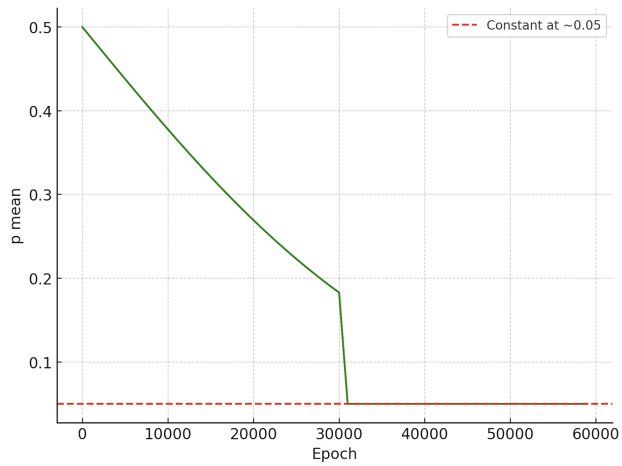


Figure 1: Mean of p across various epochs when the regularization used is $|mean(p) - \frac{s}{d}|$. In this particular experiment, $\frac{s}{d} = 0.05$.

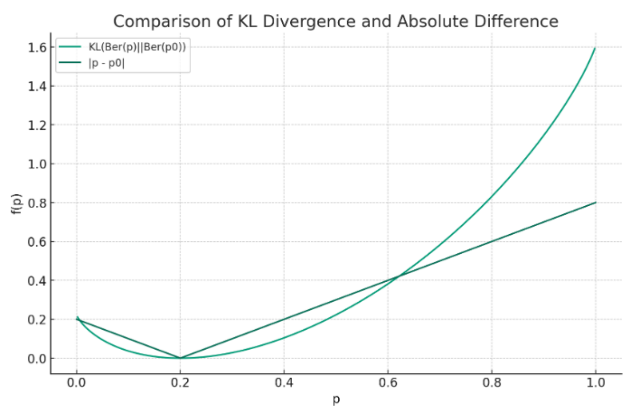
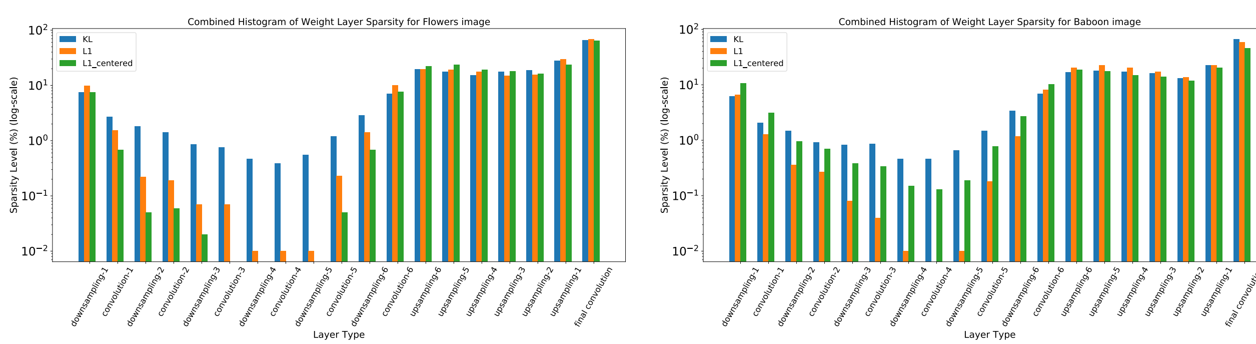


Figure 2: Comparison for KL regularization and pointwise centre ℓ_1 regularization for a scalar value. Around the prior value p_0, the KL is much smoother than L1 regularizer. $f'(p) = \log(\frac{\frac{p}{1-p}}{\frac{p_0}{1-p_0}})$ for KL, which is very close to 0 when $p \in [p_0 - \epsilon, p_0 + \epsilon]$. However, for L1 regularization, $f'(p) = 1$ or -1 for all points except $p = p_0$.



(a) Flowers

(b) Baboon

Figure 3: Layerwise architecure pruning (sparsity percentage in log-scale) by OES at initialziaiton using three different choices of regularization, KL, ℓ_1 and centered ℓ_1 for Baboon image and Flowers image in Set-14 dataset. Centered ℓ_1 means the centered mean regularizer.

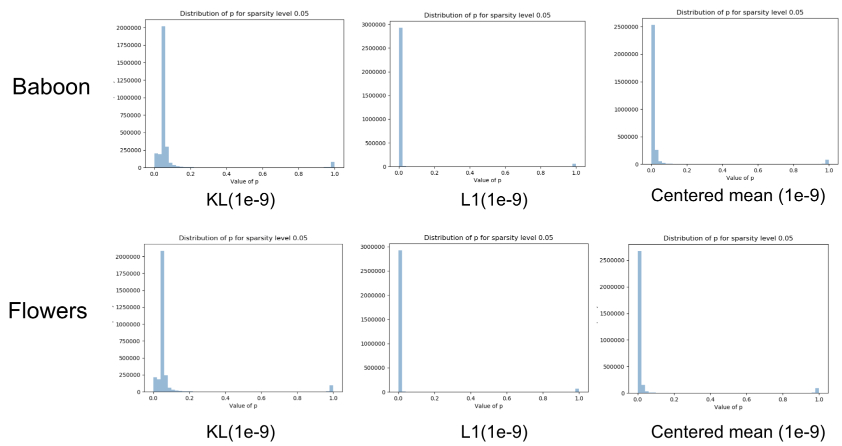


Figure 4: Histogram of logits of p when OES is ran across images with KL, ℓ_1 and centered mean regularizer. In our implementation we minimize $|\sum_i p_i - (\frac{s}{d} * numel(p))|$, to both ℓ_1 regularization and centered mean regularizer on the same scale. $numel(p)$ denotes the total number of elements in p.

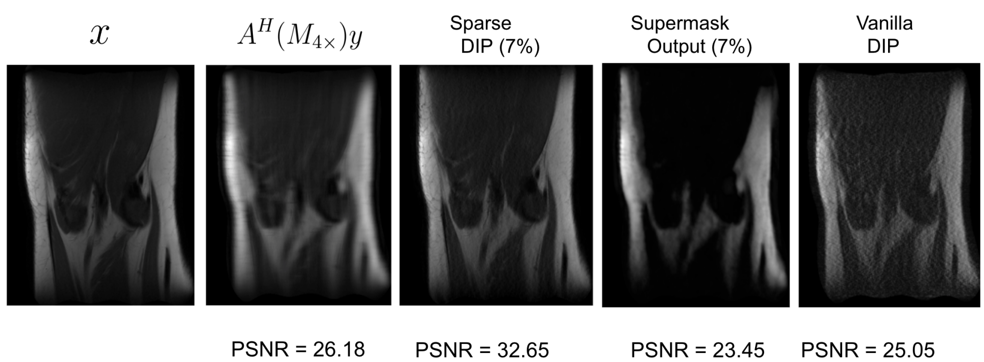


Figure 5: MRI reconstruction comparison with Sparse-DIP and Vanilla Dense DIP without early stopping. Sparse-DIP removes aliasing artifacts and preserves the important details of the images when compared to the ground-truth $\mathbf{x}$. Vanilla dense DIP overfits to the aliasing artifacts and requires careful early stopping (See Figure 7a). Supermasked output at network initialization still manages to capture some important image details. Sparse DIP 7% indicates only 7% weights are non-zero.

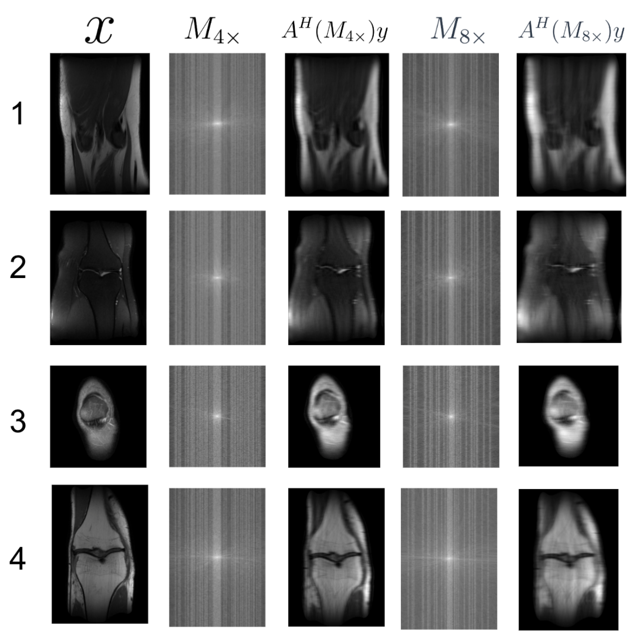
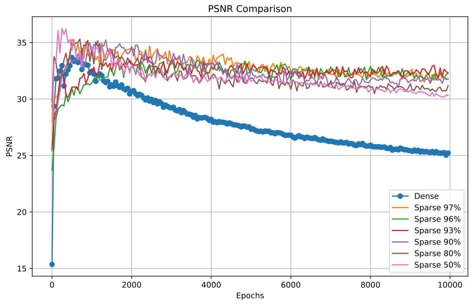
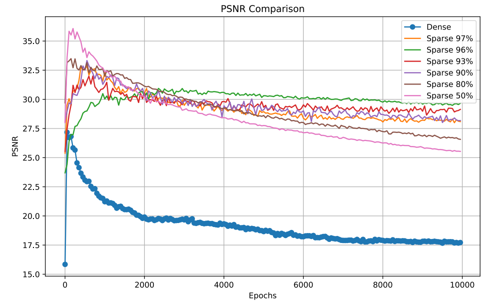


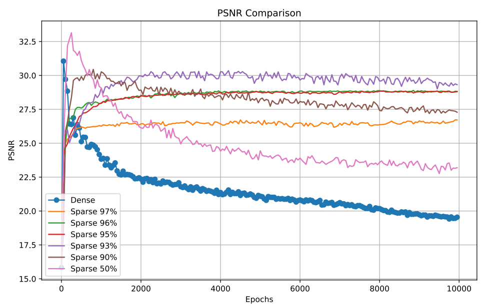
Figure 6: The 4 MRI ground-truth and measurement used in this experiment. $\mathbf{x}$ denotes the ground-truth image or full-kspace reconstruction. $\mathbf{M}_{4\times}$ and $\mathbf{M}_{8\times}$ denote the k-space undersampling masks. $A^H(\mathbf{M}_{4\times})\mathbf{y}$ and $A^H(\mathbf{M}_{8\times})\mathbf{y}$ denotes the zero-filling reconstruction that produces aliased artifacts.



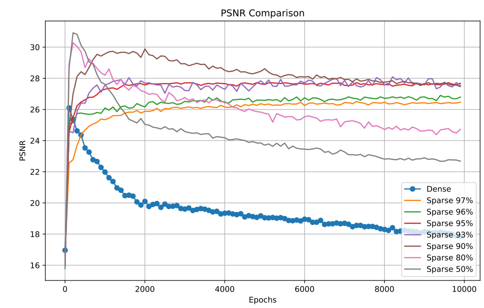
(a) Self: $P(A_1(M_{4\times}), y_1)$



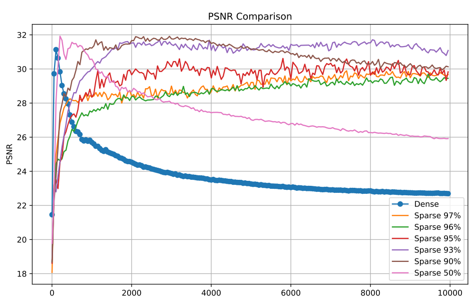
(b) Self: $P(A_1(M_{8\times}), y_1)$



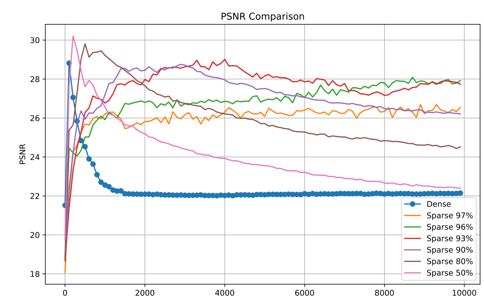
(c) Cross: $P(A_2(M_{4\times}), y_2)$



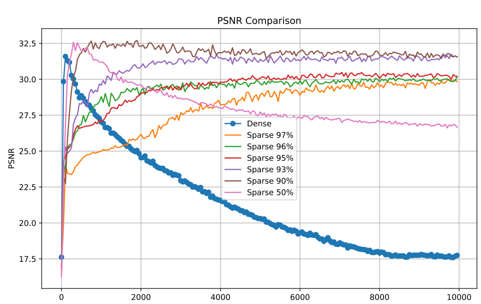
(d) Cross: $P(A_2(M_{8\times}), y_2)$



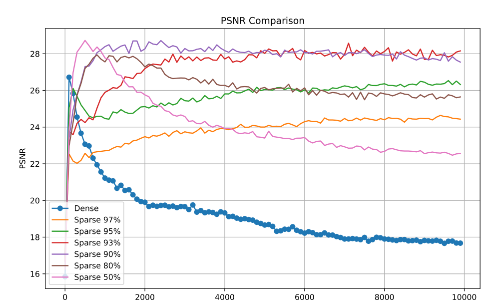
(e) Cross:$P(A_3(M_{4\times}), y_3)$



(f) Cross: $P(A_3(M_{8\times}), y_3)$



(g) $P(A_4(M_{4\times}), y_4)$



(h) Cross: $P(A_4(M_{8\times}), y_4)$

Figure 7: Performance of OES subnetworks for MRI reconstruction from $4\times$ (left column) and $8\times$ (right column) undersampled k-space measurements. In Figures(a,b) (self), the OES network mask m^* was learned from pair $((A_1(M_{4\times}), y_1))$ and then the subnetwork was used to reconstruct image from $((A_1(M_{4\times}), y_1))$. In Figures (c-h) (cross), different measurements $((A_2, y_2), (A_3, y_3), (A_4, y_4))$ are used across 2 different undersampling rate.