Supplementary Material for
Robust Equivariant Imaging: a fully unsupervised framework for learning to image from noisy and partial measurements

Dongdong Chen*
School of Engineering
University of Edinburgh
d.chen@ed.ac.uk

Julián Tachella*
School of Engineering
University of Edinburgh
julian.tachella@ed.ac.uk

Mike E. Davies
School of Engineering
University of Edinburgh
mike.davies@ed.ac.uk

1. SURE-based losses

We derive the SURE loss presented in the main paper for the Gaussian, Poisson and Poisson-Gaussian noise models. We follow the derivation in [2]. In all cases, the goal is to obtain an unbiased estimator of the supervised mean squared error (MSE) of the clean measurement u from the noisy measurement y:

$$\sum_{i=1}^{N} \frac{1}{m} \|u_i - h_\theta(y_i)\|^2 \tag{1}$$

with denoiser $h_\theta : \mathbb{R}^m \mapsto \mathbb{R}^m$ defined as

$$h_\theta = A \circ f_\theta \tag{2}$$

where $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ denotes the forward operator and $f_\theta : \mathbb{R}^m \mapsto \mathbb{R}^n$ is the (trainable) reconstruction network. The expectation of (1) with respect to the pairs (y, u) can be decomposed as

$$\mathbb{E}_{y,u}\{\sum_{i=1}^{N} \frac{1}{m} \|u_i - h_\theta(y_i)\|^2\}$$
$$= \mathbb{E}_u \sum_{i=1}^{N} \frac{1}{m} \mathbb{E}_{y|u} \|u_i - h_\theta(y_i)\|^2. \tag{3}$$

The inner expectation can be further decomposed as

$$\mathbb{E}_{y|u} \|u_i - h_\theta(y_i)\|^2$$
$$= \mathbb{E}_{y|u}\{\|u_i\|^2 + \|h_\theta(y_i)\|^2 - 2u_i^\top h_\theta(y_i)\}$$
$$= \mathbb{E}_{y|u}\{u_i^\top y_i\} + \mathbb{E}_{y|u}\{\|h_\theta(y_i)\|^2\} - 2\mathbb{E}_{y|u}\{u_i^\top h_\theta(y_i)\}$$

where we used that $\mathbb{E}_{y|u}\{y_i\} = u_i$ for all noise models. An unbiased estimator of the second term is simply $\|h_\theta(y_i)\|^2$, which does not require clean measurements u. The terms

$$\mathbb{E}_{y|u}\{u_i^\top h_\theta(y_i)\} \tag{4}$$

and

$$\mathbb{E}_{y|u}\{u_i^\top y_i\} \tag{5}$$

depend on u and require a noise-dependent analysis, which is presented in the following subsections.

1.1. Gaussian

We begin with the Gaussian noise model. In this case, we have $y \sim \mathcal{N}(u, \sigma^2 I)$ where I is an $m \times m$ identity matrix. The terms (4) and (5) can be computed in an unsupervised way (without clean u) using the following lemma:

***Lemma* 1 (Lemma 2 in [5])** *Let* $y \in \mathbb{R}^m$ *such that* $y \sim \mathcal{N}(u, \sigma^2 I)$ *be a random variable and let* $\phi : \mathbb{R}^m \mapsto \mathbb{R}^m$ *be a weakly differentiable function such that* $\mathbb{E}_{y|u}\{|\delta\phi_j(y)/\delta y_j|\} < \infty$ *for all j and input y. Then,*

$$\mathbb{E}_{y|u}\{u^\top \phi(y)\} = \mathbb{E}_{y|u}\{y^\top \phi(y) - \sigma^2 \nabla \cdot \phi(y)\} \tag{6}$$

where $\nabla \cdot \phi(y) = \sum_j \frac{\delta\phi_j(y)}{\delta y_j}$ *denotes the divergence of ϕ.*

Applying Lemma 1 to (4) with $\phi = h_\theta$ and to (5) with $\phi = \text{identity}$, we get the following unbiased estimator of the MSE:

$$\sum_{i=1}^{N} \frac{1}{m}\{\|y_i\|^2 - \sigma^2 m + \|h_\theta(y_i)\|^2 - 2y^\top h_\theta(y) +$$
$$+ 2\sigma^2 \nabla \cdot h_\theta(y_i)\}$$

$$= \sum_{i=1}^{N} \frac{1}{m}\|y_i - h(y_i)\|^2 - \sigma^2 + \frac{2\sigma^2}{m}1^\top \delta h_\theta(y_i)$$

where $\delta h_\theta(y) = [\frac{\delta h_\theta(y)}{\delta y_1}, \dots, \frac{\delta h_\theta(y)}{\delta y_m}]^\top$ is the gradient of h_θ with respect to y.

Using a Monte Carlo approximation of the last term (c.f., Theorem 2 of the main paper), and decomposing $h_\theta = A \circ f_\theta$, we obtain the unsupervised loss used in the main paper:

$$\mathcal{L}_{\text{SURE}}(\theta) = \sum_{i=1}^{N} \frac{1}{m} \|y_i - A(f_\theta(y_i))\|_2^2 - \sigma^2 \qquad (7)$$
$$+ \frac{2\sigma^2}{m\tau} b_i^\top \left(A(f_\theta(y_i + \tau b_i)) - A(f_\theta(y_i)) \right)$$

where $b_i \sim \mathcal{N}(0, I)$ and τ is a small positive number.

1.2. Poisson

In the Poisson noise case, the noisy measurements are modeled as $y = \gamma z$ with $z \sim \text{Poisson}(\frac{u}{\gamma})$. The following lemma provides unsupervised expressions for (4) and (5).

***Lemma* 2 (Lemma 1.2 in [2])** *Let $z \in \mathbb{R}^m$ such that $z \sim$ Poisson(u) be a random variable and let $\phi : \mathbb{R}^m \mapsto \mathbb{R}^m$ be a function such that $\mathbb{E}_{z|u}\{|\phi_j(z)|\} < \infty$ for all j.*

$$\mathbb{E}_{z|u}\{u^\top \phi(z)\} = \mathbb{E}_{z|u}\{z^\top \phi^{[-1]}(z)\} \qquad (8)$$

where the jth entry of the vector $\phi^{[-\alpha]}(z)$ is given by $\phi_j(y - \alpha e_j)$ and e_j is the jth canonical vector.

Let $\phi(z) = h_\theta(y) = h_\theta(\gamma z)$, then term (4) is given by

$$\mathbb{E}_{y|u}\{u_i^\top h_\theta(y_i)\} = \gamma \mathbb{E}_{z|u}\{(\frac{u_i}{\gamma})^\top \phi(z_i)\} \qquad (9)$$
$$= \gamma \mathbb{E}_{z|u}\{z_i^\top \phi^{[-1]}(z_i)\} \qquad (10)$$
$$= \mathbb{E}_{y|u}\{y_i^\top h_\theta^{[-\gamma]}(y_i)\} \qquad (11)$$

and an unbiased estimator is simply

$$y_i^\top h_\theta^{[-\gamma]}(y_i) \qquad (12)$$

Now let $\phi(z) = y = \gamma z$ be the identity function, term (5) is given by

$$\mathbb{E}_{y|u}\{u_i^\top y_i\} = \mathbb{E}_{y|u}\{y_i^\top (y_i - 1\gamma)\} \qquad (13)$$

where 1 denotes a vector of m ones. An unbiased estimator is given by

$$y_i^\top (y_i - 1\gamma). \qquad (14)$$

Thus, an unbiased estimator of the MSE is given by

$$\sum_{i=1}^{N} \frac{1}{m} \{\|y_i\|^2 - \gamma 1^\top y_i + \|h_\theta(y_i)\|^2 - 2y_i^\top h_\theta^{[-\gamma]}(y_i)\}$$
$$\approx \sum_{i=1}^{N} \frac{1}{m} \{\|y_i\|^2 - \gamma 1^\top y_i + \|h_\theta(y_i)\|^2 + 2\gamma y_i^\top \delta h_\theta(y_i)\}$$
$$= \sum_{i=1}^{N} \frac{1}{m} \|y_i - h_\theta(y_i)\|^2 - \frac{\gamma}{m} 1^\top y_i + \frac{2\gamma}{m} y_i^\top \delta h_\theta(y_i)$$

where we used a Taylor expansion to approximate $h_\theta^{[-\gamma]}(y) \approx h_\theta(y) - \gamma \delta h_\theta(y)$ [2]. Using a Monte Carlo estimate of the last term [2] (similar to Theorem 2 in the main paper) and decomposing $h_\theta = A \circ f_\theta$, we get the unsupervised loss of the main paper

$$\mathcal{L}_{\text{SURE}}(\theta) = \sum_{i=1}^{N} \frac{1}{m} \|y_i - A(f_\theta(y_i))\|_2^2 - \frac{\gamma}{m} 1^\top y_i \qquad (15)$$
$$+ \frac{2\gamma}{m\tau} (b_i \odot y_i)^\top \left(A(f_\theta(y_i + \tau b_i)) - A(f_\theta(y_i)) \right)$$

where b is an i.i.d. random vector following a Bernoulli distribution.

1.3. Poisson-Gaussian

Noisy measurements under a Poisson-Gaussian noise model are defined by

$$y = \gamma z + \epsilon \quad \text{with} \quad \begin{cases} u = A(x) \\ z \sim \text{Poisson}\left(\frac{u}{\gamma}\right) \\ \epsilon \sim \mathcal{N}(0, \sigma^2 I) \end{cases} \qquad (16)$$

where $\gamma > 0$ controls the Poisson noise contribution and $\sigma > 0$ controls the Gaussian noise contribution. An unsupervised equivalent of (4) and (5) can be obtained using Lemmas 1 and 2. Let $\phi_\epsilon(z) = h_\theta(y) = h_\theta(\gamma z + \epsilon)$, then (4) is given by

$$\mathbb{E}_{y|u}\{u_i^\top h_\theta(y_i)\} = \gamma \mathbb{E}_{\epsilon|u}\mathbb{E}_{z|u}\{(\frac{u_i}{\gamma})^\top \phi_\epsilon(z_i)\} \qquad (17)$$
$$= \gamma \mathbb{E}_{y|u}\{z_i^\top \phi_\epsilon^{[-1]}(z_i)\} \qquad (18)$$
$$= \gamma \mathbb{E}_{y|u}\{\frac{y_i - \epsilon}{\gamma}^\top h_\theta^{[-\gamma]}(y_i)\} \qquad (19)$$
$$= \mathbb{E}_{y|u}\{y_i^\top h_\theta^{[-\gamma]}(y_i)\} - \mathbb{E}_{y|u}\{\epsilon_i^\top h_\theta^{[-\gamma]}(y_i)\} \qquad (20)$$
$$= \mathbb{E}_{y|u}\{y_i^\top h_\theta^{[-\gamma]}(y_i) - \sigma^2 \nabla \cdot h_\theta^{[-\gamma]}(y_i)\} \qquad (21)$$

where we used Lemma 2 in (18) and Lemma 1 in (21). Thus,

$$y_i^\top h_\theta^{[-\gamma]}(y_i) - \sigma^2 \nabla \cdot h_\theta^{[-\gamma]}(y_i) \qquad (22)$$

is an unbiased estimator of (4). Similarly, setting $\phi_\epsilon(z) = y = \gamma z + \epsilon$, (5) is equal to

$$\mathbb{E}_{y|u}\{u_i^\top y_i\} = \mathbb{E}_{y|u}\{y_i^\top (y_i - 1\gamma) - \sigma^2 m\} \qquad (23)$$

and

$$y_i^\top (y_i - 1\gamma) - \sigma^2 m \qquad (24)$$

is an unbiased estimator of (5). Thus, an unbiased estimator of the MSE is given by

$$\sum_{i=1}^{N} \frac{1}{m} \{\|y_i\|^2 - \gamma 1^\top y_i - \sigma^2 m + \|h_\theta(y_i)\|^2 -$$
$$- 2y_i^\top h_\theta^{[-\gamma]}(y_i) + 2\sigma^2 \nabla \cdot h_\theta^{[-\gamma]}(y_i)\}$$

$$\approx \sum_{i=1}^{N} \frac{1}{m} \|y_i - h_\theta(y_i)\|^2 - \frac{\gamma}{m} \mathbf{1}^\top y_i - \sigma^2 +$$
$$+ \frac{2}{m}(\gamma y_i^\top + \sigma^2 \mathbf{1}^\top)\delta h_\theta(y) - \frac{2\sigma^2 \gamma}{m} \mathbf{1}^\top \delta^2 h_\theta(y)$$

where $\delta^2 h_\theta(y) = [\frac{\delta^2 h_\theta(y)}{\delta y_1^2}, \ldots, \frac{\delta^2 h_\theta(y)}{\delta y_m^2}]^\top$ is a vector of second derivatives of h_θ with respect to y. Using a Monte Carlo approximation of the last two terms [2] (similar to Theorem 2 in the main paper) and decomposing $h_\theta = A \circ f_\theta$, we obtain the unsupervised loss in the main paper:

$$\mathcal{L}_{\text{SURE}}(\theta) = \sum_{i=1}^{N} \frac{1}{m} \|y_i - A(f_\theta(y_i))\|_2^2 - \frac{\gamma}{m} \mathbf{1}^\top y_i - \sigma^2$$
$$+ \frac{2}{m\tau}(b_i \odot (\gamma y_i + \sigma^2 I))^\top (A(f_\theta(y_i + \tau b_i)) - A(f_\theta(y_i)))$$
$$+ \frac{2\gamma\sigma^2}{m\tau} c_i^\top (A(f_\theta(y_i + \tau c_i)) + A(f_\theta(y_i - \tau c_i)) \ldots$$
$$- 2A(f_\theta(y_i)))$$

$$(25)$$

where $b_i \sim \mathcal{N}(0, I)$, c_i are i.i.d. random variables that follow a Bernoulli distribution.

2. Training Details

We first provide the details of the network architectures and hyperparameters used in simulations for Figures 1, 4-7 and Tables 1-3 of the main paper. We implemented the algorithms and operators (e.g., `radon` and `iradon`) in Python with PyTorch 1.6 and trained the models on NVIDIA 1080ti and 2080ti GPUs. We used the network architecture in [1] for defining G_θ. Figure 1 illustrates the architecture of the residual U-Net [4] used in our paper. The training details for each task are as follows:

Accelerated MRI. For the $4\times$ accelerated MRI task, we used Adam with a batch size of 2 and an initial learning rate of 5×10^{-4}. The weight decay is 10^{-8}. We trained the networks for 500 epochs, keeping the learning rate constant for the first 300 epochs and then shrinking it by a factor of 0.1. We set $\alpha = 1$ and $\tau = 0.01$.

Inpainting. For the inpainting task, we also used Adam but with a batch size of 1 and an initial learning rate of 10^{-4}. The weight decay is 10^{-8}. We trained the networks for 500 epochs, shrinking the learning rate by a factor of 0.1 every 100 epochs. We set $\alpha = 1$ and $\tau = 0.01$.

Sparse-view CT. For the 50-views CT task, we used the Adam optimizer with a batch size of 2 and an initial learning rate of 5×10^{-4}. The weight decay is 10^{-8}. We trained the networks for 3000 epochs, shrinking the learning rate by a
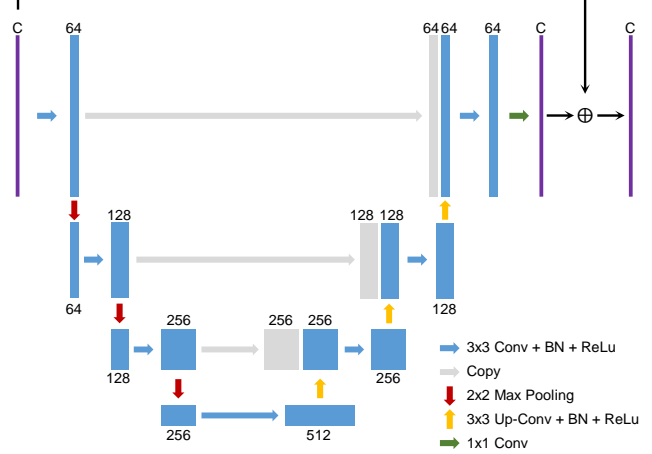

Figure 1. The residual U-Net [4] used in the paper. The number of input and output channels is denoted as C, with $C = 1, 2, 3$ in the CT, MRI and inpainting task, respectively.
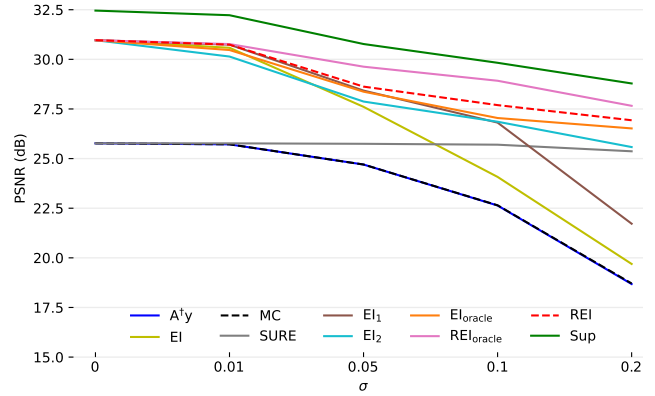

Figure 2. Reconstruction performance (PSNR) as a function of noise level σ for different training losses on the MRI image reconstruction task with $4\times$ compression and noisy k-space measurements.

factor of 0.1 every 1000 epochs. In particular, because the CT model is nonlinear and involves an exponential mapping such that the consistency loss for y is very large, we scale the $\mathcal{L}_{\text{SURE}}$ (in Equation (14) of main paper) accordingly by a factor of 10^{-5}. We set $\alpha = 1000$ and $\tau = 10$.

3. More results

3.1. Ablation study on training loss

We first compared different variants of loss function for training the REI models. In particular, we consider the following 9 training loss variants:

- **MC**: $\mathcal{L}_{\text{MC}}(\theta)$

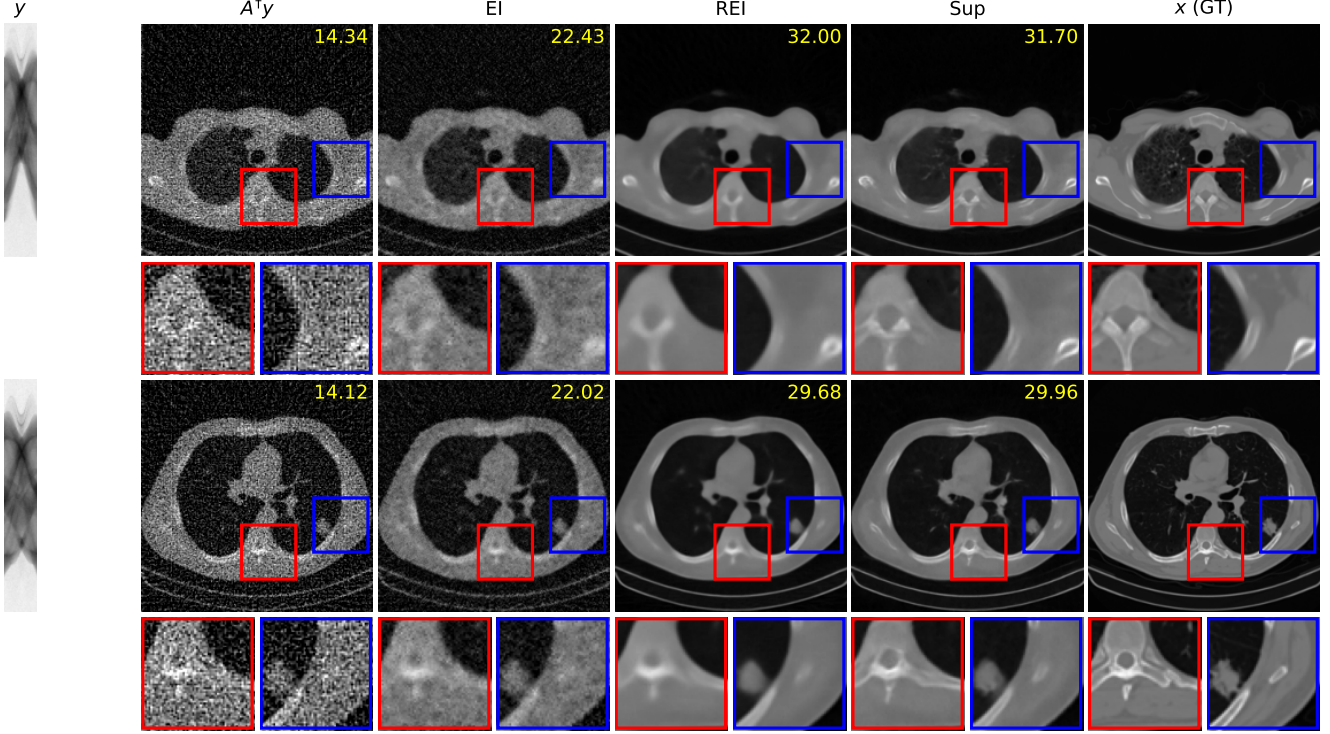- **SURE**: $\mathcal{L}_{\text{SURE}}(\theta)$

3

Figure 3. *Low-dose* CT image reconstruction (50 views) on the test observations with mixed Poisson-Gaussian noise, $I_0 = 10^4$, $\sigma = 50$, $\gamma = 1$. PSNR values are shown in the top right corner of the images.

- **EI**: $\mathcal{L}_{\text{MC}}(\theta) + \alpha\mathcal{L}_{\text{EQ}}(\theta)$

- **EI$_1$**: $\mathcal{L}_{\text{MC}}(\theta) + \alpha\mathcal{L}_{\text{REQ}}(\theta)$

- **EI$_2$**: $\mathcal{L}_{\text{SURE}}(\theta) + \alpha\mathcal{L}_{\text{EQ}}(\theta)$

- **EI$_{\text{oracle}}$**: $\sum_{i=1}^{N} \frac{1}{m}\|u_i - Af_\theta(y_i)\|^2 + \alpha\mathcal{L}_{\text{EQ}}(\theta)$

- **REI$_{\text{oracle}}$**: $\sum_{i=1}^{N} \frac{1}{m}\|u_i - Af_\theta(y_i)\|^2 + \alpha\mathcal{L}_{\text{REQ}}(\theta)$

- **REI**: $\mathcal{L}_{\text{SURE}}(\theta) + \alpha\mathcal{L}_{\text{REQ}}(\theta)$

- **Sup**: $\sum_{i=1}^{N} \frac{1}{n}\|x_i - f_\theta(y_i)\|^2$

Figure 2 shows the reconstruction performance (PSNR) as a function of noise level σ for different variants of training loss on the $4\times$ accelerated MRI image reconstruction task. Recall the oracle losses include additional oracle access to the clean measurements u_i. This provides us with a way to study the contributions of the equivariant and SURE losses.

From the figure we can observe that: (i) MC fails to learn anything except to converge to the linear reconstruction $A^\dagger y$, while SURE achieves a stable estimation to the clean measurement consistency. (ii) REI outperforms EI$_1$ and EI$_2$ which demonstrates the effectiveness of the proposed training loss (Equation (14) in the main paper). (iii)

REI performs better than EI$_{\text{oracle}}$ demonstrating the benefits of using a noisy input in our proposed robust Equivariance loss. (iv) Both Sup and REI$_{\text{oracle}}$ outperform REI due to having access to the ground truth clean images and measurements, respectively. However, as noted in the main paper the REI performance lies close to that for REI$_{\text{oracle}}$, suggesting that the SURE loss is doing a reasonable job of estimating the (oracle) clean measurement consistency loss.

3.2. Effect of the equivariance hyperparameter α

We consider the optimal value and sensitivity of the hyperparameter α on the inpainting task. Table 1 shows the REI reconstruction performance for different equivariance strength values (α in Equation (14) of the main paper) and different Poisson noise levels γ. We see that an optimal value here is around $\alpha = 1$, although the performance is generally good over the range $0.1 \leq \alpha \leq 1$ indicating that REI is not sensitive to the precise value of α. However, when α is either too small or too large we do observe deterioration in performance. Consistent observations were also found in the MRI task. Therefore, we set $\alpha = 1$ for inpainting and MRI tasks.

In the nonlinear CT task, since the measurement values are substantially larger than those in the MRI and inpainting experiments we scaled the SURE loss by 10^{-5} and set $\alpha = 1000$. We leave further optimization of α in this task for

future research.

γ	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 1$	$\alpha = 10$	$\alpha = 100$
0.01	19.5 ± 1.1	19.66 ± 1.7	21.0 ± 1.3	16.2 ± 1.6	10.6 ± 1.4
0.05	17.0 ± 1.0	18.1 ± 1.0	18.2 ± 1.2	13.5 ± 1.7	7.4 ± 2.0
0.1	11.2 ± 1.1	16.0 ± 1.0	16.6 ± 1.3	12.3 ± 1.5	8.0 ± 1.6

Table 1. Effect of the equivariance strength α on the reconstruction performance (PSNR) in the inpainting reconstruction (Urban 100 dataset) task with different noise level γ.

3.3. Effect of the small positive number τ

We follow the suggestion of [3] and select the value of τ to be around $\max(y)/1000$. Table 2 presents REI reconstruction performance (PSNR) on the MRI task with different magnitudes of τ where the Gaussian noise level is fixed as $\sigma = 0.1$. It shows that REI works best when $\tau = 10^{-2}$. In the MRI and inpainting experiments, we set $\tau = 10^{-2}$ and set $\tau = 10$ to the CT task to account for the larger values of y.

σ	$\tau = 10^{-1}$	$\tau = 10^{-2}$	$\tau = 10^{-3}$	$\tau = 10^{-4}$
0.1	24.6 ± 2.2	27.7 ± 2.0	26.9 ± 2.3	25.3 ± 2.4

Table 2. Sensitivity of the small positive number τ on the reconstruction performance (PSNR) in the $4\times$ accelerated MRI (fastMRI dataset) task with a fixed noise level $\sigma = 0.1$.

	FBP	EI	REI	REI$_{\text{oracle}}$	Sup
CT	12.9 ± 1.1	21.6 ± 0.7	30.5 ± 1.0	30.5 ± 1.1	30.7 ± 1.0

Table 3. Reconstruction performance (PSNR) of *low-dose* 50-views CT (CT100 dataset) task for different methods on the test noisy measurements. $I_0 = 10^4$ and $\sigma = 50$.

3.4. Low-dose Sparse-view MPG CT

We have shown in the main paper an initial demonstration that REI can learn a competitive CT reconstruction with MPG noisy measurements in the nonlinear CT task. Here we are interested in further testing whether REI can handle the more challenging low-dose (i.e., higher Poisson noise) case. To do that, we set $I_0 = 10^4$ and $\sigma = 50$, Table 3 and Figure 3 show some preliminary results. We have the following observations: (i) both FBP and EI fail to learn the reconstruction due to the high MPG noise in the measurements. (ii) In some examples (e.g., the first row in Figure 3), REI enjoy a slightly higher PSNR than that of Sup, but visually the reconstruction result of Sup is sharper. This may indicate that the Sup network is overfitting and that REI enjoys better generalization on the test measurements due to the equivariance constraint, c.f EI in [1]. (iii) As shown in Table 3, REI achieved comparable performance (PSNR) to both Sup and REI$_{\text{oracle}}$. Moreover, our model outperforms FBP by more than 17 dB and enjoys a 9 dB gain against EI. Both these results and those in our main paper here suggest that robust equivariant imaging is a powerful learning paradigm. We plan to explore the nonlinear CT further in future research, e.g., a wider range of noise levels, hyperparameter optimisation, etc.

References

[1] Dongdong Chen, Julián Tachella, and Mike E Davies. Equivariant imaging: Learning beyond the range space. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. 3, 5

[2] Yoann Le Montagner, Elsa D Angelini, and Jean-Christophe Olivo-Marin. An unbiased risk estimator for image denoising in the presence of mixed Poisson-Gaussian noise. *IEEE Transactions on Image processing*, 23(3):1255–1268, 2014. 1, 2, 3

[3] Christopher A Metzler, Ali Mousavi, Reinhard Heckel, and Richard G Baraniuk. Unsupervised learning with Stein's unbiased risk estimator. *arXiv preprint arXiv:1805.10531*, 2018. 5

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3

[5] Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pages 1135–1151, 1981. 1