Supplementary Material for
Equivariant Imaging: Learning Beyond the Range Space

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. Training Details

We first provide the details of the network architectures and hyperparameters of Figs. 4-6 and Table 1 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. Figure 1 illustrates the architecture of the residual U-Net used [2] in our paper.

For the sparse-view CT task, we used the Adam optimizer with a batch size of 2 and an initial learning rate of 0.0005. The weight decay is 10^{-8}. The distribution strength β is 10^{-8} for EI_{adv}. We trained the networks for 5000 epochs, keeping the learning rate constant for the first 2000 epochs and then shrinking it by a factor of 0.1 every 1000 epochs. More reconstruction examples are presented in Figure 3.

For the inpainting task, we also used Adam but with a batch size of 1 and an initial learning rate of 0.001. The weight decay is 10^{-8}. The distribution strength β is 10^{-8} for EI_{adv}. We trained the networks for 2000 epochs, shrinking the learning rate by a factor of 0.1 every 500 epochs. Figure 2 shows the peak signal-to-noise ratio (PSNR) of the reconstructions on the training and test measurements. Again, the generalization error of EI is smaller than for the supervised model. More reconstruction examples are presented in Figure 4.

α	0	1	10	100	1000
50-views CT	31.01	36.78	36.88	36.94	33.31

α	0	0.1	1	10
Inpainting	5.84	23.42	25.14	22.96

Table 1: Effect of the equivariance hyperparameter α on the reconstruction performance (PSNR) in the 50-views CT reconstruction (CT100 dataset) and image inpainting (Urban100 dataset) tasks.

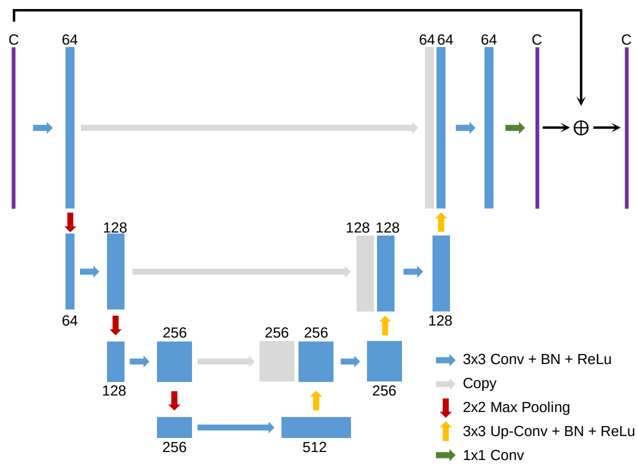

Figure 1: The residual U-Net [2] used in the paper. The number of input and output channels is denoted as C, with $C = 1$ in the CT task and $C = 3$ in the inpainting task.

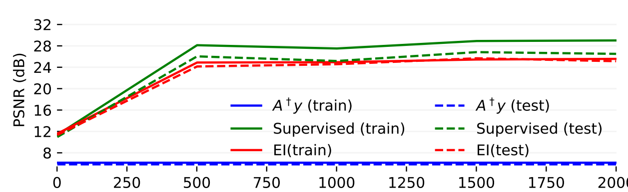

Figure 2: Reconstruction performance (PSNR) as a function of training epoch for the supervised model [1] and our method (no ground truth) on inpainting task measurements for training and testing.

2. More results

Effect of the equivariance hyperparameter α Table 1 shows EI reconstruction performance (PSNR) with different equivariance strength values (α in Eqn. (6) of the main paper). It performs reasonably well when $\alpha = 100$ for the

CT task and $\alpha = 1$ for the inpainting task. When α is too small, the performance drops considerably; at the extreme of no equivariance ($\alpha = 0$), the model fails to learn. These results support our motivation of equivariant imaging.

Effect of the networks' inductive bias In the deep image prior (DIP) paper, the authors showed that some specific convolutional networks can be trained to fit a single image by only enforcing measurement consistency [3]. The DIP approach relies heavily on the choice of the network architecture (generally an autoencoder), and does not work with various popular architectures (*e.g.* those with skip-connections). Moreover, this approach is constrained to a single image and cannot incorporate additional training data.

In contrast, we show that our method can learn beyond the range space without heavily relying on the inductive bias of an specific autoencoder architecture. Moreover, we show that EI outperforms the best DIP architecture as it leverages the full compressed training dataset. We compare our method with the DIP on the 50-views CT image reconstruction task. For our method, we use the same residual U-Net as in the other experiments. We build the DIP using two architectures: the same residual U-Net used in EI (which we denote DIP-1) and the best autoencoder network suggested in [3] (which we denote DIP-2). Following [3], we input iid Gaussian noise to both DIP-1 (1 channel) and DIP-2 (32 channels). Our model is trained using the hyperparameters for sparse-view CT (see Section 1). We train DIP-1 and DIP-2 using 5000 training iterations and a learning rate of 0.001. As shown in Figure 5, our method outperforms the DIP methods. DIP-2 performs significantly better than DIP-1 due to the inductive bias of that autoencoder architecture. In contrast, our method works very well even with the residual U-Net. Moreover, our model also outperforms DIP-2 by 5 dB.

Equivariant imaging using a single training image We are interested in whether the proposed method works for single image reconstruction, *i.e.* reconstructing a single compressed measurement. Here we provide some preliminary results. As an example, we compared our method with the DIP on the inpainting task for single image reconstruction. We trained all 3 models (EI, DIP-1, DIP-2) using 5000 training iterations and a learning rate of 0.001 on a single measurement input. The results are presented in Figure 6. We observe that our method works very well for this single image reconstruction task and outperforms both DIP-1 and DIP-2. In addition, DIP-1 performs worse than DIP-2 due to the residual architecture with skip-connections. Again, our model is not so dependent on the inductive bias of network and works well when using the residual connections. We note that although our method is able to learn with a single measurement, the role of equivariance in this scenario needs to be explored more, and we leave this for future work.

References

[1] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017. 1, 3, 4

[2] 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. 1

[3] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9446–9454, 2018. 2, 4, 5

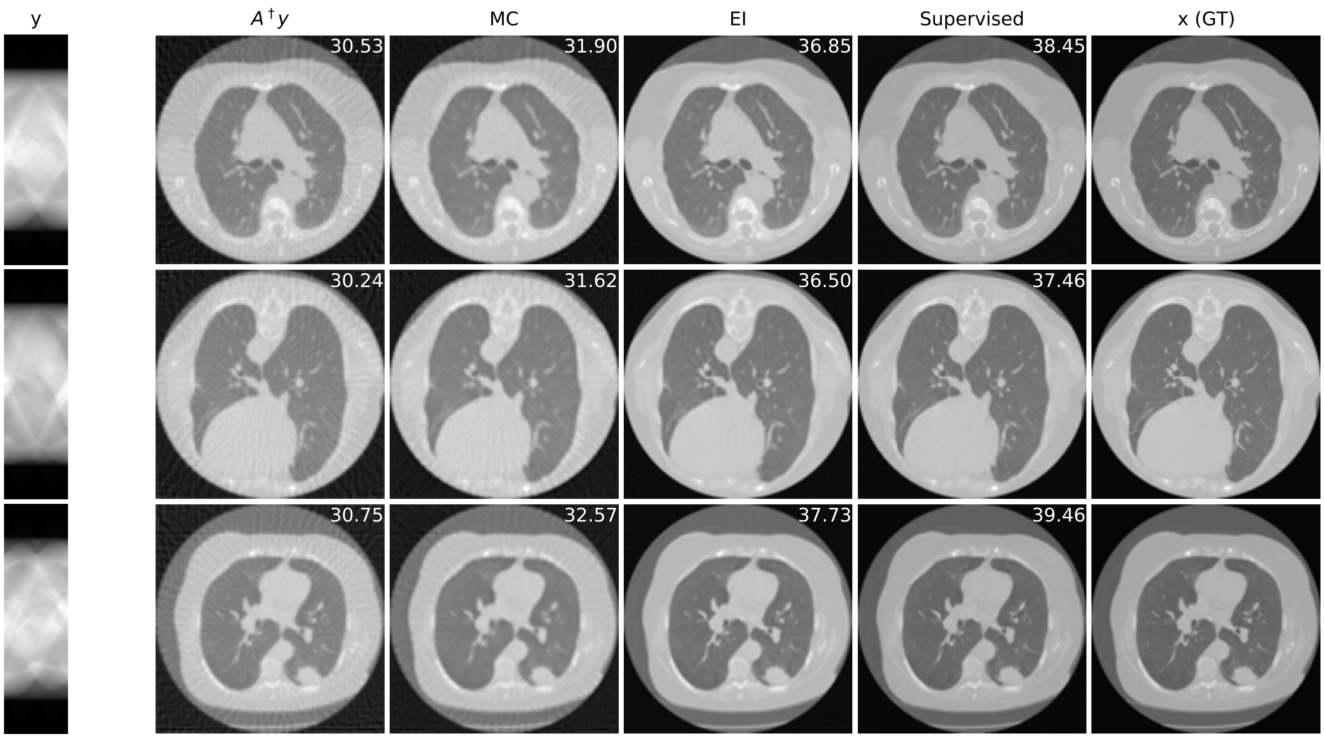

Figure 3: More examples of sparse-view CT image reconstruction on the unseen test measurements. We train the supervised model (FBPConvNet [1]) with measurement/ground truth pairs while we train the equivariance learned model with measurements alone. We adopt *random rotations* as the transformation T for our equivariance learning. We obtained results comparable to supervised learning in artifacts-removal. Corresponding PSNR are shown in images.

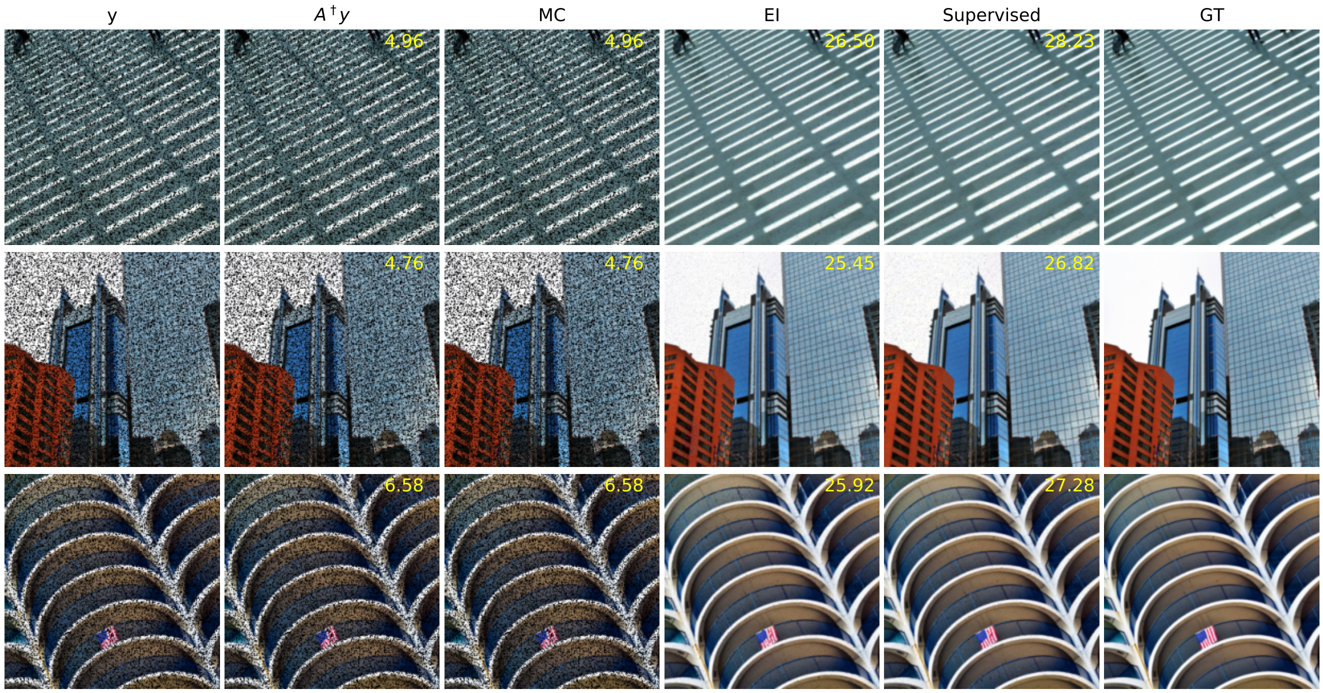

Figure 4: More examples of image inpainting reconstruction on the unseen test measurements. We train the supervised model [1] with measurement/ground truth pairs while we train the equivariance learned model with measurements alone. We adopt *random shifts* as the transformation T for our equivariance learning. We obtained results comparable to supervised learning in recovering missing pixels. Corresponding PSNR are shown in images.

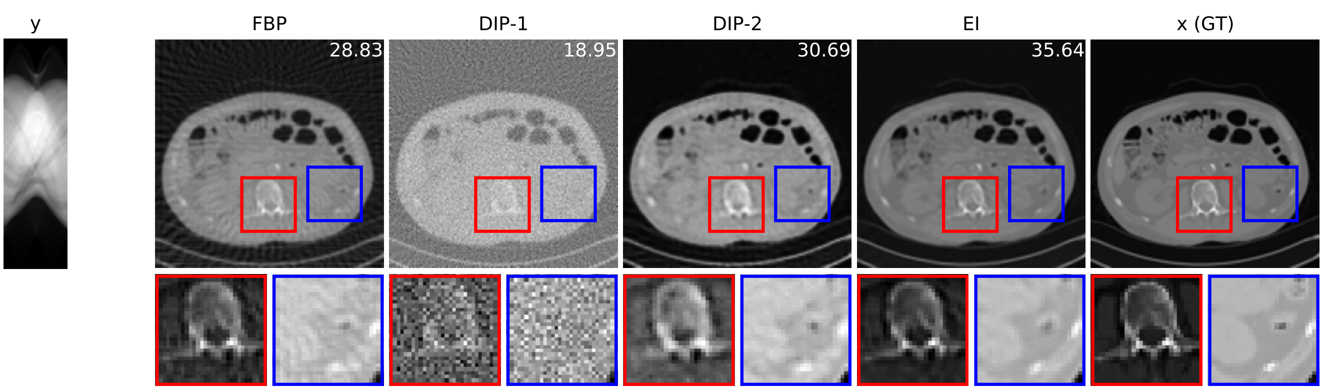

Figure 5: Comparison between EI and DIP on 50-views CT reconstruction. We denote DIP-1 and DIP-2 as the DIP learned models trained with residual U-Net (same as EI) and Encoder-Decoder (the best architecture for DIP as suggested in [3]), respectively. We trained EI on a measurement set and direct apply the trained model on the given new measurement here. Both DIP methods are trained using the given measurement here. Corresponding PSNR are shown in images.

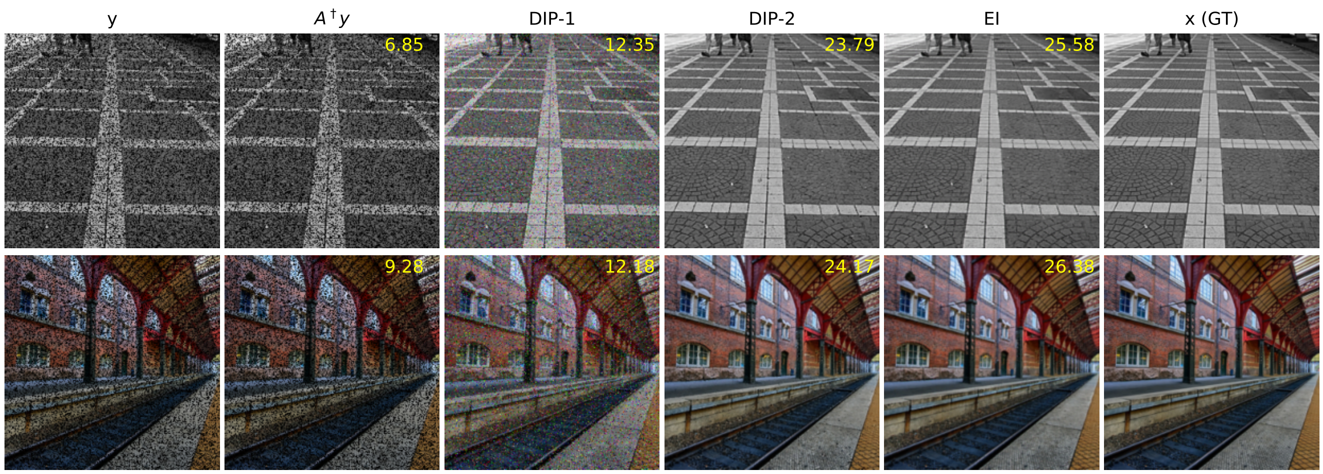

Figure 6: Comparison between EI and DIP for single image reconstruction on the inpainting task. We denote DIP-1 and DIP-2 as the DIP learned models trained with residual U-Net (same as EI) and Encoder-Decoder (the best architecture for DIP as suggested in [3]), respectively. All the models are trained with the given single compressed measurement data y. Corresponding PSNR are shown in images.