Supplementary Materials

Yuhui Quan[1], Xinran Qin[1], Tongyao Pang[2], and Hui Ji[2]

[1] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[2] Department of Mathematics, National University of Singapore, 119076, Singapore
csyhquan@scut.edu.cn, csqinxinran@mail.scut.edu.cn, matpt@nus.edu.sg, matjh@nus.edu.sg

1 Algorithm Description of Test-Time Model Adaption

Algorithm 1 Inference with Test-Time Model Adaption

Input: Measurement sample $\boldsymbol{y}^\star$; Sensing matrix $\boldsymbol{\Phi}^*$; Pre-trained parameters $\widehat{\boldsymbol{\omega}}$
Parameter: Learning rate τ; Epoch number T
Output: Reconstructed image $\boldsymbol{x}^\star$
1. Initialize $\boldsymbol{\omega}^\star$ with $\widehat{\boldsymbol{\omega}}$.
2. for $i = 1, \cdots, T$, update $\boldsymbol{\omega}^\star$ on test sample:
3. $\quad\boldsymbol{\omega}^\star := \boldsymbol{\omega}^\star - \tau \nabla_{\boldsymbol{\omega}^\star} \mathcal{L}^{\text{Dual}}_{\boldsymbol{y}^\star}(\boldsymbol{\omega}^\star)$.
4. return $\boldsymbol{x}^\star = f_{\boldsymbol{\Phi}^*}(\boldsymbol{y}^\star; \boldsymbol{\omega}^\star)$.

2 Proof of Proposition 1

Here we only provide the proof regarding the connection between $\mathcal{L}^{\text{Measure}}$ and $\mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\gamma}} \|\boldsymbol{\Phi} f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\gamma}) - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2$. The proof regarding the connection between $\mathcal{L}^{\text{Image}}$ and $\mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\gamma}} \|f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}(\boldsymbol{z}+\boldsymbol{r})) - \boldsymbol{x}\|_2^2$ is the same. Firstly, rewrite $\mathcal{L}^{\text{Measure}}$ by

$$\begin{aligned}\mathcal{L}^{\text{Measure}} =&\mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\gamma}} \big[\|\boldsymbol{\Phi} f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\gamma}) - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \\ &+ 2\overline{(\boldsymbol{\gamma}-\boldsymbol{\epsilon})}^\top \boldsymbol{\Phi}(f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\gamma}) - \boldsymbol{x}) + \overline{(\boldsymbol{\gamma}-\boldsymbol{\epsilon})}^\top (\boldsymbol{\gamma}-\boldsymbol{\epsilon})\big],\end{aligned} \tag{1}$$

where the last term $\mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\gamma}} \overline{(\boldsymbol{\gamma}-\boldsymbol{\epsilon})}^\top (\boldsymbol{\gamma}-\boldsymbol{\epsilon})$ is a constant irrelevant to the value of the NN parameters $\boldsymbol{\omega}$. Since $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ conditioned on $\boldsymbol{x}$ are independent and follow the same distribution $P_1(\cdot|\boldsymbol{x})$, we have

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\gamma}} \overline{\boldsymbol{\epsilon}}^\top \boldsymbol{\Phi}(f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\gamma}) - \boldsymbol{x}) \\ =&\mathbb{E}_{\boldsymbol{x}} \mathbb{E}_{\boldsymbol{\epsilon}|\boldsymbol{x}} \mathbb{E}_{\boldsymbol{\gamma}|\boldsymbol{x}} \overline{\boldsymbol{\epsilon}}^\top \boldsymbol{\Phi}(f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}\boldsymbol{x}+\boldsymbol{\epsilon}+\boldsymbol{\gamma}) - \boldsymbol{x}) \\ =&\int_{\boldsymbol{x}}\int_{\boldsymbol{\epsilon}|\boldsymbol{x}}\int_{\boldsymbol{\gamma}|\boldsymbol{x}} p_{\boldsymbol{x}}(\boldsymbol{x}) P_1(\boldsymbol{\gamma}|\boldsymbol{x}) P_1(\boldsymbol{\epsilon}|\boldsymbol{x}) \overline{\boldsymbol{\epsilon}}^\top \boldsymbol{\Phi}(f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}\boldsymbol{x}+\boldsymbol{\epsilon}+\boldsymbol{\gamma}) - \boldsymbol{x}) \\ =&\int_{\boldsymbol{x}}\int_{\boldsymbol{\epsilon}|\boldsymbol{x}}\int_{\boldsymbol{\gamma}|\boldsymbol{x}} p_{\boldsymbol{x}}(\boldsymbol{x}) P_1(\boldsymbol{\epsilon}|\boldsymbol{x}) P_1(\boldsymbol{\gamma}|\boldsymbol{x}) \overline{\boldsymbol{\gamma}}^\top \boldsymbol{\Phi}(f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}\boldsymbol{x}+\boldsymbol{\gamma}+\boldsymbol{\epsilon}) - \boldsymbol{x}) \\ =&\mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\gamma}} \overline{\boldsymbol{\gamma}}^\top (f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}\boldsymbol{x}+\boldsymbol{\epsilon}+\boldsymbol{\gamma}) - \boldsymbol{x}).\end{aligned} \tag{2}$$

Thus the second term on the right hand side of Eqn. (1) is zero, which leads to

$$\mathcal{L}^{\text{Measure}} = \mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\gamma}}\|\boldsymbol{\Phi} f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\gamma}) - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + const.$$

The proof is done.

3 Proof of $\mathbb{E}(\boldsymbol{r}^\top \boldsymbol{e}) = 0$

As $\mathbb{E}\boldsymbol{s}_i = 0$ and $\boldsymbol{s}$ is independent from $\boldsymbol{e}$ and $\boldsymbol{e}'$, we can obtain

$$\mathbb{E}(\boldsymbol{r}^\top \boldsymbol{e}) = \mathbb{E}(\boldsymbol{e}' \odot \boldsymbol{s})^\top \boldsymbol{e} = \sum_i \mathbb{E}\boldsymbol{s}_i \boldsymbol{e}'_i \boldsymbol{e}_i = \sum_i (\mathbb{E}\boldsymbol{s}_i)(\mathbb{E}\boldsymbol{e}'_i \boldsymbol{e}_i) = 0. \tag{3}$$

4 Proof of Proposition 2

Since $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}'$ are i.i.d. Gaussian noise of zero mean and independent from $\boldsymbol{x}$, we have

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}(\boldsymbol{\epsilon}')^\top \boldsymbol{y} = \mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},\boldsymbol{\epsilon}'}(\boldsymbol{\epsilon}')^\top(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\epsilon}) = 0.$$

It yields that

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}(\mathcal{L}^{\text{SURE+}}(\boldsymbol{\omega}) - \mathcal{L}^{\text{Measure}}) \\
=&\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big\{2\sigma^2 \text{tr}\Big(\boldsymbol{\Phi}^{\text{H}} \frac{\partial f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})}{\partial \boldsymbol{y}}\Big) - 2(\boldsymbol{\epsilon}')^\top\big(\boldsymbol{\Phi} f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})\big) + 2(\boldsymbol{\epsilon}')^\top \boldsymbol{y} - (\boldsymbol{\epsilon}')^\top \boldsymbol{\epsilon}'\Big\} \\
=&\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big\{2\sigma^2 \text{tr}\Big(\boldsymbol{\Phi}^{\text{H}} \frac{\partial f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})}{\partial \boldsymbol{y}}\Big) - 2(\boldsymbol{\epsilon}')^\top\big(\boldsymbol{\Phi} f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})\big)\Big\} - M\sigma^2.
\end{aligned} \tag{4}$$

Thus, we only need to prove that

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\sigma^2 \text{tr}\Big(\boldsymbol{\Phi}^{\text{H}} \frac{\partial f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})}{\partial \boldsymbol{y}}\Big) = \mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}(\boldsymbol{\epsilon}')^\top\big(\boldsymbol{\Phi} f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})\big). \tag{5}$$

For ease of notation, we denote $\boldsymbol{g}(\boldsymbol{y}+\boldsymbol{\epsilon}') = \boldsymbol{\Phi} f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})$, and we have

$$\text{div}_{\boldsymbol{\epsilon}'}\boldsymbol{g} = \text{div}_{\boldsymbol{y}}\boldsymbol{g} = \text{tr}\Big(\boldsymbol{\Phi}^{\text{H}} \frac{\partial f_{\boldsymbol{\Phi}}(\boldsymbol{y}+\boldsymbol{\epsilon}';\boldsymbol{\omega})}{\partial \boldsymbol{y}}\Big).$$

Then we can rewrite (5) as

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\sigma^2 \text{div}_{\boldsymbol{\epsilon}'}\boldsymbol{g} = \mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}(\boldsymbol{\epsilon}')^\top \boldsymbol{g}. \tag{6}$$

It is enough to prove that

$$\mathbb{E}_{\boldsymbol{\epsilon}'_i}\sigma^2 \nabla_{\boldsymbol{\epsilon}'_i}\boldsymbol{g}_i = \mathbb{E}_{\boldsymbol{\epsilon}'_i}\boldsymbol{\epsilon}'_i \boldsymbol{g}_i,\ \forall i \in \{1, 2, \ldots, M\}. \tag{7}$$

Let $\psi_\sigma(\cdot) : \mathbb{R} \to \mathbb{R}$ denote the probability distribution function of univariate normal distribution of variance σ^2. It is known that

$$\nabla \psi_\sigma(x) = -\frac{1}{\sigma^2} x \psi_\sigma(x). \tag{8}$$

By integration by parts, we can obtain

$$\mathbb{E}_{\boldsymbol{\epsilon}_i'}\sigma^2\nabla_{\boldsymbol{\epsilon}_i'}\boldsymbol{g}_i = \int \sigma^2\nabla_{\boldsymbol{\epsilon}_i'}\boldsymbol{g}_i\psi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' = \sigma^2\boldsymbol{g}_i\psi_\sigma(\boldsymbol{\epsilon}_i')|_{-\infty}^{+\infty} - \int \sigma^2\boldsymbol{g}_i\nabla\psi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i'$$
$$= \sigma^2\boldsymbol{g}_i\psi_\sigma(\boldsymbol{\epsilon}_i')|_{-\infty}^{+\infty} + \int \boldsymbol{g}_i\boldsymbol{\epsilon}_i'\psi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' = \int \boldsymbol{g}_i\boldsymbol{\epsilon}_i'\psi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i = \mathbb{E}_{\boldsymbol{\epsilon}_i'}\boldsymbol{g}_i\boldsymbol{\epsilon}_i'. \tag{9}$$

Note that $\sigma^2\boldsymbol{g}_i\psi_\sigma(\boldsymbol{\epsilon}_i')|_{-\infty}^{+\infty} = 0$, as the exponential decay of ψ_σ is faster than the polynomial growth of $\boldsymbol{g}_i$. The proof is done.

5 Visual Comparison on More Samples

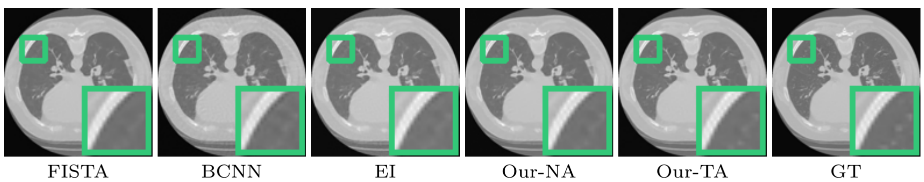

Fig. 1. Results of CT reconstruction.

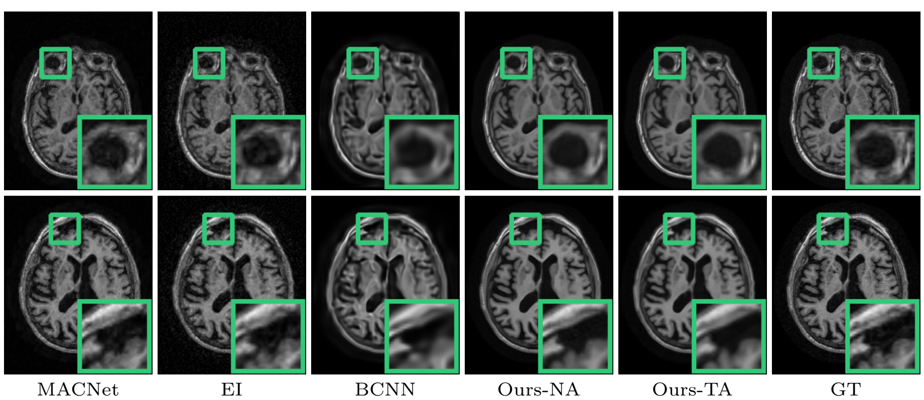

Fig. 2. Results of noisy MRI reconstruction with the radial mask of CS ratio 25%.

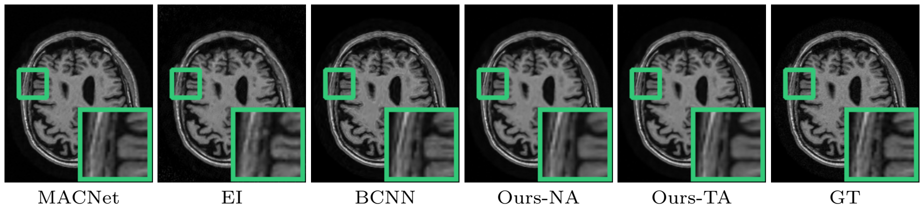

Fig. 3. Results of noiseless MRI reconstruction with the radial mask of CS ratio 25%.

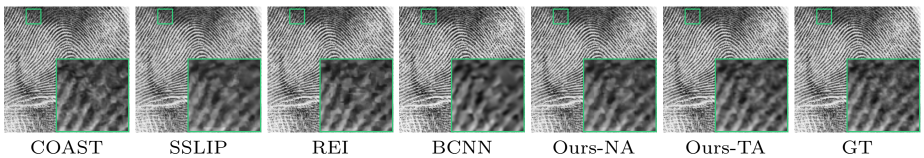

Fig. 4. Results of noisy NIR from Gaussian measurements of CS ratio 40%.

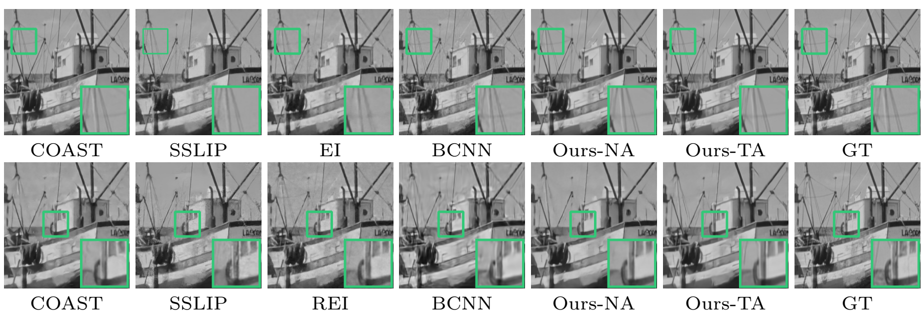

Fig. 5. Results of NIR from Gaussian measurements of CS ratio 25%. The upper row is for the noiseless setting and the bottom row for the noisy setting.

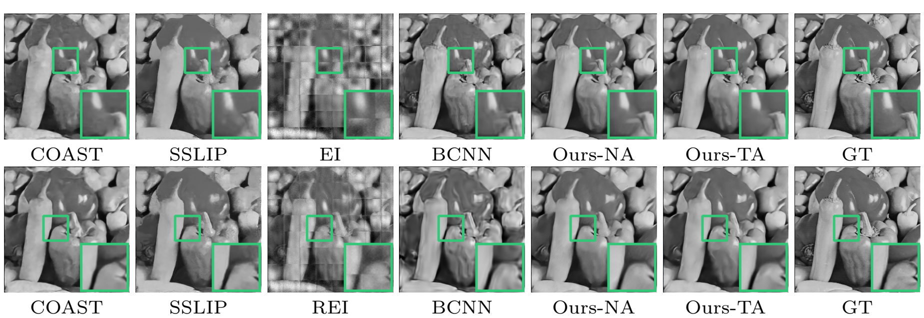

Fig. 6. Results of NIR from Gaussian measurements of CS ratio 10%. The upper row is for the noiseless setting and the bottom row for the noisy setting.