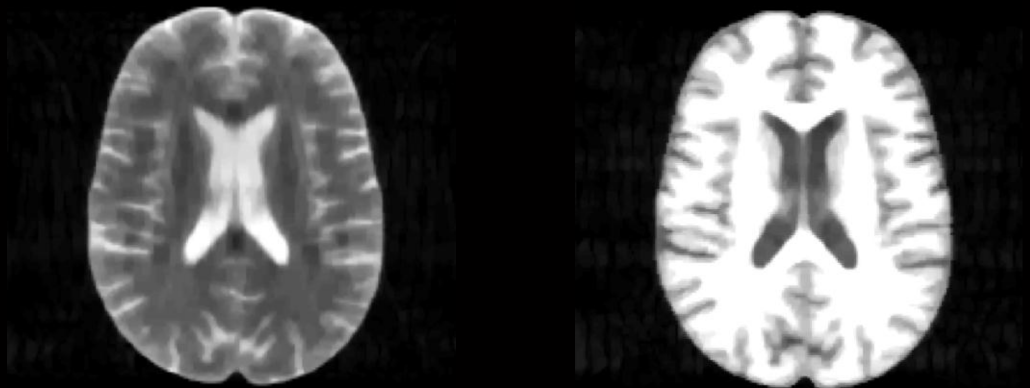
error: scaled 10×

SRI24 atlas

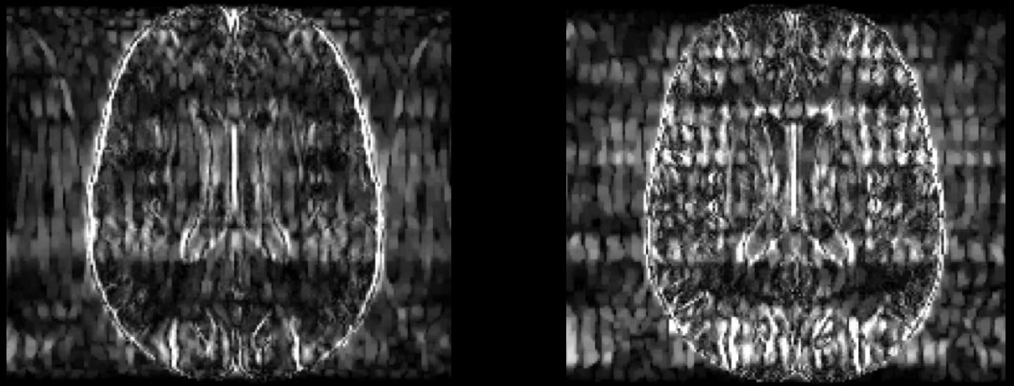

proton density
fully-sampled prior
T2 weighted
undersampled
T1 weighted
undersampled
R = 4
sampling pattern

SRI24 atlas

sparseMRI[1]: Total Variation



9.5% RMSE



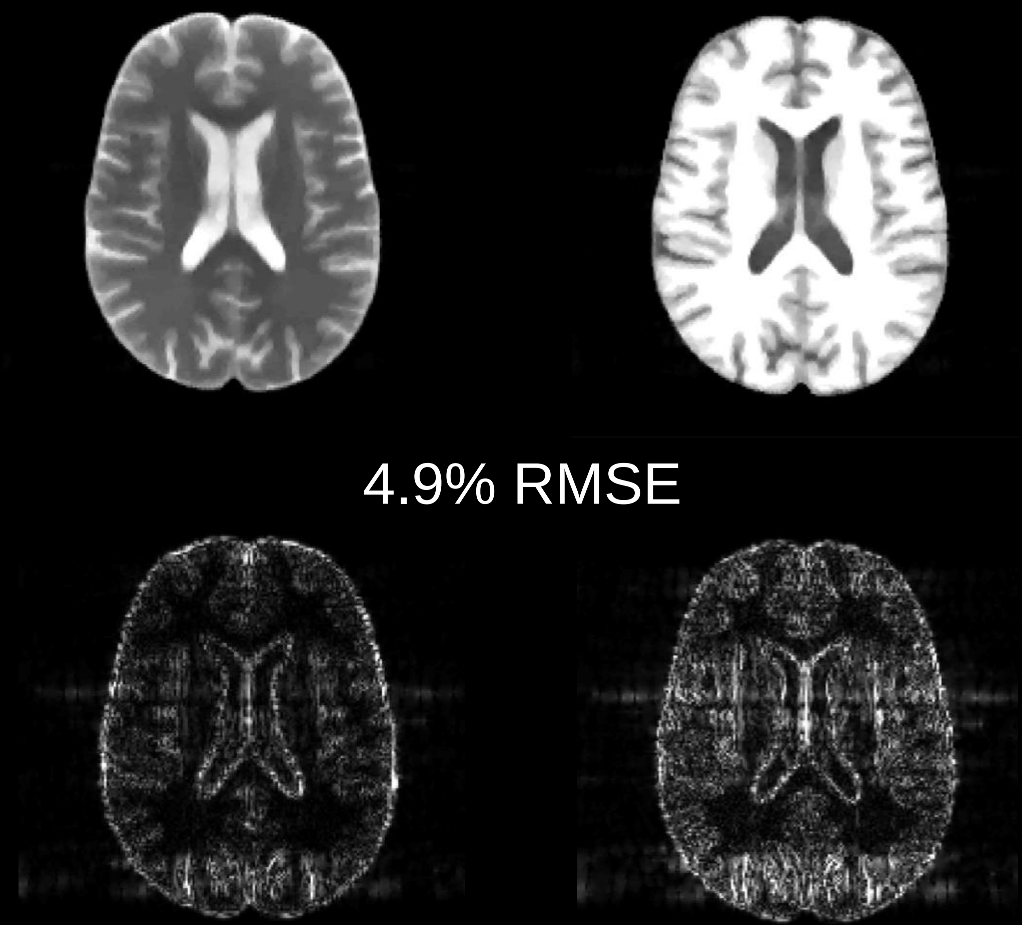
error: scaled 10×

sparseMRI:  9.5% RMSE

[1] Lustig et al. MRM, 2007

SRI24 atlas

Joint Bayesian CS[1]



4.9% RMSE

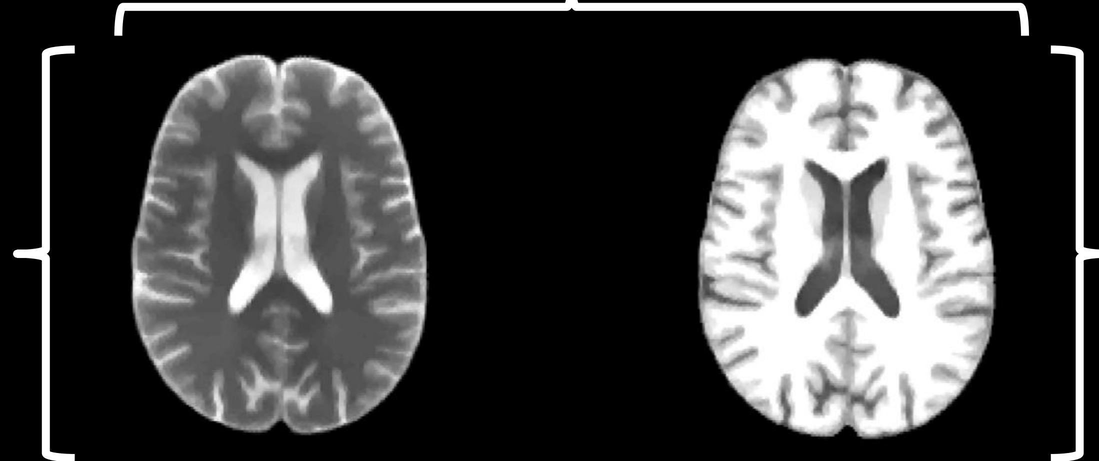
error: scaled 10×

[1] Bilgic et al. MRM, 2011

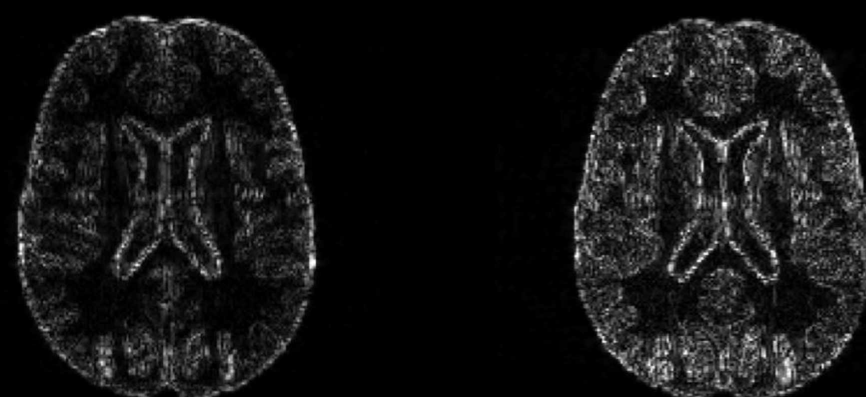
SRI24 atlas

proton density
fully-sampled prior

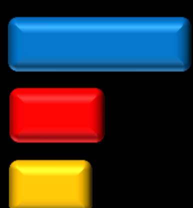
Joint Bayesian CS w/ prior



4.3% RMSE



error: scaled 10×



Conclusion

- In a multi-contrast scan, one of the acquisitions may be much faster than the others (e.g. AutoAlign)
- When the fast contrast is fully-sampled, we use it as prior information to help recover the undersampled contrasts
- Our method uses the prior image only to initialize Bayesian CS iterations, hence imposes the prior in a soft manner

Conclusion

- In a multi-contrast scan, one of the acquisitions may be much faster than the others (e.g. AutoAlign)
- When the fast contrast is fully-sampled, we use it as prior information to help recover the undersampled contrasts
- Our method uses the prior image only to initialize Bayesian CS iterations, hence imposes the prior in a soft manner
- Acknowledgements:

  Siemens Healthcare, Siemens-MIT Alliance,

  CIMIT-MIT Medical Engineering Fellowship, R01 EB 007942