Noise-Tolerant Self-Supervised Inversion

Hirofumi Kobayashi [*1] **Ahmet Can Solak** [*1] **Joshua Batson** [*1] **Loic Royer** [*1]

Abstract

We propose a general framework for solving inverse problems in the presence of noise that requires no prior on the signal, no estimate of the noise, and no clean training data. The only assumption is that the forward model is available and that the noise exhibits statistical independence across different dimensions of the measurement. We build upon the theory of "$\mathcal{J}$-invariant" functions (Batson & Royer, 2019) and show how self-supervised denoising *à la* Noise2Self is a special case of learning a noise-tolerant pseudo-inverse of the identity. Given any forward model g we give a theoretical framework on how to learn a noise-robust pseudo-inverse f. We demonstrate our approach by showing how a convolutional neural network can be taught in a self-supervised manner to deconvolve images.

1. Introduction

2. Related Work

Inverse problems in imaging have been investigated for a long time. A classical approach is total variation minimisation. Several algorithms have been proposed to efficiently solve the total variation minimization problem (Chambolle & Pock, 2011) [@hiro: please check Chinmay;s review and other refs if needed]. In recent years, deep convolutional neural networks (CNNs) have shown promising potential to solve inverse problems in various imaging applications including denoising (Zhang et al., 2017), deconvolution (Xu et al., 2014), compressive sensing (Mousavi & Baraniuk, 2017) and super-resolution (Dong et al., 2014) [can add more if needed]. It is even shown that a single CNN model can solve all of these linear inverse problems (Rick Chang et al., 2017). However, these methods are all based on supervised learning, leading to a need for paired training data and ground truth.

[@hiro: please first discuss self-supervision in general, then self-supervision for denosing, and then for inversion] More recently, self-supervised learning methods[] have shown equivalent or even better performance in solving inverse problems than supervised learning. By incorporating denoiser-approximate message passing algorithm and Stein's unbiased risk estimator, a CNN model can achieve compressive sensing recovery and denoising altogether without using ground truth (Zhussip et al., 2019). Alternatively, by leveraging the pixel-wise stochasticity and independence of noise, a CNN model can denoise images that need linear transformation (e.g. X-ray CT) (Hendriksen et al., 2020). Another approach in self-supervised learning is to use adversarial training. A generative adversarial network (GAN) that is only trained on corrupted training data can generate clean images (Pajot et al., 2018). A more recent work shows that by combining individual GAN model trained on blurred, noisy or compressed images can generate images free of blur, noise and compression artifacts (Kaneko & Harada, 2020). Since these approach use generative models, they cannot solve the inversion problem on a given input image.

3. Theory

Problem statement. Consider a measurement of a system with forward model g and stochastic noise n. We desire to recover the unknown state x from the observation $y = n \circ g(x)$. In the case where there is no noise, i.e., n is the identity function, this reduces to finding a (pseudo)-inverse for g. In the case where g is the identity, this reduces to finding a denoising function for the noise distribution n. One solving strategy is optimization-based, where a prior on x manifests as a regularizer W, and one seeks to minimize a total loss $\|g(x) - y\|^2 + W(x)$. This requires one to solve an optimisation problem for each observation, and also requires an arbitrary choice of the strength and class of the prior W. One might want to learn a noise-tolerant pseudo-inverse of g, but in the absence of training data y it is not clear how. In particular, if one naively optimizes a self-consistency loss $\|g(f(y)) - y\|^2$, then f may learn to invert g while leaving in the effects of the noise n, producing a noisy reconstruction. For example, if g represents the blurring induced by a microscope objective (convolution with the point-spread-function), then setting f to be the corresponding sharpening filter (convolution with the the Fourier-domain reciprocal of g) will greatly amplify the noise in y while producing a self-consistency loss of 0. We propose a modification of this loss which rewards both noise suppression and inversion.

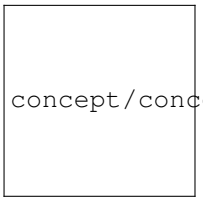

Figure 1. Noise-Tolerant Self-Supervised Inversion. [Concept figure describing the theory]

Proposal. We extend the $\mathcal{J}$-invariance framework introduced in (Batson & Royer, 2019) for denoising in cases where the noise is statistically independent across different dimensions of the measurement. Recall that a function $f : \mathbb{R}^m \to \mathbb{R}^m$ is $\mathcal{J}$-invariant with respect to a partition $\mathcal{J} = \{J_1, \ldots, J_r\}$ of $\{1, \ldots, n\}$ if for any $J \in \mathcal{J}$, the value of $f(x)_J$ does not depend on the value of x_J [1]. If a noise function n is independent across the partition, i.e., $n(x)_J$ and $n(x)_{J^c}$ are independent conditional on x, and $\mathbb{E}[n(x)] = x$, then we show in (Batson & Royer, 2019) that following holds for any $\mathcal{J}$-invariant f:

$$\mathbb{E}\left\|f(n(x)) - n(x)\right\|^2 = \mathbb{E}\left\|f(n(x)) - x\right\|^2 + \mathbb{E}\left\|x - n(x)\right\|^2. \quad (1)$$

That is, the self-supervised loss, given by the difference between the noisy data and the denoised data is equal to the ground-truth loss, up to a constant independent of the denoiser f. Now let's consider the case where the noise is applied after a known forward model g such that: $y = n \circ g(x)$. We are now interested in the following generalised self-supervised loss:

$$\mathbb{E}\left\|g(f(y))) - y\right\|^2 \quad (2)$$

However, in order to decompose Eq. 2 in the same way as done with Eq. 1 we would need to $g \circ f$ to be $\mathcal{J}$-invariant. Unfortunately, in the general case, it is difficult to specify properties of f that would guarantee the $\mathcal{J}$-invariance of $g \circ f$. This makes the strategy of explicit $\mathcal{J}$-invariance pursued in (Laine et al., 2019) difficult. However, a simple masking procedure can turn any function into a $\mathcal{J}$-invariant function, which will allow us to take advantage of $\mathcal{J}$-invariance when computing a training loss, even if the final function we use at prediction time is not $\mathcal{J}$-invariant.

Given the partition $\mathcal{J}$, we choose some family of masking functions m_J. For example, m_J could replace coordinates in J with zeros, by random values, or by some interpolation of coordinates outside of J. Then, for any function f and our fixed forward model g, we then compute the following loss:

$$\mathbb{E}\sum_J \left\|(g \circ f \circ m_J)(y)_J - y_J\right\|^2. \quad (3)$$

Because the composite function h defined by

$$h_J = (g \circ f \circ m_J)_J \quad (4)$$

is $\mathcal{J}$-invariant, Equation 1 applies, and the loss is equal to

$$\mathbb{E}\sum_J \left\|(g \circ f \circ m_J)(y)_J - g(x)_J\right\|^2 + \mathbb{E}\left\|y - g(x)\right\|^2. \quad (5)$$

Which means that the generalised self-supervised loss is equal to the ground-truth loss, up to a constant independent of the pseudo-inverse f.

Differential learning. Now assume that we use f_θ, a θ-parameterized family of differentiable functions f from which we aim to find the best noise-tolerant inverse f^*. Since the loss in Eq. 5 is defined in terms of h_J and not in terms of f we need a scheme to optimise f_θ through the fixed forward model g. Assuming that the forward model is also differentiable, we propose to solve this optimisation problem by stochastic optimisation via backpropagation of the differential model $h_{J,\theta} = (g \circ f_\theta \circ m_J)_J$.

Retrospectively, we find that learning the denoising function in (Batson & Royer, 2019) is the special case of learning a noise-tolerant inverse of the identity function.

4. Application

Deconvolving noisy images. To demonstrate our framework we apply it to the standard inverse problem of image deconvolution. In this case the forward model g is the convolution of the true image x with a blur kernel k. The observed image y is thus:

$$y = n(k * x)$$

In the case that the noise function n is the identity, the problem can be solved perfectly[2] by using the inverse filter k^{-1}. However, in general and in practice the noise function n

[1] where x_J denotes x restricted to dimensions in J

[2] Assuming compact support for k and infinite numerical precision.

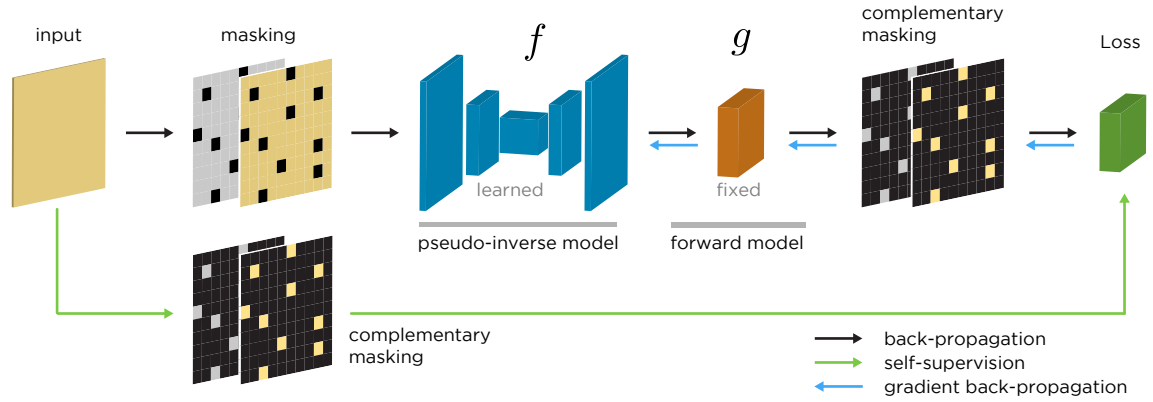

Figure 2. Training strategy for Self-Supervised Inversion. We train a composite model $f \circ g$ to be a noise-tolerant identity using our generalised self-supervised loss (Eq. 3). We feed observed images y as input and force the network to learn to return back y. First, the model must take the observation y as input and return the deconvolved image x. Second, this candidate deconvolved image is passed to the fixed forward model g to return back the observation y. After successful training, f must be the pseudo-inverse of g. At inference time, we can denoise an image y simply by applying f to x. In the absence of masking, training becomes sensitive to noise and the deconvolved image suffer from noise.

is not the identity and in fact captures many measurement imperfections such as measurement quantisation as well as signal dependent and non signal dependent noises. In the following we consider a Poisson-Gaussian noise model augmented with 'salt&pepper' – a good model for low signal-to-noise observations on camera detectors. Without loss of generality we restrict ourselves to the 2D case.

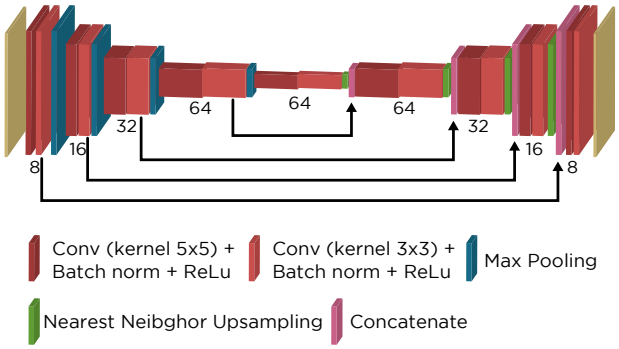

Figure 3. Detailed model architecture for the UNet used for f.

Training strategy and model architecture. Fig. 2 explains the self-supervised training strategy. Instead of a single trainable model f as in Noise2Self (Batson & Royer, 2019) we train the composition of a trainable inverse f followed by the fixed forward model g under the generalised self-supervised loss in Eq. 3. As shown in Fig. 3 we implement f with a standard UNet (Ronneberger et al., 2015). Once the function $f \circ g$ has been trained, we can use f as pseudo inverse to deconvolve the blurred and noisy image y. Because of the use of a masking scheme the learned pseudo-inverse f is noise-tolerant. However, since the forward model g is typically a low-pass filter, it is conceivable for the model f to produce a deconvolved image with erroneous high frequencies that would then be suppressed by g and thus never seen nor penalised by the loss. However, in practice, we don't observe issues probably because of the combination of the convolutional bias(Ulyanov et al., 2018) induced by the UNet model and our usage of weight regularisation that penalises the generation of unsubstantiated details (Both L_1 and L_2 regularisation, see code for implementation details).

5. Results

Benchmark dataset. We tested the deconvolution performance of our model on a diverse set of 22 two-dimensional monochrome images ranging in size between 512×512 and 2592×1728 pixels. The 22 images are normalised within $[0, 1]$ and have 8 bit precision. For each image we apply a Gaussian-like blur kernel[3] followed by a Poisson-Gaussian noise model augmented by salt-and-pepper noise:

$$n(z) = s_p(z + \eta(z)N) \tag{6}$$

Where $\eta(z) = \sqrt{\alpha z + \sigma^2}$, α is the Poisson term and σ is the standard deviation of the Gaussian term, and N is the independent normal Gaussian noise. Function s applies 'salt-and-pepper' noise by replacing a proportion p of pixels with a random value choosen uniformly within $[0, 1]$. In our

[3]Corresponding to the optical point-spread-function of a $0.8NA$ $16\times$ microscope objective with 0.406×0.406 micron pixels.

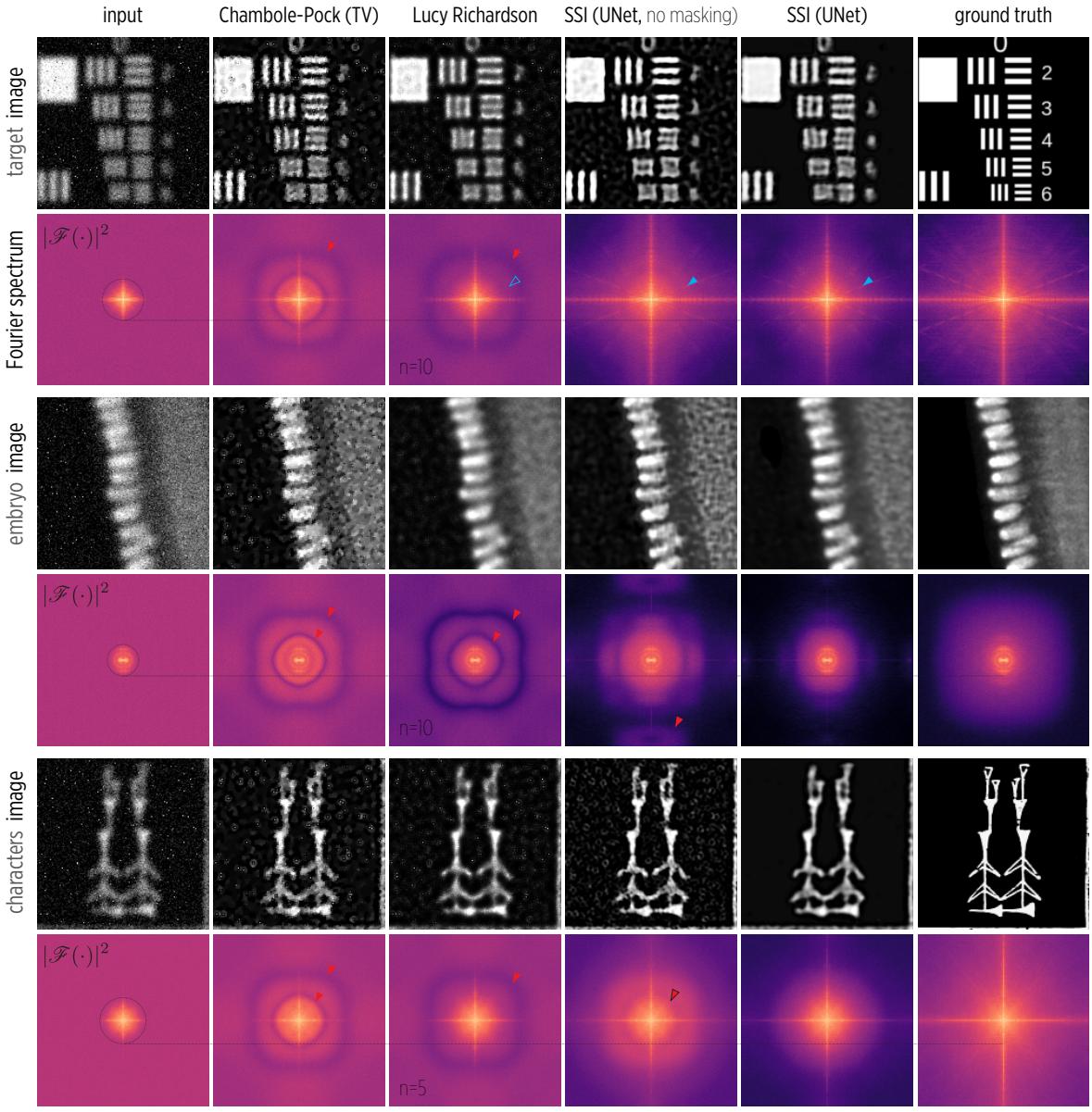

Figure 4. Performance of classic, and self-supervised inversion methods on natural images on three selected images. We show crops (80x80 pixels) as well as whole image spectra for four approaches: Chambole-Pock with a TV prior, Lucy-Richardson (LR), and Self-Supervised Inversion (SSI) via UNet with, and without, masking. SSI reconstructions achieve a good trade-off between noise reduction and high spatial frequency fidelity. The classic methods (CP, LR) are more likely to introduce distortion at high frequencies (red pointers) whereas SSI spectra a rather clean at high frequencies. Moreover, SSI spectra have a low noise-floor which corresponds to good noise reduction. However, the masking procedure which introduces a *blind-spot* in the receptive field leads to an attenuation at very high frequencies. The number of iterations for LR is indicated, for each of the three images we choose the number of iterations with the best performance.

experiments we choose a strong noise regime with: $\alpha = 0.001$, $\sigma = 0.1$, $= 0.01$. Finally, the images are quantised with 10 bits of precision.

Single image training and inference. In true self-supervised fashion, we decided to train one model per image and not use any additional images for training. Adding more *adequate* training instances or simply training on larger images would certainly help as the deep learning literature

attests (Cho et al., 2015). However, here we are interested in the baseline performance in the purely self-supervised case.

Comparison with classic approaches. We compare our Self-Supervised Inversion (SSI) approach with standard inversion algorithms such as Lucy-Richardson (LR) deconvolution(Richardson, 1972), Conjugate Gradient optimisation with TV prior(Chambolle & Pock, 2011), and Chambole-Pock primal-dual inversion also with a TV prior(Chambolle & Pock, 2011). In the case of LR deconvolution we evaluate three different number of iterations (5, 10, and 20) to explore the trade-off between noise amplification and sharpening (See Fig. 5).

Results. Table. 1 gives averages for four image comparison metrics: Peak Signal to Noise Ratio (PSNR)(Wang Yuanji et al., 2003), Structural Similarity (SSIM)(Wang et al., 2003), Mutual Information (MI)(Russakoff et al., 2004), and Spectral Mutual Information (SMI). The SMI metric rational is directly measure fidelity in frequency domain: it computes the mutual information in the frequency domain by taking the Discrete Cosine Transform (DCT 2) of both images and then computing the mutual information of these two images.

Example deconvolved images are shown in Fig. 4 with crops for the three images: *target*, *embryo*, *characters*. We show the images and their Fourier spectra and compare the methods: Chambole-Pock, Lucy-Richardson, Self-Supervised Inversion, and a control: Self-Supervised Inversion without masking. Overall, we find that Self-Supervised Inversion achieves the best performance across all metrics evaluated. The second best is Lucy-Richardson with 5 iterations. However, visual inspection of the corresponding images and spectra shows that while these images have little noise they also lack sharpness (see Fig. 5).

Table 2 lists the average training time and inference for the different methods. Some classical optimisation methods don't require training but are often very slow. As expected, differential learning based methods have long training times but are much faster at inference time.

Table 1. Average deconvolution performance per method for a benchmark set of 22 images. We evaluate image fidelity between the ground truth and: blurry, blurry&noisy, and restored images. We compute the Peak Signal to Noise Ratio (PSNR) (Wang Yuanji et al., 2003), Structural Similarity (SSIM) (Wang et al., 2003), Mutual Information (MI) (Russakoff et al., 2004), and Spectral Mutual Information (SMI). For all metrics, higher is better. The metrics SSIM, MI, and SMI are always within $[0, 1]$ with 0 being the worst value, and 1 attained when the two images are identical. For all fidelity metrics image deconvolution by Self-Supervised Inversion performed best.

	PSNR	SSIM	MI	SMI
blurry	23.1	0.77	0.17	0.38
blurry&noisy (input)	17.8	0.29	0.07	0.18
Conjugate Gradient TV	19.4	0.41	0.09	0.21
Chambole Pock TV	18.7	0.40	0.07	0.23
Lucy Richardson $n = 5$	22.2	0.59	0.12	0.25
Lucy Richardson $n = 10$	21.1	0.52	0.10	0.25
Lucy Richardson $n = 20$	18.5	0.38	0.08	0.19
SSI UNet *no masking*	17.7	0.38	0.07	0.14
SSI UNet	**22.5**	**0.61**	**0.14**	**0.27**

Table 2. Average inversion speed per method for a benchmark set of 22 images.

method	training time (s)	inference time (s)
Conjugate Gradient TV	0.00	95.74
Chambole Pock TV	0.00	306.60
Lucy Richardson $n = 5$	0.00	0.23
Lucy Richardson $n = 10$	0.00	0.10
Lucy Richardson $n = 20$	0.00	0.17
SSI UNet *no masking*	222.67	0.03
SSI UNet	249.01	0.03

6. Discussion

We have

7. Code

Python implementation of Self-Supervised Inversion in PyTorch(Paszke et al., 2017) with examples can be found at github.com/royerlab/ssi-code.

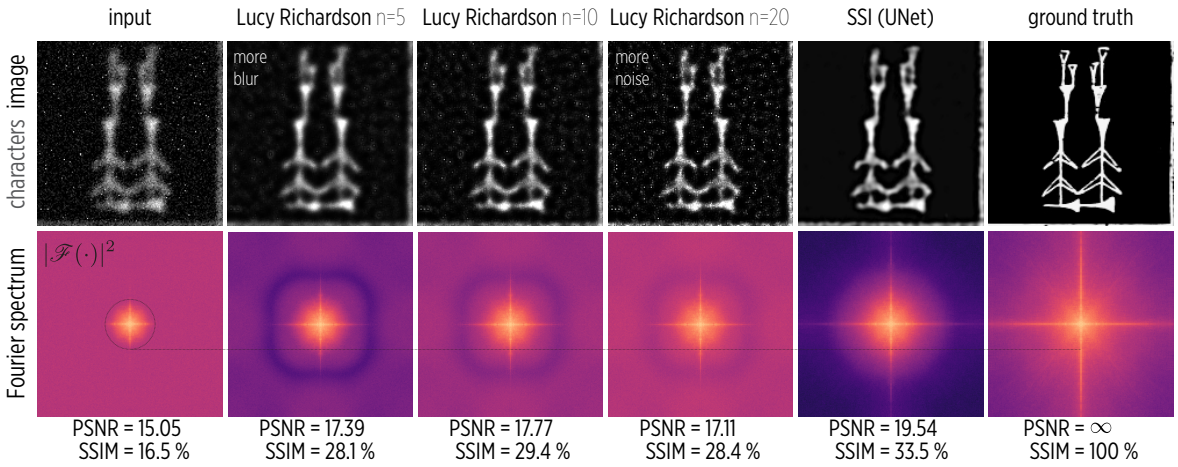

Figure 5. Lucy-Richardson deconvolution blurr-noise tradeoff. Lucy-Richardson deconvolution [ref] is an iterative algorithm that first reconstruct the low frequencies of an image and then incrementally refines the reconstruction with higher frequency components. It follows that low-iteration reconstructions are less sensitive to noise whereas high-iteration reconstructions are sharper but also noisier – hence a trade-off between sharpness and noise. In contrast, our self-supervised inversion approach is both insensitive to noise and sharpens the image.

Acknowledgements

Thank you to the Chan Zuckerberg Biohub for financial support. Loic A. Royer thanks his wife Zana Vosough for letting him finish this work on a week-end ;-)

References

Batson, J. and Royer, L. Noise2self: Blind denoising by self-supervision. *arXiv preprint arXiv:1901.11365*, 2019.

Chambolle, A. and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

Cho, J., Lee, K., Shin, E., Choy, G., and Do, S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy. *arXiv: Learning*, 2015.

Dong, C., Loy, C. C., He, K., and Tang, X. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pp. 184–199. Springer, 2014.

Hendriksen, A. A., Pelt, D. M., and Batenburg, K. J. Noise2inverse: Self-supervised deep convolutional denoising for linear inverse problems in imaging. *arXiv preprint arXiv:2001.11801*, 2020.

Kaneko, T. and Harada, T. Blur, noise, and compression robust generative adversarial networks. *arXiv preprint arXiv:2003.07849*, 2020.

Laine, S., Karras, T., Lehtinen, J., and Aila, T. High-quality self-supervised deep image denoising. In *Advances in Neural Information Processing Systems*, pp. 6968–6978, 2019.

Ljosa, V., Sokolnicki, K. L., and Carpenter, A. E. Annotated high-throughput microscopy image sets for validation. *Nature Methods*, 9(7):637–637, July 2012.

Mousavi, A. and Baraniuk, R. G. Learning to invert: Signal recovery via deep convolutional networks. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 2272–2276. IEEE, 2017.

Pajot, A., de Bezenac, E., and Gallinari, P. Unsupervised adversarial image reconstruction. 2018.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in PyTorch. In *NIPS-W*, 2017.

Richardson, W. H. Bayesian-based iterative method of image restoration. *JoSA*, 62(1):55–59, 1972.

Rick Chang, J., Li, C.-L., Poczos, B., Vijaya Kumar, B., and Sankaranarayanan, A. C. One network to solve them all–solving linear inverse problems using deep projection models. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5888–5897, 2017.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *arXiv:1505.04597 [cs]*, May 2015.

Russakoff, D. B., Tomasi, C., Rohlfing, T., and Maurer, C. R. Image similarity using mutual information of regions. pp. 596–607, 2004.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., Burdziak, C., Moon, K. R., Chaffer, C. L., Pattabiraman, D., Bierie, B., Mazutis, L., Wolf, G., Krishnaswamy, S., and Pe'er, D. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3): 716–729.e27, July 2018.

Wang, Z., Simoncelli, E. P., and Bovik, A. C. Multiscale structural similarity for image quality assessment. *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, 2:1398–1402 Vol.2, 2003.

Wang Yuanji, Li Jianhua, Lu Yi, Fu Yao, and Jiang Qinzhong. Image quality evaluation based on image weighted separating block peak signal to noise ratio. *International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003*, 2:994–997 Vol.2, 2003.

Xu, L., Ren, J. S., Liu, C., and Jia, J. Deep convolutional neural network for image deconvolution. In *Advances in neural information processing systems*, pp. 1790–1798, 2014.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, July 2017.

Zhussip, M., Soltanayev, S., and Chun, S. Y. Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10255–10264, 2019.