rle AT MIT

Joint Bayesian Compressed Sensing for Multi-contrast Reconstruction

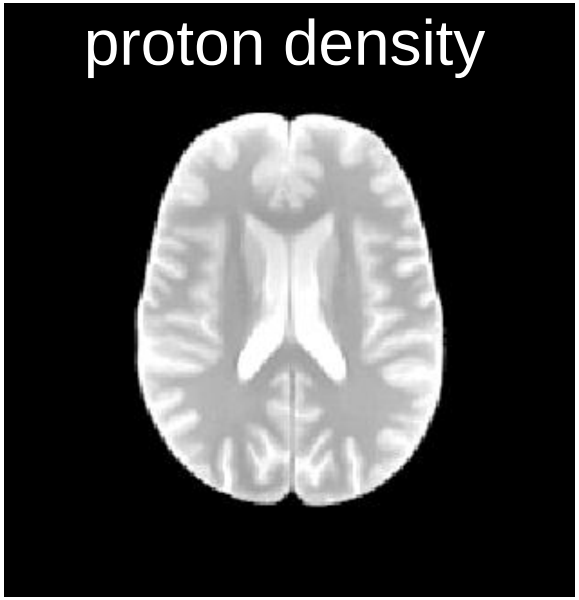
Berkin Bilgic[1], Vivek K. Goyal[1], Elfar Adalsteinsson[1,2]

[1]EECS, MIT, Cambridge, MA, United States

[2]Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, United States

Multi-contrast data acquisition

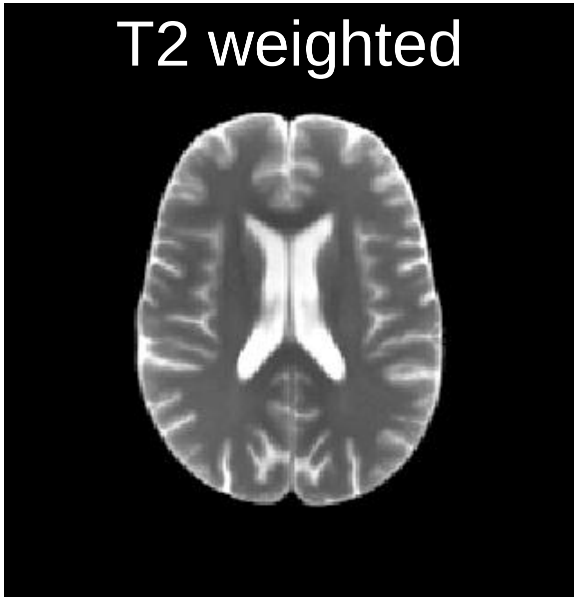
- In clinical MRI, it is common to image the same region of interest under multiple contrast settings
- This aims to increase the diagnostic power of MRI as tissues exhibit different characteristics under different contrasts
- For instance, SRI24 atlas[1] contains such multi-contrast data,



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

Multi-contrast data acquisition

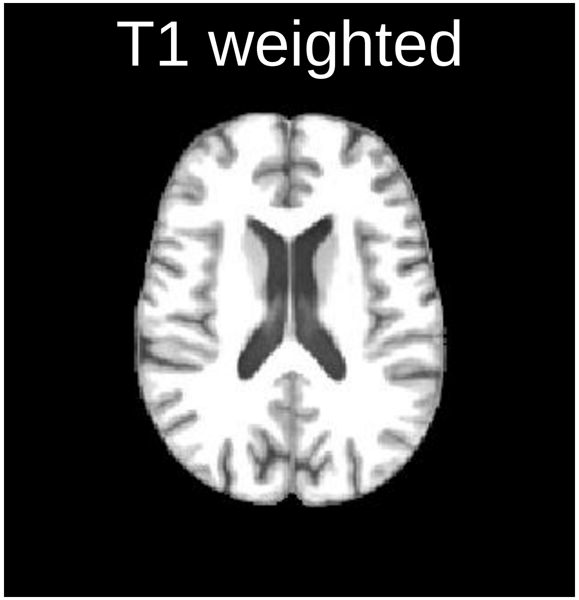
- In clinical MRI, it is common to image the same region of interest under multiple contrast settings
- This aims to increase the diagnostic power of MRI as tissues exhibit different characteristics under different contrasts
- For instance, SRI24 atlas[1] contains such multi-contrast data,



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

Multi-contrast data acquisition

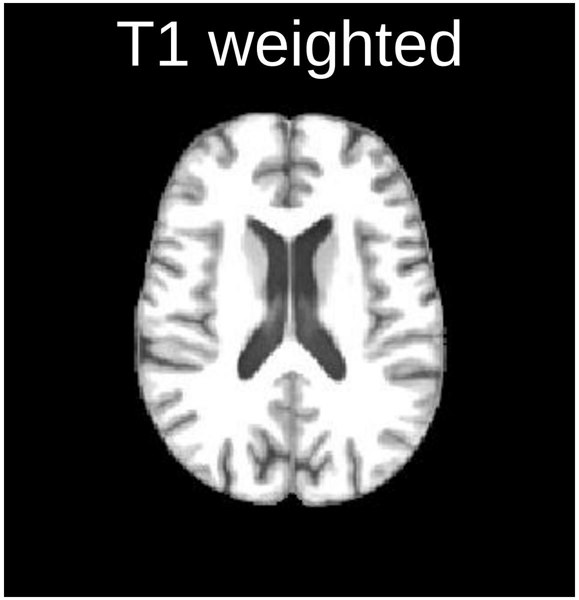
- In clinical MRI, it is common to image the same region of interest under multiple contrast settings
- This aims to increase the diagnostic power of MRI as tissues exhibit different characteristics under different contrasts
- For instance, SRI24 atlas[1] contains such multi-contrast data,



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

Multi-contrast data acquisition

- In clinical MRI, it is common to image the same region of interest under multiple contrast settings
- This aims to increase the diagnostic power of MRI as tissues exhibit different characteristics under different contrasts
- For instance, SRI24 atlas[1] contains such multi-contrast data,



Undersampling the *k*-space

- To reduce data acquisition time, it is possible to collect a subset of *k*-space frequencies below the Nyquist rate due to

$$\boldsymbol{y} = \mathbf{F}_{\Omega}\boldsymbol{x} + \boldsymbol{n}$$

$\boldsymbol{y} \in \mathbb{C}^{\boldsymbol{K}}$ is the undersampled k-space data,

$\mathbf{F}_{\Omega} \in \mathbb{C}^{\boldsymbol{K} \times \boldsymbol{M}}$ is the undersampled 2D-DFT matrix, with $\boldsymbol{K} < \boldsymbol{M}$

$\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{M}}$ is the spatial image and,

$\boldsymbol{n} \in \mathbb{C}^{\boldsymbol{K}}$ is the noise in k-space

Undersampling the *k*-space

- To reduce data acquisition time, it is possible to collect a subset of *k*-space frequencies below the Nyquist rate due to

$$\boldsymbol{y} = \mathbf{F}_{\Omega}\boldsymbol{x} + \boldsymbol{n}$$

$\boldsymbol{y} \in \mathbb{C}^{K}$ is the undersampled k-space data,

$\mathbf{F}_{\Omega} \in \mathbb{C}^{K \times M}$ is the undersampled 2D-DFT matrix, with $K < M$

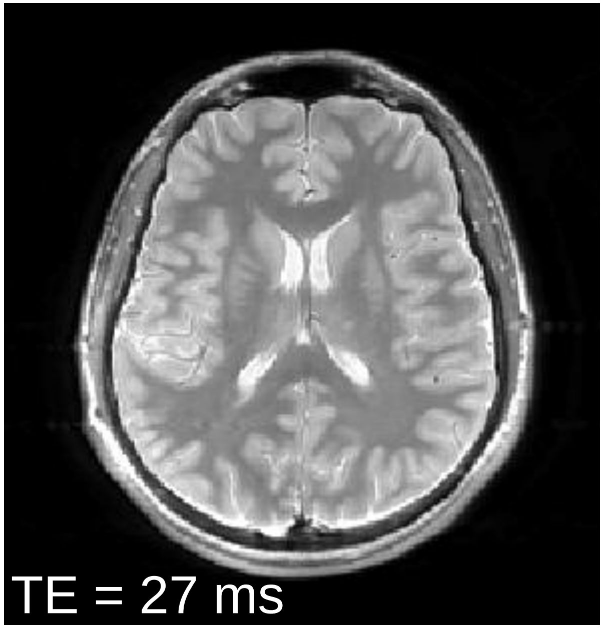
$\boldsymbol{x} \in \mathbb{R}^{M}$ is the spatial image and,

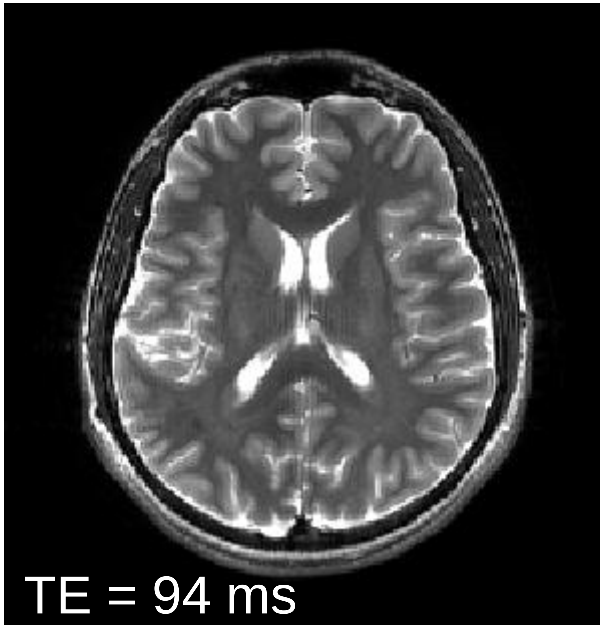
$\boldsymbol{n} \in \mathbb{C}^{K}$ is the noise in k-space

- This work aims to reconstruct multi-contrast data from undersampled acquisitions by making use of
  - Bayesian Compressed Sensing theory and,
  - The similarity between the different contrast images.

Similarity of multi-contrast images

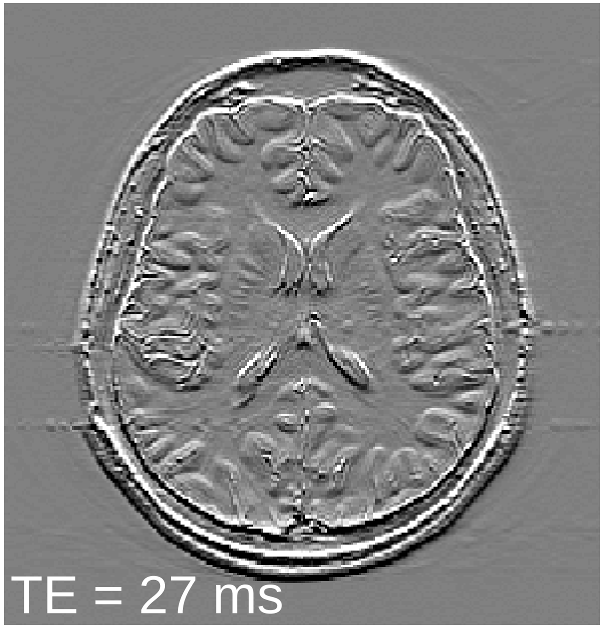
❖ Multi-contrast images possess unique properties, e.g. intensity levels at a given voxel

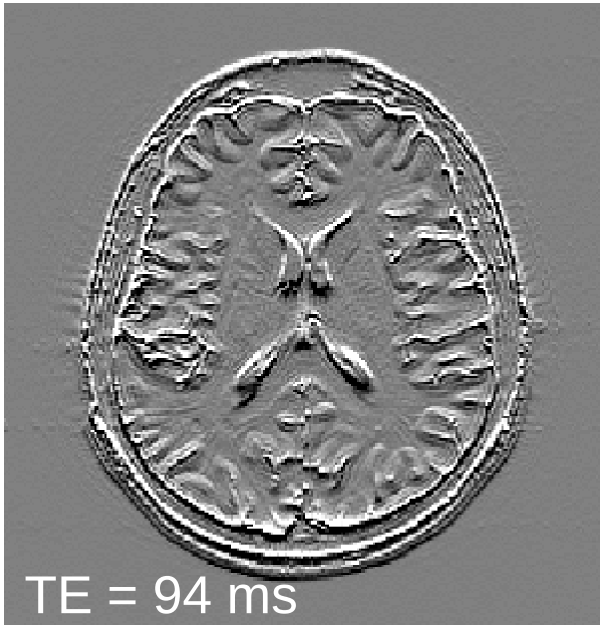




Similarity of multi-contrast images

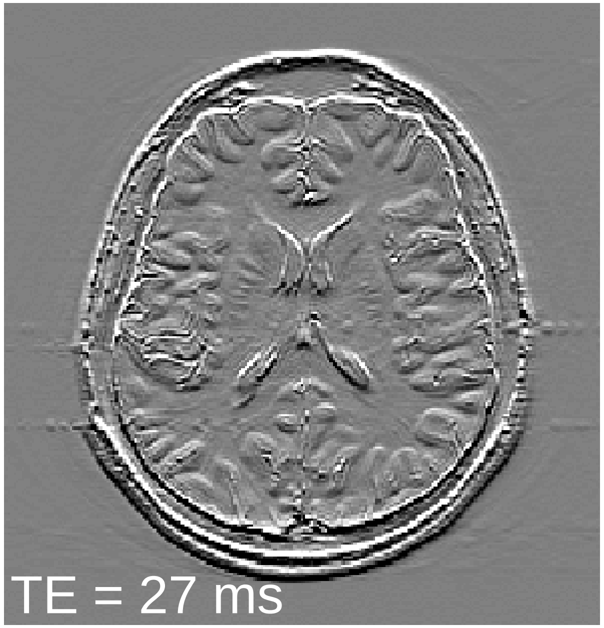
- Multi-contrast images possess unique properties, e.g. intensity levels at a given voxel
- At the same time exhibit common features. We make use of the similarity in sparsity support under gradient transform

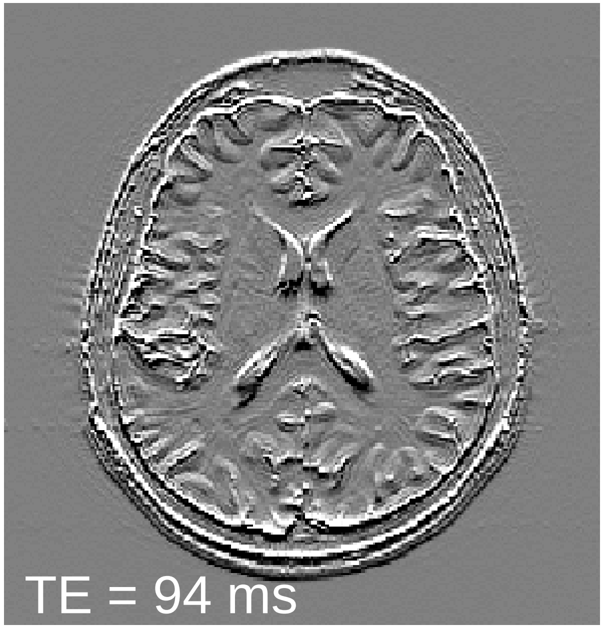




Similarity of multi-contrast images

- Multi-contrast images possess unique properties, e.g. intensity levels at a given voxel
- At the same time exhibit common features. We make use of the similarity in sparsity support under gradient transform

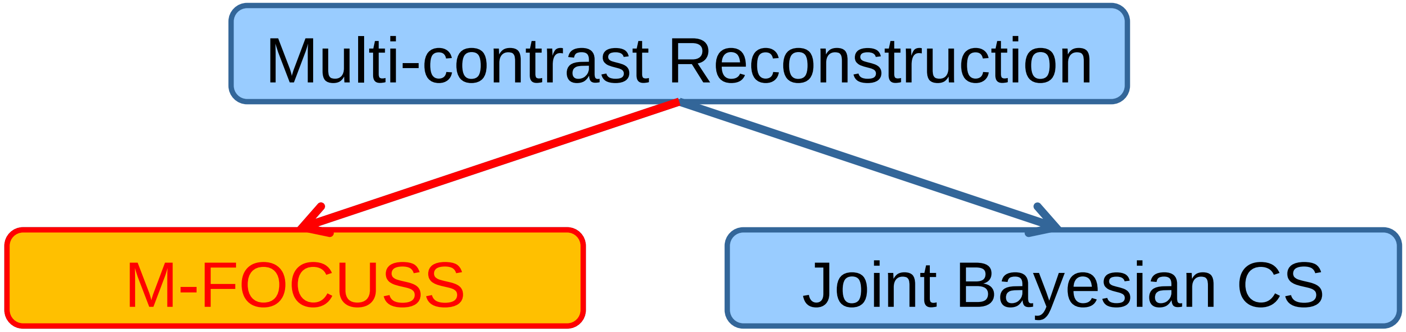




- Positions of non-zero coefficients are similar, even though there is no perfect overlap

Joint reconstruction algorithms

- We consider two joint reconstruction algorithms,



- And first introduce the M-FOCUSS method.

M-FOCUSS algorithm

- First approach is based on using an existing algorithm, **M-FOCUSS**[1] (**M**ultiple-**FOC**al **U**nderdetermined **S**ystem **S**olver) for joint reconstruction

[1] Cotter *et al*. T Signal Proces, 2005

M-FOCUSS algorithm

- First approach is based on using an existing algorithm, **M-FOCUSS**[1] (**M**ultiple-**FOC**al **U**nderdetermined **S**ystem **S**olver) for joint reconstruction
- M-FOCUSS places an ℓ_1 norm penalty on the gradient coefficients of each image, and an ℓ_2 norm penalty across the multi-contrast images

$$\min_{\boldsymbol{x}_i} \sum_{\boldsymbol{i}=1}^{\boldsymbol{L}} \left\| \mathbf{F}_{\Omega} \boldsymbol{x}_{\boldsymbol{i}} - \boldsymbol{y}_{\boldsymbol{i}} \right\|_2^2 + \lambda \cdot \sum_{\boldsymbol{j}=1}^{\boldsymbol{M}} \left(\sum_{\boldsymbol{i}=1}^{\boldsymbol{L}} \left| \partial \mathbf{x}_{\boldsymbol{i},\boldsymbol{j}} \right|^2 \right)^{1/2}$$

[1] Cotter *et al*. T Signal Proces, 2005

M-FOCUSS algorithm

- First approach is based on using an existing algorithm, **M-FOCUSS**[1] (**M**ultiple-**FOC**al **U**nderdetermined **S**ystem **S**olver) for joint reconstruction
- M-FOCUSS places an ℓ_1 norm penalty on the gradient coefficients of each image, and an ℓ_2 norm penalty across the multi-contrast images

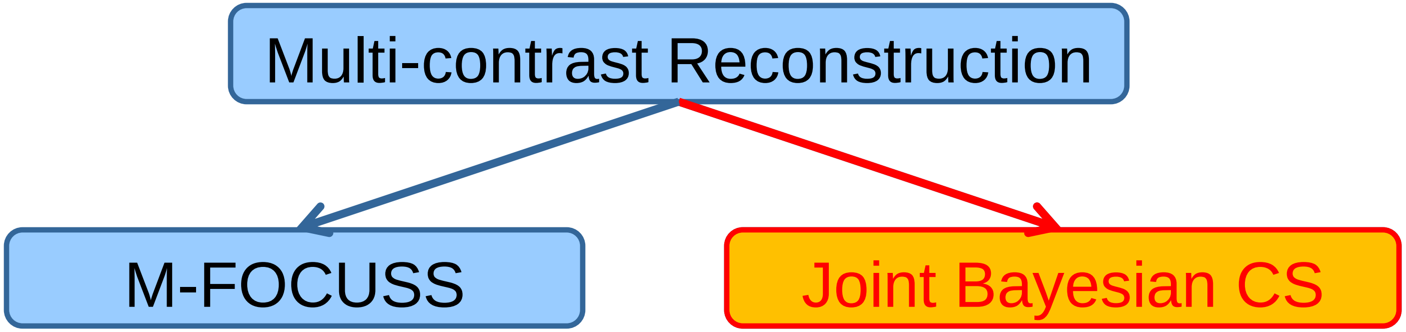
$$\min_{x_i} \sum_{i=1}^{L} \left\| \mathbf{F}_{\Omega} \boldsymbol{x}_i - \boldsymbol{y}_i \right\|_2^2 + \lambda \cdot \sum_{j=1}^{M} \left(\sum_{i=1}^{L} \left| \partial \boldsymbol{x}_{i,j} \right|^2 \right)^{1/2}$$

- As proposed, it is constrained to use the same undersampling pattern for each image
- And makes the strict assumption that the sparsity supports of the images are the same.

[1] Cotter *et al*. T Signal Proces, 2005

Joint reconstruction algorithms

- We consider two joint reconstruction algorithms,



- Next, we introduce our joint Bayesian reconstruction method.

Sparse representation and data likelihood

- To obtain a sparse representation of the images $\left\{\boldsymbol{x}_i\right\}_{i=1}^{L}$ with L different contrasts, we augment the undersampled *k*-space data $\left\{\boldsymbol{y}_i\right\}_{i=1}^{L}$ as

$$\left(1-e^{-2\pi j\omega/n}\right)\cdot \boldsymbol{y}_i(\omega,\upsilon) = \mathbf{F}_{\Omega_i}\,\boldsymbol{\delta}_i^x \equiv \boldsymbol{y}_i^x$$

$\delta_i^x \in \mathbb{R}^M$ is i^{th} vertical image gradient

$\boldsymbol{y}_i^x \in \mathbb{C}^{K_i}$ is the undersampled k-space data of δ_i^x

Sparse representation and data likelihood

- To obtain a sparse representation of the images $\{\boldsymbol{x}_i\}_{i=1}^{L}$ with L different contrasts, we augment the undersampled *k*-space data $\{\boldsymbol{y}_i\}_{i=1}^{L}$ as

$$\left(1-e^{-2\pi j\omega/n}\right)\cdot \boldsymbol{y}_i(\omega,\upsilon)=\mathbf{F}_{\Omega_i}\,\boldsymbol{\delta}_i^x \equiv \boldsymbol{y}_i^x$$

$\delta_i^x \in \mathbb{R}^M$ is $\boldsymbol{i}^{\text{th}}$ vertical image gradient

$\boldsymbol{y}_i^x \in \mathbb{C}^{K_i}$ is the undersampled k - space data of δ_i^x

- Modeling the *k*-space noise to be Gaussian with zero mean and variance σ^2, the likelihood of observing the data becomes

$$\left.\begin{aligned} \boldsymbol{Y}_i^x &= \left[\mathcal{R}e\left(\boldsymbol{y}_i^x\right),\mathcal{I}m\left(\boldsymbol{y}_i^x\right)\right]^T \\ \Phi_i &= \left[\mathcal{R}e\left(\mathbf{F}_{\Omega_i}\right),\mathcal{I}m\left(\mathbf{F}_{\Omega_i}\right)\right]^T \end{aligned}\right\} \boldsymbol{p}\left(\boldsymbol{Y}_i^x \mid \delta_i^x,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-K_i}\exp\left(-\left\|\boldsymbol{Y}_i^x-\Phi_i\delta_i^x\right\|_2^2/2\sigma^2\right)$$

Bayesian analysis for joint inference

- Next, we would like to impose a sparsity promoting prior distribution over the image gradients $\left\{\delta_i^x\right\}_{i=1}^{L}$ and $\left\{\delta_i^y\right\}_{i=1}^{L}$,
- And compute their posterior distribution with the Bayes' rule using this prior, the likelihood term and the observed *k*-space data $\left\{\boldsymbol{Y}_i^x\right\}_{i=1}^{L}$ and $\left\{\boldsymbol{Y}_i^y\right\}_{i=1}^{L}$
- At the same time, we would like to enable information sharing across the multi-contrast images.

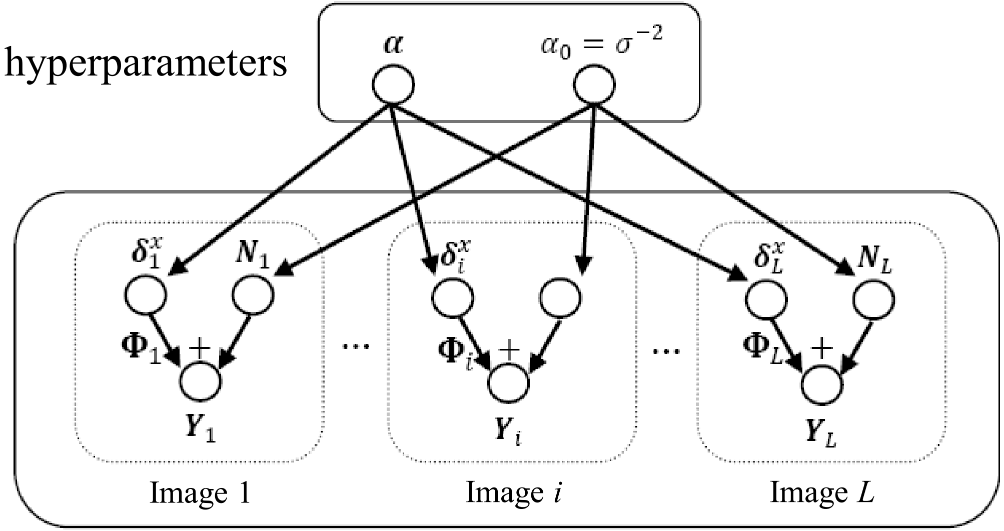
Bayesian analysis for joint inference

- Next, we would like to impose a sparsity promoting prior distribution over the image gradients $\left\{\delta_i^x\right\}_{i=1}^L$ and $\left\{\delta_i^y\right\}_{i=1}^L$,
- And compute their posterior distribution with the Bayes' rule using this prior, the likelihood term and the observed *k*-space data $\left\{\boldsymbol{Y}_i^x\right\}_{i=1}^L$ and $\left\{\boldsymbol{Y}_i^y\right\}_{i=1}^L$
- At the same time, we would like to enable information sharing across the multi-contrast images.
- To this end, we carry out the inference within a hierarchical Bayesian model[1]

[1] Ji *et al*. IEEE T Signal Proces, 2009

Hierarchical Bayesian Model for joint inference

- At the bottom layer, we have the undersampled *k*-space observations, which are jointly parameterized by the hyperparameters on the layer above.

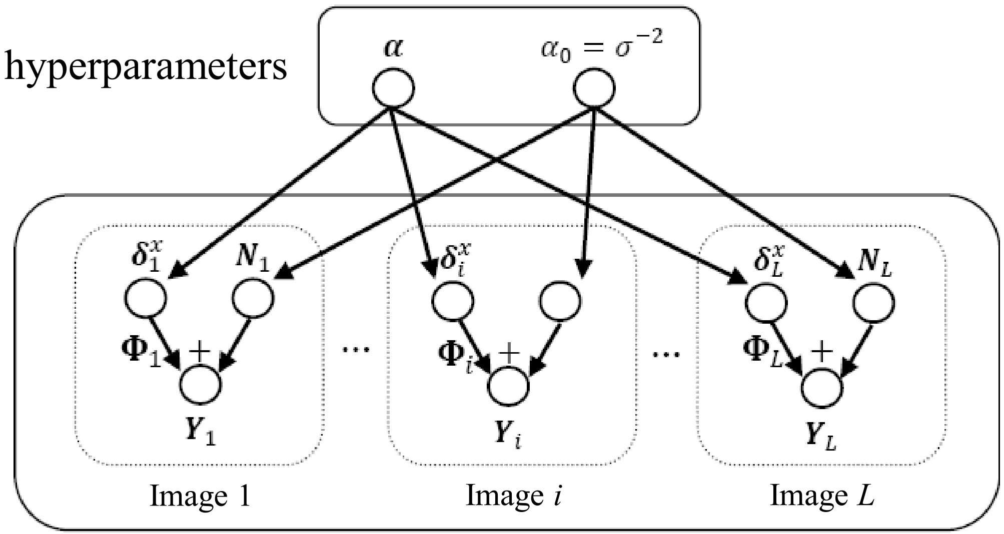


coupled by hyperparameters $\boldsymbol{\alpha}$ and $\alpha_0 = \sigma^{-2}$

k-space observations

Hierarchical Bayesian Model for joint inference

- At the bottom layer, we have the undersampled *k*-space observations, which are jointly parameterized by the hyperparameters on the layer above.
- We capture the similarity in the gradient domain by defining the hyperparameters $\boldsymbol{\alpha}$ over the L gradient images

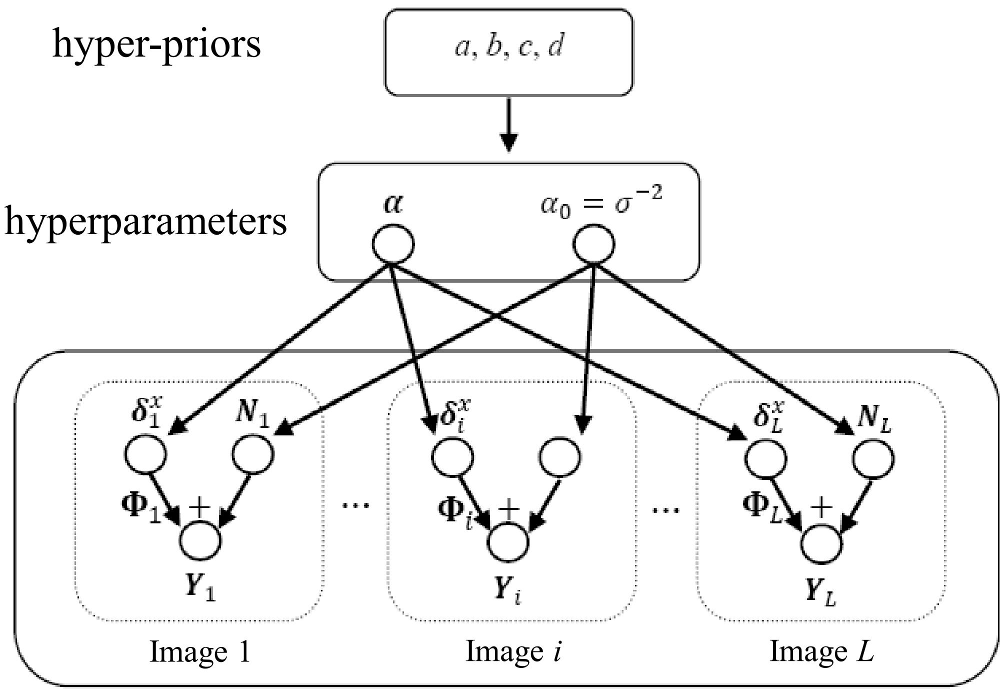


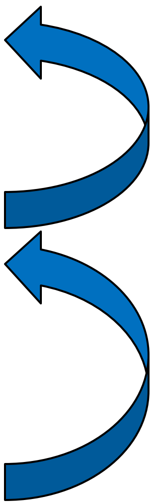
coupled by hyperparameters $\boldsymbol{\alpha}$ and $\alpha_0 = \sigma^{-2}$

k-space observations

Hierarchical Bayesian Model for joint inference

- At the bottom layer, we have the undersampled *k*-space observations, which are jointly parameterized by the hyperparameters on the layer above.
- We capture the similarity in the gradient domain by defining the hyperparameters $\boldsymbol{\alpha}$ over the L gradient images
- The hyperparameters are in turn controlled by the hyper-priors at the top level.





controlled by
a, b, c, d

coupled by hyperparameters
$\boldsymbol{\alpha}$ and $\alpha_0 = \sigma^{-2}$

k-space observations

Prior on the signal coefficients

- The gradient coefficients are modeled to be drawn from a product of zero mean Gaussians

$$p\left(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha}\right) = \prod_{j=1}^{M} \mathcal{N}\left(\delta_{i,j}^x \mid 0, \alpha_j^{-1}\right)$$

and the precisions of the Gaussians are determined by $\boldsymbol{\alpha} \in \mathbb{R}^M$

- And Gamma priors are placed over the hyperparameters $\boldsymbol{\alpha}$

$$p(\boldsymbol{\alpha} \mid c, d) = \prod_{j=1}^{M} Ga(\alpha_j \mid c, d)$$

where $Ga\left(\alpha_j \mid c,d\right) = \frac{d^c}{\Gamma(c)} \alpha_j^{c-1} exp\left(-d\alpha_j\right)$

Prior on the signal coefficients

- The gradient coefficients are modeled to be drawn from a product of zero mean Gaussians

$$p\left(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha}\right)=\prod_{j=1}^{M}\mathcal{N}\left(\delta_{i,j}^x \mid 0,\alpha_j^{-1}\right)$$

and the precisions of the Gaussians are determined by $\boldsymbol{\alpha}\in\mathbb{R}^M$

- And Gamma priors are placed over the hyperparameters $\boldsymbol{\alpha}$

$$p(\boldsymbol{\alpha}\mid c,d)=\prod_{j=1}^{M}Ga(\alpha_j\mid c,d)$$ where $$Ga\left(\alpha_j|c,d\right)=\frac{d^c}{\Gamma(c)}\alpha_j^{\,c-1}exp\left(-d\alpha_j\right)$$

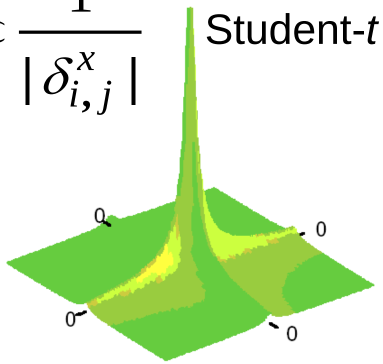
- We can marginalize over the hyperparameters $\boldsymbol{\alpha}$ and obtain the *marginal* prior that enforces sparsity

$$p(\delta_{i,j}^x)=\int p(\delta_{i,j}^x|\alpha_j)p(\alpha_j\mid c,d)d\alpha_j$$

sharp peak at 0 → $c,d=0$

$$p(\delta_{i,j}^x)\propto\frac{1}{|\delta_{i,j}^x|}$$

Student-*t*



Computing the posterior for the signals

- Since the data likelihood and the signal prior are both Gaussian, the posterior for the gradient coefficients is also in the same family,

$$p(\boldsymbol{\delta}_i^x \mid \mathbf{Y}_i^x, \boldsymbol{\alpha}, \alpha_0) = \frac{p(\mathbf{Y}_i^x \mid \boldsymbol{\delta}_i^x, \alpha_0)\, p(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha})}{p(\mathbf{Y}_i^x \mid \boldsymbol{\alpha}, \alpha_0)}$$

Computing the posterior for the signals

- Since the data likelihood and the signal prior are both Gaussian, the posterior for the gradient coefficients is also in the same family,

posterior

$$\boxed{p(\boldsymbol{\delta}_i^x \mid \mathbf{Y}_i^x, \boldsymbol{\alpha}, \alpha_0)} = \frac{p(\mathbf{Y}_i^x \mid \boldsymbol{\delta}_i^x, \alpha_0)\, p(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha})}{p(\mathbf{Y}_i^x \mid \boldsymbol{\alpha}, \alpha_0)}$$

Computing the posterior for the signals

- Since the data likelihood and the signal prior are both Gaussian, the posterior for the gradient coefficients is also in the same family,

likelihood

posterior

$$p(\boldsymbol{\delta}_i^x \mid \mathbf{Y}_i^x, \boldsymbol{\alpha}, \alpha_0) = \frac{p(\mathbf{Y}_i^x \mid \boldsymbol{\delta}_i^x, \alpha_0)\, p(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha})}{p(\mathbf{Y}_i^x \mid \boldsymbol{\alpha}, \alpha_0)}$$

Computing the posterior for the signals

- Since the data likelihood and the signal prior are both Gaussian, the posterior for the gradient coefficients is also in the same family,

posterior, likelihood, prior

$$\underbrace{p(\boldsymbol{\delta}_i^x \mid \mathbf{Y}_i^x, \boldsymbol{\alpha}, \alpha_0)}_{\text{posterior}} = \frac{\overbrace{p(\mathbf{Y}_i^x \mid \boldsymbol{\delta}_i^x, \alpha_0)}^{\text{likelihood}}\,\overbrace{p(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha})}^{\text{prior}}}{p(\mathbf{Y}_i^x \mid \boldsymbol{\alpha}, \alpha_0)}$$

Computing the posterior for the signals

- Since the data likelihood and the signal prior are both Gaussian, the posterior for the gradient coefficients is also in the same family,

Gaussian Gaussian

$$p(\boldsymbol{\delta}_i^x \mid \mathbf{Y}_i^x, \boldsymbol{\alpha}, \alpha_0) = \frac{p(\mathbf{Y}_i^x \mid \boldsymbol{\delta}_i^x, \alpha_0)\, p(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha})}{p(\mathbf{Y}_i^x \mid \boldsymbol{\alpha}, \alpha_0)}$$

Computing the posterior for the signals

- Since the data likelihood and the signal prior are both Gaussian, the posterior for the gradient coefficients is also in the same family,

$$\underbrace{p(\boldsymbol{\delta}_i^x \mid \mathbf{Y}_i^x, \boldsymbol{\alpha}, \alpha_0)}_{\text{also Gaussian}} = \frac{\overbrace{p(\mathbf{Y}_i^x \mid \boldsymbol{\delta}_i^x, \alpha_0)}^{\text{Gaussian}}\ \overbrace{p(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha})}^{\text{Gaussian}}}{p(\mathbf{Y}_i^x \mid \boldsymbol{\alpha}, \alpha_0)}$$

Computing the posterior for the signals

- Since the data likelihood and the signal prior are both Gaussian, the posterior for the gradient coefficients is also in the same family,

also Gaussian — Gaussian — Gaussian

$$\underbrace{p(\boldsymbol{\delta}_i^x \mid \mathbf{Y}_i^x, \boldsymbol{\alpha}, \alpha_0)}_{\text{also Gaussian}} = \frac{\overbrace{p(\mathbf{Y}_i^x \mid \boldsymbol{\delta}_i^x, \alpha_0)}^{\text{Gaussian}}\,\overbrace{p(\boldsymbol{\delta}_i^x \mid \boldsymbol{\alpha})}^{\text{Gaussian}}}{p(\mathbf{Y}_i^x \mid \boldsymbol{\alpha}, \alpha_0)}$$

$$\boldsymbol{\delta}_i^x \approx \mathcal{N}\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)$$

$$\boldsymbol{\mu}_i = \alpha_0\,\boldsymbol{\Sigma}_i\,\boldsymbol{\Phi}_i^T\,\mathbf{Y}_i^x$$

$$\boldsymbol{\Sigma}_i = (\alpha_0\,\boldsymbol{\Phi}_i^T\,\boldsymbol{\Phi}_i + \mathbf{A})^{-1}$$

$$\mathbf{A} = diag(\alpha_1, \alpha_2, \ldots, \alpha_M)$$

We only need to estimate the α_i's

Maximum Likelihood estimation of hyperparameters

- We seek point estimates for the hyperparameters $\boldsymbol{\alpha}$ and α_0 in a maximum likelihood framework.

$$\max_{\boldsymbol{\alpha},\alpha_0} \mathcal{L}(\boldsymbol{\alpha},\alpha_0) = \max_{\boldsymbol{\alpha},\alpha_0} \sum_{i=1}^{L} \log\ p\left(\mathbf{Y}_i^x \mid \boldsymbol{\alpha},\alpha_0\right)$$

- Summation over the L images enables **information sharing** while estimating the hyperparameters.

- Once the hyperparameters are estimated, the posterior for the gradient coefficients $\boldsymbol{\delta}_i^x$ is determined based only on its related *k*-space data $\boldsymbol{Y}_i^x$ due to,

$$\boldsymbol{\mu}_i = \alpha_0 \boldsymbol{\Sigma}_i \boldsymbol{\Phi}_i^T \boldsymbol{Y}_i^x$$

Reconstructing the images from their gradients

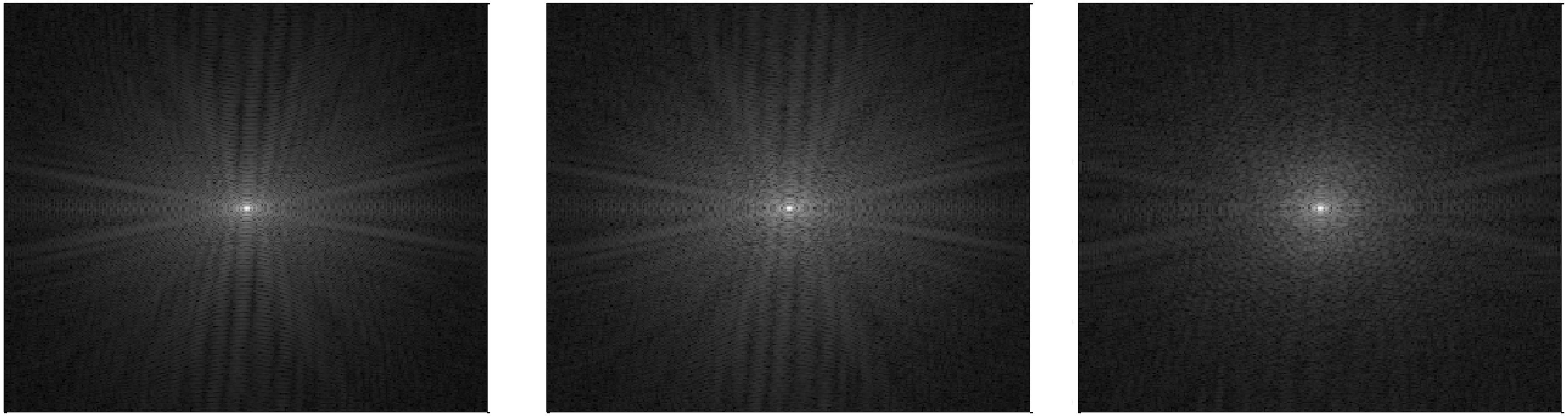
- After estimating the vertical and horizontal gradients $\left\{\delta_i^x\right\}_{i=1}^{L}$ and $\left\{\delta_i^y\right\}_{i=1}^{L}$, we seek the images $\left\{\boldsymbol{x}_i\right\}_{i=1}^{L}$ consistent with these and the *k*-space data $\left\{\boldsymbol{y}_i\right\}_{i=1}^{L}$ in a Least Squares setting,

$$\hat{\mathbf{x}}_i = \underset{\boldsymbol{x}_i}{argmin} \left\| \partial_x \boldsymbol{x}_i - \boldsymbol{\delta}_i^x \right\|_2^2 + \left\| \partial_y \boldsymbol{x}_i - \boldsymbol{\delta}_i^y \right\|_2^2 + \lambda \left\| \mathbf{F}_{\Omega_i} \boldsymbol{x}_i - \boldsymbol{y}_i \right\|_2^2$$

$$for\ i = 1,\dots,L$$

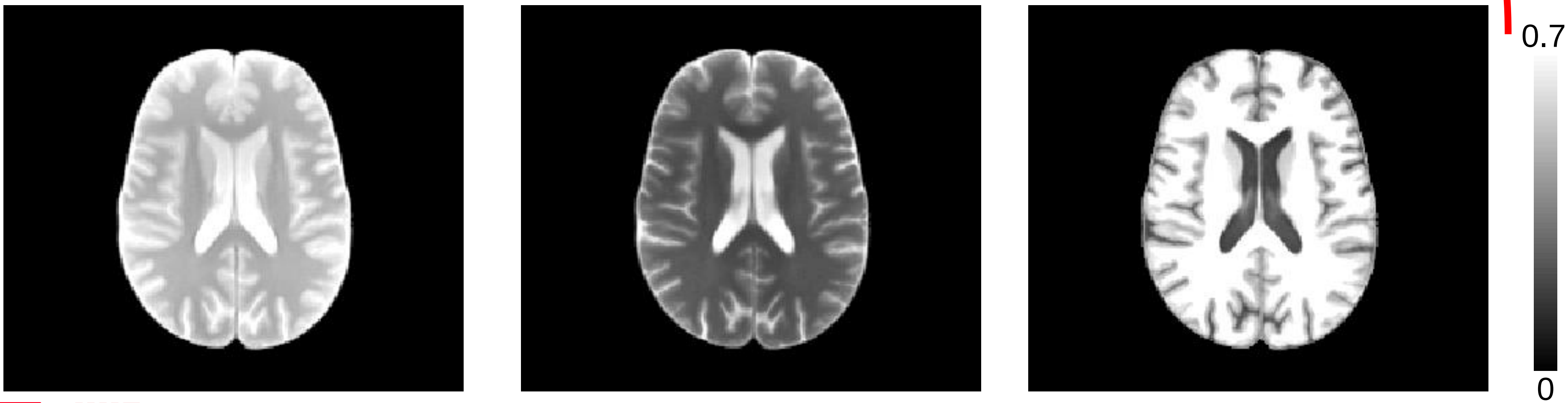
where ∂_x and ∂_y are vertical and horizontal gradient operators

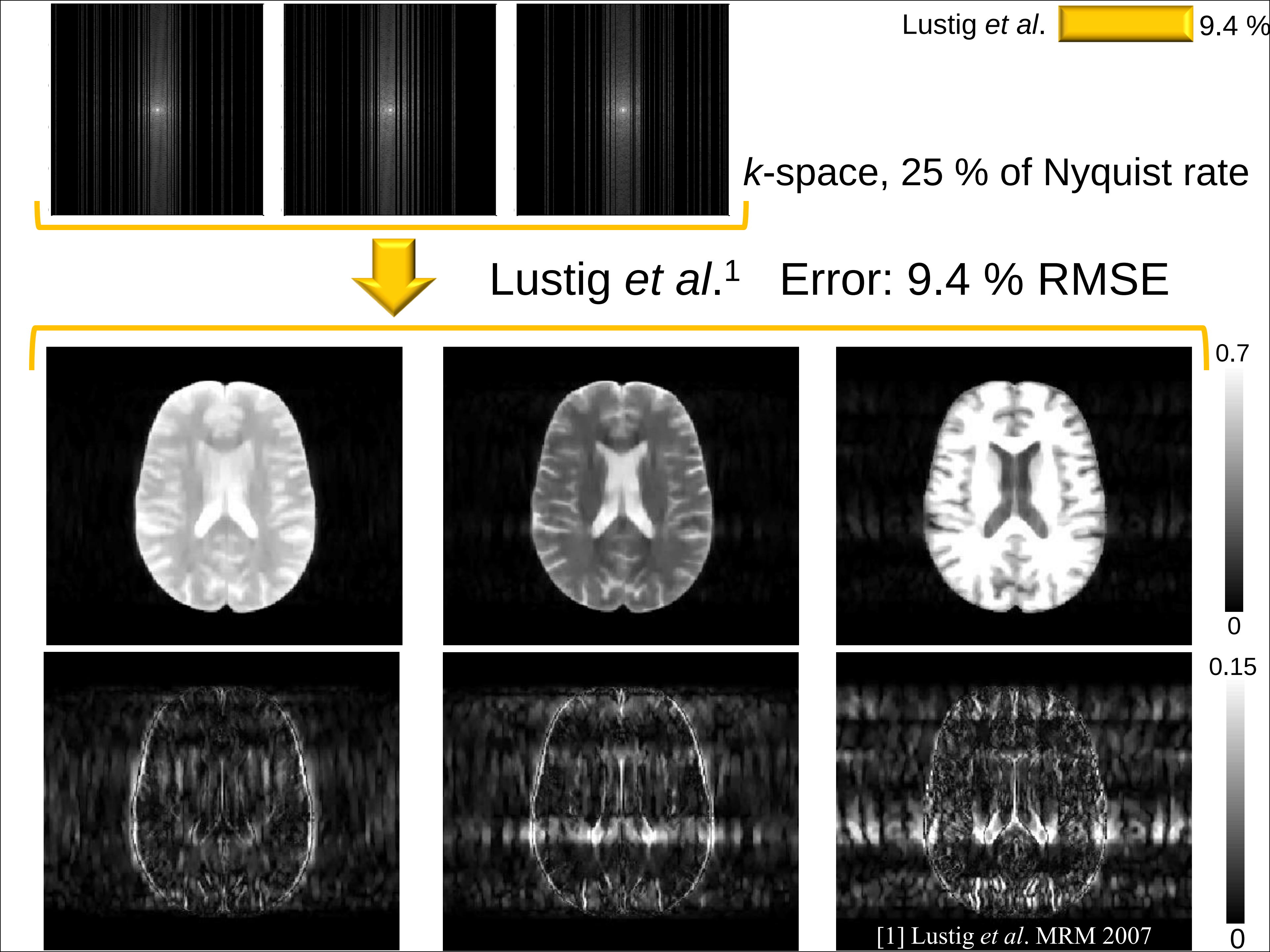
SRI24 Atlas

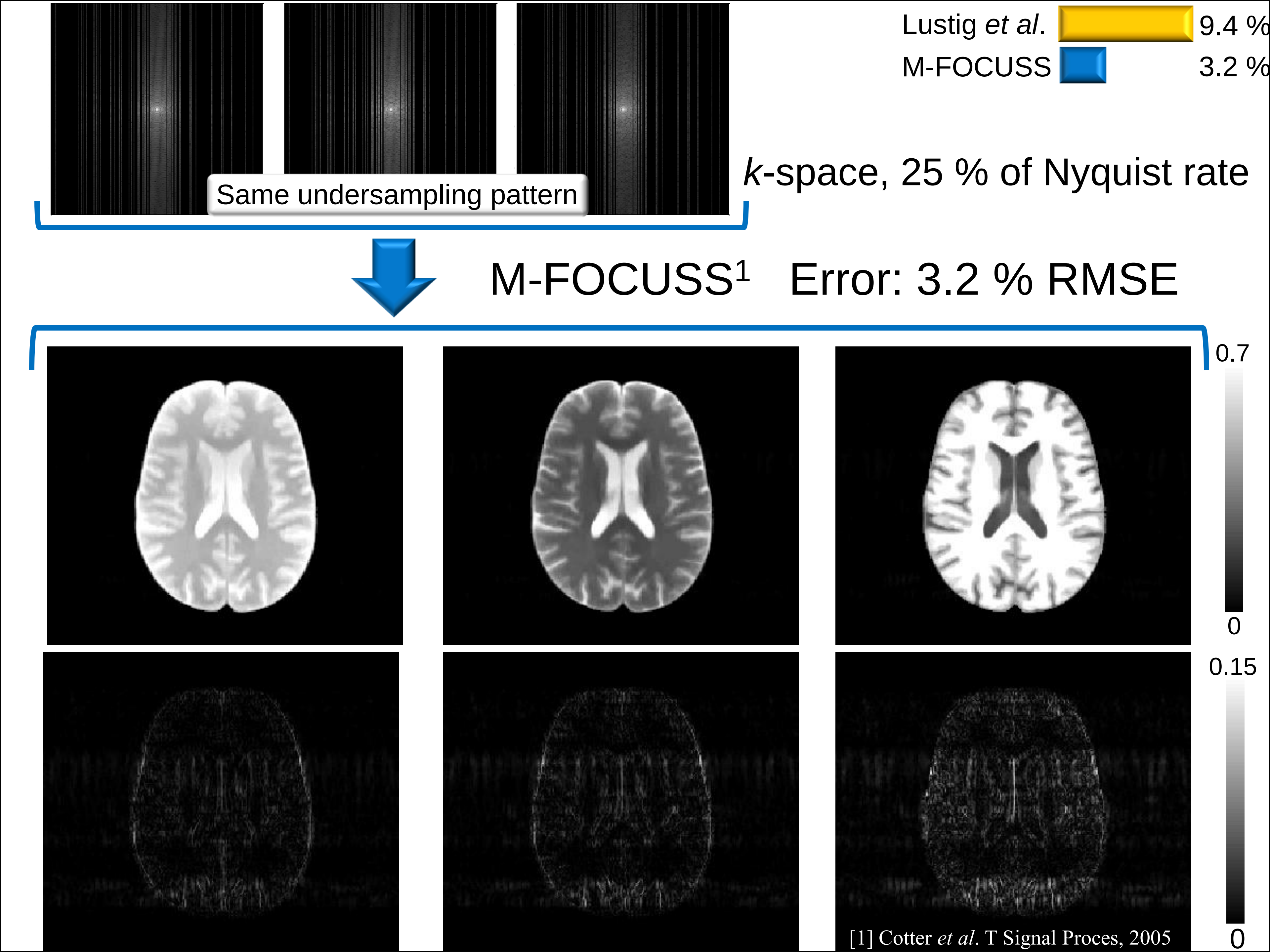


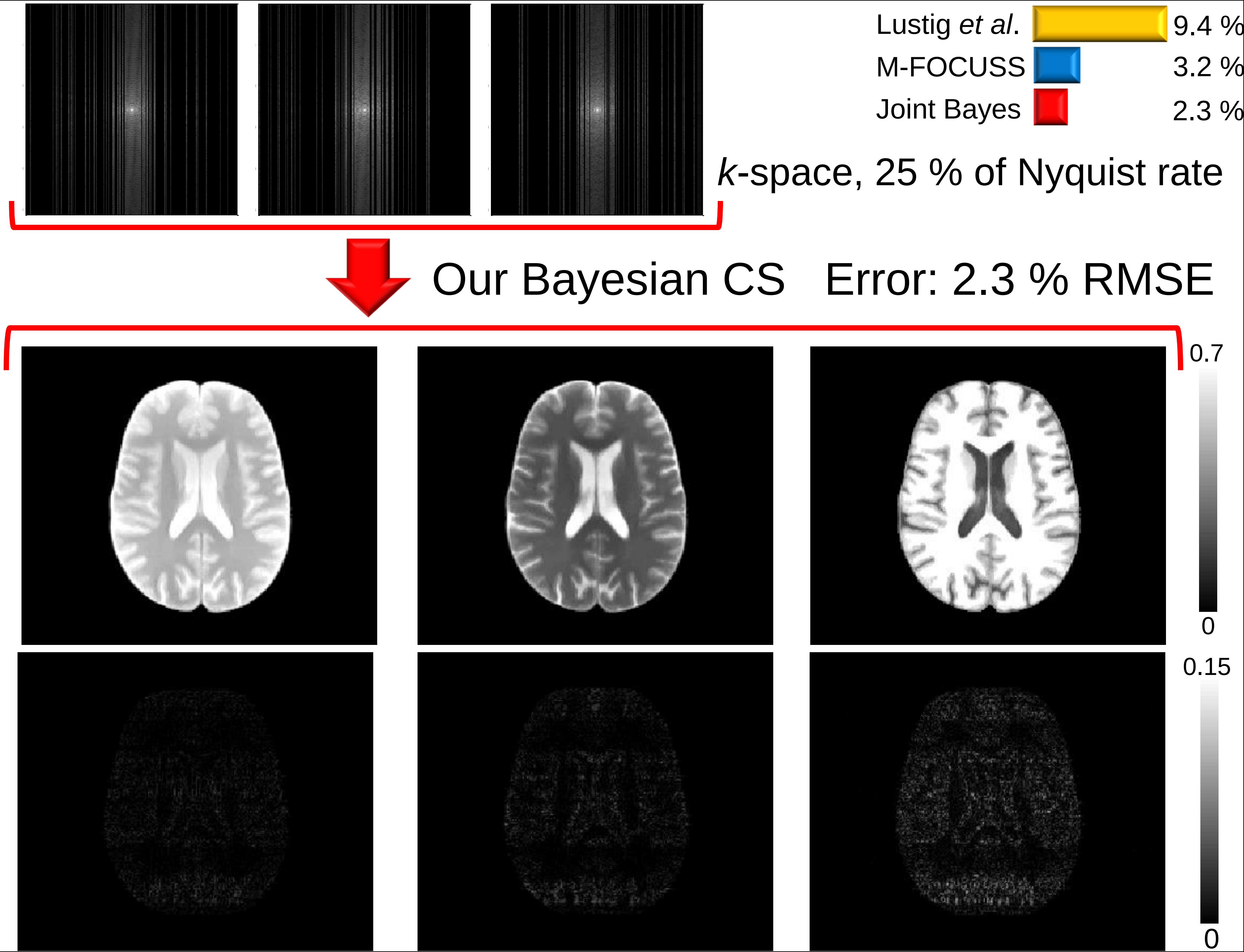
k-space, 100 % of Nyquist rate

Inverse FFT   Error: 0 % RMSE




Lustig et al.
9.4 %
k-space, 25 % of Nyquist rate
Lustig et al.[1] Error: 9.4 % RMSE
0.7
0
0.15
0
[1] Lustig et al. MRM 2007

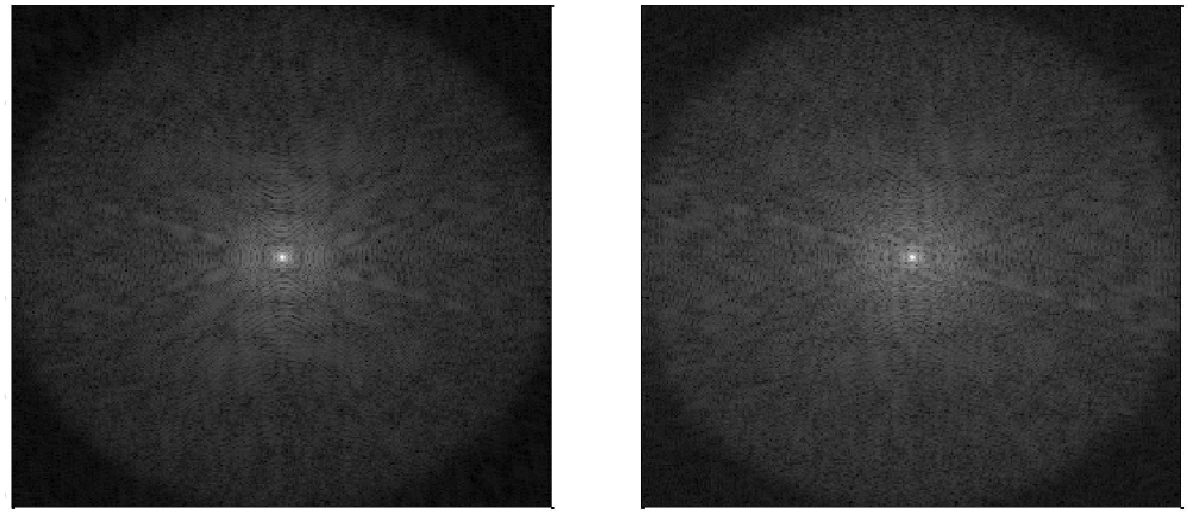

Lustig et al.
9.4 %
M-FOCUSS
3.2 %
Same undersampling pattern
k-space, 25 % of Nyquist rate
M-FOCUSS[1] Error: 3.2 % RMSE
0.7
0
0.15
0
[1] Cotter et al. T Signal Proces, 2005



Lustig et al.
9.4 %
M-FOCUSS
3.2 %
Joint Bayes
2.3 %
k-space, 25 % of Nyquist rate
Our Bayesian CS
Error: 2.3 % RMSE
0.7
0
0.15
0

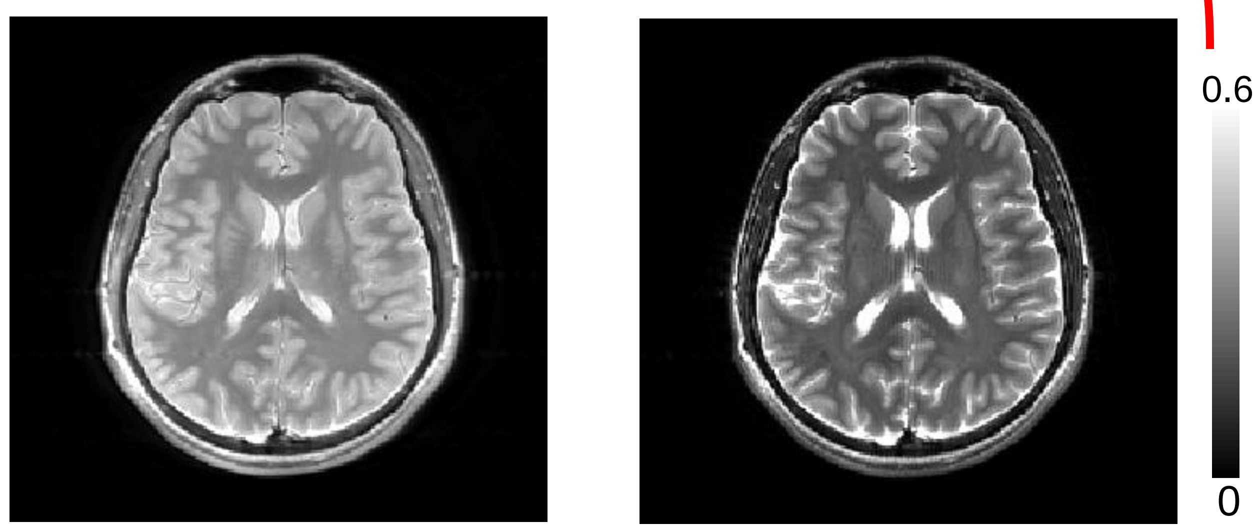
TSE Scans : *in vivo* acquisition

k-space
100 % of Nyquist rate



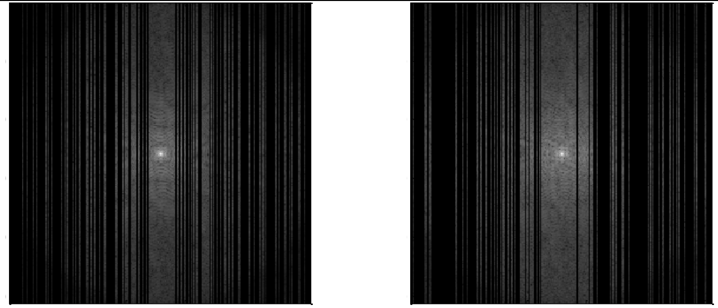
Inverse FFT

Error: 0 % RMSE



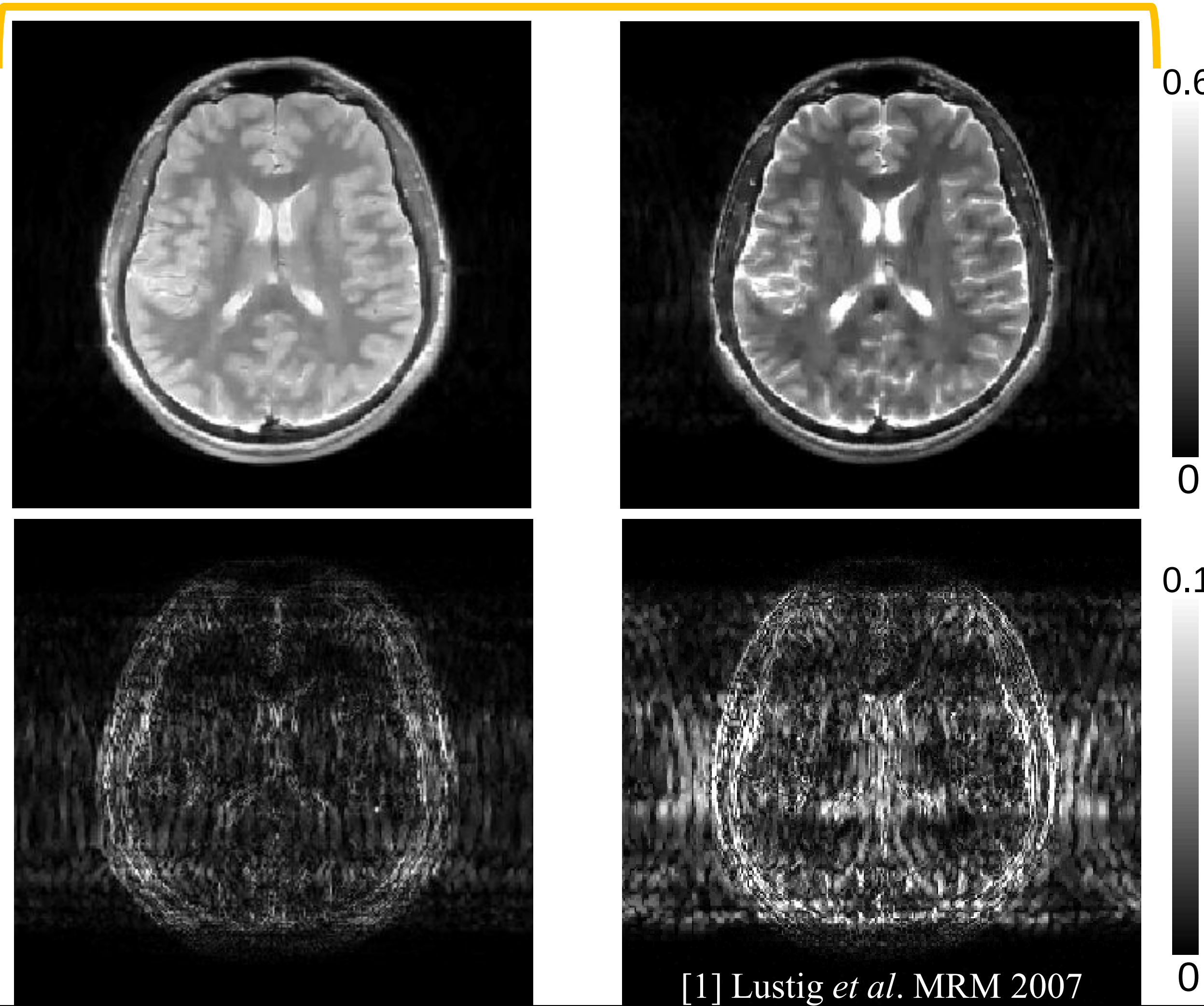
Lustig *et al*. 9.4 %

k-space, 40 % of Nyquist rate



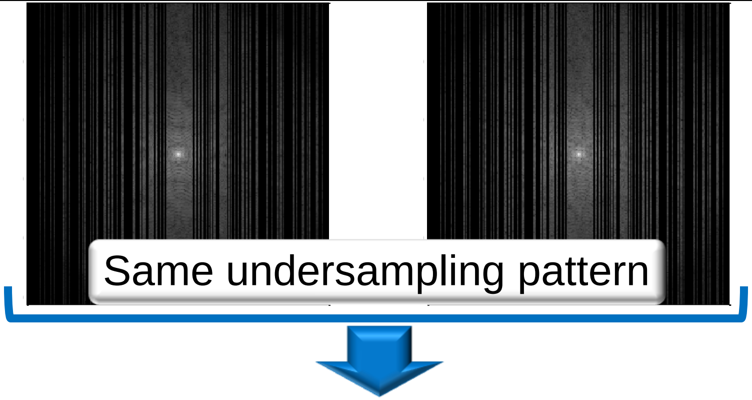
Lustig *et al.*[1]

Error: 9.4 % RMSE



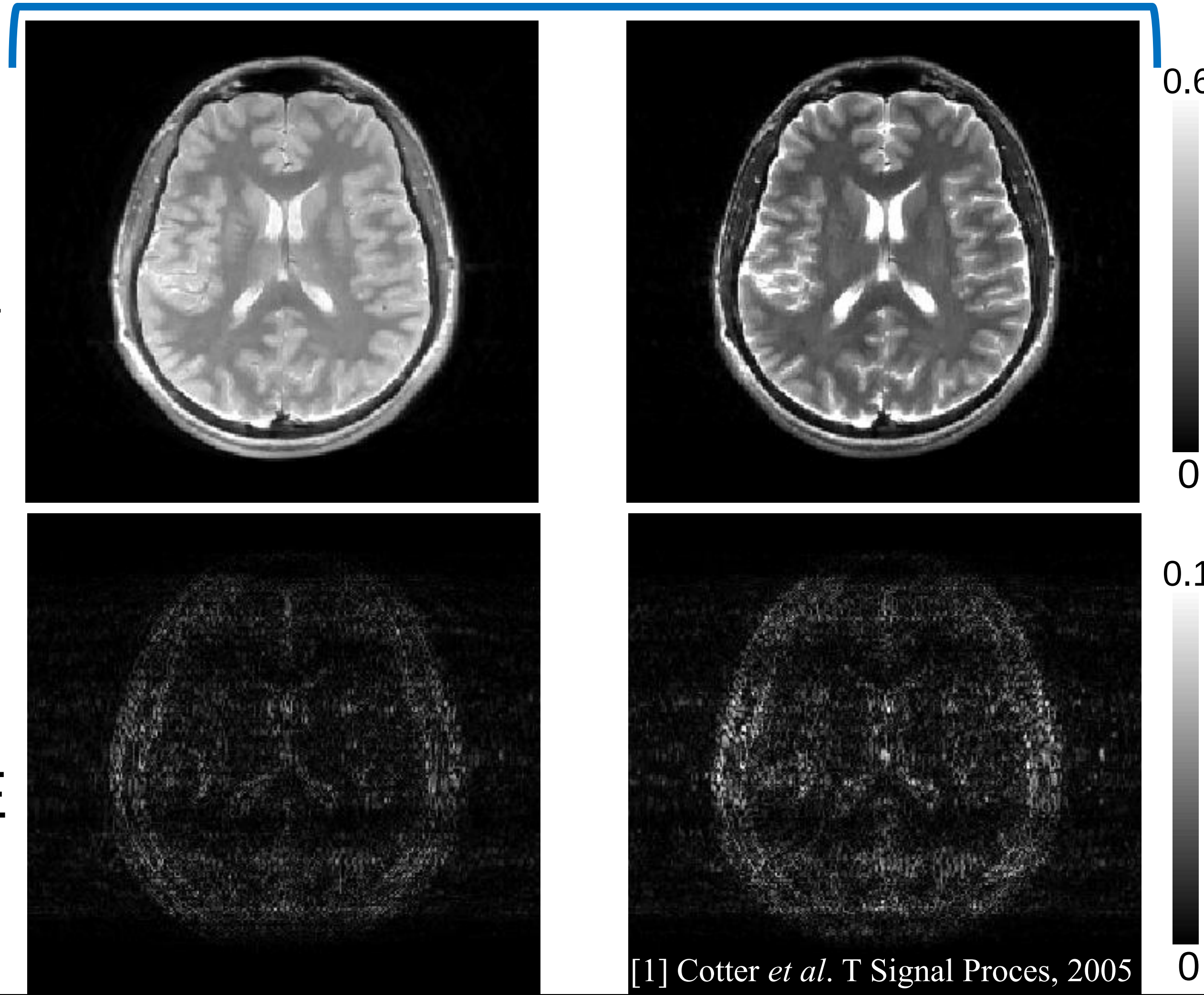
[1] Lustig *et al*. MRM 2007

k-space, 40 % of Nyquist rate

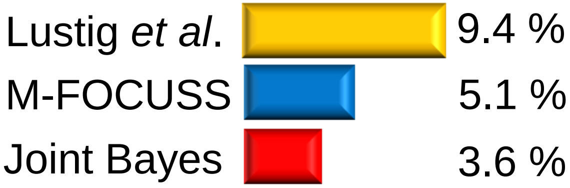


M-FOCUSS[1]

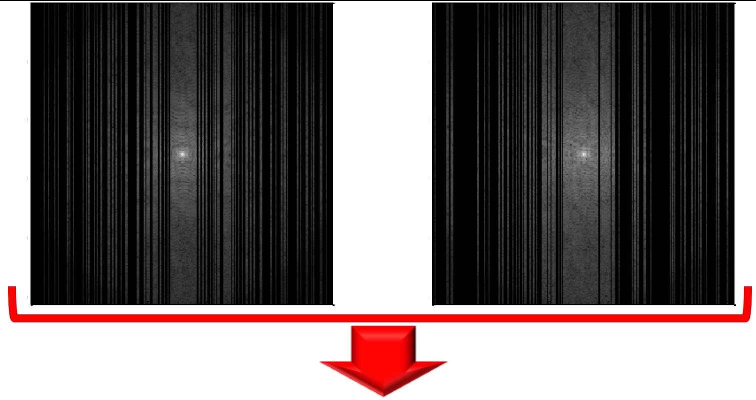
Error: 5.1 % RMSE



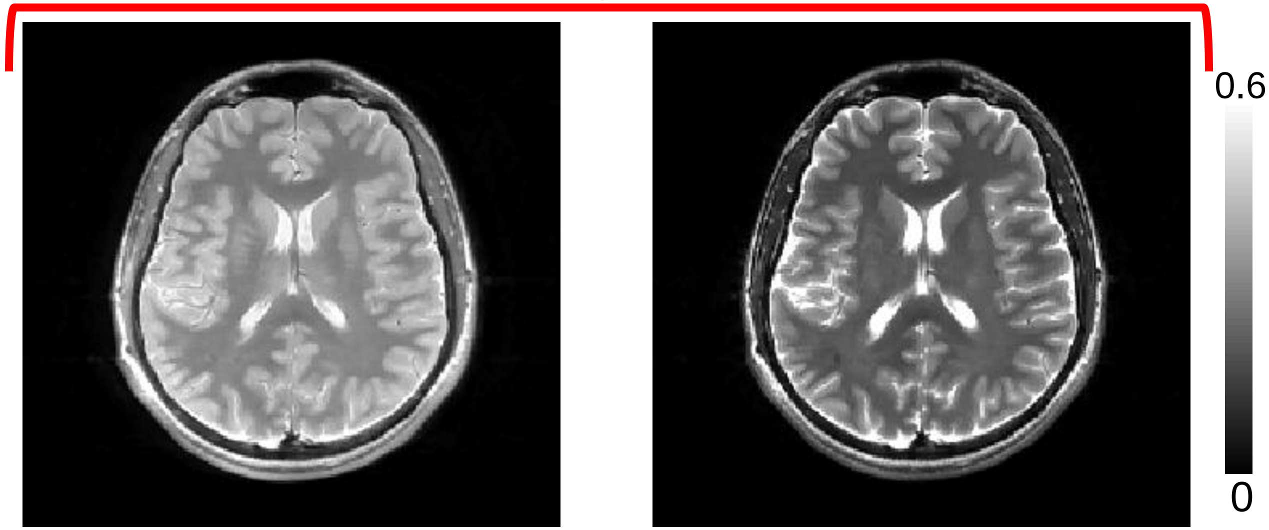
[1] Cotter *et al*. T Signal Proces, 2005



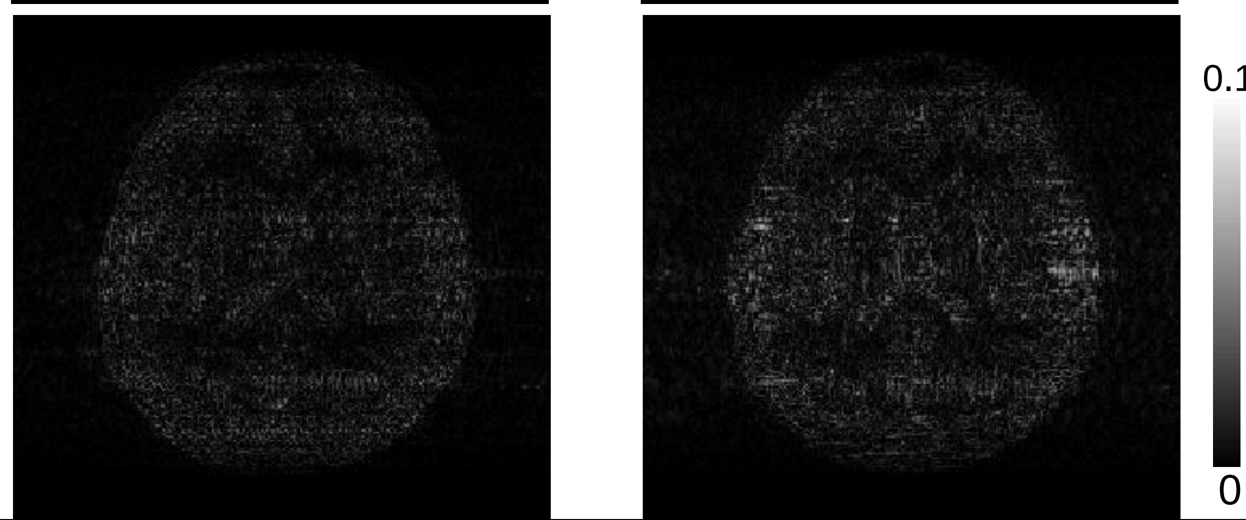
k-space, 40 % of Nyquist rate



Our Bayesian CS



Error: 3.6 % RMSE



Extensions and Limitations

- We assumed the multi-contrast images to be real-valued. Extension to complex-valued images is possible by using a mirror-symmetric sampling pattern and separating real and imaginary parts of the images.

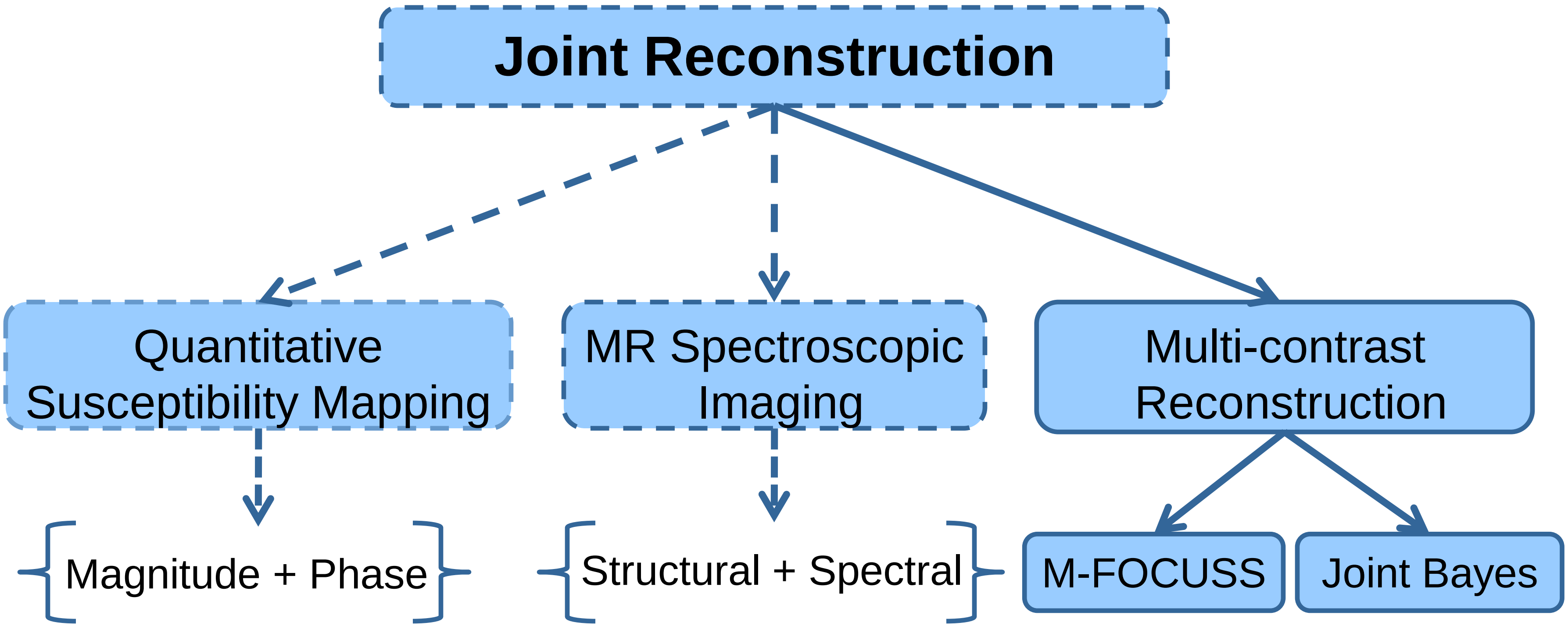
Extensions and Limitations

- We assumed the multi-contrast images to be real-valued. Extension to complex-valued images is possible by using a mirror-symmetric sampling pattern and separating real and imaginary parts of the images.

- Whereas the other two methods take under an hour, the Bayesian method takes about **20 hours** with this initial implementation.

- Current work is on increasing the reconstruction speed using
  - Graphics Processing Cards (GPUs) on the hardware front, and
  - Employing variational Bayesian analysis on the algorithm front

Other applications of joint reconstruction



Conclusion

- We presented two joint reconstruction algorithms, M-FOCUSS and joint Bayesian CS, that significantly improved reconstruction quality of multi-contrast images from undersampled data.
- While joint Bayesian method reduced reconstruction errors by up to 4 times relative to a popular CS algorithm[1], current implementation suffers from long reconstruction times.
- M-FOCUSS is a notable candidate that trades off reconstruction quality and processing speed.

[1] Lustig *et al*. MRM 2007