Subsampled Fourier Ptychography via Pretrained Invertible and Untrained Network Priors

Fahad Shamshad, Asif Hanif, and Ali Ahmed
Department of Electrical Engineering, Information Technology University, Lahore, Pakistan
{fahad.shamshad, msee18003, ali.ahmed}@itu.edu.pk

Abstract

Recently pretrained generative models have shown promising results for subsampled Fourier Ptychography (FP) in terms of quality of reconstruction for extremely low sampling rate and high noise. However, one of the significant drawbacks of these pretrained generative priors is their limited representation capabilities. Moreover, training these generative models requires access to a large number of fully-observed clean samples of a particular class of images like faces or digits that is prohibitive to obtain in the context of FP. In this paper, we propose to leverage the power of pretrained invertible and untrained generative models to mitigate the representation error issue and requirement of a large number of example images (for training generative models) respectively. Through extensive experiments, we demonstrate the effectiveness of proposed approaches in the context of FP for low sampling rates and high noise levels.

1 Introduction

Resolution loss in long distance imaging can primarily be attributed to the diffraction blur, that is caused by limited aperture of the imaging system [1]. To mitigate the effects of the diffraction blur, recently an emerging computational imaging technique known as Fourier Ptychography (FP) has shown promising results [2, 3]. The objective of FP is to recover a high-resolution image from multiple diffraction-limited low-resolution images. In this paper, we consider recovering the signal $\boldsymbol{x} \in \mathbb{R}^n$ captured via forward acquisition model of FP, given by:

$$\boldsymbol{y}_\ell = |\mathcal{A}_\ell(\boldsymbol{x})| + \boldsymbol{n}_\ell, \quad \text{for} \;\; \ell = 1, 2, ..., L, \tag{1}$$

where $\boldsymbol{y}_\ell \in \mathbb{R}^m$ is diffraction-limited image corresponding to ℓ^{th} camera, $\mathcal{A}_\ell : \mathbb{R}^n \to \mathbb{C}^m$ is the linear operator representing the forward acquisition model, and $\boldsymbol{n}_\ell \in \mathbb{R}^m$ denotes noise perturbation. For ℓ^{th} camera, the linear operator $\mathcal{A}_\ell$ has the form $\mathcal{F}^{-1}\mathcal{P}_\ell \circ \mathcal{F}$, where $\mathcal{F}$ denotes 2D Fourier transform, $\mathcal{P}_\ell$ is a pupil mask that acts as a bandpass filter in the Fourier domain, and $\circ$ represents the Hadamard product. Specifically, FP works by iteratively stitching together a sequence of frequency limited low-resolution images $\boldsymbol{y}_\ell$ in Fourier domain to recover the high-resolution true image $\boldsymbol{x}$. Since optical sensors can measure only the magnitude of the signal [4], phase information is lost during the acquisition process — making the FP problem highly ill-posed.

To make the FP problem well-posed, generally additional measurements (i.e. $m \gg n$) are acquired in the form of high overlapping frequency bands in the frequency domain [3]. Although effective, these redundant measurements can pose severe limitations in terms of high computational cost. Recently, by devising realistic sampling strategies, prior information (sparsity and structured sparsity [5]) about the true signal has been leveraged to reduce the number of measurements (subsampling) in FP setup. However, it has been observed that these conventional signal priors often fail to capture the rich structure that many natural signals exhibit [6]. Priors learned from huge datasets can effectively capture this rich structure using the power of deep neural networks [7]. These deep neural networks or deep learning based end-to-end approaches have not yet been explored for reducing the number

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

of measurements in FP. Moreover, even a slight change in the parameters of FP forward acquisition model such as number of cameras or overlap would require costly retraining of these models.

To bridge the gap between deep learning based approaches (that can take advantage of the powerful learned priors) and conventional *hand-designed* priors such as sparsity (that are flexible enough to handle a variety of model parameters), recently pretrained deep generative models have emerged as an impressive alternative for subsampled FP problem [8]. However, one of the significant drawbacks of these pretrained generative priors is their limited representation capabilities. That is once trained, these pretrained generative models are incapable of producing any target image that lies outside their range[1]. Moreover, training of these generative models require access to a large number of fully-observed clean samples of a particular class of images like faces or digits. Unfortunately, obtaining multiple high-resolution samples can be expensive or impractical for many applications including FP.

In this paper, we aim to handle the aforementioned issues in context of subsampled FP by leveraging upon two recent works related to solving inverse imaging problems via pretrained invertible [9] and untrained generative model [10, 11, 12]. Our first contribution is to leverage the power of pretrained invertible generative models to mitigate the representation error issue of conventional generative models for non-convex and non-linear inverse problem of subsampled FP. We refer our first approach as *Invertible Ptych*. Our second contribution is to relax the requirement of massive amount of training data for training of generative models which can be prohibitively expensive to obtain in domains such as FP. We refer to our second approach throughout this paper as *Untrained Ptych*. Through numerical simulations, we demonstrate that proposed approaches can get better reconstructions, both qualitatively and quantitatively, at low subsampling ratios and high noise perturbations.

2 Problem Formulation

We refer interesting readers to related work of [13] for details of forward acquisition model of subsampled FP. As shown in [5, 13], in order to reduce the sample complexity, we can discard some observations in the sensor plane. This can be treated as applying a subsampling operator ($\mathcal{M}_\ell(\cdot)$), to effectively reduce the number of measurements. Mathematically, observation $\boldsymbol{y}_\ell$ for ℓth camera can be modeled as

$$\boldsymbol{y}_\ell = |\mathcal{M}_\ell(\mathcal{A}_\ell(\boldsymbol{x}))| + \boldsymbol{n}_\ell, \tag{2}$$

where $\mathcal{A}_\ell = \mathcal{F}^{-1}\mathcal{P}_\ell \circ \mathcal{F}$ is the measurement model prior to optical sensor acquisition step, $\circ$ denotes the Hadamard product, and $\mathcal{M}_\ell$ is the subsampling operator. Subsampling operator when applied to measurements $\boldsymbol{y}$, randomly picks a fraction of samples (f) discarding the others [5]. We define the subsampling ratio as the fraction of samples retained by $\mathcal{M}_\ell$ divided by the total number of observed samples i.e.

$$\text{Subsampling Ratio (\%)} = \frac{\text{Fraction of samples retained } (f) \times 100}{\text{Total observed samples } (nL)}.$$

The subsampling mask resembles the operation of a binary matrix having entries 1's and 0's. The mask has been element-wise multiplied with the observations in such a way that pixels corresponding to 1's are retained and those corresponding to 0's are discarded. Hence subsampling ratio governs the percentage of samples that will be retained. Without assuming any prior information about the true image $\boldsymbol{x}$, we can minimize the FP measurement loss as

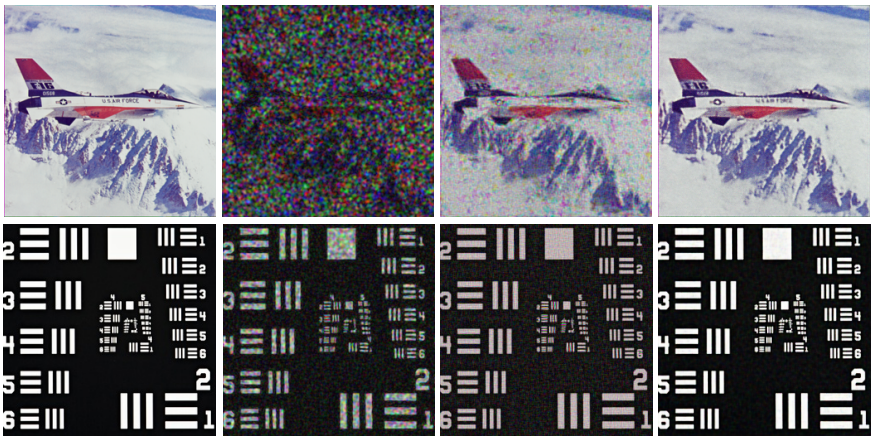
$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min} \sum_{\ell=1}^{L} \|\boldsymbol{y}_\ell - |\mathcal{M}_\ell\mathcal{A}_\ell(\boldsymbol{x})|\|_2^2, \tag{3}$$

to find the estimate of the true image $\boldsymbol{x}$. Note that without assuming any prior information about $\boldsymbol{x}$, (3) is notoriously difficult to solve as infinitely many solutions satisfy (3).

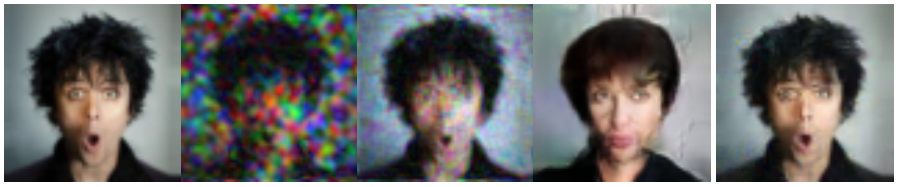
3 Proposed Approach

In this section, we formally introduce our proposed approaches *Untrained Ptych* and *Invertible Ptych*. Specifically, we denote by $G_\theta(z)$ as a deterministic function representing generative model that takes input $z \in \mathbb{R}^k$ and parameterized by the set of weights $\theta \in \mathbb{R}^d$ to produce output $G_\theta(z) \in \mathbb{R}^n$.

[1]Range of the pretrained generator can be defined as the set of all the images that can be generated by the that generator.



(a) Original (b) IERA (c) CoPRAM (d) Untrained Ptych



(a) Original (b) IERA (c) CoPRAM (d) DCGAN (e) Invertible

Figure 1: First two rows show subsampled FP reconstruction results of *Untrained Ptych* along with baseline methods. First row shows results with 2% subsampling ratio in noiseless setting. Second row shows result with 5% noise for subsampling ratio of 20%. Last row shows results of *Invertible Ptych* reconstructions for sampling rate of 2% in noiseless setting. Reconstructions of proposed approaches are visually appealing and relatively less corrupted with artifacts as compared to that of CoPRAM and IERA.

Untrained Ptych: Following the work of [10], we propose to use the structure of untrained convolutional generative models as a prior to obtain estimate of true image $\boldsymbol{x}$. Specifically, *Untrained Ptych* approach aims to find the set of weights of convolutional generative model (initialized randomly), $\hat{\theta}$, that produce output $G_{\hat{\theta}}(z)$ which best matches with the subsampled FP measurements $\boldsymbol{y}$ while obeying forward acquisition model (3). Specifically, we solve the following optimization problem

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^d}{\arg\min} \sum_{\ell=1}^{L} \|\boldsymbol{y}_\ell - |\mathcal{M}_\ell \mathcal{A}_\ell (G_\theta(\boldsymbol{z}))|\|_2^2. \tag{4}$$

The estimated image $\hat{\boldsymbol{x}}$ is acquired by a forward pass of the z through the generator $G_{\hat{\theta}}$ as $\hat{\boldsymbol{x}} = G_{\hat{\theta}}(z)$.

Invertible Ptych: For *Invertible Ptych*, we assume that we have access to a pretrained invertible generative model $G_\theta()$ (θ denotes pretrained weights) that has been trained on a specific class of natural images like face dataset. In this work, we use invertible architecture GLOW [14]. For *Invertible Ptych*, we solve the following optimization program via gradient descent

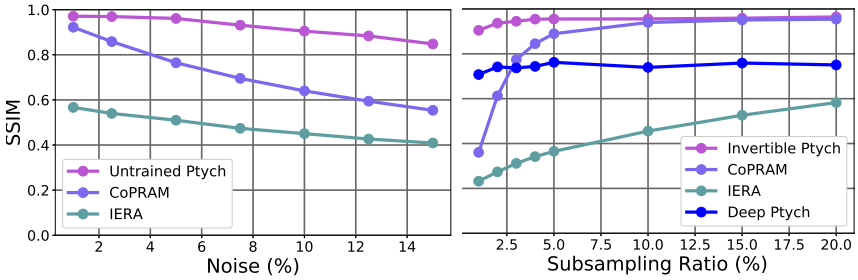
$$\hat{z} = \underset{z \in \mathbb{R}^k}{\arg\min} \sum_{\ell=1}^{L} \|\boldsymbol{y}_\ell - |\mathcal{M}_\ell \mathcal{A}_\ell (G_\theta(\boldsymbol{z}))|\|_2^2. \tag{5}$$

The estimated image $\hat{\boldsymbol{x}}$ is acquired by a forward pass of the $\hat{z}$ through the pretrained generator G_θ as $\hat{\boldsymbol{x}} = G_\theta(\hat{z})$.

4 Numerical Simulations

In this section, we evaluate the performance of proposed approaches on trained and untrained generative networks for different subsampling ratios and noise levels[2]. To quantitatively evaluate the performance of our algorithm, we use two metrics Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

[2]Noise of 1%, for image scaled between 0 to 1, translates to Gaussian noise with zero mean and a standard deviation of 0.01



(a) Untrained Ptych (b) Invertibe Ptych

Figure 2: Average SSIM plots of *Untrained Ptych* and *Invertible Ptych* for different noise levels and subsampling ratios.

Table 1: PSNR (dB) for different subsampling ratios and noise levels for *Untrained Ptych* and *Invertible Ptych*

	Subsamp Ratio (%)					**Noise (%)**				
	1	**2**	**3**	**5**	**10**	**1**	**2.5**	**5**	**7.5**	**10**
IERA	7.92	8.83	9.3	10.47	12.45	14.43	14.62	14.66	14.39	13.96
CoPRAM	12.07	18.08	21.61	22.94	24.94	24.57	22.7	22.42	21.39	20.32
Untrained	**23.53**	**25.53**	**28.46**	**30.01**	**30.67**	**33.99**	**33.83**	**33.01**	**30.2**	**29.51**
IERA	9.00	9.89	10.68	11.51	13.37	15.51	15.16	15.38	15.17	14.80
CoPRAM	12.07	16.33	19.57	21.94	23.29	24.27	23.61	22.55	21.43	20.87
Deep Ptych	19.92	20.61	20.29	20.91	20.44	20.72	20.87	20.87	20.99	20.79
Invertible	**25.41**	**27.42**	**28.36**	**29.23**	**29.71**	**30.60**	**28.64**	**27.73**	**28.01**	**27.29**

Comparison Methods We consider IERA (Iterative Error Reduction Algorithm) [2], CoPRAM [5] and Deep Ptych [13] as our baseline methods for evaluating the performance of our approaches for different subsampling ratios and noise levels.

Architectures For untrained generative model, we use U-Net based architecture with skip connections as in [10]. For invertible generative model, we use the Glow architecture [14], though our framework could be used with other invertible architectures as well. Let K be the number of steps of flow before a splitting layer, and L be the number of times the splitting is performed. To train over CelebA[7], we choose the network to have $K = 32$, $L = 4$ and affine coupling, and train it with a learning rate 0.0001.The model was trained over 8-bit images with 10,000 warmup iterations as in [14]. We refer the reader to [14] for specific details on the operations performed in each of the network layer.

Datasets For experiments with natural images, we perform subsampling and noise experiments on 5 standard test images of Aeroplane, Cameraman, Lena, USAF chart, and Boat each with size of 256×256. For face images, Glow model has been trained on CelebA dataset, center cropped to size 64×64 We use 30,000 images for training Glow and DCGAN(for Deep Ptych) based generative models [15] .

4.1 Qualitative and Quantitative results

In this section we provide the qualitative and quantitative results of subsampled FP for *invertible-ptych* and *untrained-ptych* by varying the subsampling ratio and noise level.

Qualitative results for *untrained-ptych* for subsampling ratio of 1% (in noiseless setting) and noise level of 5% (for 20% subsampling ratio) are shown in Figure 1. It can be seen that reconstructed images via *untrained ptych* are visually appealing as compared to those of IERA and CoPRAM that contain artifacts. We observe similar trend for *invertible ptych* as shown in Figure 1 (bottom row). Note that the results of DCGAN, though sharp, are constrained to lie in the range of the pretrained generator. On the other hand, *invertibe ptych* has no such limitation (by design).

Quantitative results, in terms of PSNR and SSIM, for *Untrained Ptych* and *Invertible Ptych* are shown in Figure 2 and Table 1 respectively. The results are averaged over 5 images for *Untrained Ptych* and 15 CelebA test images for *Invertible Ptych*. As shown in Figure 2, *Untrained Ptych* is able to acheive higher SSIM values for all noise levels as compared to IERA and CoPRAM. Similar performance gain has been observed for *Invertibe Ptych* for roustness against noise as compared to baseline methods. PSNR results in Table 1 indicate that proposed approaches are able to acheive considerable gain for low subsampling ratios and high noise levels as compared to baseline methods.

Acknowledgement

We gratefully acknowledge the support of the NVIDIA Corporation for the donation of NVIDIA TITAN Xp GPU for this research.

References

[1] Jason Holloway, Yicheng Wu, Manoj K Sharma, Oliver Cossairt, and Ashok Veeraraghavan. Savi: Synthetic apertures for long-range, subdiffraction-limited visible imaging using fourier ptychography. *Science advances*, 3(4):e1602564, 2017.

[2] Jason Holloway, M Salman Asif, Manoj Kumar Sharma, Nathan Matsuda, Roarke Horstmeyer, Oliver Cossairt, and Ashok Veeraraghavan. Toward long-distance subdiffraction imaging using coherent camera arrays. *IEEE Transactions on Computational Imaging*, 2(3):251–265, 2016.

[3] Guoan Zheng, Roarke Horstmeyer, and Changhuei Yang. Wide-field, high-resolution fourier ptychographic microscopy. *Nature photonics*, 7(9):739, 2013.

[4] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[5] Gauri Jagatap, Zhengyu Chen, Chinmay Hegde, and Namrata Vaswani. Sub-diffraction imaging using fourier ptychography and structured sparsity. In *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc.(ICASSP)*, 2018.

[6] Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. *arXiv preprint arXiv:1705.07576*, 2017.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[8] Fahad Shamshad, Farwa Abbas, and Ali Ahmed. Deep ptych: Subsampled fourier ptychography using generative priors. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7720–7724. IEEE, 2019.

[9] Muhammad Asim, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: mitigating representation error and dataset bias. *arXiv preprint arXiv:1905.11672*, 2019.

[10] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[11] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. *arXiv preprint arXiv:1810.03982*, 2018.

[12] Reinhard Heckel. Regularizing linear inverse problems with convolutional neural networks. *arXiv preprint arXiv:1907.03100*, 2019.

[13] Fahad Shamshad, Farwa Abbas, and Ali Ahmed. Deep ptych: Subsampled fourier ptychography using generative priors. *arXiv preprint arXiv:1812.11065*, 2018.

[14] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[15] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.