Supplementary Information
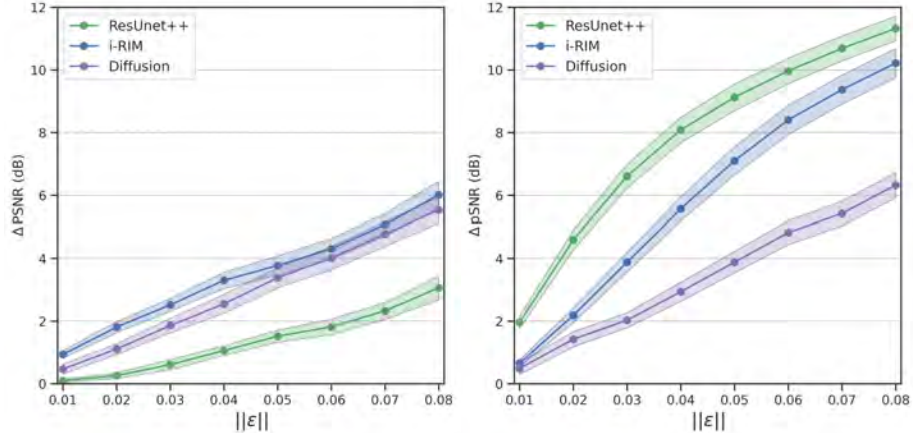

Fig. S1: Comparisons of the pSNR change of all models against l_2-bounded adversaries with increasing perturbation amplitudes in the frequency space. **left,** A worst-case perturbation was generated by using the trained i-RIM model. Comparisons in the left subplot correspond to our experiments outlined in Fig. 2a. **Right,** A worst-case perturbation was generated by using the ResUnet++ model. Comparisons in the right subplot correspond to our experiments outlined in Fig. 2b.
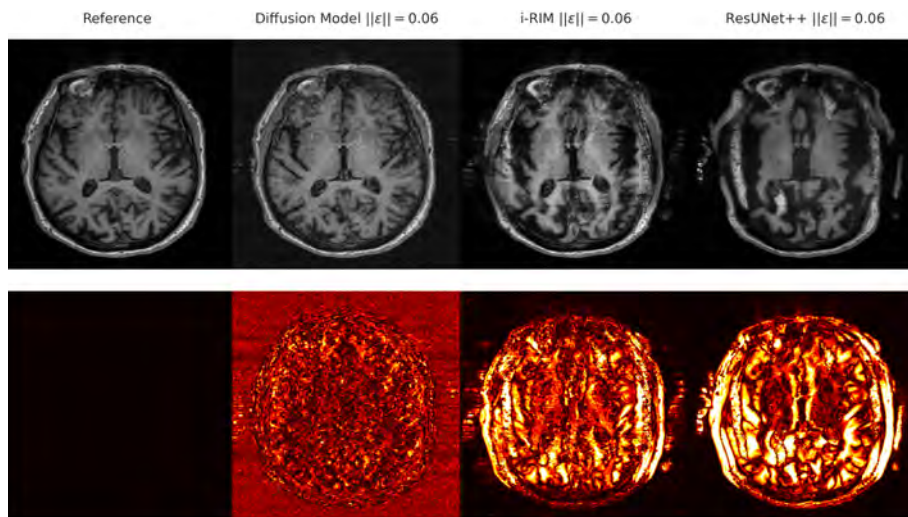
Fig. S2: **Top,** Reconstructions of both supervised and unsupervised methods under the mid-level perturbation generated from ResUnet++. **Bottom,** Mean absolute error map of all reconstructions. The diffusion model, trained like a denoiser, exhibits an unstable reconstruction with a distinct noisy pattern that is difficult to distinguish from the clean reconstruction.