Diffusion-Weighted Imaging with Learned Nonlinear Latent Space Modeling and Self-Supervised Reconstruction (DeepDWI)

Zhengguo Tan, Julius Glaser, Patrick A Liebig, Annika Hofmann, Frederik B Laun, Florian Knoll

***Abstract*—The code is publicly available at: `https://github.com/ZhengguoTan/DeepDWI`.**

***Index Terms*—Diffusion-weighted imaging, Image reconstruction, Generative AI, Latent space, Self-supervised learning**

I. Introduction

HIGH-dimensional magnetic resonance imaging (HD-MRI) has been an emerging and flourishing field, which has achieved substantial improvements in terms of spatiotemporal fidelity. Instead of the conventional two-dimensional static single-contrast-weighted imaging, HD-MRI acquires and reconstructs multi-dimensional information. For instance, Brown et al. [?] proposed magnetic resonance spectroscopic imaging (MRSI), which uses multiple readout gradients to acquire multiple echo images for the computation of spatially resolved metabolic distribution. Le BiHan et al. [?] and Merboldt et al. [?] proposed diffusion-weighted imaging (DWI), which utilizes spatially and angularly varying diffusion encoding gradients in combination with fast echo-planar imaging (EPI) readouts [?] to obtain multi-contrast diffusion-weighted images as a probe into tissue microstructure. Ma et al. [?] proposed magnetic resonance fingerprinting (MRF) which consists of a T_1- and T_2-prepared pseudo-randomized sequence to acquire time-resolved transient-state images and a Bloch-equation-based dictionary matching algorithm [?] for simultaneous quantitative T_1 and T_2 mapping.

HD-MRI, however, conventionally requires long scan time. Advances in parallel imaging [?], [?], [?], [?], [?] and compressed sensing [?], [?], [?] have enabled accelerated acquisition for HD-MRI. In particular, the low-rank modeling and regularization [?] has been a powerful tool in reducing the dimensionality of high-dimensional data, which enables accelerated acquisition and high spatiotemporal-resolution reconstruction. Usually, singular value decomposition (SVD) is used to learn a truncated temporal basis function from a large-scale physics-informed dictionary [?], [?], [?]. The temporal basis function is then integrated with the MRI forward model, i.e. the sensitivity encoding operator [?], for joint reconstruction of the corresponding spatial basis images. In addition, low-rank regularization can be employed in the joint reconstruction [?].

Beyond the low-rank technique, advanced neural networks, e.g. autoencoder [?], have been explored for HD-MRI reconstruction and proven to supply more accurate representations of high-dimensional data than SVD. Lam et al. [?] and Mani et al. [?] proposed to first learn a denoising autoencoder (DAE) model from a physics-informed simulated dictionary and then incorporate the learned DAE model as a regularizer in the alternating direction method of multipliers (ADMM) [?] unrolling reconstruction. Pioneered by Gregor and LeCun [?], algorithm unrolling enables the use of learned deep *prior* as regularization and faster inference than iterative reconstruction with hand-crafted regularization functions [?]. Algorithm unrolling has been introduced to accelerated MRI reconstruction and employed in various scenarios: supervised learning with fully sampled reference images [?], [?], self-supervised learning with only undersampled data available for training [?], [?].

Deep neural networks are capable of learning not only regularization functions, but also MR-physics forward operators. Zhu et al. [?] proposed the automated transform by manifold approximation (AUTOMAP), which learns the mapping between the sensor and the image domain for data-driven supervised image reconstruction. Liu et al. [?] proposed the reference-free T_1 parameter maps extraction (RELAX) self-supervised deep learning reconstruction, which learns the mapping from T_1 parameter maps to undersampled multi-coil

This work was supported in part by German Research Foundation (DFG) under projects 513220538 and 512819079, project 500888779 in the Research Unit RU5534 for MR biosignatures at UHF, and by the National Institutes of Health (NIH) under grants R01 EB024532 and P41 EB017183. In addition, scientific support and HPC resources were provided by the Erlangen National High Performance Computing Center (NHR) of Friedrich-Alexander-University Erlangen-Nuremberg (FAU) under the NHR project b143dc. NHR is funded by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by DFG under project 440719683.

Z. Tan was with the Department Artificial Intelligence in Biomedical Engineering (AIBE), FAU, Erlangen, Germany. He is now with the Michigan Institute for Imaging Technology and Translation (MIITT), Department of Radiology, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: zgtan@med.umich.edu).

J. Glaser was with the Department of Medical Engineering, FAU, Erlangen, Germany. He is now with the Institute of Radiology, University Hospital Erlangen, FAU, Erlangen, Germany (e-mail: julius.glaser@fau.de).

P. A. Liebig is with Siemens Healthcare GmbH, Erlangen, Germany (e-mail: patrick.liebig@siemens-healthineers.com).

A. Hofmann is with the Department AIBE, FAU, Erlangen, Germany (e-mail: annika.ah.hofmann@fau.de).

F. B. Laun is with the Institute of Radiology, University Hospital Erlangen, FAU, Erlangen, Germany (e-mail: Frederik.Laun@uk-erlangen.de).

F. Knoll is with the Department AIBE, FAU, Erlangen, Germany (e-mail: florian.knoll@fau.de).

multi-contrast k-space data. Arefeen et al. [?] proposed to replace the conventional SVD-based linear subspace modeling [?] by the latent decoder model within DAE for improved T_2-weighted image reconstruction.

Several challenges exist when adopting deep learning to DWI reconstruction. First, the capability of DAE to learn diffusion MRI models is open to questions. DAE is composed of sequential fully connected layers with nonlinear activation functions. This simple architecture may fail to learn complicated functions. DWI signal is such an example. The standard diffusion tensor model [?] consists of six tensor elements, and forms DWI signals based on the multiplication of exponential functions. Second, it is rather difficult to acquire fully-sampled data for the training of a regularization functional. On the one side, fully sampled DWI requires a longer echo train in EPI, which not only elongates the scan time but also increases off-resonance-induced geometric distortion. On the other side, there exists a wide range of diffusion acquisition modes, thereby requiring a larger dataset than the two-dimensional imaging scenario.

To overcome these challenges, we aim to develop a generalized DWI reconstruction framework with learned nonlinear latent space modeling and self-supervised reconstruction, dubbed DeepDWI.

II. Related Work

A. Variational Autoencoder (VAE)

Autoencoders (AEs) are neural network type trained in an unsupervised manner. They utilize two smaller networks an encoder and a decoder, connected through a latent space. The trained encoder acts as a dimensionality reducer, producing a compact sparse representation of the data in the latent space. The decoder is trained to reconstruction the input data from the encoder produced latent representation. Because the the standard AE training scheme does not constrain the latent space in any way, the network has problems to generalize. The Variational Autoencoder (VAE) was first introduced by Kingma and Welling [?] and uses the principle of variational inference. By that the VAE is forced to learn the probability distributions of the latent variables and a second term is added to the loss function of the network, the result of the Kullbeck Leibler Divergence, which describes how much the learned latent distributions diverge from the real distributions by that regularizing the reconstruction training. The latent space in VAE is not consisting of simple neurons but resambles distributions of the latent activations and samples from these. The sampling operation is not backpropagateable, however, so the network makes use of the reparameterization trick - the VAE learns a mean and a variance for each latent variable and these are used to scale the drawn samples from a gaussian distribution (mean=0, variance=1). In Figure 1 (A) the used VAE model is shown.

B. Multi-Band Multi-Shot DWI Acquisition & Modeling

Figure 1 (B) illustrates the joint k-q-slice forward forward operator for multi-band multi-shot DWI acquisition [?]. This operator can be understood as an extended sensitivity encoding (SENSE) operator [?], which maps the multi-slice multi-diffusion-weighted images ($\tilde{\mathbf{x}}$) to their corresponding k-space,

$$\mathcal{A}(\tilde{\mathbf{x}}) = \mathbf{P\Sigma\Theta FS\Phi}\tilde{\mathbf{x}} \tag{1}$$

Here, the images $\tilde{\mathbf{x}}$ are point-wise multiplied with the pre-computed shot-to-shot phase variation maps ($\mathbf{\Phi}$) and coil sensitivity maps ($\mathbf{S}$). The output images are then converted to k-space via two-dimensional fast Fourier transform ($\mathbf{F}$), point-wise multiplied with the multi-band phases ($\mathbf{\Theta}$), summed along the slice dimension ($\mathbf{\Sigma}$), and then multiplied by the undersampling mask ($\mathbf{P}$).

With the operator $\mathcal{A}$, the inverse problem in DWI reads,

$$\underset{\tilde{\mathbf{x}}}{\operatorname{argmin}} \, \|\mathbf{y} - \mathcal{A}(\tilde{\mathbf{x}})\|_2^2 + \lambda \mathcal{R}(\tilde{\mathbf{x}}) \tag{2}$$

where $\mathbf{y}$ is the measured k-space data, and $\mathcal{R}(\hat{x})$ is the the regularization function with the regularization strength λ. When using the Tikhonov regularization, i.e. $\mathcal{R}(\tilde{\mathbf{x}}) = \|\tilde{\mathbf{x}}\|_2^2$, Equation (2) can be solved via the conjugate gradient (CG) method.

C. Algorithm Unrolling for Image Reconstruction

The regularization function in Equation (2) can be nonlinear, e.g. the sparsity [?] or the low-rankness [?] constraint. In this scenario, algorithms such as the fast iterative shrinkage thresholding (FISTA) [?] and the alternating direction method of multipliers (ADMM) [?] are often employed. These algorithms consist of a substep that transforms $\tilde{\mathbf{x}}$ to a sparsifying domain or a specialized matrix format (e.g., the spatial-diffusion matrix in our previous work [?]) and then performs nonlinear thresholding to promote sparsity or low-rankness. This substep shares similarities to deep neural networks, and inspires the seminal work on algorithm unrolling by Gregor and LeCun [?]. Instead of a hand-crafted regularization function, algorithm unrolling learns deep *prior* via the use of deep neural networks as the regularization function. This enables the learning of true image *prior* during the training process and much faster inference than iterative reconstruction with hand-crafted regularization functions. An excellent review of algorithm unrolling has been provided by Monga et al. [?].

In the area of image reconstruction for accelerated MRI,

III. Methods

IV. Results

A. VAE enables robust & accurate learning of DWI signal

B. Zero-shot learning enables motion-robust DWI

C. Zero-shot learning: model generalization

D. VAE modeling with zero-shot learning reconstruction

V. Discussion

VI. Conclusion

Acknowledgment

Z. T. thanks to Ms. Soundarya Soundarresan for her work and discussion on denoising autoencoder. Z. T. thanks to Dr. Xiaoqing Wang for the discussion on self-supervised learning.

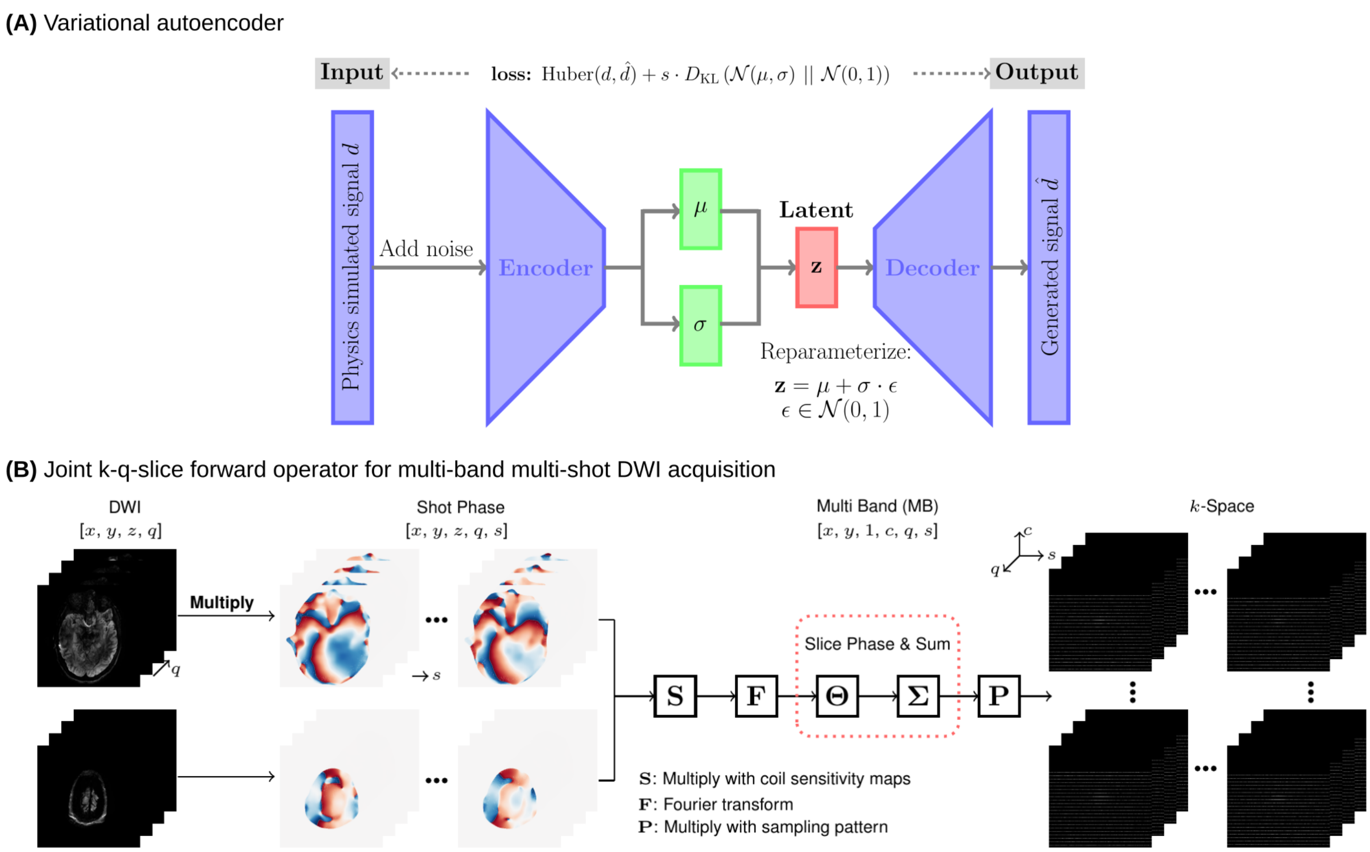

Fig. 1. **(A)** The architecture of a variational autoencoder. **(B)** An illustration of the joint $\boldsymbol{k}$-$\boldsymbol{q}$-slice forward operator for multi-band multi-shot DWI acquisition. $[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{q}]$ denotes the shape of input DWI ($\tilde{\mathbf{x}}$), with $\boldsymbol{x}$ and $\boldsymbol{y}$ as the image size, $\boldsymbol{z}$ as the number of slices, and $\boldsymbol{q}$ as the number of diffusion encodings. The operator outputs multi-dimensional $\boldsymbol{k}$-space with the shape $[\boldsymbol{x}, \boldsymbol{y}, \mathbf{1}, \boldsymbol{c}, \boldsymbol{q}, \boldsymbol{s}]$, with $\boldsymbol{c}$ as the number of receiver coils, $\boldsymbol{s}$ as the number of shots.