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

Appendix A. Knowledge distillation for MRI Super-Resolution

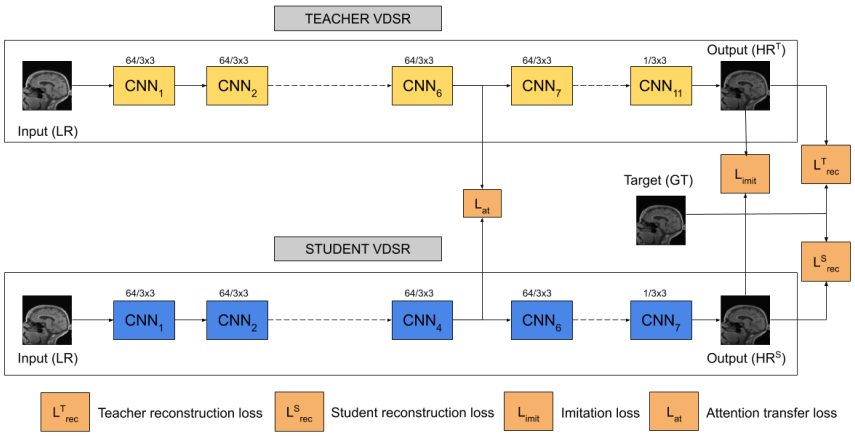

Figure 1: Teacher VDSR: 11 convolution layers. Student VDSR: 7 convolution layers. Attention Transfer Loss: Loss between sixth convolution layer of teacher and fourth convolution layer of student VDSR. Imitation Loss: Loss between reconstructed output of teacher and student VDSR.

A.1. MRI Super-Resolution architecture

MRI Super-Resolution involves the reconstruction a high-resolution (HR) image from a low-resolution (LR) image. Interpolation methods fail to recover the loss of high frequency information like fine edges of objects. So, deep learning architectures like VDSR (Kim et al., 2016) were proposed. VDSR architecture consists of n blocks of convolution and ReLU with a residual connection between the input and the output. LR image is interpolated to match the dimension of HR image.

A.2. Proposed knowledge distillation framework

The overview of knowledge distillation designed for MRI super resolution architecture VDSR is depicted in Figure 1.

1. **Teacher VDSR**: VDSR with with $n = 11$ is chosen as teacher
2. **Student VDSR**: VDSR with $n = 7$ is chosen as student
3. **Attention transfer**: Attention transfer is done using the middle layers of both Teacher and Student VDSR (between 7^{th} and 4^{th} layer).
4. **Imitation Loss**: Imitation loss is calculated between outputs of Teacher and student VDSR.

A.3. Dataset Preparation

We use Calgary-Campinas dataset (Souza et al., 2018) and prepared the fully sampled train and valid MRI slices as done in (Souza et al., 2019). Low resolution MRI (4x) images are created using the procedure followed in (Chen et al., 2018).

Table 1: Quantitaive comparison of Teacher, Student, KD VDSR

	PSNR	SSIM
ZF	27.9 +/- 1.13	0.8117 +/- 0.01
Teacher (333K)	30.14 +/- 1.33	0.8659 +/- 0.01
Student (185K)	29.87 +/- 1.31	0.8596 +/- 0.01
KD (185K)	30.08 +/- 1.33	0.8639 +/- 0.01

A.4. Results and discussion

In Table 1, the quantitative metrics and parameter count of Teacher VDSR, Student VDSR, KD VDSR are presented. It can be seen that, KD VDSR an equivalent architecture of Student VDSR provides improved reconstruction compared to Student VDSR. Qualitative comparison is depicted in 2. KD VDSR provides 44% parameter reduction compared to Teacher VDSR. The validation loss comparison of Teacher, Student, KD VDSR is shown in Figure 3. From the graph, it can be inferred that KD VDSR has less validation error compared to Student VDSR and is near to Teacher VDSR.

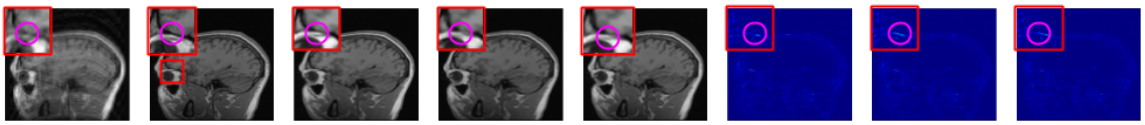

Figure 2: From Left to Right: Undersampled, Target, Teacher, Student, Ours(KD), Teacher Residue, Student Residue, KD Residue. As with MRI Reconstruction, in addition to lower reconstruction errors the distilled model is able to retain finer structures better when compared to the student.

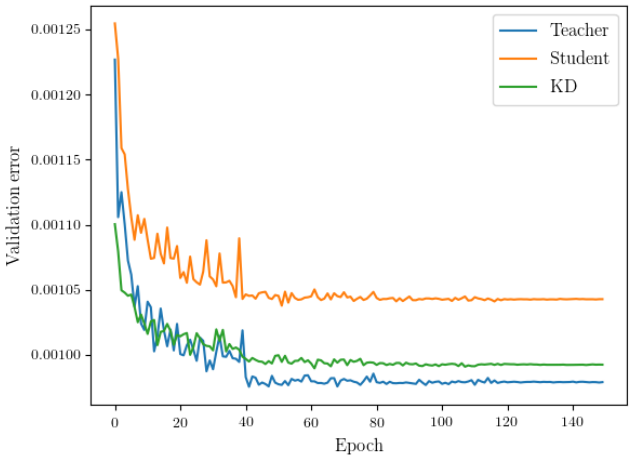

Figure 3: Validation loss comparison of teacher, student, kd vdsr

Appendix B. Choice of Feature Distillation Position

	PSNR	SSIM
Teacher	28.43 +/- 3.13	0.8335 +/- 0.06
Student	27.87 +/- 3.11	0.8156 +/- 0.07
AT12	28.05 +/- 3.17	0.8217 +/- 0.07
AT22	28.09 +/- 3.18	0.8236 +/- 0.07
AT32	28.11 +/- 3.16	0.8235 +/- 0.07
AT42	28.07 +/- 3.18	0.8223 +/- 0.07

Table 2: Studying the effect of Teacher supervision obtained from different convolution layers. Across all experiments, teacher supervision was obtained from the output of the third convolution in each cascade. This decision can be corroborated by noting that teacher supervision obtained from second and third layers provide superior reconstruction. Conversely, supervision obtained from the output of the first layer and the fourth layer leads to a relatively poor reconstruction. We infer that imparting low level (first layer) or extremely complex information (penultimate layer) to the student does little in ameliorating the quality of the reconstructed image.

Appendix C. Undersampling masks for Cardiac, Brain and Knee

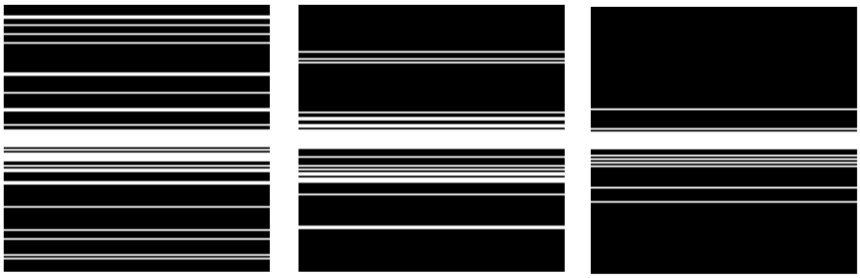

Figure 4: Cardiac MRI dataset undersampling mask. From Left to Right: 4x, 5x, 8x

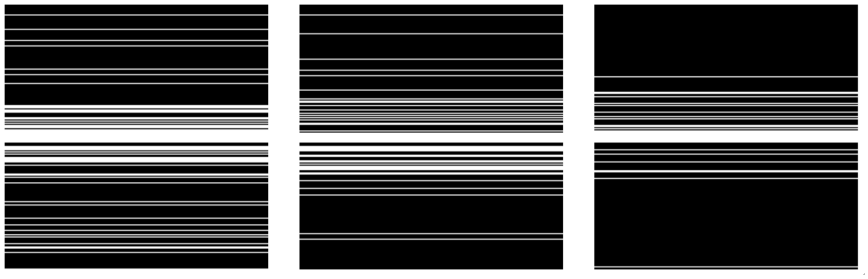

Figure 5: Brain MRI dataset undersampling mask. From Left to Right: 4x, 5x, 8x

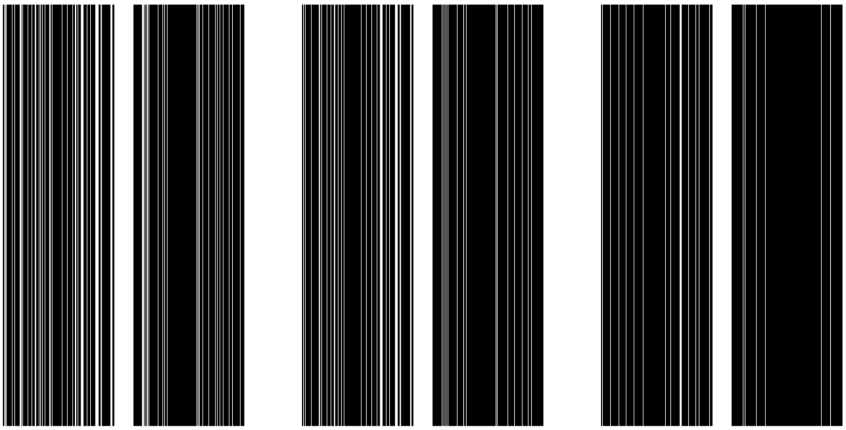

Figure 6: Knee MRI dataset undersampling mask. From Left to Right: 4x, 5x, 8x

Appendix D. Qualitative and quantitative comparison of attention map residues

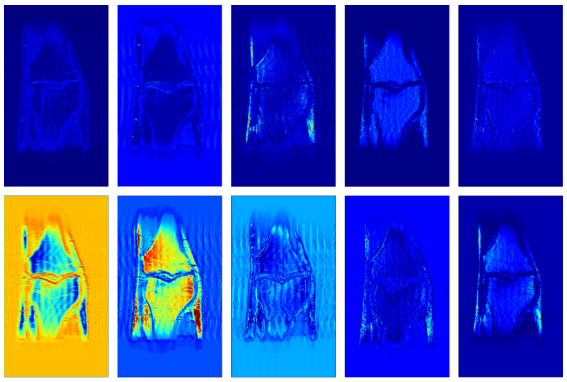

Figure 7: Top(From Left to Right): Residue computed between feature attention maps of the distilled model and feature attention maps of teacher from cascade 1 to cascade 5. Bottom(From Left to Right): Residue computed between student(pre-KD) and teacher features from cascade 1 to cascade 5. We qualitatively establish that the information distillation occurring in the first cascade of the distilled model gives it a head start, helping it mimic the teacher's attention map better. On the contrary, the student is a relatively slow learner requiring lot more levels of convolution layers before it can get reasonably close to the teacher.

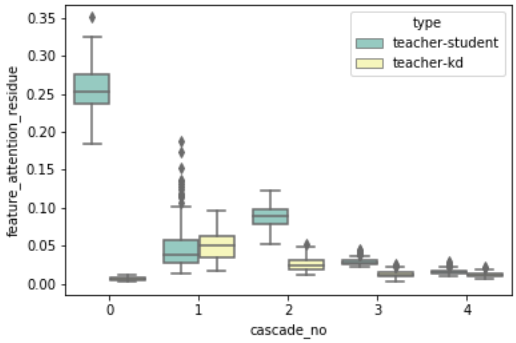

Figure 8: Box plot Feature map residues of student w.r.t teacher and distilled model w.r.t teacher across validation data in ACDC. The quantitative results strengthen the inferences drawn from the qualitative observations made in the previous section. The distilled model is able to learn quicker than student as can be ascertained from the huge difference in residues of feature maps obtained from the first cascade layer.

Appendix E. Comparative study of feature distillation methods

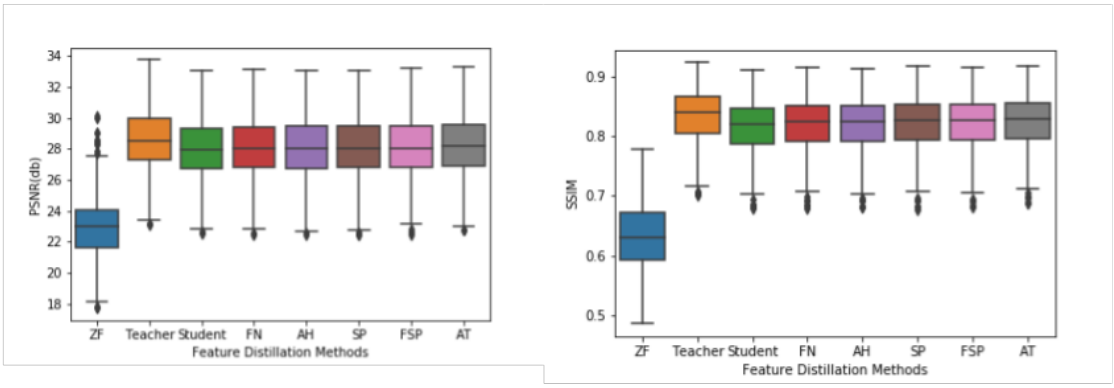

Figure 9: PSNR and SSIM across validation data in ACDC for various feature distillation methods(acc factor:8x). The acronyms used in the plots are expanded as follows: ZF-Zero Filled, FN-Fitnet, AH-Attentive Hint, SP-Similarity Preserving, FSP-Flow of Solution, AT-Attention Transfer (Ours). The plot quantitatively consolidates the superior performance of our method over other feature distillation methods.

References

Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li. Brain mri super resolution using 3d deep densely connected neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 739–742, April 2018.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral)*, June 2016.

Roberto Souza, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170:482 – 494, 2018. ISSN 1053-8119.

Roberto Souza, R. Marc Lebel, and Richard Frayne. A hybrid, dual domain, cascade of convolutional neural networks for magnetic resonance image reconstruction. In *International Conference on Medical Imaging with Deep Learning – Full Paper Track*, 08–10 Jul 2019.