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

Table of Contents

Motivation

- Magnetic Resonance Imaging (MRI) workflow consists of image acquisition, reconstruction, restoration, registration and analysis.
- Deep learning networks have shown encouraging results for every stage in the MRI workflow.
- Deep learning networks are specific to task, dataset (anatomical study, contrast).
- For MRI reconstruction, deep learning networks are also specific to type of degradation (acceleration factor, undersampling mask).
- Integration of deep learning models to MRI workflow demands larger storage and compute power.
- Development of memory-efficient model is required.

Solution

- Model compression - Deploying state-of-the-art deep networks in low-power and resource limited devices without significant drop in accuracy.
- Knowledge distillation (KD) - Develop 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.
- For MRI reconstruction and restoration, we propose:
  - Attention-based feature distillation - Student learn the intermediate representation of the teacher.
  - Imitation loss - Regularizer to the reconstruction loss.

MRI reconstruction

- MRI reconstruction
  - Transformation of Fourier space (k-space) data to image domain.
  - MRI is a slow acquisition modality, acceleration is done by under sampling k-space.
  - De-alias the artifact due to undersampling and provide reconstruction equivalent of fully sampled k-space.
- Deep Cascade - Convolution Neural Network (DC-CNN)
  - Cascade of convolutional neural networks (CNN) and a data consistency (DC) layer.
  - CNN - To learn Image-to-Image mapping, DC - To provide consistency in Fourier domain.
  - n_c cascades, every cascade has n_d convolution layers and 1 DC layer.
- Teacher DC-CNN
  - $n_c = 5, n_d = 5$
- Student DC-CNN
  - $n_c = 5,\ n_d = 3$

KD for MRI reconstruction

- Attention based feature distillation
  - Attention transfer loss for information distillation:

$$L_{AT} = \sum_{j \in I} || \frac{Q_S^j}{||Q_S^j||_2} - \frac{Q_T^j}{||Q_T^j||_2} ||_2 \tag{1}$$

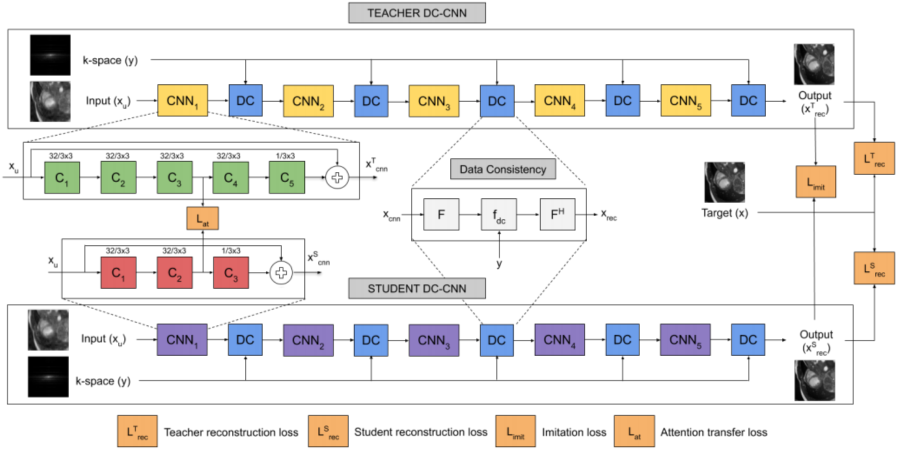
where $Q_S^j = vec(F_{sum}(A_S^j))$, $Q_T^j = vec(F_{sum}(A_T^j))$, $F_{sum}(A) = \sum_{i=1}^{C} |A_i|^2$ and I denote the set of teacher-student convolution layers which is selected for attention transfer

- Imitation Loss
  - Regularizer to the student reconstruction loss

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

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

Block Diagram



Training procedure

- Step1: Train the teacher DC-CNN f^{T}_{cnn} weights θ^{T} using teacher reconstruction loss $L^{T}_{rec} = ||x - x^{T}_{rec}||$
- Step2: Train the student DC-CNN f^{S}_{cnn} weights θ^{S} using attention transfer loss $L_{AT} = ||Q_T - Q_S||$ between teacher and student
- Step3: Load the weights θ^{S} from Step2 and re-train f^{S}_{cnn} weights θ^{S} using student reconstruction and imitation loss $L^{S}_{total} = \alpha||x - x^{S}_{rec}|| + (1-\alpha)||x^{T}_{rec} - x^{S}_{rec}||$

Quantitative comparison

		4x		5x		8x	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Cardiac	ZF	24.27 ± 3.10	0.6996 ± 0.08	23.82 ± 3.11	0.6742 ± 0.08	22.83 ± 3.11	0.6344 ± 0.09
	Teacher	32.51 ± 3.23	0.9157 ± 0.04	31.49 ± 3.32	0.9002 ± 0.04	28.43 ± 3.13	0.8335 ± 0.06
	Student	31.92 ± 3.17	0.9053 ± 0.04	30.79 ± 3.24	0.8863 ± 0.05	27.87 ± 3.11	0.8156 ± 0.07
	Ours	32.07 ± 3.21	0.9084 ± 0.04	31.01 ± 3.27	0.8913 ± 0.04	28.11 ± 3.17	0.8236 ± 0.07
Brain	ZF	31.38 ± 1.02	0.6651 ± 0.02	29.93 ± 0.80	0.6304 ± 0.02	29.37 ± 0.98	0.6065 ± 0.03
	Teacher	40.03 ± 2.00	0.9781 ± 0.00	39.03 ± 1.28	0.971 ± 0.00	35.04 ± 1.38	0.9374 ± 0.01
	Student	39.36 ± 1.82	0.9753 ± 0.00	38.58 ± 1.28	0.9674 ± 0.00	34.39 ± 1.26	0.9281 ± 0.01
	Ours	39.8 ± 1.89	0.977 ± 0.00	38.78 ± 1.24	0.9688 ± 0.00	34.83 ± 1.35	0.9337 ± 0.01
Knee	ZF	29.66 ± 3.86	0.8066 ± 0.08	29.2 ± 3.87	0.8007 ± 0.08	28.71 ± 3.88	0.7985 ± 0.08
	Teacher	37.15 ± 3.55	0.9436 ± 0.03	35.16 ± 3.46	0.9231 ± 0.03	32.53 ± 3.49	0.8887 ± 0.05
	Student	36.37 ± 3.53	0.9367 ± 0.03	34.37 ± 3.47	0.9144 ± 0.04	31.92 ± 3.58	0.8804 ± 0.05
	Ours	36.7 ± 3.52	0.9392 ± 0.03	34.71 ± 3.44	0.9181 ± 0.04	32.32 ± 3.57	0.8867 ± 0.05

Qualitative comparison

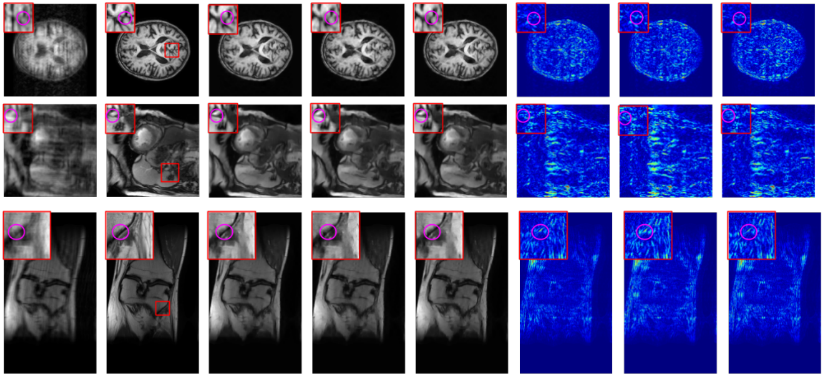


Figure: From Left to Right: Zero-filled, Target, Teacher, Student, Ours (KD-MRI), Teacher Residue, Student Residue, KD-MRI Residue

Conclusion

- We proposed a knowledge distillation (KD) framework for image to image problems in the MRI workflow in order to develop compact, low-parameter models without a significant drop in performance.
- We propose obtaining teacher supervision through a combination of attention transfer and imitation loss.
- We demonstrated its efficacy on the DC-CNN network and show consistent improvements in student reconstruction across datasets and acceleration factors.

Thank you

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