Transfer Learning Promotes Robust Parametric Mapping of Diffusion Encoded MR Fingerprinting

Alan Finkelstein[1] ALAN_FINKELSTEIN@URMC.ROCHESTER.EDU
[1] *Department of Biomedical Engineering, University of Rochester, Rochester, NY, USA*

Congyu Liao[2,3] CYLIAO@STANFORD.EDU
Xiaozhi Cao[2,3] XIAOZHIC@STANFORD.EDU
[2] *Department of Radiology, Stanford University, Stanford, CA, USA*
[3] *Department of Electrical Engineering, Stanford University, Stanford CA, USA*

Jianhui Zhong[1,4,5] JIANHUI_ZHONG@URMC.ROCHESTER.EDU
[4] *Department of Imaging Sciences, University of Rochester Medical Center, Rochester, NY, USA*
[4] *Department of Physics and Astronomy, University of Rochester, Rochester, NY, USA*

Editors: Under Review for MIDL 2022

Abstract

MR fingerprinting (MRF) is a framework to simultaneously quantify multiple tissue properties. Acquisition parameters are varied pseudo-randomly and each signal evolution is matched with a dictionary of simulated entries. However, dictionary methods are computationally and memory intensive. Deep learners (DL) are capable of mapping complex MRF signal evolutions to a quantitative parametric space, reducing the computational requirements and reconstruction time; yet fail to perform as well in the setting of noise. Drawing from natural language processing (NLP) we proposed a transfer learning (TL) model to improve MRF parametric estimates with realistic noise levels. The weights of a network trained on clean data are used to instantiate the weights of a noisy model. The model is constrained to learn noise invariant features, by freezing the last layer. Signal evolutions were modeled using a recurrent neural network (RNN) to reconstruct T1, T2, and the apparent diffusion coefficient (ADC). Compared to a model trained with noise, but without TL our approached resulted in a 15% reduction in mean squared error (MSE). Monte Carlo simulations performed at varying SNR (10-60 dB) showed our method yielded losses comparable to the clean model at higher SNRs and proved more robust in the setting of noise at lower SNRs.

Keywords: Transfer Learning, MR Fingerprinting, Recurrent Neural Network

1. Introduction

MRF is a framework for quantifying multiple biophysical properties simultaneously, rapidly, with maps inherently co-registered.(Ma et al., 2013). Parametric mapping is achieved by pseudo-randomly varying acquisition parameters and matching the MRF signal evolutions with a dictionary of simulated entries. Dictionary matching methods remain computationally and memory intensive; scale exponentially with the number of parameters to be reconstructed, and suffer from discretization. DL methods have yielded promising results in MRF (Cohen et al., 2018); yet do not perform as well in the setting of noise. Further,

© 2022 A. Finkelstein, C. Liao, X. Cao & J. Zhong.

confounders, such as the influence of ADC on T2 in MRF-FISP experiments are typically unaccounted for (Kobayashi and Terada, 2019).

In this work we modelled the MRF signal using a recurrent neural network (RNN). The diffusion encoding strength and ADC were also modelled in our approach, which has been shown to improve T2 estimation (Kobayashi and Terada, 2019). To improve performance in the setting of noise we proposed a transfer learning (TL) model, which has been shown to improve classification performance in NLP experiments (Zhang et al., 2020). The weights of a clean model were used to instantiate the weights of a noisy model. The final regression layer was frozen to constrain the model to learn noise invariant features, thereby improving the robustness of the model. We expect that this approach will help with other quantitative MRI tasks where Rician noise is a common problem, and expand MRF to include more parameters.

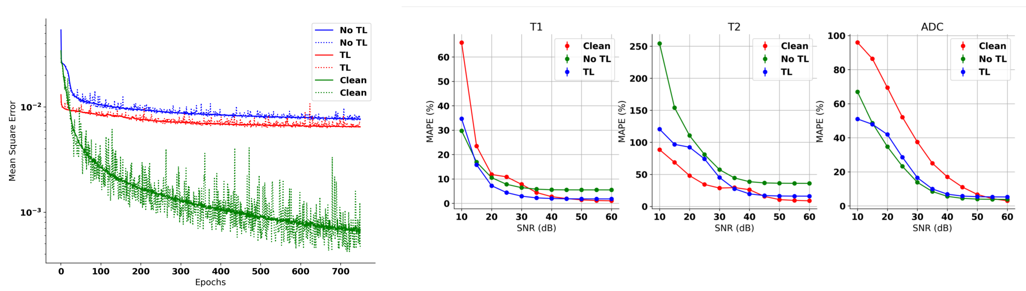
2. Methods

MRF signals were simulated for a set of combinations of T1 (10-4000 ms), T2 (10-3000 ms), and ADC (0.5-3e-9 m^2/s), for which T2 $<$ T1. 3.12e5 instances were used for training and 7.8e4 instances used for validation. Networks were trained to convergence using a supercomputing cluster at the University of Rochester, using an NVIDIA A100 GPU. All networks were based on the proposed architecture, composed of 3 stacked GRUs, with 100 hidden units in the final hidden state. A final fully connected layer was used to generate T1, T2, and ADC. All networks were trained using the Adam optimizer with an initial learning rate (lr) of 5e-5 and the L2 loss optimized.

A clean model was trained without noise and used as the base model for TL. Noisy models were trained with 1 % standard deviation complex gaussian noise. TL was implemented for the final network by instantiating the weights of the TL network with the weights of the clean network (Zhang et al., 2020). The final fc layer was frozen, while the weights of the GRU layers were modified using the same lr as the clean model. All models were evaluated using a numerical brain phantom with complex gaussian noise added at 40 dB. The normalized root mean square error (NRMSE) and mean absolute percentage error (MAPE) were calculated for each reconstruction. Monte Carlo simulations were performed across 100 iterations for variable SNR (10-60 dB), and the MAPE calculated at each SNR.

3. Results

TL resulted in a 15% reduction in MSE compared to the noisy model without TL by the end of training (Figure 1a). Parametric maps were less noisy, with lower NRMSE and MAPE for T1 and T2 maps compared to the noisy model. TL models were more robust in the setting of noise, evinced by Monte Carlo simulations and recapitulated the accuracy achieved with the clean model at higher SNR (Figure 1b).



(a) (b)

Figure 1: (a) Loss curve of training (solid line) and test (dashed line) data. (b) Results of Monte Carlo Simulations for T1, T2, and ADC across variable SNR (10-60 dB).

Table 1: NRMSE and MAPE reported for numerical brain phantom with 40 dB noise.

Model	Metric	T1	T2	ADC
Clean	NRMSE	0.004	0.038	0.012
	MAPE	0.009	0.058	0.022
Noisy without TL	NRMSE	0.015	0.200	0.044
	MAPE	0.048	0.398	0.052
Noisy with TL	NRMSE	0.007	0.183	0.047
	MAPE	0.018	0.263	0.060

4. Conclusions

We showed that TL results in improved reconstruction accuracy and robust parametric estimates, compared to DL models with noise and no TL. We also showed *in silico* that ADC may be reliably reconstructed in the presence of T1 and T2, which has be shown to be a confounder in MRF-FISP experiments in previous work (Kobayashi and Terada, 2019).

References

Ouri Cohen, Bo Zhu, and Matthew S Rosen. Mr fingerprinting deep reconstruction network (drone). *Magnetic resonance in medicine*, 80(3):885–894, 2018.

Yuta Kobayashi and Yasuhiko Terada. Diffusion-weighting caused by spoiler gradients in the fast imaging with steady-state precession sequence may lead to inaccurate t2 measurements in mr fingerprinting. *Magnetic Resonance in Medical Sciences*, 18(1):96, 2019.

Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey L Sunshine, Jeffrey L Duerk, and Mark A Griswold. Magnetic resonance fingerprinting. *Nature*, 495(7440):187–192, 2013.

Shucong Zhang, Cong-Thanh Do, Rama Doddipatla, and Steve Renals. Learning noise invariant features through transfer learning for robust end-to-end speech recognition. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7024–7028. IEEE, 2020.