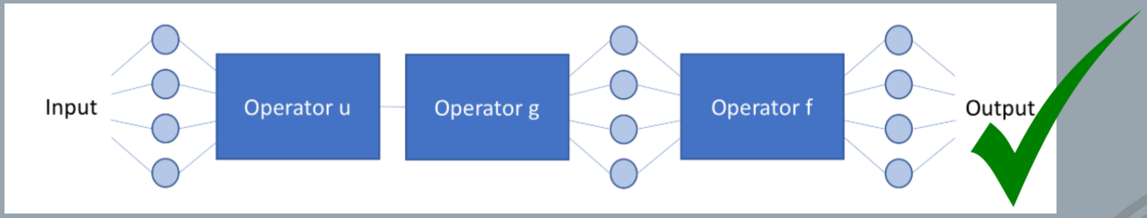
Embedding of Operators Part II

by Maximilian Rohleder

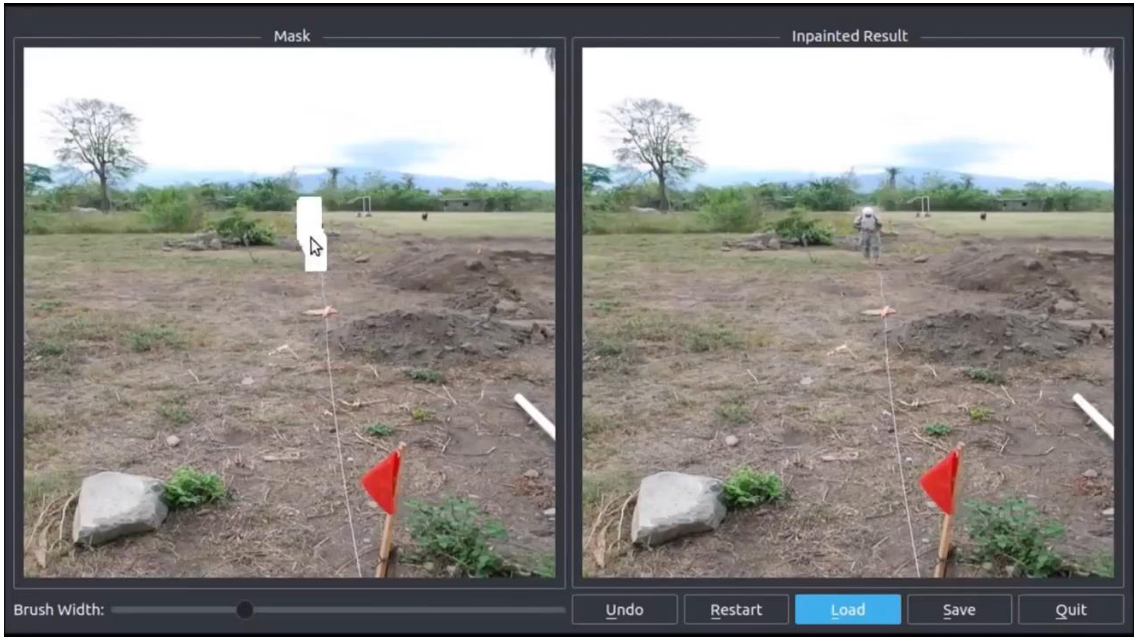


Motivation

Deep Learning has come far…

But do we know what Nets learn?

Some expert applications are well known like CT Reconstruction (1917)



Can we use this expertise to efficiently train networks?

Embedding of Operators Part II

Structure

1. What are „Variational Networks"?
2. From variational methods to networks (Example)
    1. Concept of Residual Networks
3. Examples

Roth, V., Vetter, T., Kobler, E., Klatzer, T., Hammernik, K., and Pock, T., eds., **Variational Networks: Connecting Variational Methods and Deep Learning**: Pattern Recognition: Springer International Publishing

Variational Methods

"VNs are fully learned models based on the framework of incremental proximal gradient methods." [1]

data constraints

$$\min_{x \in \mathcal{X}} F(x) := \boxed{f(x)} + \boxed{h(x)} = \sum_{c=1}^{C} f_c(x) + h(x),$$

$$x_{t+1} = \mathrm{prox}_h^{\eta_t}\left(x_t - \eta_t \nabla f_{\boxed{c(t)}}(x_t)\right)$$

$$\mathrm{prox}_h^{\eta}(z) := \arg\min_{x}\left(h(x) + \frac{1}{2\eta}\|x - z\|_2^2\right)$$

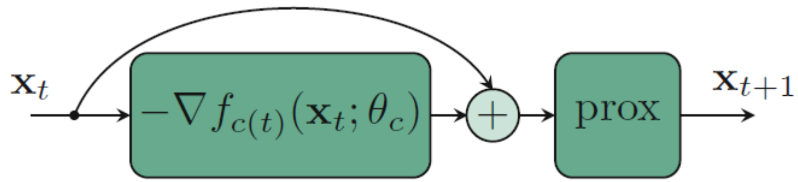
C:	Nr. Components
$\boldsymbol{x} \in \mathrm{R}^n$	data e.g. images
η:	learning rate

c(t) = mod(C, t)

Variational Networks

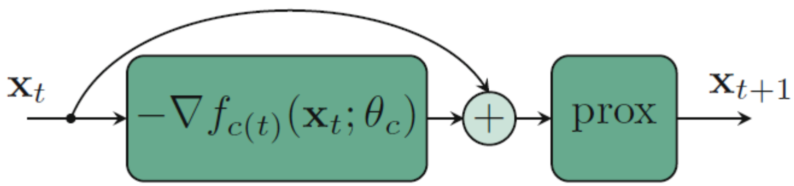
$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) := \sum_{c=1}^{C} f_c(\boldsymbol{x}; \boldsymbol{\theta}_c) + h(\boldsymbol{x})$$

$$\boldsymbol{x}_{t+1} = \text{prox}_h^{\eta_t} \left(\boldsymbol{x}_t - \eta_t \nabla f_{c(t)}(\boldsymbol{x}_t; \boldsymbol{\theta}_{c(t)}) \right)$$

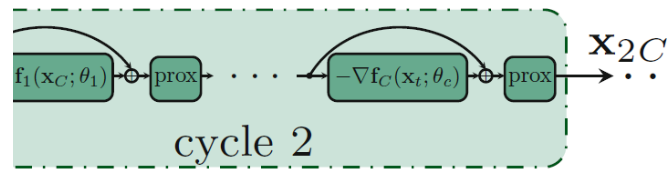


(a) Variational Unit (VU)

Variational Networks



(a) Variational Unit (VU)

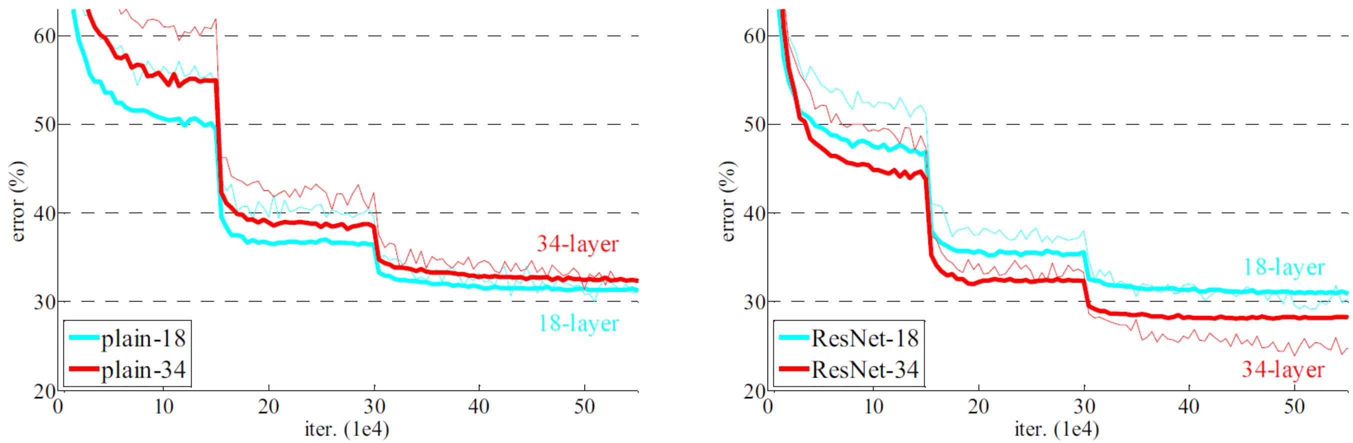


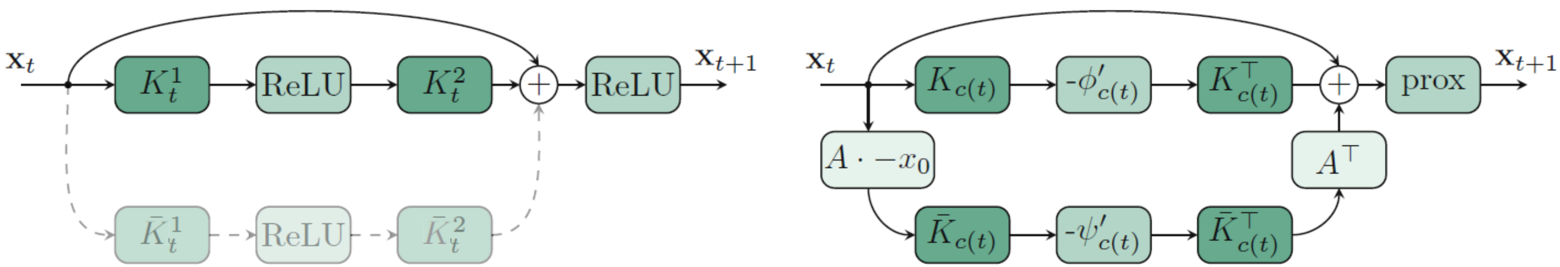
work (VN)

Variational (iterative) methods: online optimization

Deep Learning: Learn underlying information offline

Variational Networks vs. Residual Networks





(a) (multi-) Residual Unit

(b) Variational Unit

Example: *"Learning Networks for deblurring and denoising "*

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) := \sum_{c=1}^{C} f_c(\boldsymbol{x}; \boldsymbol{\theta}_c) + h(\boldsymbol{x})$$

problem specific

$$\min_{\boldsymbol{x} \in \mathcal{X}^n} F(\boldsymbol{x}) := \sum_{c=1}^{C} f_c(\boldsymbol{x}; \boldsymbol{\theta}_c) = R_c(\boldsymbol{x}; \boldsymbol{\theta}_c) + D_c(\boldsymbol{x}; \boldsymbol{\theta}_c),$$

$$R_c(\boldsymbol{x}; \boldsymbol{\theta}_c) = \sum_{i=1}^{N_r} \sum_{j=1}^{n} \phi_i^c \left((K_i^c \boldsymbol{x})_j \right) \quad D_c(\boldsymbol{x}; \boldsymbol{\theta}_c) = \sum_{i=1}^{N_d} \sum_{j=1}^{n} \psi_i^c \left(\left(\bar{K}_i^c (A\boldsymbol{x} - \boldsymbol{x}_0) \right)_j \right)$$

Fields of Experts — *Higher order statistics*

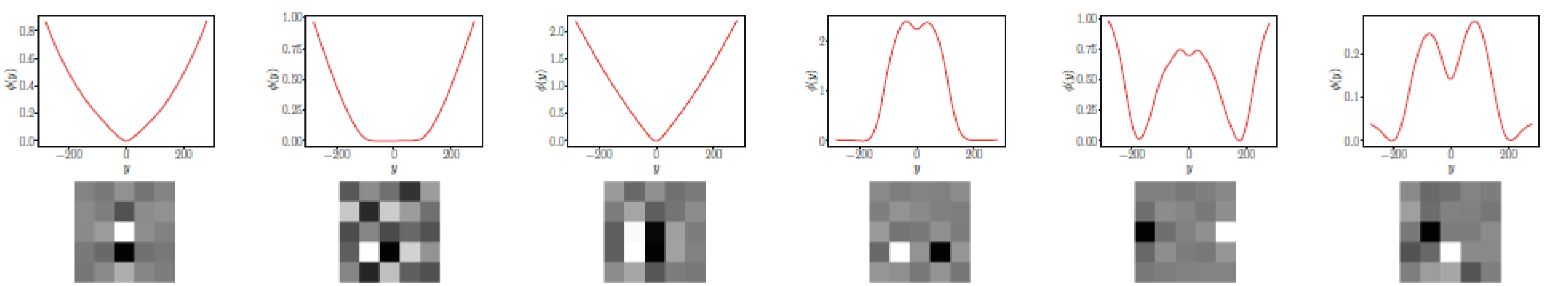
Example: *"Learning Networks for deblurring and denoising "*

$$R_c(x;\theta_c) = \sum_{i=1}^{N_r}\sum_{j=1}^{n} \phi_i^c\left((K_i^c x)_j\right) \quad D_c(x;\theta_c) = \sum_{i=1}^{N_d}\sum_{j=1}^{n} \psi_i^c\left(\left(\bar{K}_i^c(Ax - x_0)\right)_j\right)$$

$$\psi_i'^c(y) = \phi_i'^c(y) = \sum_{j=1}^{N_w} \exp\left(-\frac{(y-\mu_j)^2}{2\sigma^2}\right) w_{ij}^c$$

A : known blur kernel
A : identity for denoising

$$\theta_c = (k_1^c, w_1^c, \ldots, k_{N_r}^c, w_{N_r}^c, \bar{k}_1^c, \bar{w}_1^c, \ldots, \bar{k}_{N_d}^c, \bar{w}_{N_d}^c)$$



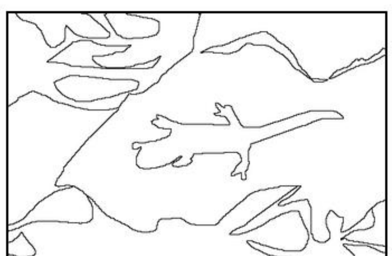
Results: *“Learning a Variational Network for Reconstruction of Accelerated MRI Data“*

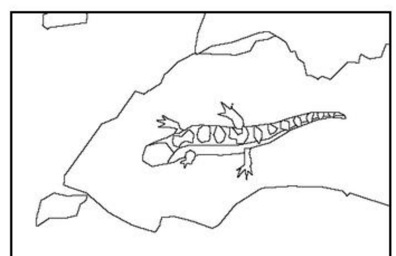
- BDS500 dataset (actually for detection and segmentation)
- deblurring and non blind denoising
- Set up different VNs (number of components, convex / non-convex)

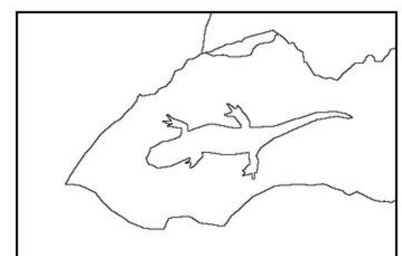
We are able to compare nets to iterative methods now!!!

Type	Corresponding scheme
$\mathrm{VN}_N^{1,t}$	Proximal gradient method (7) (energy minimization)
$\mathrm{VN}_N^{C,t}$	Proximal incremental method (5) (approximate energy minimization)
$\mathrm{VN}_N^{t,t}$	Single cycle proximal incremental method (5) (reaction diffusion)


Original Image


Subject 1


Subject 2


Subject 3

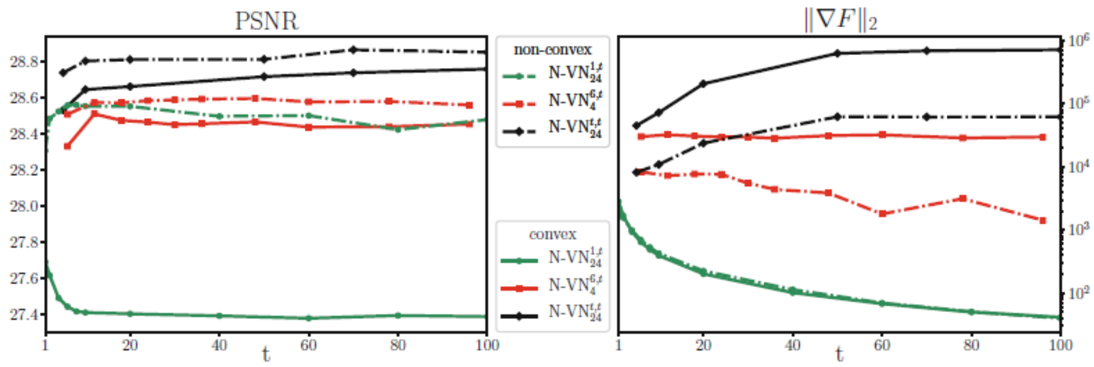


Fig. 5. Average PSNR curves on the test set of the trained VN types for Gaussian image denoising along with the gradient norm of the corresponding energy $F(\boldsymbol{x}_t)$. (Color figure online)

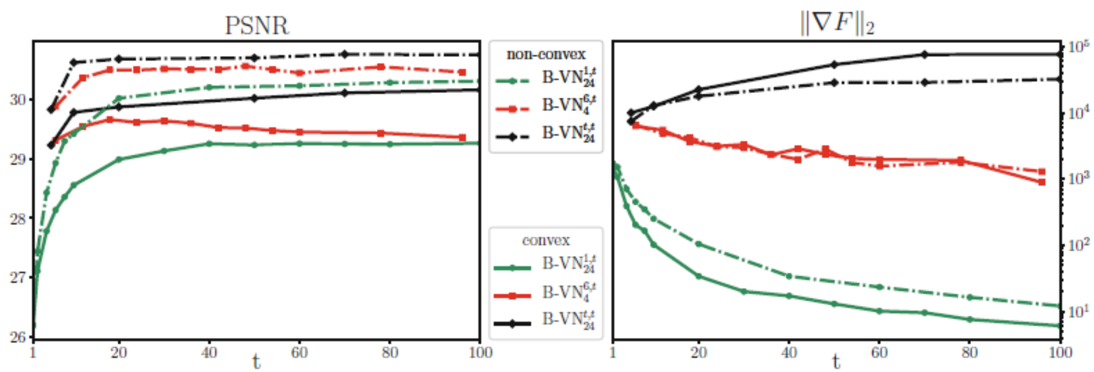


Fig. 6. Average PSNR scores and corresponding gradient norm on the test set of the different VN types for non-blind deblurring. (Color figure online)

go

Outlook

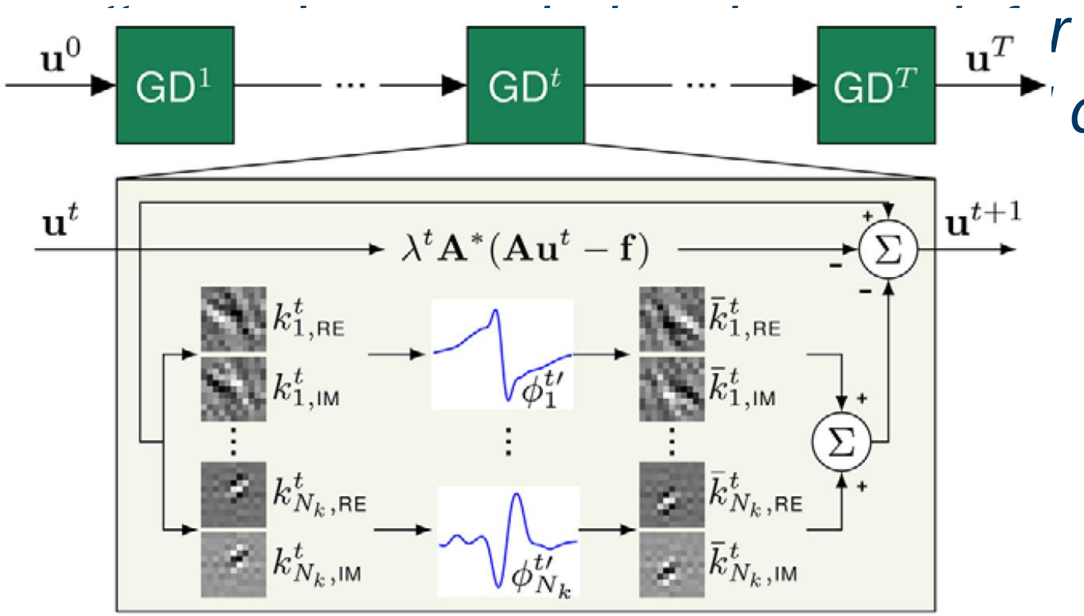
r

data”

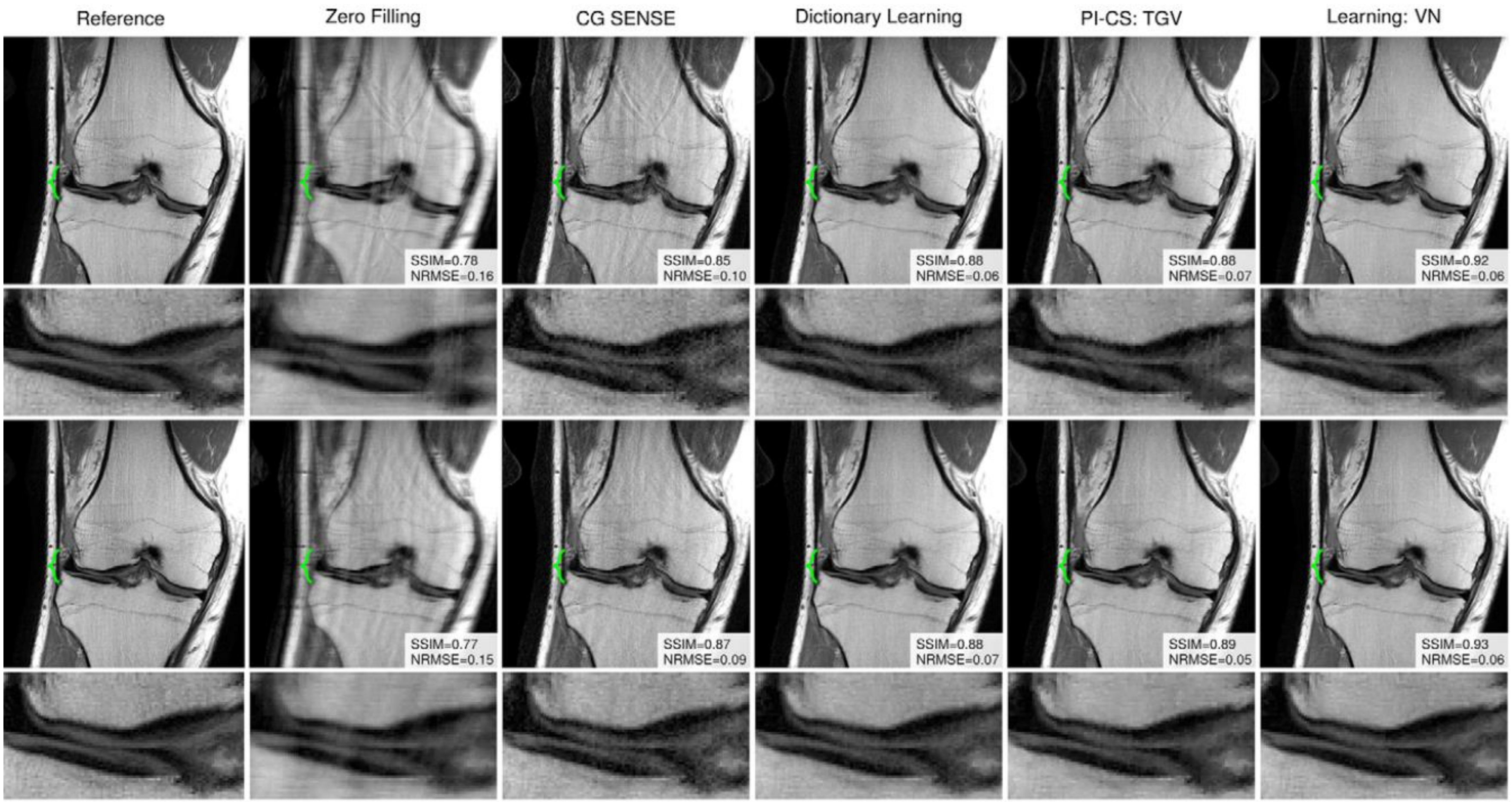
Compressed

- Sparse re
- Undersan

- Also used



$$\mathbf{u}^{t+1} = \mathbf{u}^t - \alpha^t \left(\sum_{i=1}^{N_k} (\mathbf{K}_i)^\top \Phi'_i(\mathbf{K}_i \mathbf{u}^t) + \lambda \mathbf{A}^*(\mathbf{A}\mathbf{u}^t - \mathbf{f}) \right)$$


Reference
Zero Filling
CG SENSE
Dictionary Learning
PI-CS: TGV
Learning: VN
SSIM=0.78
NRMSE=0.16
SSIM=0.85
NRMSE=0.10
SSIM=0.88
NRMSE=0.06
SSIM=0.88
NRMSE=0.07
SSIM=0.92
NRMSE=0.06
SSIM=0.77
NRMSE=0.15
SSIM=0.87
NRMSE=0.09
SSIM=0.88
NRMSE=0.07
SSIM=0.89
NRMSE=0.05
SSIM=0.93
NRMSE=0.06

[B1] Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J., “Contour Detection and Hierarchical Image Segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pp. 898–916, 2011. http://dx.doi.org/10.1109/TPAMI.2010.161.

[B2] Hammernik, K., Klatzer, T., Kobler, E., Recht, M. P., Sodickson, D. K., Pock, T., and Knoll, F., “Learning a variational network for reconstruction of accelerated MRI data,” Magnetic resonance in medicine, vol. 79, no. 6, pp. 3055–3071, 2018.

[B3] He, K., Zhang, X., Ren, S., and Sun, J., eds., Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[B4] Kingma, D. P., and Welling, M., Auto-Encoding Variational Bayes, 2014. http://arxiv.org/pdf/1312.6114.

[B5] Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B., Image Inpainting for Irregular Holes Using Partial Convolutions, 2018. http://arxiv.org/pdf/1804.07723.

[B6] Maier, A., Schebesch, F., Syben, C., Würfl, T., Steidl, S., Choi, J.-H., and Fahrig, R., Precision Learning: Towards Use of Known Operators in Neural Networks, 2017. http://arxiv.org/pdf/1712.00374.

[B7] Roth, V., Vetter, T., Kobler, E., Klatzer, T., Hammernik, K., and Pock, T., eds., Variational Networks: Connecting Variational Methods and Deep Learning: Pattern Recognition: Springer International Publishing, 2017.

[B8] Wurfl, T., Hoffmann, M., Christlein, V., Breininger, K., Huang, Y., Unberath, M., and Maier, A. K., “Deep Learning Computed Tomography: Learning Projection-Domain Weights From Image Domain in Limited Angle Problems,” IEEE transactions on medical imaging, vol. 37, no. 6, pp. 1454–1463, 2018.