Reconstructing Synthetic Lensless Images in the Low-Data Regime

Abeer Banerjee[1,2]
abeer.ceeri20a@acsir.res.in

Himanshu Kumar[1,2]
himanshu.ceeri20a@acsir.res.in

Sumeet Saurav[1,2]
sumeet@ceeri.res.in

Sanjay Singh[1,2]
sanjay@ceeri.res.in

[1] CSIR-Central Electronics Engineering Research Institute (CSIR-CEERI) Pilani-333031, India

[2] Academy of Scientific and Innovative Research (AcSIR) Ghaziabad-201002, India

Abstract

Lensless image reconstruction is a class of inverse problems in computational imaging that is gaining immense popularity in the imaging research community because of its potential to revolutionize traditional imaging systems radically. It is possible to build cameras with extremely flat form factors that enable imaging in challenging scenarios. In this paper, we reconstruct simulated lensless images using a very low number of examples where we pre-train a decoder architecture with the available examples to model the data semantics, and then solve the inverse problem of lensless image reconstruction guided by a physics-informed forward loss function. We contrastively analyze and evaluate the reconstruction performance of our model against its untrained counterpart and show a substantial improvement in the reconstruction quality and convergence time even with a few example images.

1 Introduction

Computational cameras are increasingly being adopted for scientific and commercial applications owing to their promising performance in highly constrained scenarios. Lensless computational cameras are an example where a potential to revolutionize traditional imaging systems could be observed. The complete elimination of the need for lenses brings a major prospect for camera size compactification, however, the absence of lenses requires a computational algorithm for the restoration of the images thus formed. Since the images formed without a focusing element are not directly comprehensible to the human eye, lensless imaging has a privacy-preserving attribute inherent to it.

There have been notable advancements in the field of computational imaging in recent years, particularly in the area of deep learning techniques for solving inverse problems. Deep learning methods succeed in learning highly complex patterns and relationships from large datasets which can be observed from the exceptional restoration performance compared to the traditional techniques. The research community on computational imaging is focusing

on novel techniques for solving inverse problems that include generative priors, untrained neural network priors, and unfolding networks. These advances have led to significant improvements in solving inverse problems in various applications such as medical imaging, remote sensing, and digital photography.

Generative Priors: They are probabilistic models trained to generate images that model the data distribution of a huge image dataset. These models use generative adversarial networks (GANs) [6], acting as natural image priors for image reconstruction as decoder-based latent variable models. Recent research provides a robust theory for solving inverse problems including compressed sensing [5, 8, 12, 13, 22], phase-retrieval [9, 18], and blind deconvolution [2, 7]. Despite their effectiveness, deep generative models have certain drawbacks, including the requirement of a large amount of data for training and their non-trivial representation error, as they model the natural image manifold through a low-dimensional parameterization.

Untrained Neural Networks Priors: Recent advancements in solving inverse imaging problems have revealed that randomly initialized neural networks can serve as natural image priors without prior training, unlike neural networks that require large datasets for training. In [20], the authors proposed solving inverse problems like denoising, inpainting, and super-resolution by optimizing the parameters of a convolutional neural network to fit a single image. Another study [11] proposed under-parametrized optimization for image compressing using non-convolutional networks for linear inverse imaging. [3, 4] performed untrained reconstruction of lensless images using over-parameterized and under-parameterized neural networks. Well-regularized untrained neural network models have been successful for various inverse problems without using any training data [10, 14, 21]. However, there is limited research addressing how to interpolate between no-data and high-data regimes. To address this gap, [16] proposed a low-shot learning technique employing an untrained neural network, which leverages the benefits of both and interpolates between no-data and high-data regimes which could be used for colorization and compressed sensing with moderate compression ratios.

In this paper, we propose a decoder-based neural network architecture for lensless image reconstruction in the low-data regime. Inspired by [16], we pre-train our network with a few images sharing similar data semantics by jointly optimizing the network parameters and the latent space. We solve the inverse problem of lensless image reconstruction by modeling the lensed image as the output of the generator and performing a two-step optimization guided by a physics-informed forward loss function. We contrastively evaluate the reconstruction performance of our low-shot network against its untrained counterpart and show a significant improvement in reconstruction quality and convergence time. We also show how it compares against a fully trained state-of-the-art method.

2   Methodology

In this section, we discuss the details regarding the architecture of our decoder framework. We provide the required data for performing few-shot reconstruction in restricted domains and explain the reason for the choice of these particular domains. We describe the training procedure while discussing the loss functions used for the two-step optimization procedure for solving the inverse problem of lensless image reconstruction. We also provide the implementation details for reproducing the results shown in this paper.
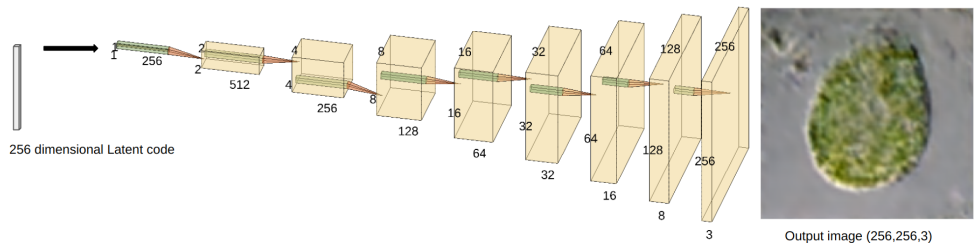
Figure 1: Decoder architecture using 256-dimensional latent code. The output image is generated at a resolution of 256×256.

2.1 Architecture

Our model has a convolutional decoder architecture that enables it to take the latent vector as the input and predict an output image at a resolution of 256×256. In Fig. 1, each block in yellow corresponds to $Conv2D \rightarrow ReLU \rightarrow Conv2D \rightarrow ReLU \rightarrow Upsample2D$ operation which increases the output resolution. Although transposed convolutions could have been used for the same task, still we prefer to use convolutional blocks followed by upsampling. This prevents checkerboard artifacts on the final reconstruction after the deconvolution process since convolutions followed by upsampling make the network biased against such artifacts.

2.2 Data

By definition, the low-shot approach does not depend on the availability of a large dataset, however, it is important for samples from a domain to share similar attributes such that the network can model the domain from the available few examples. Although datasets on lensless images are publicly available [1, 17], they are too diverse to be modeled by a low number of examples. Potential applications of lensless computational cameras could be in the research area of lensless microscopy, privacy-preserving face-verification, etc, therefore, we perform reconstructions by restricting the domain to microscopic images collected from pinterest.com, and face images obtained from FlatCam Faces Dataset [19]. The pre-training and testing images for the microscopy section can be accessed using this link: shorturl.at/pswF4.

2.3 Loss Functions and Training

The main objective of low-shot learning is to make use of the available few examples to provide an improved reconstruction compared to its untrained counterpart. In contrast to untrained neural networks that work in the no-data regime, low-shot networks work in the low-data regime where the available examples are used to pre-train the network such that the network is capable of learning the range statistics. The weight parameters of the decoder framework, θ are randomly initialized and we start with a latent code, z, sampled from the random normal distribution.

We define the pre-training step as the joint optimization of z and θ using the low shots, as shown in Fig. 2. We minimize the L_2 loss between $G(z,\theta)$ and the low shots, x_i, as suggested by [16].

The solution to the inverse problem is obtained using a two-step optimization approach as shown in Fig. 3. For solving the inverse problem, we compute the physics-informed forward loss where we use the prior knowledge of the point spread function (PSF) to simulate the forward imaging process. Our approach works on the assumption that the shift-invariance property, i.e., the angular memory effect of the PSF is globally valid. Therefore, the forward imaging process can be modeled by performing the 2D convolution of the PSF and the lensed image. It is to be noted that this assumption does not perfectly model the forward process for larger impinging angles, and the simulated lensless image is an approximation of the real lensless image. For calculating the physics-informed forward loss, we convolve the random PSF with the intermediate reconstruction obtained from the network.

The first step of the two-step optimization procedure is to find the approximate solution by optimizing over z while keeping θ fixed. We initialize the latent code by averaging the optimized individual latent codes, z_i for $i \in [S]$, where S is the number of shots, corresponding to each low-shot obtained via pre-training. Then the physics-informed forward loss between the intermediate lensless image, $G(z, \tilde{\theta})$, and the original lensless image, y_0, is optimized according to Eq. 1.

$$\min_z \frac{1}{2}||I_{PSF} * G(z, \tilde{\theta}) - y_0||^2 \tag{1}$$

Next, we use the image adaptivity approach [13] for obtaining the final solution, where the z and θ are jointly optimized subject to the physics-informed forward loss for obtaining the final reconstruction according to Eq. 2.

$$\min_{\theta,z} \frac{1}{2}||I_{PSF} * G(z, \theta) - y_0||^2 \tag{2}$$

2.4   Implementation Details

The decoder network was pre-trained using the available examples for 45,000 iterations with a learning rate of $1e-3$ to obtain the jointly optimized network parameters, θ, and latent code space, z. We used the Adam optimizer and L1 loss for the pre-training section. The lensless image was reconstructed by solving the inverse problem using a two-step optimization approach. The first step for lensless microscopic images involves optimizing the latent
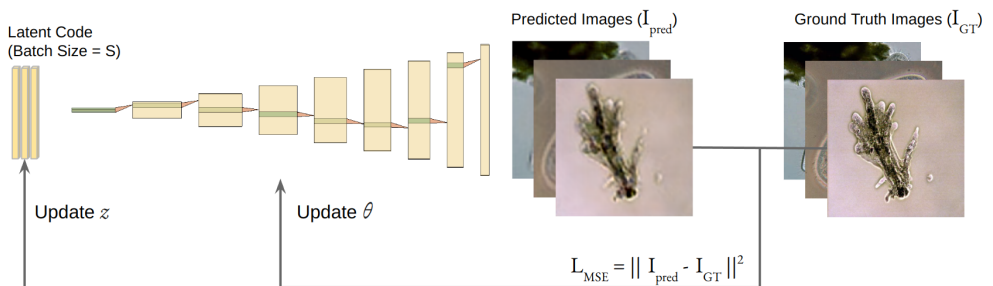

Figure 2: Pre-training stage: The network is pre-trained using the low number of examples and the optimal latent code (z) and network parameters (θ) are obtained via joint optimization.
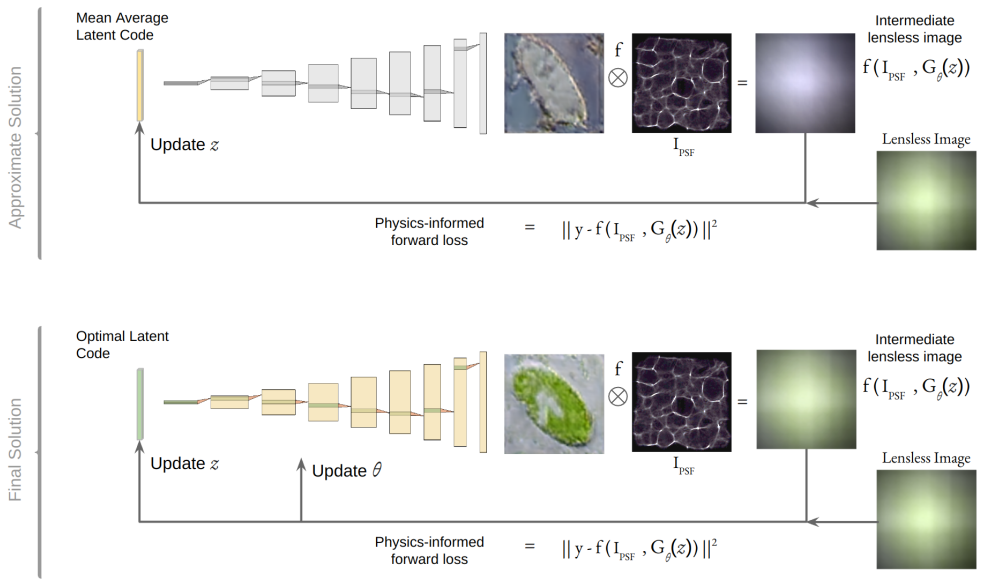
Figure 3: This is the complete optimization framework for solving the inverse problem of lensless image reconstruction.

space z using Adam for 500 iterations with a learning rate of $1e-1$, by keeping the network parameters fixed, while for lensless face images, we had to perform 1200 iterations. In the next step, the network parameters were also allowed to update with a learning rate of $1e-4$ for around 1500 iterations which led to a converged model for both cases.

3 Results

In this section, we show the reconstructed outputs generated using the decoder pre-trained using the low shots. We compare the reconstruction performance against the untrained counterpart and show that our model actually benefits from the low-shots provided. In Fig. 4, the variation in L1 loss during the untrained iterative optimization step for solving the inverse problem is plotted for both networks. It can be observed that the low(10)-shot model starts with a very low L1 loss due to the pre-training step, and hence is able to converge within 1.5k iterations.

3.1 Quantitative Comparison

We compare the reconstructed images generated by using our low-shot network against the untrained reconstructions. The untrained model, however, takes about 15k iterations to converge to a convincing output, therefore we also provide the intermediate reconstruction performance corresponding to the convergence time of our low-shot model. Since quantitative evaluation metrics like Peak Signal-to-Noise-Ratio (PSNR) and Structural SIMilarity (SSIM) index often do not outline the complete picture, we also provide visual comparative reconstructions for a better interpretation. Image-specific reconstruction results are provided in Table 1, where the image number corresponds to the test-image filenames provided in the
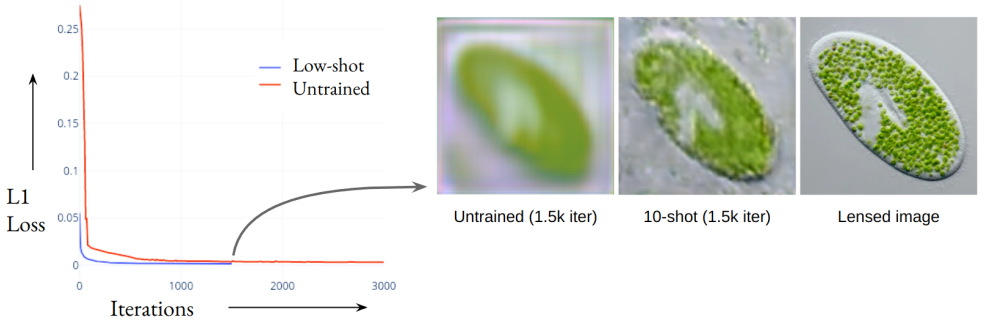
Figure 4: Comparison of loss curves obtained using an untrained iterative approach (red) and the low(10)-shot approach (blue).

link in Sec. 2.2. It can be observed from Table 1 that it required 10 shots to obtain a PSNR of 20dB in the case of lensless microscopic image reconstruction. However, for face reconstruction, it took a minimum of 25 shots to obtain a similar performance within the same convergence time.

Image	Untrained (1.5k)		Untrained (15k)		S-shot (1.5k)	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Micro-1	17.78	0.56	22.66	0.70	25.00	0.88
Micro-2	19.34	0.59	23.88	0.78	25.93	0.90
Micro-3	20.69	0.63	23.03	0.71	26.18	0.93
Face-1	11.13	0.37	14.67	0.49	20.04	0.60
Face-2	13.05	0.39	17.46	0.66	23.03	0.79
Face-3	9.96	0.28	14.23	0.42	18.44	0.56

Table 1: PSNR (in dB) and SSIM comparison of our low-shot model against the untrained model. S=10 for microscopic images, and S=25 for face images.

3.2   Visual Comparison

In Fig. 5, we have provided a detailed comparison of reconstructions using the untrained method contrasted against our low-shot method. For a fair comparison, we also provide the intermediate reconstructions of the untrained model corresponding to the convergence time of the low-shot network. It can be observed that the pre-training step with the available low shots can greatly improve the understanding of data semantics thus translating to faster convergence and improved reconstruction quality. In Fig. 6, we provide reconstruction results of our method trained on a limited set of 10 domain-restricted images, alongside the fully trained state-of-the-art (SOTA) model [15] that effectively illustrates the performance gap in reconstructions. Emphasizing the data limitations, our reconstructions stem from just 10 images, without additional perceptual enhancements, in contrast to the data-intensive nature of the fully trained GAN-based method.
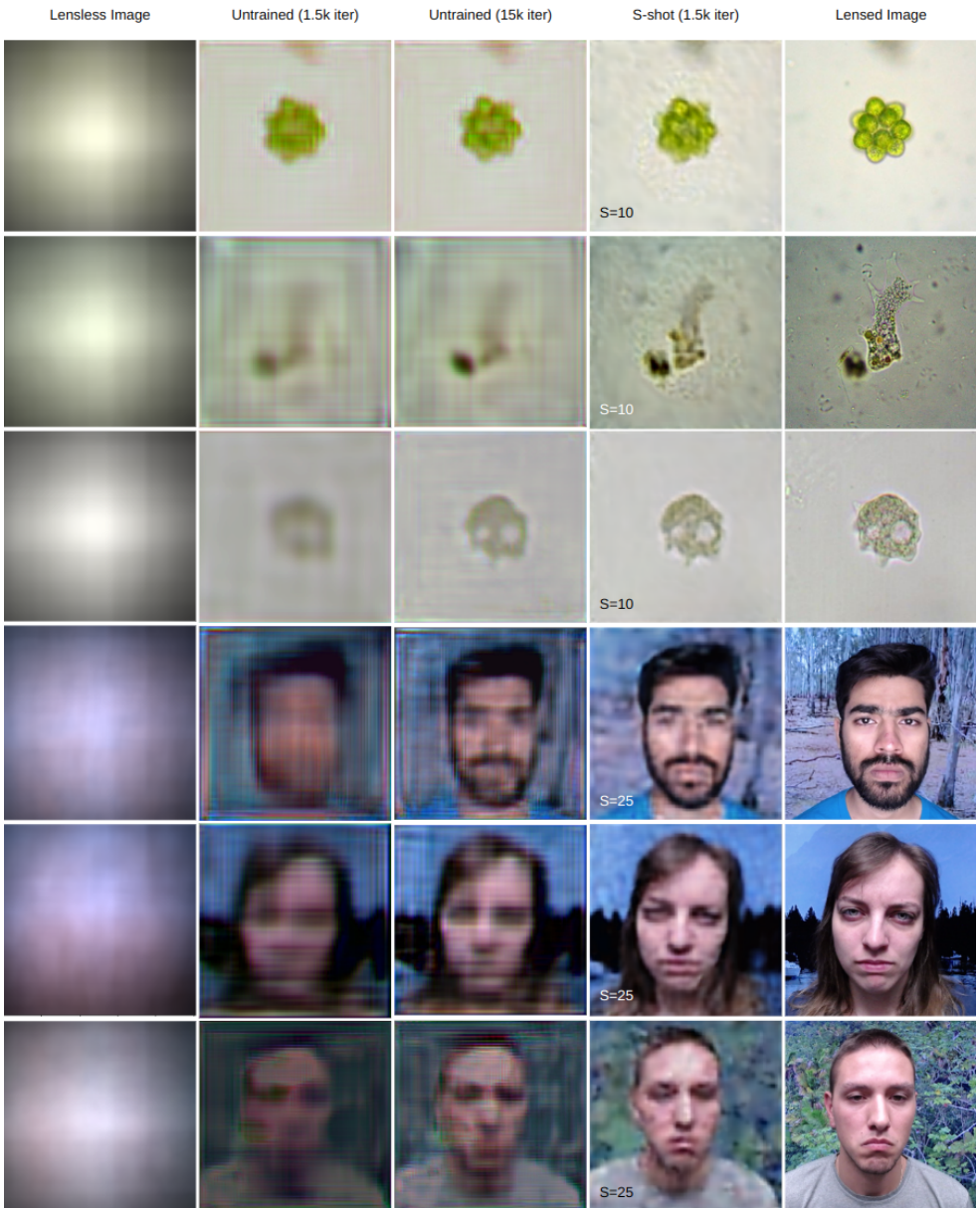
Figure 5: Visual comparison of reconstruction results obtained using the low(S)-shot network converging after ∼1.5k iterations, against the untrained network converging after ∼15k iterations.

4 Conclusion

We presented an approach for reconstructing lensless images in the low-data regime. The reconstructions have improved quality over their untrained counterpart and also converge sig-

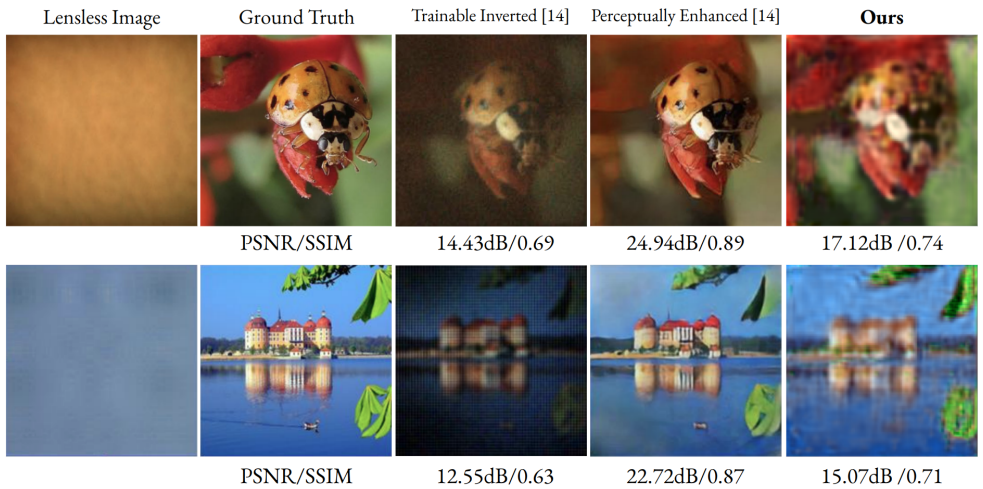| Lensless Image	Ground Truth	Trainable Inverted [14]	Perceptually Enhanced [14]	**Ours**
PSNR/SSIM		14.43dB/0.69	24.94dB/0.89	17.12dB /0.74
PSNR/SSIM		12.55dB/0.63	22.72dB/0.87	15.07dB /0.71

Figure 6: Visual comparison of our method (last column) using only 10 domain-restricted images, against a fully trained PSF-agnostic GAN-based state-of-the-art method [14]. The third and fourth columns represent the output of the trainable inversion layer and the GAN-based perceptual enhancement layer in [14], respectively.

nificantly faster. We performed a comparative performance evaluation of our model against its untrained counterpart with multiple standard evaluation metrics and also provided a visual comparison for a better interpretation of results. We have also performed a comparative analysis against reconstructions generated by fully trained neural networks to show the gap in performance. In this paper, we observed that by restricting the domain, it is possible to model the data semantics even with as few as 25 images. We performed low-shot reconstructions with restricted domains of microscopic images and face images and observed both qualitative and quantitative improvements in the reconstruction quality with the added advantage of faster convergence. We expect this work to be a step toward physics-informed lensless image reconstruction. Modeling diverse data distribution using a smaller number of examples would improve the generalizability of this approach which remains a problem to be solved, which would contribute to increasing the practicality of lensless computational cameras.

References

[1] M. Salman Asif, Ali Ayremlou, Aswin C. Sankaranarayanan, Ashok Veeraraghavan, and Richard Baraniuk. FlatCam: Thin, lensless cameras using coded aperture and computation. *IEEE Trans. Computational Imaging*, 3(3):384–397, 2017.

[2] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Blind image deconvolution using deep generative priors. *IEEE Transactions on Computational Imaging*, 6:1493–1506, 2018.

[3] Abeer Banerjee, Himanshu Kumar, Sumeet Saurav, and Sanjay Singh. Lensless image

reconstruction with an untrained neural network. In *International Conference on Image and Vision Computing New Zealand*, pages 430–441. Springer, 2022.

[4] Abeer Banerjee, Sumeet Saurav, and Sanjay Singh. Physics-informed deep deblurring: Over-parameterized vs. under-parameterized. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 1615–1619. IEEE, 2023.

[5] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 537–546. JMLR.org, 2017.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[7] Paul Hand and Babhru Joshi. Global guarantees for blind demodulation with generative priors. *ArXiv*, abs/1905.12576, 2019.

[8] Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. *CoRR*, abs/1705.07576, 2017. URL http://arxiv.org/abs/1705.07576.

[9] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/1bc2029a8851ad344a8d503930dfd7f7-Paper.pdf.

[10] Reinhard Heckel. Regularizing linear inverse problems with convolutional neural networks. *ArXiv*, abs/1907.03100, 2019.

[11] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. *International Conference on Learning Representations*, 2019.

[12] Wen Huang, Paul Hand, Reinhard Heckel, and Vladislav Voroninski. A provably convergent scheme for compressive sensing under random generative priors. *Journal of Fourier Analysis and Applications*, 27:1–34, 2018.

[13] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Image-adaptive gan based reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3121–3129, 2020.

[14] Gauri Jagatap and Chinmay Hegde. Algorithmic guarantees for inverse imaging with untrained network priors. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/831b342d8a83408e5960e9b0c5f31f0c-Paper.pdf.

[15] Salman S Khan, VR Adarsh, Vivek Boominathan, Jasper Tan, Ashok Veeraraghavan, and Kaushik Mitra. Towards photorealistic reconstruction of highly multiplexed lensless images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7860–7869, 2019.

[16] Oscar Leong and Wesam Sakla. Low shot learning with untrained neural networks for imaging inverse problems. *arXiv preprint arXiv:1910.10797*, 2019.

[17] Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Learned reconstructions for practical mask-based lensless imaging. *Optics express*, 27(20):28075–28090, 2019.

[18] Fahad Shamshad and Ali Ahmed. Robust compressive phase retrieval via deep generative priors. *ArXiv*, abs/1808.05854, 2018.

[19] Jasper Tan, Li Niu, Jesse K Adams, Vivek Boominathan, Jacob T Robinson, Richard G Baraniuk, and Ashok Veeraraghavan. Face detection and verification using lensless cameras. *IEEE Transactions on Computational Imaging*, 5(2):180–194, 2018.

[20] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. *International Journal of Computer Vision*, 128:1867–1888, 2017.

[21] Dave Van Veen, Ajil Jalal, Eric Price, Sriram Vishwanath, and Alexandros G. Dimakis. Compressed sensing with deep image prior and learned regularization. *ArXiv*, abs/1806.06438, 2018.

[22] Yan Wu, Mihaela Rosca, and Timothy Lillicrap. Deep compressed sensing. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6850–6860. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/wu19d.html.