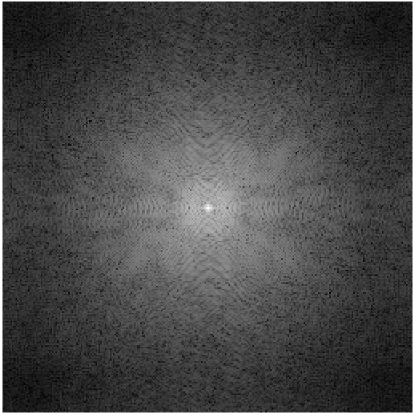
Learning a Variational Network for Reconstruction of Accelerated MRI Data

Kerstin Hammernik[1–3], Erich Kobler[1], Michael P Recht[2,3], Daniel K Sodickson[2,3], Thomas Pock[1,4], Florian Knoll[2,3]

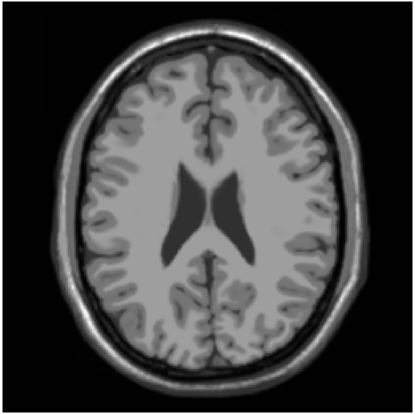
[1]Institute of Computer Graphics and Vision, Graz University of Technology, Austria
[2]Center for Biomedical Imaging, Department of Radiology, NYU School of Medicine
[3]Center for Advanced Imaging Innovation and Research, NYU School of Medicine
[4]AIT Austrian Institute of Technology GmbH, Vienna, Austria

Deep Reconstruction Workshop 2017

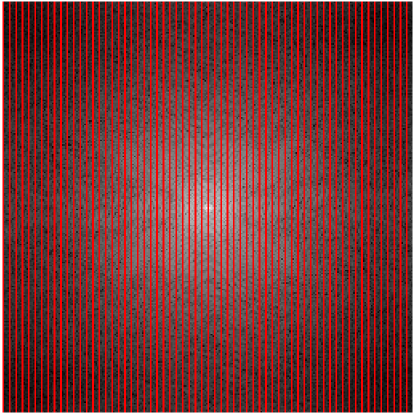
MRI reconstruction in a nutshell



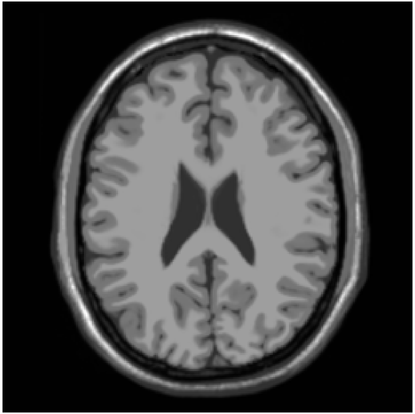
$$\xrightarrow{\mathcal{FT}^{-1}}$$



MRI reconstruction in a nutshell

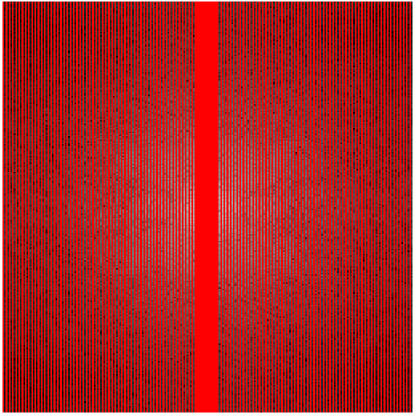


$$\xrightarrow{\mathcal{FT}^{-1}}$$



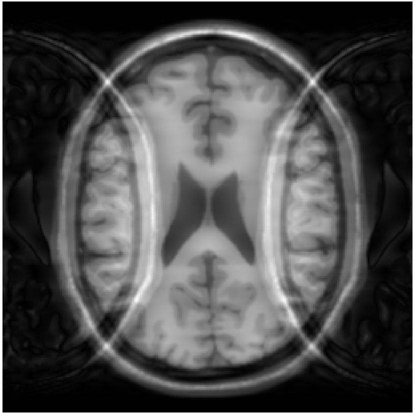
Cartesian undersampling

Accelerated MRI in a nutshell



$\xrightarrow{\mathcal{FT}^{-1}}$

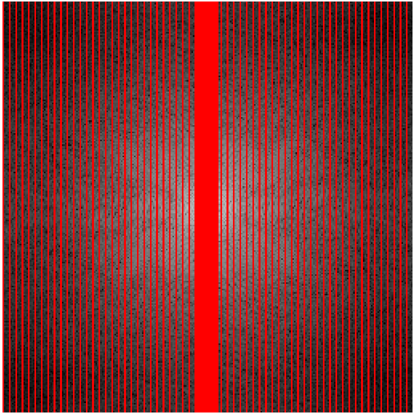
$R = 2$



Nyquist-Shannon sampling theorem is violated $\rightarrow$ backfolding artifacts!

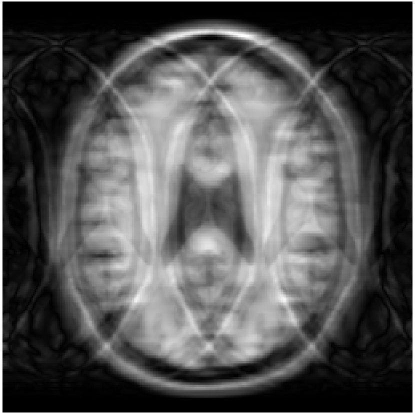
Cartesian undersampling

Accelerated MRI in a nutshell



$$\xrightarrow{\mathcal{F}\mathcal{T}^{-1}}$$

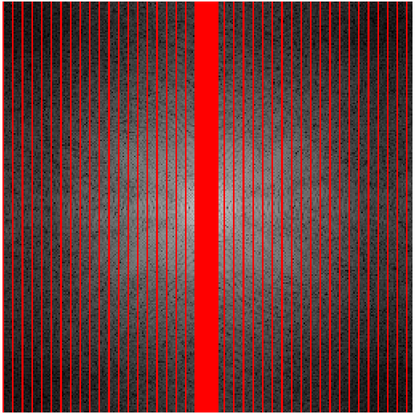
$$R = 4$$



Nyquist-Shannon sampling theorem is violated $\rightarrow$ backfolding artifacts!

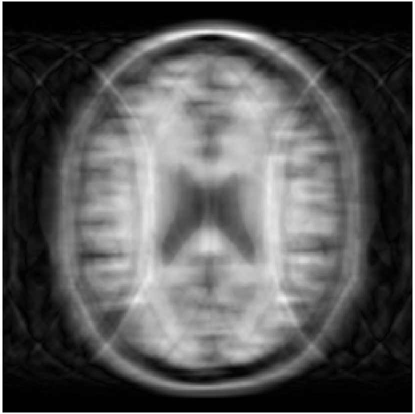
Cartesian undersampling

Accelerated MRI in a nutshell



$\xrightarrow{\mathcal{FT}^{-1}}$

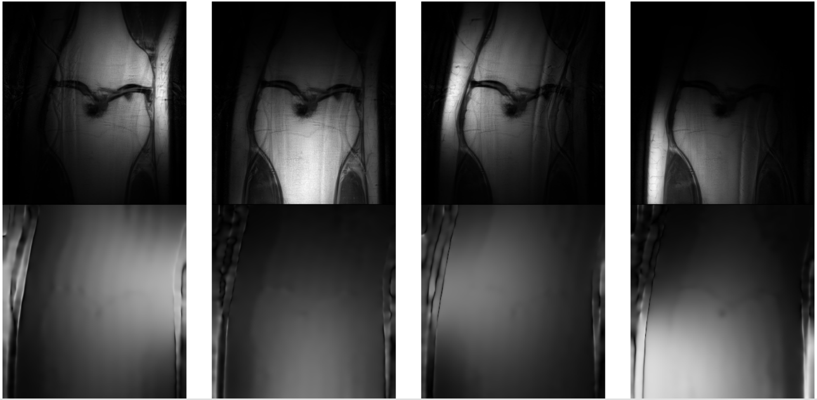
$R = 6$



Nyquist-Shannon sampling theorem is violated $\rightarrow$ backfolding artifacts!

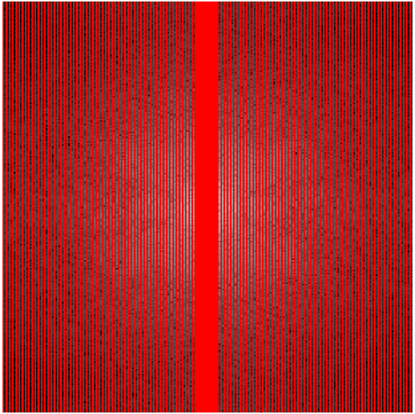
Parallel Imaging (PI) in a nutshell

- Combine data from **multiple** receiver coils
- Each coil is sensitive only in a certain spatial region
- Used for accelerated MRI



Linear reconstruction (SENSE)

[Pruessmann 1999]



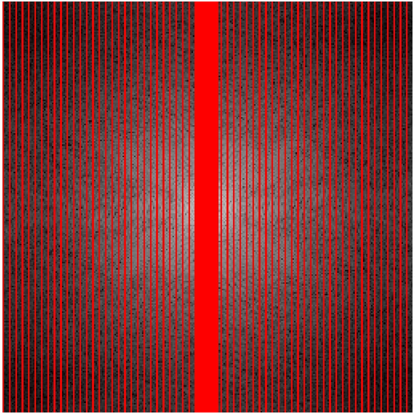
SENSE
$\longrightarrow$

$R = 2$

Clinical standard

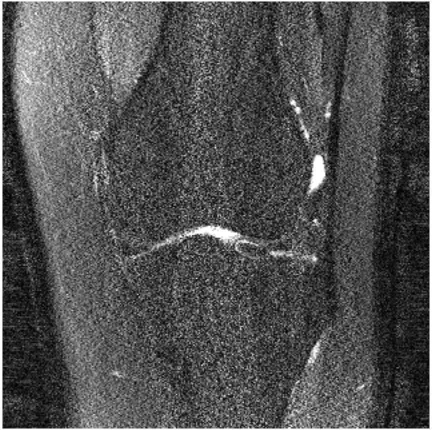
Linear reconstruction (SENSE)

[Pruessmann 1999]



SENSE
$\longrightarrow$

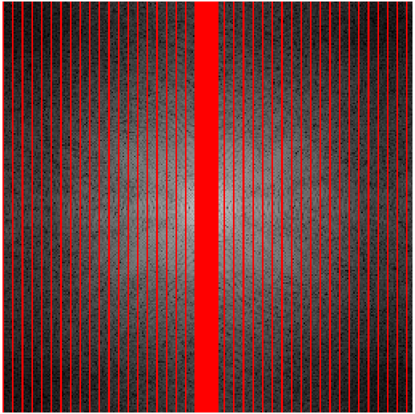
$R = 4$



Noise amplification

Linear reconstruction (SENSE)

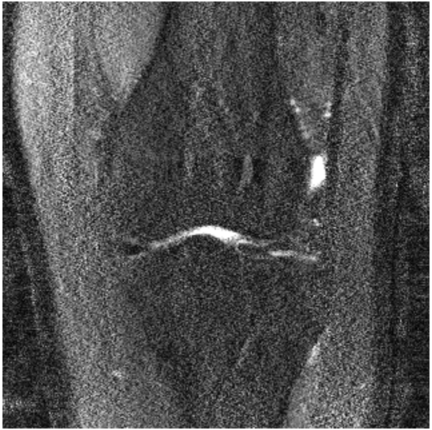
[Pruessmann 1999]



SENSE

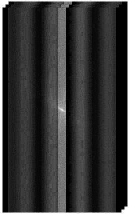
$\longrightarrow$

$R = 6$



Noise amplification $\rightarrow$ Add prior knowledge!

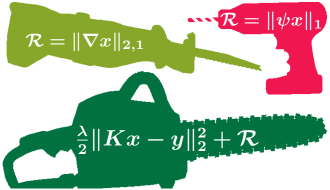
Compressed Sensing (CS) accelerated MRI



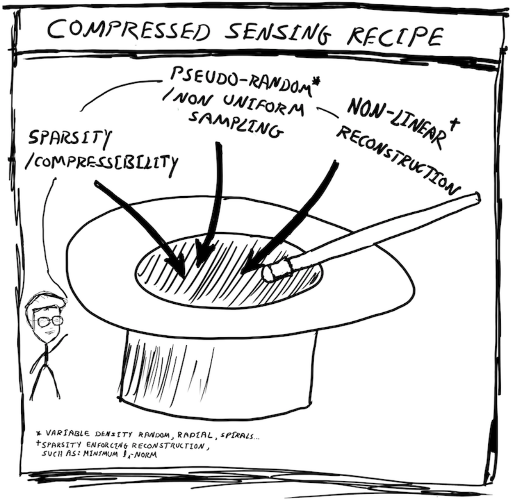
Undersampled rawdata

→ Feature engineering → Iterative reconstruction algorithm →

Reconstruction



Compressed Sensing (CS) accelerated MRI

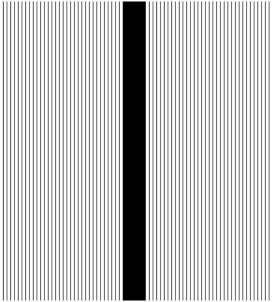


© M. Lustig

Application of CS to clinical routine exams?

CS in clinical practice?

- Cartesian sampling
- Long reconstruction times
- Handcrafted sparsifying transforms
- Unnatural appearance of images
- Parameter tuning



Challenges in CS

Variational model

$$\min_u E = \min_u \underbrace{\frac{\lambda}{2} \|Au - f\|_2^2}_{\text{data consistency}} + \underbrace{\sum_{i=1}^{N} \phi_i(K_i u)}_{\text{regularization}}$$

- u ... reconstructed image
- f ... undersampled k-space data
- A ... linear multi-coil sampling operator

Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

- How to choose proper regularization term?

Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

- How to choose proper regularization term?
  - ϕ_i ... potential function, e.g.: $\|\cdot\|_1$
  - K_i ... sparsifying transform, e.g.: ∇, ψ

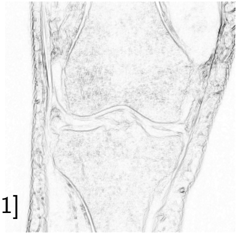
Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

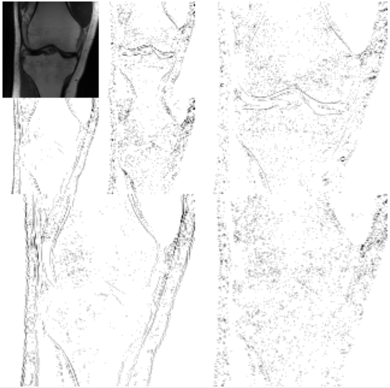
- How to choose proper regularization term?
  - ϕ_i ... potential function, e.g.: $\|\cdot\|_1$
  - K_i ... sparsifying transform, e.g.: ∇, ψ

Total Variation (TV)
$\phi(Ku) = \|\nabla u\|_{2,1}$
sparse edges
[Block 2007]

TGV
[Bredies 2010, Knoll 2011]



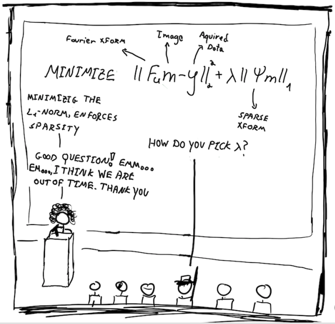
Wavelet transform
$\phi(Ku) = \|\psi u\|_1$
sparse coefficients
[Lustig 2007]



Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

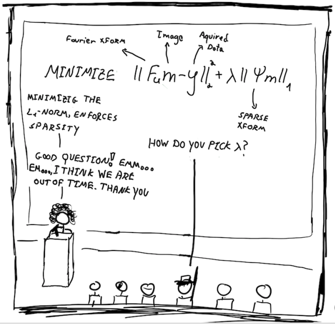
- How to choose proper regularization term?
- How to choose proper regularization parameter λ?



Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

- How to choose proper regularization term?
- How to choose proper regularization parameter λ?

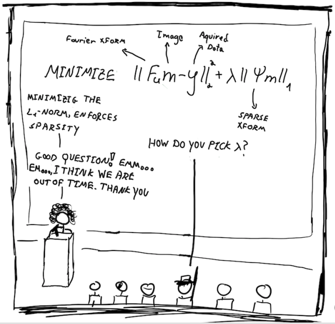


© M. Lustig

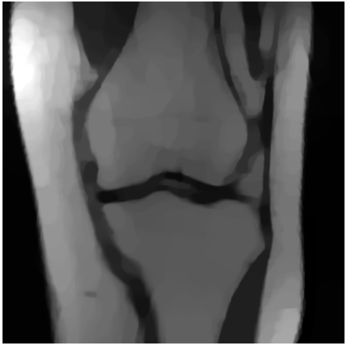
Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

- How to choose proper regularization term?
- How to choose proper regularization parameter λ?



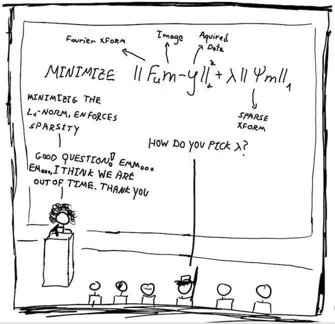
© M. Lustig

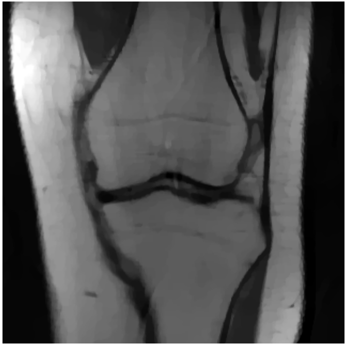


Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

- How to choose proper regularization term?
- How to choose proper regularization parameter λ?





Challenges in CS

$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

- How to choose proper regularization term?
- How to choose proper regularization parameter λ?
- Optimization algorithm?
  - non-linear CG, split Bregman, (F)ISTA, primal-dual,...

Challenges in CS

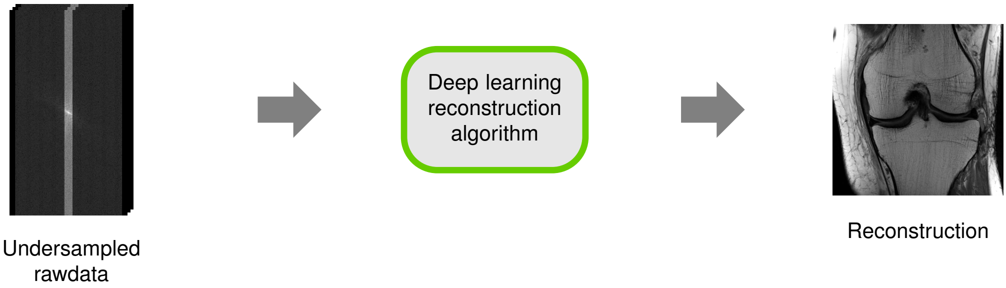
$$\min_u E = \min_u \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N} \phi_i(K_i u)$$

- How to choose proper regularization term?
- How to choose proper regularization parameter λ?
- Optimization algorithm?

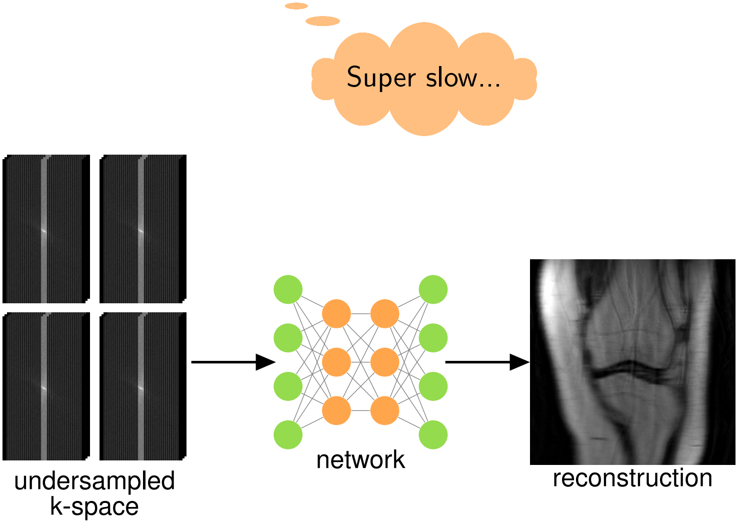
Our approach: CS meets Deep Learning

Learning optimal regularization and reconstruction algorithm
for the reconstruction of undersampled data.

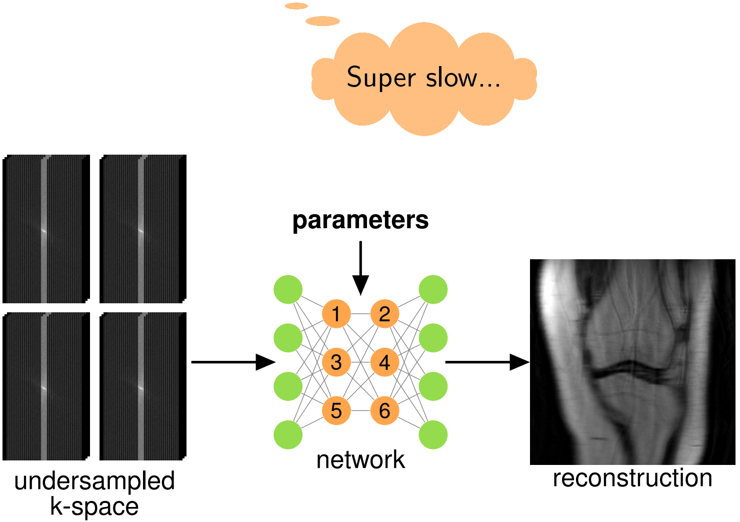
Our approach: CS meets Deep Learning



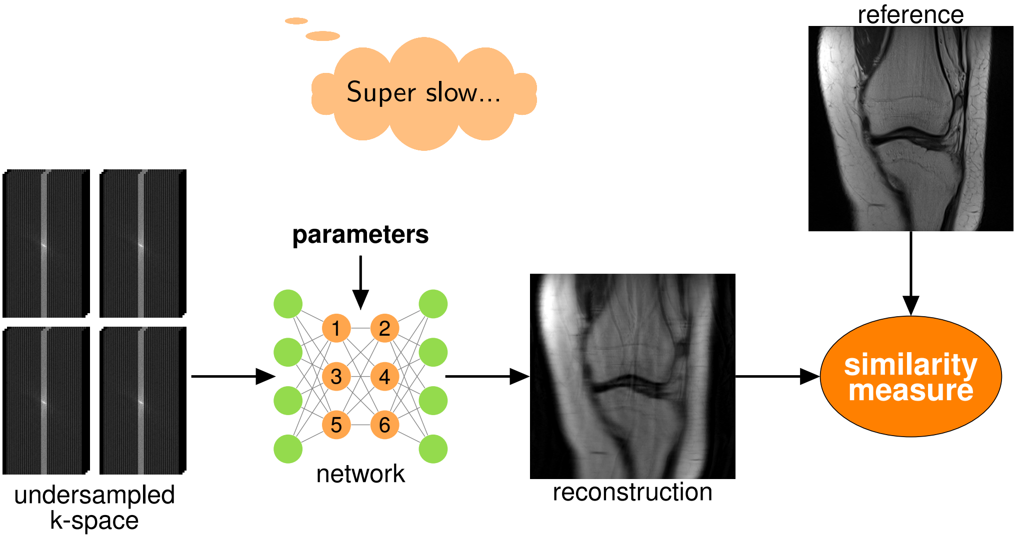
Supervised Learning in a Nutshell



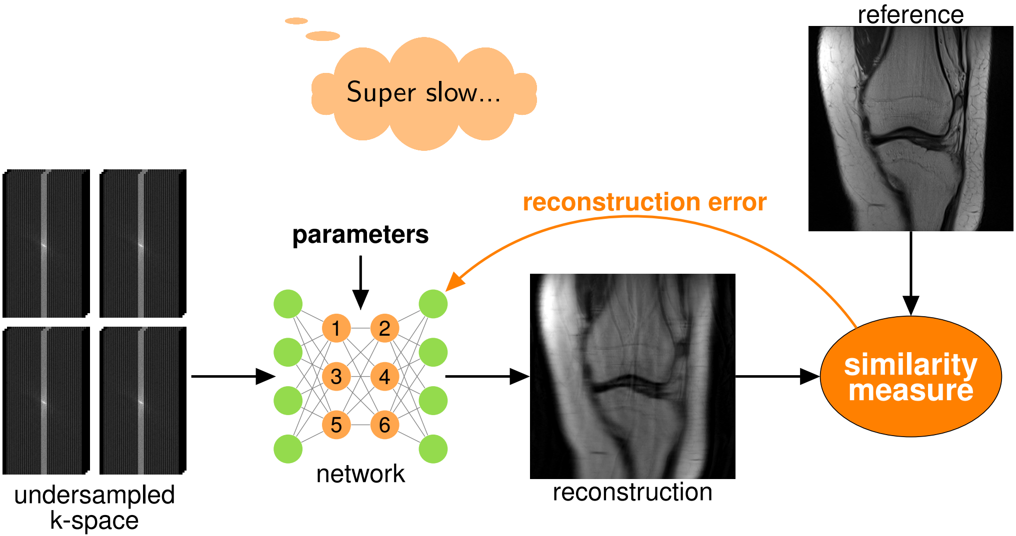
Supervised Learning in a Nutshell



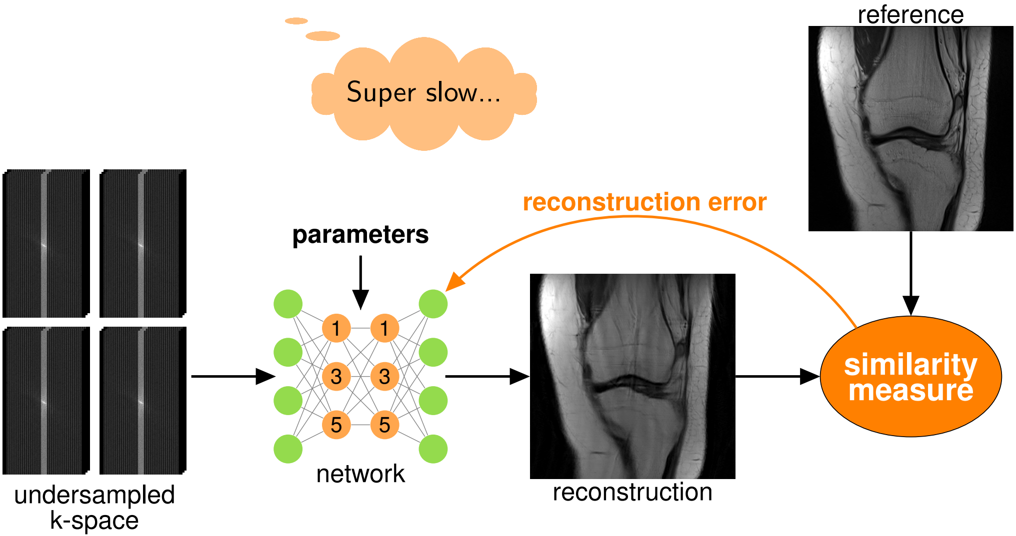
Supervised Learning in a Nutshell



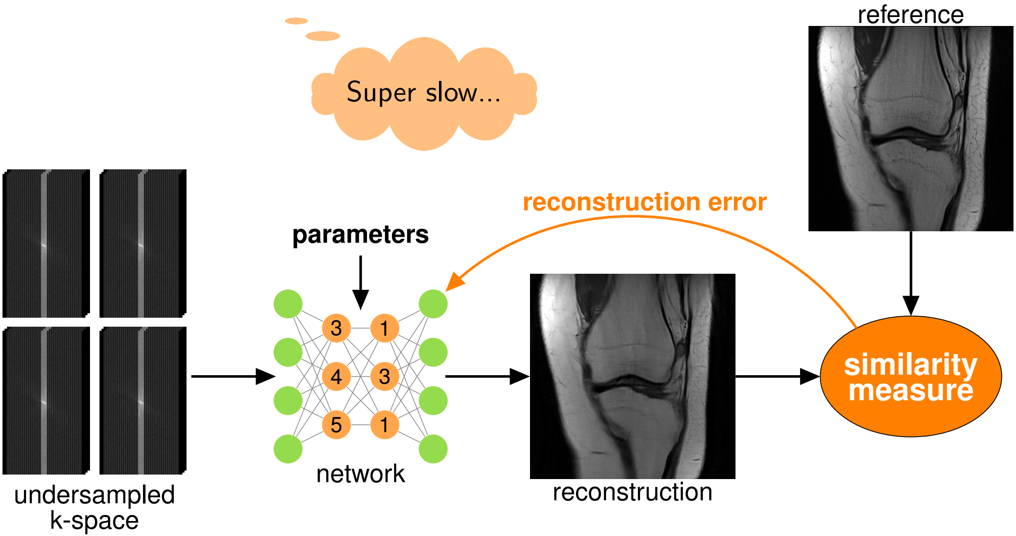
Supervised Learning in a Nutshell



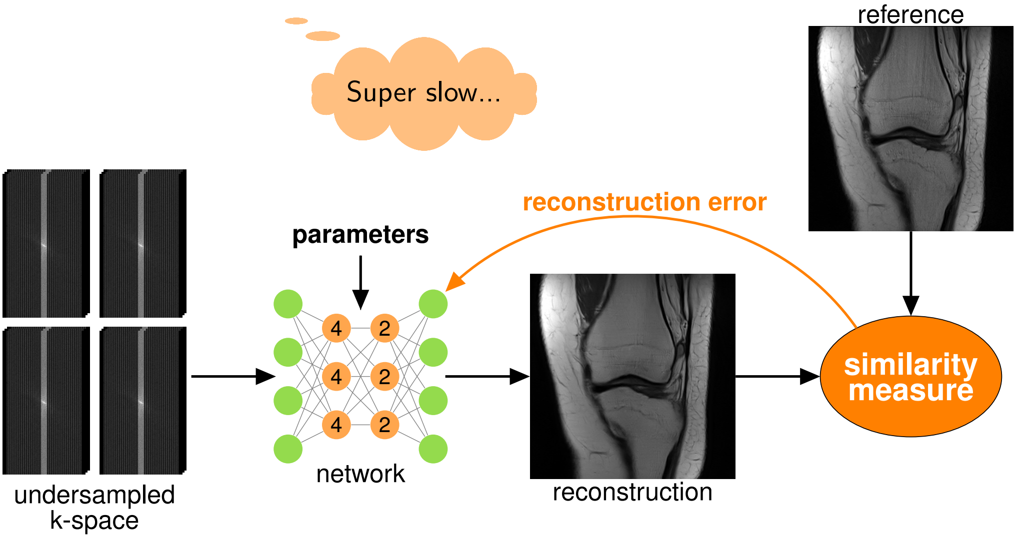
Supervised Learning in a Nutshell



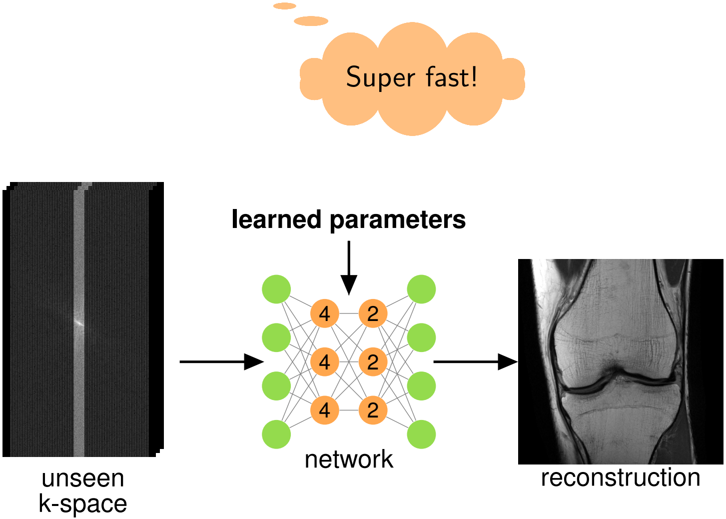
Supervised Learning in a Nutshell

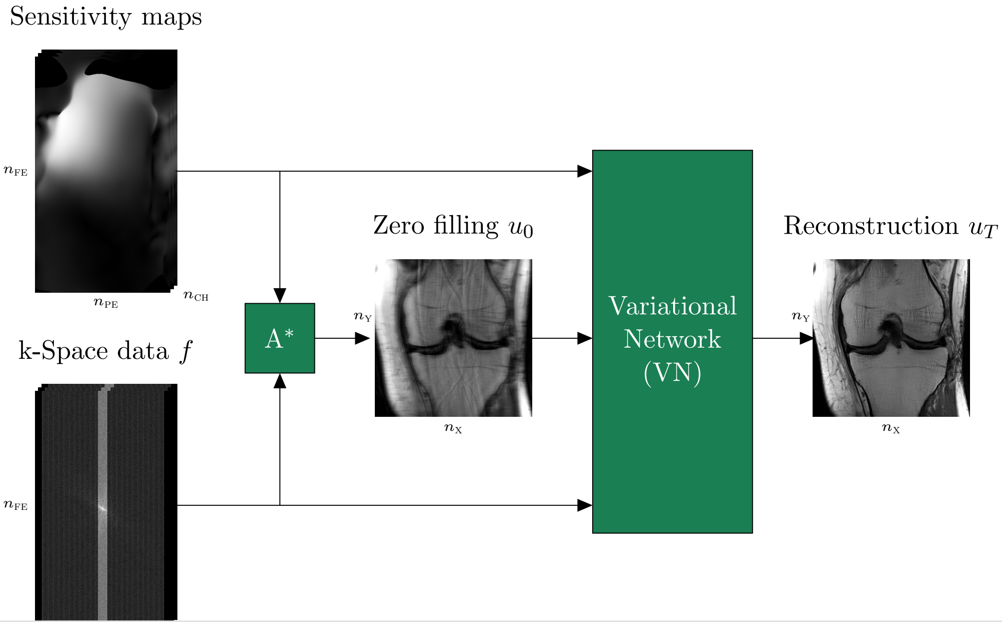


Supervised Learning in a Nutshell



Inference / Testing on new unseen data





Sensitivity maps
n_{FE}
n_{PE}
n_{CH}
k-Space data f
n_{FE}
A*
Zero filling u_0
n_{Y}
n_{X}
Variational Network (VN)
Reconstruction u_T
n_{Y}
n_{X}

Variational Network

Unrolled Gradient Descent Scheme

Learn T gradient descent (GD) steps

$$u^{t+1} = u^t - \frac{\partial}{\partial u} E(u^t)$$

$$E(u) = \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N_k} \phi_i(K_i u)$$

Variational Network
Unrolled Gradient Descent Scheme

Learn T gradient descent (GD) steps

$$u^{t+1} = u^t - \frac{\partial}{\partial u} E(u^t)$$

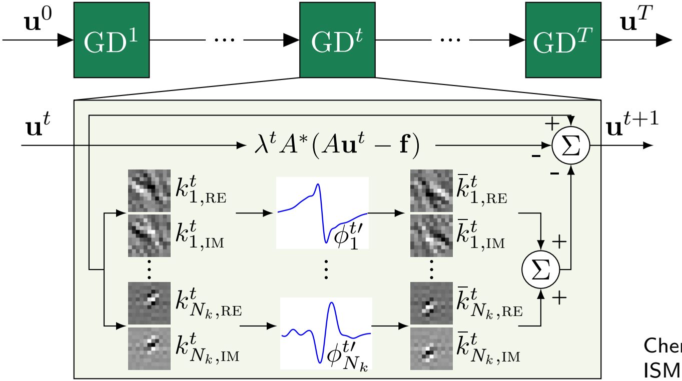
$$E(u) = \frac{\lambda}{2} \|Au - f\|_2^2 + \sum_{i=1}^{N_k} \phi_i(K_i u)$$

Relation to Compressed Sensing

$$N_k = 1, \quad \phi_1 = \|\cdot\|_1, \quad K_1 = \nabla$$

Variational Network

Unrolled Gradient Descent Scheme



$$u^{t+1} = u^t - \sum_{i=1}^{N_k} (K_i^t)^\top \phi_i^{t\prime}(K_i^t u^t)$$

Chen et al. CVPR 2015
ISMRM 2016 (1088)
ISMRM 2017 (644, 645, 687)
Kobler et al. GCPR 2017
Hammernik et al. MRM 2017

Network Parameters

Network Parameters

- **Filter kernels** k_i^t:

$$\left\|k_i^t\right\|_2 \leq 1, \mu\left(k_i^t\right) = 0$$

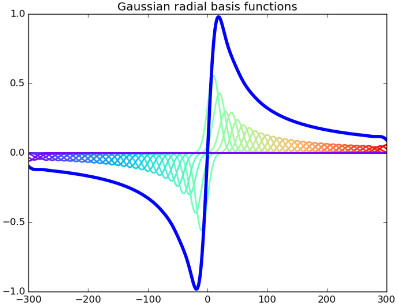
Network Parameters

- **Filter kernels** k_i^t:

$$\left\| k_i^t \right\|_2 \leq 1, \mu\left(k_i^t\right) = 0$$

- **Activation functions** $\phi_i^{t\prime}(z)$:
  Gaussian radial basis functions

$$\phi_i^{t\prime}(z) = \sum_{j=1}^{M} w_{ij}^t \exp\left(-\frac{(z-\mu_j)^2}{2\gamma^2}\right)$$



Network Parameters

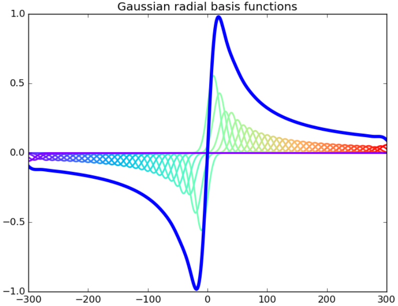
- **Filter kernels** k_i^t:

$$\left\|k_i^t\right\|_2 \leq 1, \mu\left(k_i^t\right) = 0$$

- **Activation functions** $\phi_i^{t\prime}(z)$:
  Gaussian radial basis functions

$$\phi_i^{t\prime}(z) = \sum_{j=1}^{M} w_{ij}^t \exp\left(-\frac{(z-\mu_j)^2}{2\gamma^2}\right)$$

- **Regularization parameter** λ



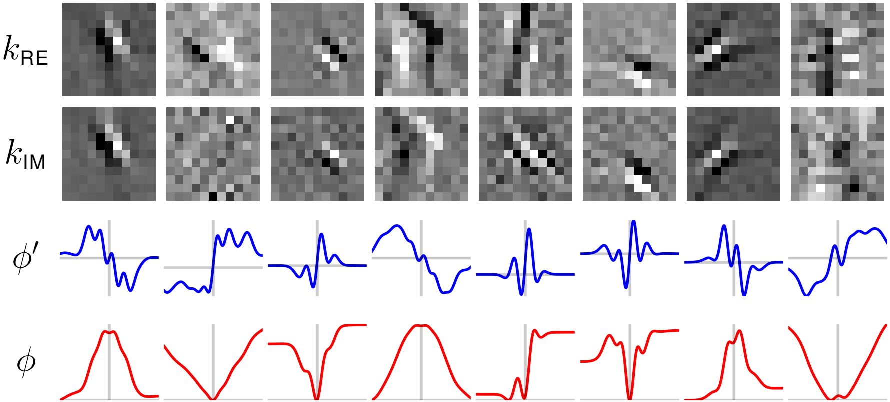
Experimental setup
Acquisition

- 20 patients
- 3T clinical scanner (Siemens Magnetom Skyra), 15-channel knee coil
- Regular clinical 2D TSE sequence with turbo factor 4
- 24 calibration lines
- Full clinical protocol with 5 sequences:
  - Coronal PD
  - Coronal fat-saturated PD
  - Axial fat-saturated T_2
  - Sagittal PD
  - Sagittal T_2

Experimental setup
Variational network

- 10 patients for training (200 slices)
- 10 patients for testing
- Similarity measure: Mean-squared error (MSE)
- Optimizer: Inertial Incremental Proximal Gradient Algorithm (IIPG)
- Stages / gradient steps: 10
- Filter kernels: 48
- Filter kernels size: 11×11
- Total network parameters: 131,050

Learned Network Parameters



Results

Coronal PD R=4

Zero filling

RMSE=0.16
SSIM=0.78

Results

Coronal PD R=4

CG SENSE

RMSE=0.10
SSIM=0.85

Results
Coronal PD R=4

PI-CS: Total Generalized Variation

RMSE=0.06
SSIM=0.88

Knoll et al. MRM 2011

Results
Coronal PD R=4

Dictionary Learning

RMSE=0.07
SSIM=0.88

Ravishankar et al. TMI 2011

Results

Coronal PD R=4

Variational Network

RMSE=0.06
SSIM=0.92

Results

Coronal PD R=4

Reference

Results

Coronal fat-sat. PD R=4

Zero filling

RMSE=0.17
SSIM=0.85

Results

Coronal fat-sat. PD R=4

CG SENSE

RMSE=0.16
SSIM=0.86

Results

Coronal fat-sat. PD R=4

PI-CS: Total Generalized Variation

RMSE=0.12
SSIM=0.89

Results

Coronal fat-sat. PD R=4

Dictionary Learning

RMSE=0.12
SSIM=0.90

Results

Coronal fat-sat. PD R=4

Variational Network

RMSE=0.11
SSIM=0.91

Results

Coronal fat-sat. PD R=4

Reference

Results for prospectively undersampled data

Axial fat-sat. T_2 R=4

PI-CS: Total Generalized Variation

Results for prospectively undersampled data

Axial fat-sat. T_2 R=4

Dictionary Learning

Results for prospectively undersampled data

Axial fat-sat. T_2 R=4

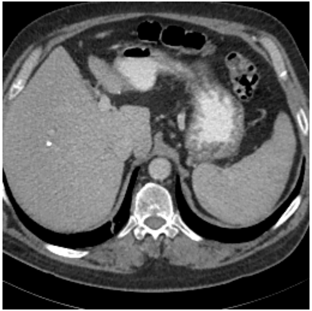
Variational Network

Going beyond accelerated MRI reconstruction

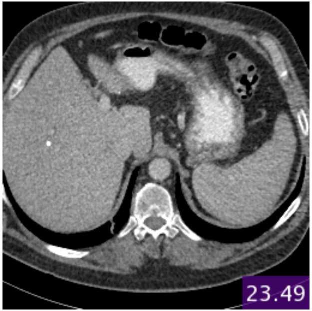
Low-Dose 3D Computed Tomography

X-ray tube current reduction

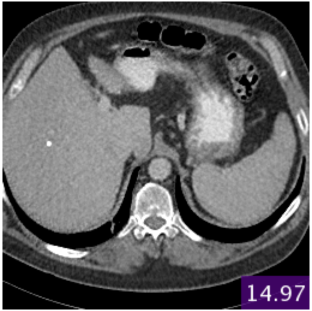
Reference



SAFIRE (Siemens)



Variational Network

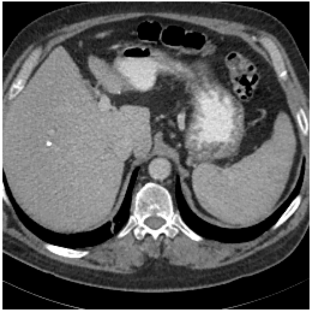


$4\times$ dose reduction

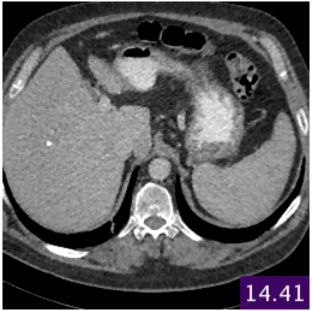
Low-Dose 3D Computed Tomography

X-ray beam interruption

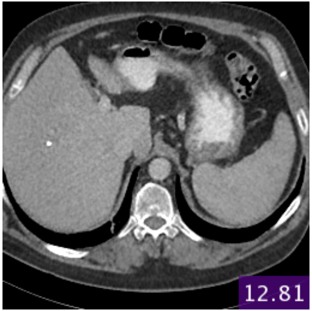
Reference



Total Variation



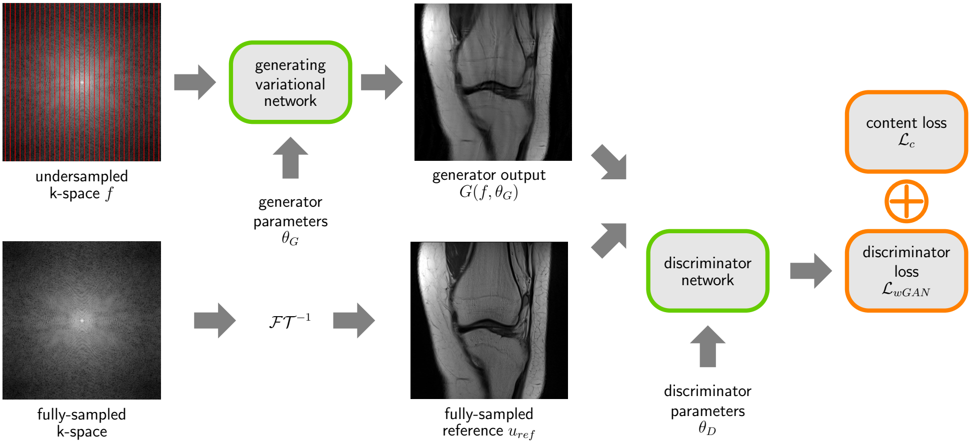
Variational Network



4× dose reduction

Outlook: How can we improve the image quality of low SNR images?

Outlook: Adversarial Training



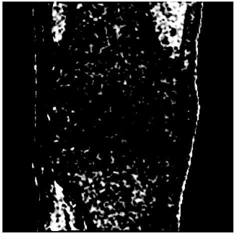
Outlook: Discriminator Training

Training iteration: 1000

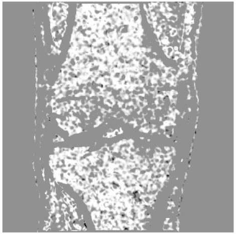
Generator output of variational network $G(f, \theta_G)$

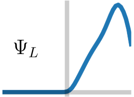
Discriminator output for generated image $D\left(G(f, \theta_G), \theta_D\right)$

Discriminator output for reference $D(u_{ref}, \theta_D)$



Difference of discriminator outputs $D\left(G(f, \theta_G), \theta_D\right) - D(u_{ref}, \theta_D)$

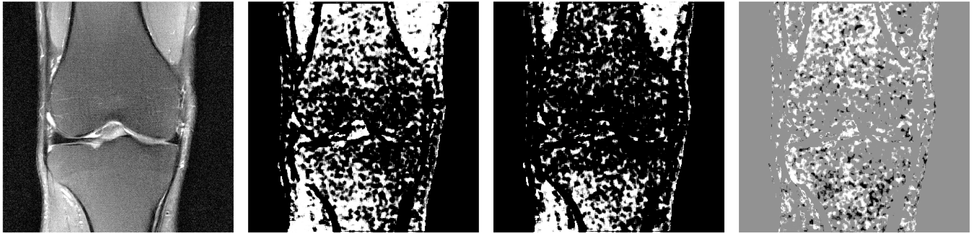


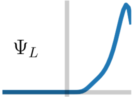


Outlook: Discriminator Training

Training iteration: 5000

Generator output of variational network $G(f,\theta_G)$	Discriminator output for generated image $D\left(G(f,\theta_G),\theta_D\right)$	Discriminator output for reference $D(u_{ref},\theta_D)$	Difference of discriminator outputs $D\left(G(f,\theta_G),\theta_D\right) - D(u_{ref},\theta_D)$

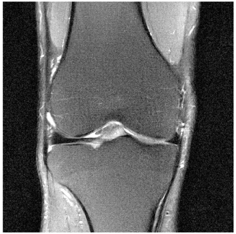




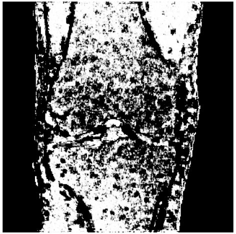
Outlook: Discriminator Training

Training iteration: 10000

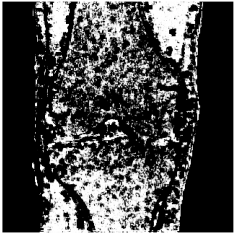
Generator output of variational network $G(f, \theta_G)$



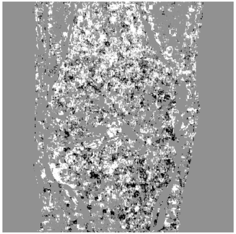
Discriminator output for generated image $D\left(G(f, \theta_G), \theta_D\right)$

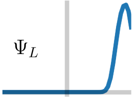


Discriminator output for reference $D(u_{ref}, \theta_D)$



Difference of discriminator outputs $D\left(G(f, \theta_G), \theta_D\right) - D(u_{ref}, \theta_D)$





Outlook: Discriminator Training

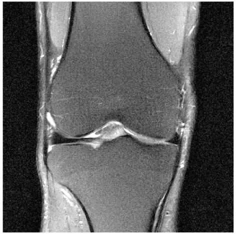
Training iteration: 15000

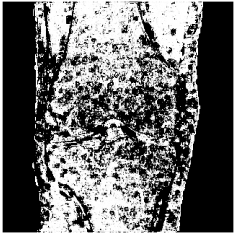
Generator output of variational network $G(f, \theta_G)$

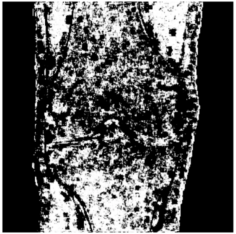
Discriminator output for generated image $D\left(G(f, \theta_G), \theta_D\right)$

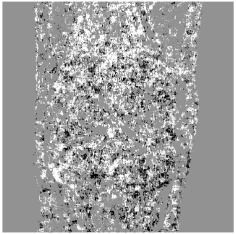
Discriminator output for reference $D(u_{ref}, \theta_D)$

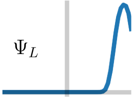
Difference of discriminator outputs $D\left(G(f, \theta_G), \theta_D\right) - D(u_{ref}, \theta_D)$







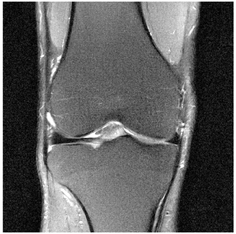




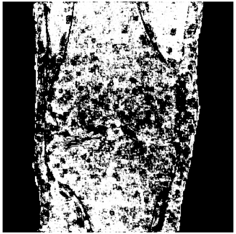
Outlook: Discriminator Training

Training iteration: 20000

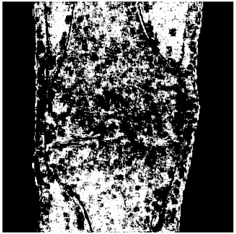
Generator output of variational network $G(f, \theta_G)$



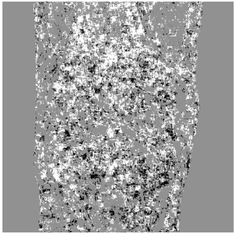
Discriminator output for generated image $D\left(G(f, \theta_G), \theta_D\right)$

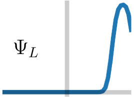


Discriminator output for reference $D(u_{ref}, \theta_D)$



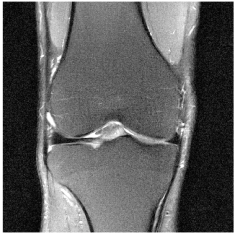
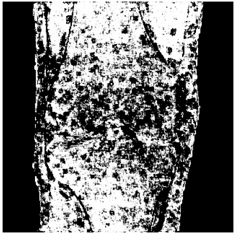
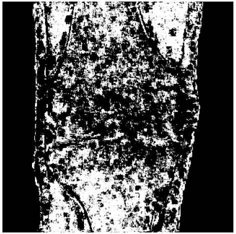
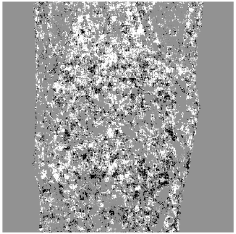
Difference of discriminator outputs $D\left(G(f, \theta_G), \theta_D\right) - D(u_{ref}, \theta_D)$

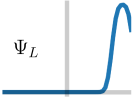




Outlook: Discriminator Training

Training iteration: 25000

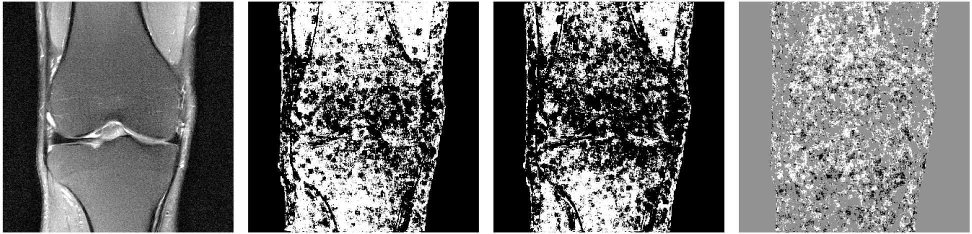
Generator output of variational network $G(f, \theta_G)$	Discriminator output for generated image $D(G(f,\theta_G),\theta_D)$	Discriminator output for reference $D(u_{ref},\theta_D)$	Difference of discriminator outputs $D(G(f,\theta_G),\theta_D) - D(u_{ref},\theta_D)$
			

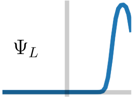


Outlook: Discriminator Training

Training iteration: 30000

Generator output of variational network $G(f, \theta_G)$	Discriminator output for generated image $D\left(G(f, \theta_G), \theta_D\right)$	Discriminator output for reference $D(u_{ref}, \theta_D)$	Difference of discriminator outputs $D\left(G(f, \theta_G), \theta_D\right) - D(u_{ref}, \theta_D)$





Preliminary Results

Coronal fat-sat. PD R=4

Reference

Variational Network

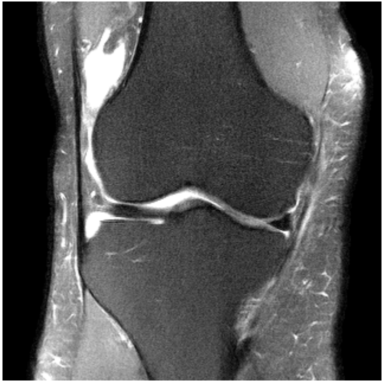
Preliminary Results

Coronal fat-sat. PD R=4

Reference

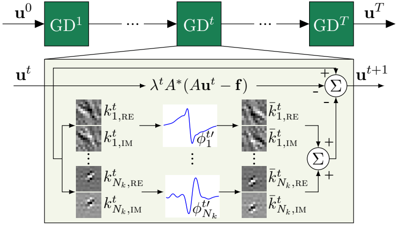
Variational Network

Variational Adversarial Network



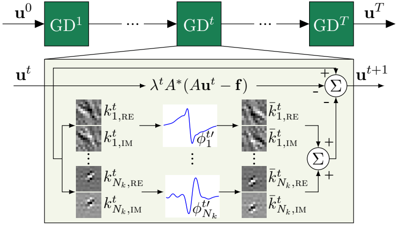
Conclusion

- Variational networks: Connecting variational models and deep learning



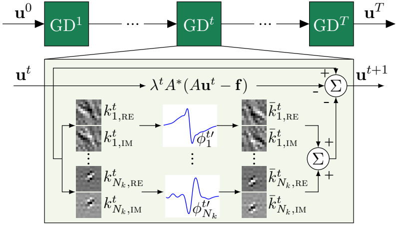
Conclusion

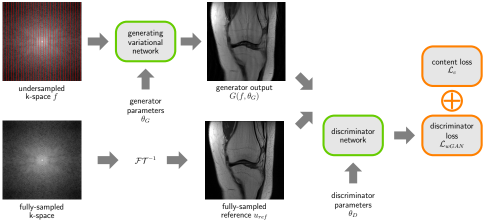
- Variational networks: Connecting variational models and deep learning
- Training environment with real multi-coil patient data



Conclusion

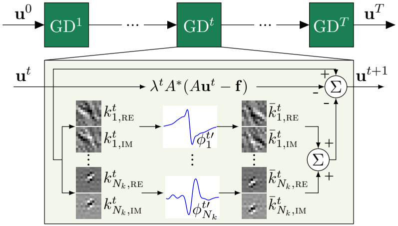
- Variational networks: Connecting variational models and deep learning
- Training environment with real multi-coil patient data
- Outlook: Variational Adversarial Networks - Inspired by adversarial training

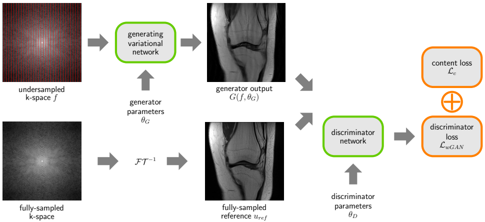




Conclusion

- Variational networks: Connecting variational models and deep learning
- Training environment with real multi-coil patient data
- Outlook: Variational Adversarial Networks - Inspired by adversarial training
- Coming soon: Tensorflow source code





Acknowledgments

Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria


Thomas Pock


Erich Kobler


Teresa Klatzer

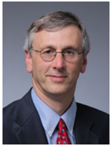

Martin Urschler

Grants:
NIH R01 EB000447
NIH P41 EB017183
FWF Y729

Department of Radiology, NYU School of Medicine, New York, USA


Florian Knoll


Daniel Sodickson


Michael Recht


Thomas Benkert


Mary Bruno


Tobias Block