The truth matters: A brief discussion on MVUE vs. RSS in MRI reconstruction

Marius Arvinte[1] (arvinte@utexas.edu) and Jon Tamir[1] (jtamir@utexas.edu)
[1]Electrical and Computer Engineering, University of Texas at Austin, USA

Introduction: The goal of this study is to raise awareness of the impact of choosing ground-truth images when numerically evaluating MRI reconstruction algorithms, and to the fact that misaligned comparisons can lead to unfair penalties (e.g., some types of algorithms may be unfairly penalized in challenge leaderboards).

Methods: We consider the following multicoil MRI reconstruction algorithms: MoDL[1] and parallel imaging and compressed sensing (PICS) with L1-Wavelet regularization[2]. MoDL is a supervised deep learning method in which the target output image can be chosen for training. This is not the case for PICS, which outputs a complex valued solution to the inverse problem. In this work, we evaluate these methods against two fully-sampled magnitude-only images: the root sum-of-squares (RSS) and the absolute value of the minimum variance unbiased estimate (MVUE). We use ESPIRiT[3] maps to estimate the MVUE and train two MoDL models, one for each target signal and evaluate at $R=\{4,8\}$ according to the fastMRI protocol. Our code is available online[4].

Results and Discussion: *Table 1* shows the average SSIM (PSNR) on 18 validation samples of T2 brain scans from the fastMRI[5] database with equispaced acceleration *R*. MoDL has good performance on its matched supervised training signal, but suffers from a severe quantitative loss when evaluated on the opposite signal. PICS only obtains good quantitative scores on MVUE images, since equivalent RSS maps are impossible to obtain from undersampled data. Unfairly penalized methods are highlighted in **red**. Notably, the scores for PICS and MoDL-MVUE are even worse than those of a zero-filled RSS reconstruction at R=4 when evaluated against RSS. This result also shows that PICS is competitive with MoDL at *R=4*, which is inconsistent with the fastMRI public leaderboards, where evaluation is always done against RSS images.

	MoDL-MVUE		**MoDL-RSS**		**PICS (L1-Wavelet)**		**Zero-Filled**	
	R = 4	*R = 8*	*R = 4*	*R = 8*	*R = 4*	*R = 8*	*R = 4*	*R = 8*
Test on MVUE	**0.950 (38.3)**	**0.891 (31.3)**	**0.775 (33.1)**	**0.716 (29.5)**	**0.929 (37.6)**	**0.757 (26.8)**	**0.780 (27.0)**	**0.631 (22.5)**
Test on RSS	**0.782 (33.7)**	**0.723 (30.0)**	**0.945 (37.4)**	**0.895 (31.2)**	**0.751 (33.2)**	**0.668 (26.7)**	**0.793 (33.2)**	**0.662 (26.7)**

Table 1. Average test SSIM (PSNR in parentheses) for the considered methods on fastMRI T2 brain scans.

Figure 1 shows ground truth RSS, MoDL trained on MVUE, PICS, and zero-filled RSS reconstruction, respectively. The key result here is that while both reconstructions look near-identical, are of high quality and likely diagnostically equivalent, the quantitative SSIM scores decrease dramatically if the comparison is done with the RSS image and both are seemingly outperformed by the zero-filled reconstruction, even though it introduces considerable artifacts.

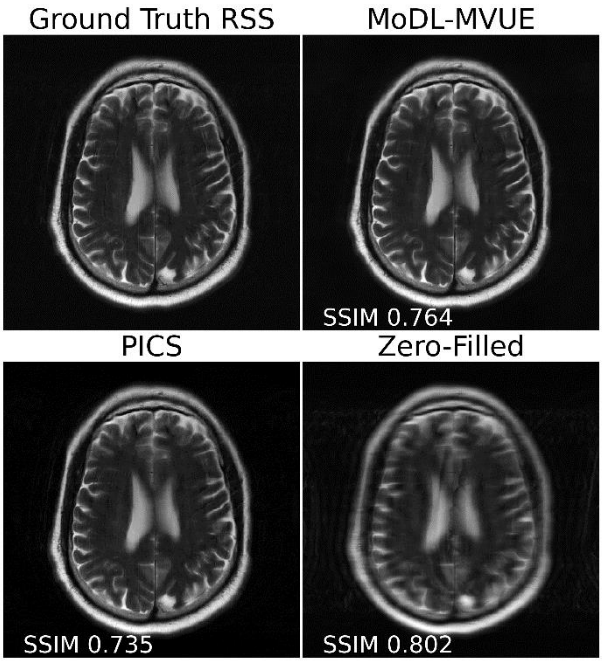

Figure 1. Ground truth, MoDL (trained on MVUE), PICS, and zero-filled RSS reconstructions of a test brain scan at R=4. Annotated scores are evaluated against RSS. Even though MoDL and PICS are both qualitatively superior, the scores suffer a penalty due to evaluating on a mismatched ground-truth image. While this is fixable for MoDL, it is not addressable for PICS or other algorithms which explicitly solve the inverse problem.

Conclusion: Research studies should clearly define the training and test signals. When using PICS-based methods, MVUE is preferred since this allows for a fair comparison. In some cases models cannot be “retrofitted” to the desired signal. For example, if the estimated sensitivity maps are normalized across the coil dimension then applying the forward and RSS operators to the complex image output of MoDL-MVUE leads to *exactly* the same magnitude image, thus is still penalized against RSS.

References: 1. Aggarwal et al. "MoDL: Model-based deep learning architecture for inverse problems." *IEEE TMI* (2018).
2. Lustig et al. "Sparse MRI: The application of compressed sensing for rapid MR imaging." *Magnetic Resonance in Medicine* (2007).
3. Uecker et al. "ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA." *Magnetic resonance in medicine* (2014).
4. https://github.com/utcsilab/mvue-vs-rss-mri
5. Zbontar et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." *arXiv preprint arXiv:1811.08839* (2018).