

KD-MRI: A knowledge distillation framework for image reconstruction and image restoration in MRI workflow

Balamurali Murugesan, Sricharan Vijayarangan, Kaushik Sarveswaran,
Keerthi Ram, Mohanasankar Sivaprakasam

Healthcare Technology Innovation Centre,
Indian Institute of Technology Madras, India

July, 2020

Motivation and Solution

- Integration of deep learning models to MRI workflow demands larger storage and compute power.
- Knowledge distillation (KD) - A type of model compression aids in development of compact models with ease of deployment.
- KD - student model (memory efficient, lower performance network) learns from teacher model (memory intensive, higher performance network) to improve the student's accuracy.

KD for MRI reconstruction

■ Attention based feature distillation

- Attention transfer loss for information distillation:

$$L_{AT} = \sum_{j \in I} \left\| \frac{Q_S^j}{\|Q_S^j\|_2} - \frac{Q_T^j}{\|Q_T^j\|_2} \right\|_2 \quad (1)$$

where $Q_S^j = \text{vec}(F_{\text{sum}}(A_S^j))$, $Q_T^j = \text{vec}(F_{\text{sum}}(A_T^j))$, $F_{\text{sum}}(A) = \sum_{i=1}^C |A_i|^2$ and I denote the set of teacher-student convolution layers which is selected for attention transfer

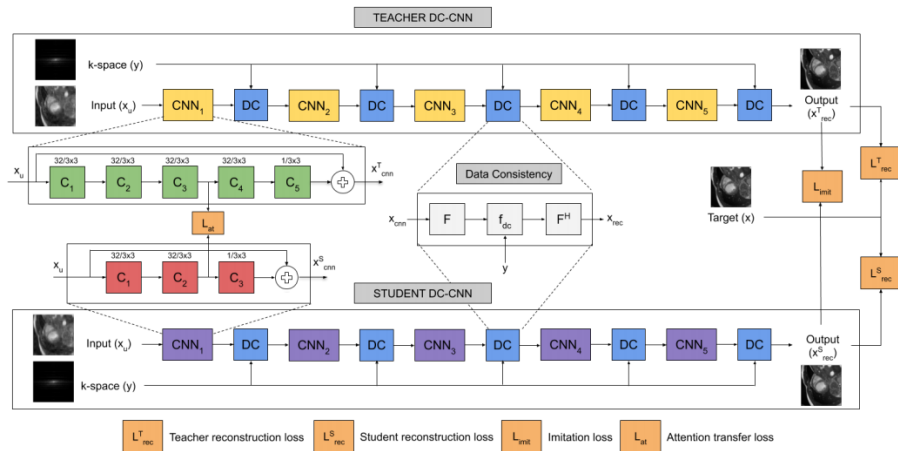
■ Imitation Loss

- Regularizer to the student reconstruction loss

$$L_{\text{total}}^S = \alpha L_{\text{rec}}^S + (1 - \alpha) L_{\text{imit}} \quad (2)$$

where $L_{\text{rec}}^S = \|x - x_{\text{rec}}^S\|$ is the loss between student prediction and target, $L_{\text{imit}} = \|x_{\text{rec}}^T - x_{\text{rec}}^S\|$ is the imitation loss between teacher and student prediction

Block Diagram



Results

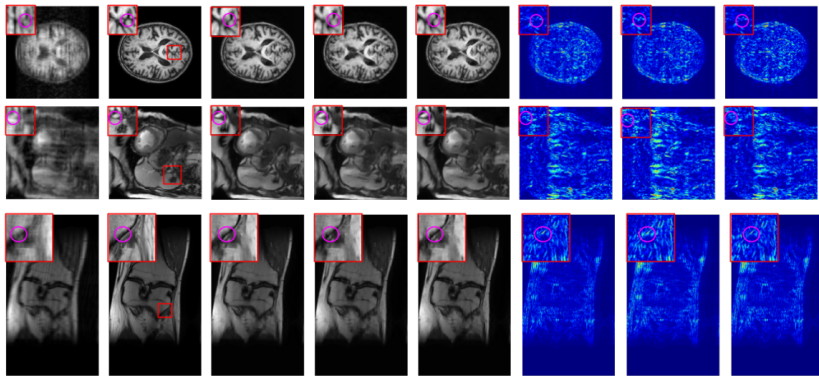


Figure: From Left to Right: Zero-filled, Target, Teacher, Student, Ours (KD-MRI), Teacher Residue, Student Residue, KD-MRI Residue

Thank you

- Paper - <https://arxiv.org/abs/2004.05319>
- Code - <https://github.com/Bala93/KD-MRI>
- Contact - balamurali@htic.iitm.ac.in

