Adaptive Ptych: Leveraging Image Adaptive Generative Priors for Subsampled Fourier Ptychography

Fahad Shamshad[1], Asif Hanif[1], Farwa Abbas[1], Muhammad Awais[2], Ali Ahmed[1]
[1]Information Technology University, Lahore, Pakistan [2]Kyung Hee University, Korea
{fahad.shamshad,msee18003,farwa.abbas,ali.ahmed}@itu.edu.pk, awais@khu.ac.kr

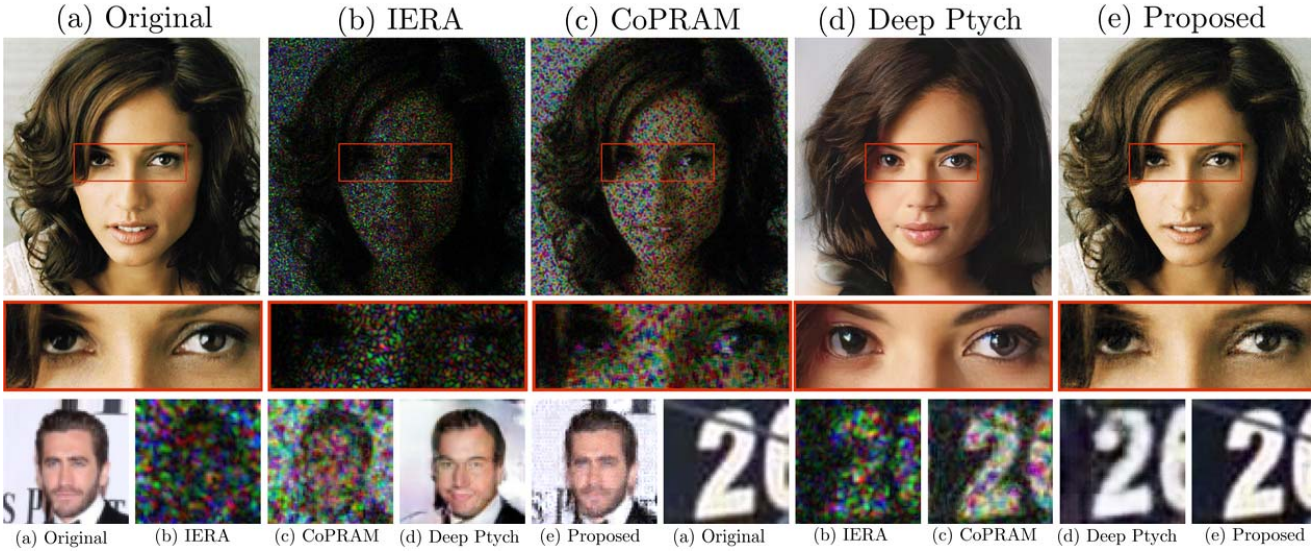


Figure 1: Reconstructing faithful estimate of original image (a) from subsampled Fourier ptychography (FP) measurements are challenging for conventional FP approaches of IERA (b) and CoPRAM (c). Generative prior based Deep Ptych approach produces visually appealing results as shown in (d) but suffers due to its limited representation capability. Our proposed approach (e) learns to reconstruct realistic results with fine details.

Abstract

Recently pretrained generative models have shown promising results for subsampled Fourier Ptychography (FP) in terms of quality of reconstruction for extremely low sampling rates. However, the representation capabilities of these pretrained generators do not capture the full distribution for complex classes of images, such as human faces or numbers, resulting in representation error. Moreover, recent studies have shown that these pretrained generative priors struggle at high-resolution in imaging inverse problems for reconstructing a faithful estimate of the true image, potentially due to mode collapse issue. To mitigate the issue of representation error of pretrained generative models for subsampled FP, we propose to make pretrained generator image adaptive by modifying it to better represent a single image (at test time) that is consistent with the subsampled FP measurements. Our experimental results demonstrate the superiority of the proposed approach over recent subsampled FP methods in terms of both quantitative metrics and visual quality.

1. Introduction

To mitigate the effects of the diffraction blur in long-distance imaging, recently an emerging computational imaging technique known as Fourier Ptychography (FP) has gained much attention in optical and signal processing community [1]. FP replaces the traditional interferometric methods, that require complex and expensive hardware, with computational algorithms [2, 3]. Specifically, FP works by iteratively stitching together a set of diffraction-limited and low-resolution images in the Fourier domain to recover the high-resolution true image. The forward acquisition model of FP for the true signal $\boldsymbol{x} \in \mathbb{R}^n$ can be written as

$$\boldsymbol{y} = |\mathcal{A}(\boldsymbol{x})| + \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ are observations, $\mathcal{A} : \mathbb{R}^n \to \mathbb{C}^m$ is a forward operator (more details in subsequent sections), and $\boldsymbol{\eta} \in \mathbb{R}^m$ denotes noise perturbation. The image formation at the sensing plane in FP is typically complex in nature, and since conventional optical sensors can measure only the magnitude of the signal, phase information is lost [4]. This

2473-9944/19/$31.00 ©2019 IEEE
DOI 10.1109/ICCVW.2019.00476

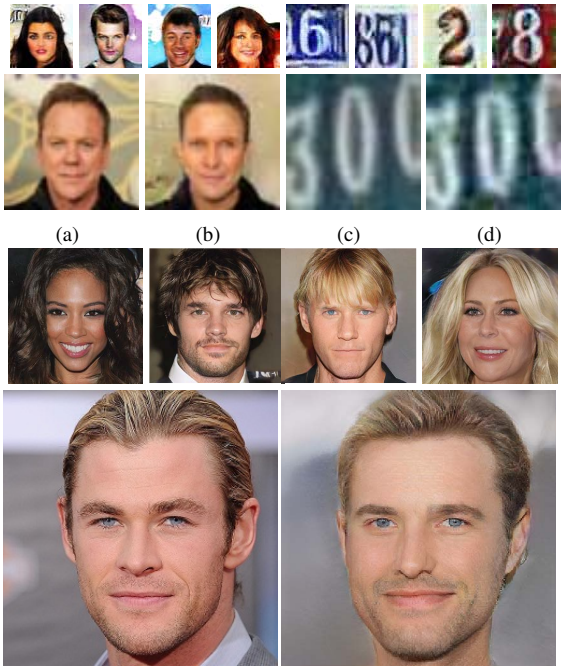


Figure 2: Illustration of limited representation capabilities of pretrained generative models. Samples of low-resolution (64×64) face images and digits are shown in top row, generated by pretrained generators trained on respective datasets. Second row shows inability of these pretrained generators to faithfully reconstruct the estimates ((b) and (d)) of the corresponding test images ((a) and (c)) due to their limited representational power. High-resolution (1024×1024) samples of face images generated by Progressive GAN (ProGAN) are shown in third row. Despite generating visually realistic samples, ProGAN fails to find the faithful estimate (e) of the corresponding true image (f) in its range (fourth row).

makes (1) highly ill-posed and difficult to solve.

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 [5]. 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 [6]) 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 [7]. Priors learned from huge datasets can effectively capture this rich structure using the power of deep neural networks [8]. These deep neural networks or deep learning based end-to-end approaches have not yet been explored for reducing the number 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 [9]. During the training of these generative models, they are encouraged to produce high dimensional samples, from low dimensional latent code, that resemble with that of the training set $\mathcal{X}$, of which true signal $\boldsymbol{x}$ is assumed to be a member i.e. $\boldsymbol{x} \in \mathcal{X}$ (we defer further details of training generative models to next section). Due to their power of capturing the high dimensional image distributions, these pretrained generative models have been shown to obtain significant performance gains over non-learning based FP methods, including sparsity, at very low sampling rates.

One of the significant drawback of these pretrained generative priors is their limited representation capabilities. That is once trained, they are incapable of producing any target image that lies outside their range[1]. Adding redundancy in the form of additional measurements is ineffective in driving reconstruction error to zero (even if we ignore the error due to non-convex optimization and noise). This shortcoming of the generative models in effectively learning the distribution of the true image dataset hampers their ability to reliably reproduce the estimate of given new (test) sample from its range as shown in Figure 2. Although, recently these generative models have been shown to produce high-resolution (1024×1024) and extremely photo-realistic images, but as illustrated in Figure 2 that even for these powerful generative models when one searches over the range of the generator for an image that is closest to the original, one is expected to get a significant mismatch. This indicates the severe mode collapse (inability of the generative model to capture the entire distribution of the true dataset) at higher resolutions.

In this paper, we aim to mitigate the limited representation capabilities of pretrained generative model in the context of subsampled FP. Specifically, we propose to make pretrained generator image adaptive by modifying it to better represent a single image that is consistent with the subsampled FP observations $\boldsymbol{y}$. We achieve it by jointly optimizing over the latent code and pretrained weights at the test time during minimization of the reconstruction loss. In this way, we not only leverage the rich semantic knowledge learned by the generator during training (in form of pretrained weights) but also allowing it to search solution beyond its range (by updating pretrained weights) while obeying the forward FP acquisition model. Through ex-

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

tensive simulations, we show that this simple strategy effectively eliminates the representation error inherited in the pretrained generators for subsampled FP problem. We further demonstrate its effectiveness at higher resolutions face images of size 1024×1024.

The rest of the paper is organized as follows. In Section 2, we briefly highlight the training procedure of GANs. Section 3 gives overview of the related work, followed by Section 4 that briefly recaps the FP setup. We formulate problem and give proposed solution in Section 5. Section 6 gives simulation results followed by limitations of proposed approach and concluding remaks in Section 7 and Section 8 respectively.

2. Training the Generative Model

In this section, we briefly recap the training process of GANs, as our approach requires a pretrained generator of the specific image class $\mathcal{X}$. GANs, originally introduced by [10], consists of two neural networks, generator ($\mathcal{G}$) and discriminator ($\mathcal{D}$). These networks compete with one another in an adversarial fashion where the generator tries to generate fake samples as real as possible, and discriminator aims to distinguish between generated samples and real samples. In these models, the generative part ($\mathcal{G}$), learns a mapping from low dimensional latent space $\boldsymbol{z} \in \mathbb{R}^k$ to a high dimensional sample space $\mathcal{G}(\boldsymbol{z}) \in \mathbb{R}^n$ where $k \ll n$. During training, these generative models are encouraged to produce samples that resemble that of training data $\mathcal{X}$. A well-trained generator, given by a deterministic function $\mathcal{G} : \mathbb{R}^k \to \mathbb{R}^n$ with a distribution P_Z over $\boldsymbol{z}$ (usually random normal or uniform), is therefore capable of generating fake data indistinguishable from the real data, it has been trained on. The min-max objective for training generative models can be written as

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{\boldsymbol{x} \sim p_{\mathcal{X}}(\boldsymbol{x})} \log \mathcal{D}(\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{z} \sim p(z)} \log(1 - \mathcal{D}(\mathcal{G}(\boldsymbol{z}))),$$

After training, we employ the trained $\mathcal{G}$ as a regularizer in our subsampled FP reconstruction algorithm.

3. Related Work

FP literature has mainly focused on the merits of experimental setup [1, 5, 11, 12] and on improving the quality of the reconstructed images [13, 14]. Relatively little attention has been given to the challenge of large measurements complexity. Exploiting low dimensional structures in the context of FP has not been explored until very recently. Zhang et al. leverage the sparsity and group sparsity priors to improve the reconstruction quality of image via threshold based gradient descent [15, 16]. However, they do not investigate the FP problem in the context of subsampled measurements. We also note work of [17] and [18] that use denoisers as plug-and-play priors in FP setup for improving reconstruction.

Recently priors learned from large datasets by exploiting the power of deep neural networks have shown promising results for faithful reconstruction of the true image in FP setup [19, 20]. Specifically, these deep learning based approaches invert the forward acquisition model of FP via end-to-end training of deep neural networks in a supervised manner [8]. However, even a slight change in the parameters of the FP measurement model, such as aperture diameter, number of cameras, or overlap would require costly retraining of these deep learning based models. Moreover, these deep learning based approaches have not been explored for the case of subsampled FP setup.

To the best of our knowledge, the first work that leverages the sparsity prior to reduce the sampling rate in FP is that of Jagatap et al. [6]. They further devise realistic subsampling strategies that can be readily implemented in conventional FP setups. Their algorithm (CoPRAM - Compressive Phase Retrieval using Alternative Minimization) significantly reduces the number of samples required for faithful reconstruction of the true signal. Chen et al. [21] leverage low-rank structure for dynamic and time-varying targets (videos) in FP to reduce the number of measurements. However, recently it has been shown in [7] that sparsity and low-rank priors often fail to capture the complex structure that many natural signals exhibit resulting in unrealistic signals also fitting the sparse and low-rank prior modeling assumption.

Generative models such as Generative Adversarial Networks (GANs) have been shown to produce promising results in various imaging inverse problems [22, 23, 24, 25, 26]. Inspired from their success in imaging inverse problems, recently [9] leverages the power of pretrained generative models for FP to reconstruct faithful estimates of the true image from far fewer samples. More precisely, their approach Deep Ptych, aims to find the latent code $\boldsymbol{z}$ via gradient descent algorithm such that $\mathcal{G}(\boldsymbol{z})$ is as similar as possible to true image $\boldsymbol{x}$. However, a significant drawback of the pretrained generative priors in solving subsampled FP problem is their limited representation capability, that might stem from mode collapse [27] or architectural choices of generator and discriminator. This shortcoming of the pretrained generative models in effectively learning the distribution of image datasets, especially at higher resolutions, hampers their ability to reliably reproduce the estimate of any given new (test) example that does not belong to its range of the generator.

4. Fourier Ptychography Setup

In this section, we briefly describe the forward acquisition model of FP as shown in Figure 3. In long-distance imaging, the resolution of the image is limited by

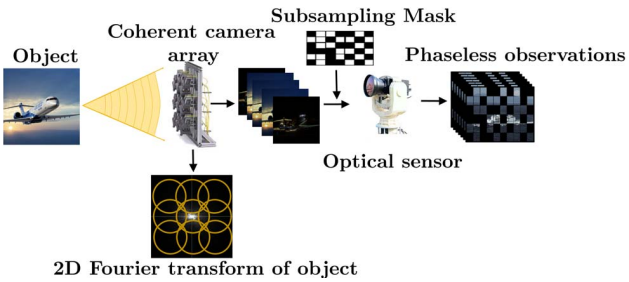


Figure 3: Fourier Ptychography forward acquistion model. The object has been illuminated using coherent light source. A coherent camera array captures illumination field from the object. The bandlimited signal is then focused to an image plane and a subsampling operator is applied. Subsequently, an optical sensor measures the magnitude while discarding the phase of signal

diffraction-limit of the imaging system. To handle this issue, multiple cameras are usually arranged in a square grid, forming a coherent camera array. Coherent camera array is placed in the far-field of the object ($\boldsymbol{x}$) to satisfy the Fraunhofer approximation. Under this approximation, when coherent illumination field emerging from the object is intercepted by the thin lens of camera array, it can be approximated as 2D Fourier transform $\mathcal{F}$. Each camera in the array have a finite aperture denoted by $\mathcal{P}_\ell$ ($\ell = 1, 2, ..., L$, where L denotes total number of cameras in the coherent camera array) that acts as a bandpass filter covering different parts of Fourier domain of the true image as shown in Figure 3. The bandlimited signal is then focused to an image plane where due to a second phase shift from the acquisition camera lens it undergoes an inverse Fourier transform ($\mathcal{F}^{-1}$). Subsequently, the complex spatial domain image is captured by an optical sensor that measures only the magnitude while discarding the phase information [4]. In order to reduce the sample complexity, we discard some observations at 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 an image, randomly picks a fraction of samples (f) discarding the others [6]. 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 f governs the percentage of samples that will be retained.

5. Problem Formulation and Proposed Solution

In this section, we formally introduce our proposed approach that we dubbed as *Adaptive Ptych*. We start by formulating the problem and stating the objective function of the pretrained generator based Deep Ptych approach [9]. Further, we show how our proposed approach circumvents the representation error inherited in the Deep Ptych.

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 $\boldsymbol{x}$. For the rest of the paper, we will call (3) as *reconstruction constraint*. Note that without assuming any prior information about $\boldsymbol{x}$, (3) is notoriously difficult to solve as infinitely many solutions satisfy the *reconstruction constraint*.

Next, we consider that the generative model, representated by the deterministic function $\mathcal{G}_\theta(\boldsymbol{z}) : \mathbb{R}^k \to \mathbb{R}^n$ where $k \ll n$, has been trained on the representative samples of specific class, $\mathcal{X}$, of the images. For instance, $\mathcal{X}$ may be set of street numbers or face images. Pretrained generator $\mathcal{G}_\theta(\boldsymbol{z})$, parameterized by the weights $\theta \in \mathbb{R}^d$, takes latent code $\boldsymbol{z}$ as input and produces output sample $\mathcal{G}_\theta(\boldsymbol{z}) \in \mathbb{R}^n$, similar to the training samples of class $\mathcal{X}$. This way $\mathcal{G}_\theta(\boldsymbol{z})$ captures the high dimensional distribution of the training set $\mathcal{X}$ and can acts as a strong prior for the true image $\boldsymbol{x}$, provided $\boldsymbol{x} \in \mathcal{X}$.

Deep Ptych approach aims to invert the pretrained generator $\mathcal{G}_\theta(\boldsymbol{z})$ by optimizing over the latent code $\boldsymbol{z}$ to best match the output of $\mathcal{G}_\theta(\boldsymbol{z})$ with the subsampled FP measurements. Specifically, Deep Ptych approach tries to find the estimate of the true image as the solution to the following non-convex optimization formulation

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

The estimated image $\hat{\boldsymbol{x}}$ is acquired by a forward pass of the $\hat{\boldsymbol{z}}$ through the generator $\mathcal{G}_\theta$ as $\hat{\boldsymbol{x}} = \mathcal{G}_\theta(\hat{\boldsymbol{z}})$. As the loss function term in (4) is non-linear and non-convex, therefore it is approximated via gradient descent algorithm started from random initialization of $\boldsymbol{z}$:

$$\hat{\boldsymbol{z}} \leftarrow \hat{\boldsymbol{z}} - \eta \frac{\partial \sum_{\ell=1}^{L} \|\boldsymbol{y} - |\mathcal{M}\mathcal{A}(\mathcal{G}_\theta(\boldsymbol{z}))\|_2^2}{\partial \boldsymbol{z}}\Bigg|_{\boldsymbol{z}=\hat{\boldsymbol{z}}}, \tag{5}$$

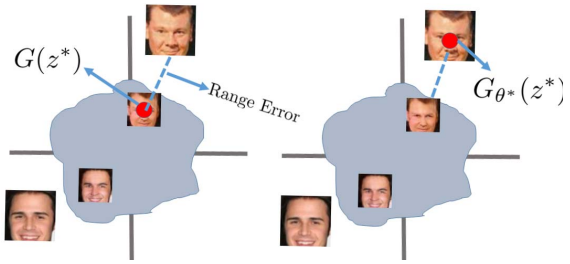


Figure 4: Illustration of Deep Ptych approach (left), where the solution (red dot) is constrained to lie inside the range (grey region) of the pretrained generator. Unlike Deep Ptych, proposed approach *Adaptive Ptych* (right) can explore solutions outside the range of the generator, while still obeying the forward FP acquisition model.

where η is the learning rate for the gradient descent algorithm. The optimization program of Deep Ptych in (4) implicitly constrains the recovered solution $\hat{\boldsymbol{x}}$ to lie in the range of the pretrained generator $\mathcal{G}_\theta$. This is illustrated in Figure 4, where we show the true images and corresponding closest images that lie in the range of the generative models trained on respective datasets. It can be seen in Figure 4 (left) that the generator trained on face dataset is not able to reconstruct corresponding true images faithfully via Deep Ptych approach due to representation/range error.

To mitigate the representation error of Deep Ptych, we propose a simple yet effective recovery algorithm that considers solutions outside the range of the pretrained generator while still leveraging the rich structure of the learned generative models. Specifically, we optimize the weights, denoted by $\boldsymbol{\theta}$, of the pretrained generator $\mathcal{G}_\theta$ along with latent code $\boldsymbol{z}$, to best match the output of the generator with the observed subsampled FP measurements $\boldsymbol{y}$. Optimization program of proposed approach is

$$\{\boldsymbol{\theta}^*, \boldsymbol{z}^*\} = \underset{\boldsymbol{\theta}, z}{\operatorname{argmin}} \sum_{\ell=1}^{L} \| \boldsymbol{y}_\ell - |\mathcal{M}_\ell \mathcal{A}_\ell(\mathcal{G}_\theta(\boldsymbol{z}))| \|_2^2, \quad (6)$$

Note that the initial value of $\boldsymbol{\theta}$ is the pretrained weights. Thus we are incorporating information captured during the test time (subsampled FP observations) with the valuable knowledge (in the form of pretrained weights) obtained during the training of the generative model. The final estimate $\hat{\boldsymbol{x}}$ will be $\hat{\boldsymbol{x}} = G_{\hat{\theta}}(\hat{\boldsymbol{z}})$. We dubbed our proposed approach as *Adaptive Ptych* as it adapts the weights of the $\mathcal{G}_\theta$ to produce an estimate of the true image that best matches with the given FP measurements $\boldsymbol{y}$.

We empirically show that proposed approach gives comparable performance to Deep Ptych at low measurements and leverage additional measurements to enhance the performance by going beyond the range of the generator and at the same time obey the forward phase retrieval model in (2). This is illustrated in Figure 4 where estimate recovered via Deep Ptych algorithm (red dot) is constrained to lie in the generator range (grey area) while proposed approach has no such limitation.

Although *Adaptive Ptych* approach gives visually appealing reconstructions, however, we numerically observe that the original high-resolution frequencies may not always be fully restored in the reconstructed images. To handle this discrepancy, we project the estimate $\hat{\boldsymbol{x}}$, obtained by (6), onto the solution space of $\boldsymbol{y}_\ell = |\mathcal{M}_\ell \mathcal{A}_\ell \boldsymbol{x}|$ by minimizing the following optimization program

$$\tilde{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \sum_{\ell=1}^{L} \| \boldsymbol{y}_\ell - |\mathcal{M}_\ell \mathcal{A}_\ell(\boldsymbol{x})| \|_2^2 + \lambda \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 \quad (7)$$

where λ balances fidelity of the approximation to subsampled measurements $\boldsymbol{y}$ and closeness of estimated solution $\tilde{\boldsymbol{x}}$ to the solution obtained via *Adaptive Ptych* algorithm $\hat{\boldsymbol{x}}$. We leverage the auto-differentiation function of Tensorflow [28] to calculate the gradient of the above optimization program to solve it via gradient descent algorithm. We take $\tilde{\boldsymbol{x}}$ from (7) as the final estimate of the high-resolution true image. The reconstructed image $\tilde{\boldsymbol{x}}$ is as close as possible to the estimate given by (6), and at the same time respecting the *reconstruction constraint* of (3). We dubbed this modified approach as *Adaptive Ptych+*.

6. Numerical Results

In this section, we evaluate the performance of the proposed approach against baseline methods through extensive experiments, both qualitatively and quantitatively. For quantitative comparison, we use peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM) [29]. All our simulations are performed by adding 1% random Gaussian noise to measurements unless stated otherwise[2]. For generative model based approaches, we report the results on a held-out test set, unseen by the generative models during training.

Datasets description: We evaluate the performance of the proposed approach on four datasets. These datasets include CelebA [30], SVHN [31], Shoes [32] and CelebA-HQ dataset [33]. CelebA dataset contains more than $200,000$ RGB face images (with $180,000$ training images) of size $218 \times 178 \times 3$ of different celebrities. We use aligned and cropped version of this dataset, where each image is of size $64 \times 64 \times 3$. SVHN consists of $32 \times 32 \times 3$ real-world images of house numbers obtained from google street view images with 73,257 images for training and 26,032 for testing. Shoes dataset consists of 50K RGB examples of different categories of shoes, resized to $64 \times 64 \times 3$. For Shoes dataset, we leave 1000 images for testing and use the rest as

[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.

Table 1: Experimental setting and relevant parameters for CelebA, SVHN, Shoes, and CelebA-HQ. We chose different parameters to show generalizability of proposed approach for different settings.

	CelebA	**SVHN/Shoes**	**CelebA-HQ**
Aperture Size of Camera	16	16	150
Frequency Bands Overlap	65%	50%	30%
Number of Cameras	81	49	9
Generator	DCGAN	DCGAN	ProGAN
Resolution	64 x 64	64 x 64	1024 x 1024

a training set. We upsample SVHN to size $64 \times 64 \times 3$ and train generative model on that dataset. Finally, the CelebA-HQ dataset consists of 30,000 face images each having size of 1024×1024. For the rest of paper, we will call CelebA, SVHN, and Shoes dataset as low-resolution datasets and CelebA-HQ as a high-resolution dataset. The representative training samples of these datasets are shown in Figure 5.

Generators Architecture: For low-resolution datasets, we use deep convolutional generative adversarial network (DCGAN) model of [34]. DCGAN uses convolutional layers in its generator and discriminator architecture to exploit the hierarchy of representations from image parts. For DCGAN, size of low dimensional latent representation $\boldsymbol{z}$ is set to 100 and is sampled from a random normal distribution. We train DCGAN model on the training set of low-resolution datasets by updating generator $\mathcal{G}$ twice and discriminator $\mathcal{D}$ once in each cycle to avoid fast convergence of $\mathcal{D}$. Each update during training use the Adam optimizer with batch size 64, $\beta_1 = 0.5$, and learning rate 0.0002. Generator, after training, is employed as a regularizer for proposed subsampled Fourier ptychography algorithm. For CelebA-HQ experiments, we use official pretrained model of progressive GAN (ProGAN) that is trained on high resolution 1024×1024 face images. The ProGAN model has a latent dimension of size 512 and is sampled from random normal distribution.

Baseline Methods: We use IERA [1], CoPRAM [6], and Deep Ptych [9] as baseline methods for qualitative and quantitative comparison. For CoPRAM, as in the original paper, we assume sparsity of the true images in Fourier basis. We use default algorithmic parameters of all baseline methods unless stated otherwise.

Setup: All simulations are performed on core-i7 computer (3.40 GHz and 16GB RAM) equipped with Nvidia Titan X GPU. We use TensorFlow library for implementing the proposed approach. Experimental FP settings for each dataset are given in Table 1.

6.1. Results on Low-Resolution Datasets

In this section, we compare the reconstruction performance of the proposed approach against baseline methods,

Figure 5: Samples of CelebA, SVHN, Shoes and CelebA-HQ datasets used for training of respective generative models. We train DCGAN model on training data of CelebA, SVHN, and Shoes. For CelebA-HQ dataset, we use official pretrained model of ProGAN [33].

both qualitatively and quantitatively, as we change the subsampling ratio. For low-resolution datasets, we use Adam optimizer to minimize the objective function in (4) with learning rate of 0.001 and 1500 steps.

Qualitative results for low-resolution datasets against different subsampling ratios are presented in Figure 6. It can be seen in Figure 6 that the reconstructed images of IERA and CoPRAM are blurry and contains artifacts for low subsampling ratio of 1%. For the same subsampling ratio, reconstructed images via Deep Ptych are visually appealing due to strong prior induced by pretrained generative models. However, due to limited representation capability of Deep ptych, its reconstructions are constrained to lie in the range of the pretrained generator that hampers its ability to produce a faithful estimate of the true image. On the other hand, as shown in Figure 6, the performance of the proposed approach is not limited by range of the generative model. It can be observed that the reconstructions of the proposed approach are visually sharp and close to the ground truth images as compared to Deep Ptych reconstructions.

Quantitative results in terms of SSIM and PSNR for *Adaptive Ptych* and baseline methods, against different subsampling ratios, are shown in Figure 7 and Table 2 respectively. These results are averaged over randomly selected 15 images taken from the test set of each dataset. As can be seen in Figure 7, *Adaptive Ptych* is able to achieve higher SSIM values at low-subsampling ratios $(1-3\%)$. For higher subsampling ratios, unlike Deep Ptych, *Adaptive Ptych* can

Figure 6: Visual comparison of the reconstructed images by proposed approach against baseline methods at different subsampling ratios. From top to bottom: Original images, IERA, CoPRAM, DCGAN, and Adaptive Ptych. Compared to the competing methods, our approach has been shown to produce superior results qualitatively.

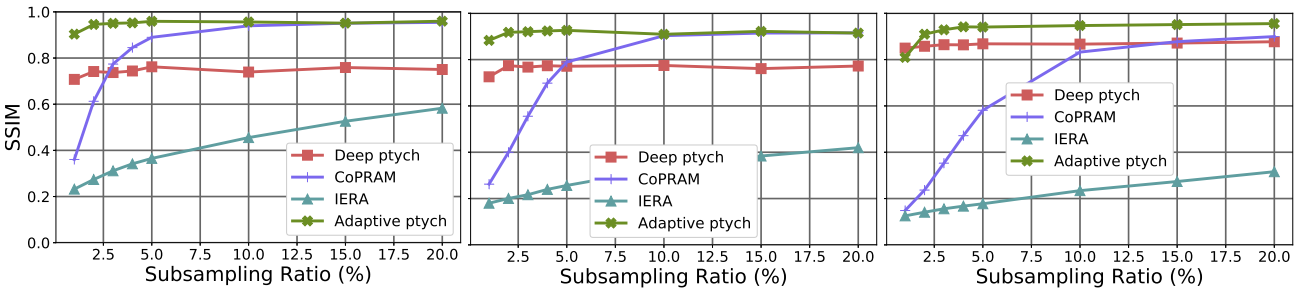


Figure 7: SSIM plots of reconstructed images via *Adaptive Ptych* against baseline methods at different subsampling ratios for CelebA, SVHN, and Shoes datasets (left to right).

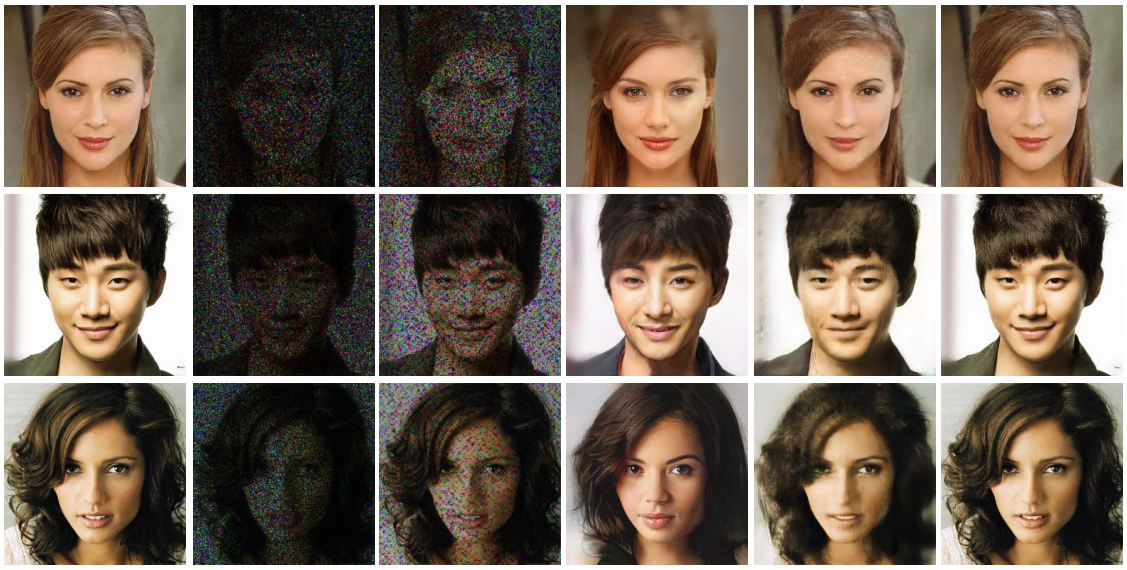
Table 2: Average PSNR results (dB) under different subsampling ratios of baseline methods and Adaptive Ptych. The best PSNR values are highlighted in bold.

	CelebA				**SVHN**				**Shoe**			
	1%	**2%**	**3%**	**5%**	**1%**	**2%**	**3%**	**5%**	**1%**	**2%**	**3%**	**5%**
IERA	9.00	9.89	14.15	10.68	8.42	9.18	9.79	10.75	3.29	3.84	4.25	4.93
CoPRAM	12.07	16.33	19.57	21.94	10.33	13.27	15.49	17.28	4.70	6.87	9.19	13.08
DCGAN	19.92	20.61	20.29	20.91	19.56	20.46	20.07	20.03	**22.11**	22.75	23.17	23.46
Adaptive Ptych	**25.67**	**28.69**	**29.68**	**30.05**	**21.37**	**22.44**	**22.33**	**22.76**	22.09	**26.97**	**27.95**	**29.74**

leverage the additional measurements for improved reconstructions that are comparable to that of CoPRAM. In Table 2, we present the average PSNR values that show a similar trend as that of SSIM metric. For instance the average PSNR of reconstructed CelebA images via *Adaptive Ptych*, at subsampling ratio of 2%, is about 12 dB and 8 dB higher as compared to that of CoPRAM and Deep Ptych.

6.2. Results on CelebA-HQ

In this section, we demonstrate the effectiveness of *Adaptive Ptych* and *Adaptive Ptych+* for high-resolution face images of CelebA-HQ. For *Adaptive Ptych*, we use Adam optimizer with 2000 steps and learning rate of 0.001 to minimize the reconstruction loss in 4. We further refine the estimate of *Adaptive Ptych* via *Adaptive Ptych+* by us-



(a) Original (b) IERA (c) CoPRAM (d) Deep Ptych (e) Adaptive Ptych (f) Adaptive Ptych+

Figure 8: Visual comparison of the reconstructed images via *Adaptive Ptych* and *Adaptive Ptych+* against baseline methods at different subsampling ratios. Compared to the competing methods, our approaches has been shown to produce superior results qualitatively.

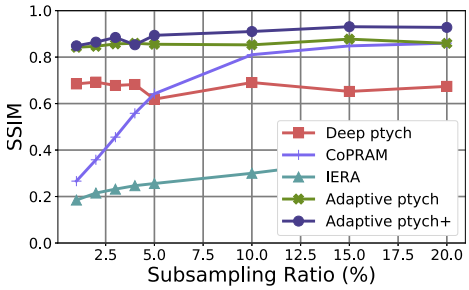


Figure 9: SSIM plot of reconstructed images via proposed approaches against baseline methods at different subsampling ratios for (a) CelebA (b) SVHN and (c) Shoes dataset.

ing Adam optimizer for 200 iterations to minimize reconstruction constraint in (7).

Qualitative results of Adaptive Ptych and Adaptive Ptych+ against baseline methods are shown in Figure 8. Once again, the results of IERA and CoPRAM contain artifacts at low subsampling ratios. Deep Ptych reconstructions, although sharp, are not satisfying due to limited representation capabilities of ProGAN. Adaptive Ptych reconstructions are visually appealing and close to the true image as compared to baseline methods. As shown in Figure 8, Adaptive Ptych may not be able to recover fine details like hair in the high-resolution face images. *Adaptive Ptych+* effectively mitigates this discrepancy as discussed in Section 5 and achieve best perceptual results among all approaches.

Quantitative results in terms of SSIM, averaged over 5 test images (due to computational and time constraints) of CelebA-HQ dataset, against different subsampling ratios are shown in Figure 9. We observe that Adaptive Ptych and Adaptive Ptych+ are able to achieve higher average SSIM values as compared to baseline methods. The average SSIM value of Deep Ptych reconstructions saturates as we increase the subsampling ratio, contrary to corresponding *Adaptive Ptych* and *Adaptive Ptych+* SSIM values that continue to increase by leveraging upon the additional measurements at higher subsampling ratios.

7. Limitations

Currently, our approach is limited by the requirement of the massive amount of training data to train the generative model accurately. This can be highly prohibitive in the context of FP, due to time and cost constraints. Our experiments show the effectiveness of the proposed approach for low noise level of 1%. Extending the proposed approach to high noise levels, that is relevant in FP setup, is challenging but promising future direction. Another limitation of pretrained generator based approaches is that they work only for specific classes of images on which they have been trained.

8. Conclusion

To conclude, we propose to mitigate the limited representation capability of pretrained generative models for subsampled Fourier ptychography problem. We demonstrate the superiority of the proposed approach against baseline methods both qualitatively and quantitatively.

References

[1] 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.

[2] 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.

[3] Guoan Zheng. Breakthroughs in photonics 2013: Fourier ptychographic imaging. *IEEE Photonics Journal*, 6(2):1–7, 2014.

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

[6] Gauri Jagatap, Zhengyu Chen, Chinmay Hegde, and Namrata Vaswani. Sub-diffraction imaging using fourier ptychography and structured sparsity. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6493–6497. IEEE, 2018.

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

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

[9] 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.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[11] Lei Tian, Xiao Li, Kannan Ramchandran, and Laura Waller. Multiplexed coded illumination for fourier ptychography with an led array microscope. *Biomedical optics express*, 5(7):2376–2389, 2014.

[12] Roarke W Horstmeyer, Yuhua Chen, Joel A Tropp, and Changhuei Yang. Ptychography imaging systems and methods with convex relaxation, February 21 2019. US Patent App. 16/171,270.

[13] Michal Odstrčil, Maxime Lebugle, Manuel Guizar-Sicairos, Christian David, and Mirko Holler. Towards optimized illumination for high-resolution ptychography. *Optics express*, 27(10):14981–14997, 2019.

[14] Michael Kellman, Emrah Bostan, Nicole Repina, and Laura Waller. Physics-based learned design: Optimized coded-illumination for quantitative phase imaging. *IEEE Transactions on Computational Imaging*, 2019.

[15] Yongbing Zhang, Pengming Song, Jian Zhang, and Qionghai Dai. Fourier ptychographic microscopy with sparse representation. *Scientific reports*, 7(1):8664, 2017.

[16] Yongbing Zhang, Ze Cui, Jian Zhang, Pengming Song, and Qionghai Dai. Group-based sparse representation for fourier ptychography microscopy. *Optics Communications*, 404:55–61, 2017.

[17] Yu Sun, Shiqi Xu, Yunzhe Li, Lei Tian, Brendt Wohlberg, and Ulugbek S Kamilov. Regularized fourier ptychography using an online plug-and-play algorithm. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7665–7669. IEEE, 2019.

[18] Zhixin Li, Desheng Wen, Zongxi Song, Gang Liu, Weikang Zhang, and Xin Wei. Sub-diffraction visible imaging using macroscopic fourier ptychography and regularization by denoising. *Sensors*, 18(9):3154, 2018.

[19] Armin Kappeler, Sushobhan Ghosh, Jason Holloway, Oliver Cossairt, and Aggelos Katsaggelos. Ptychnet: Cnn based fourier ptychography. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1712–1716. IEEE, 2017.

[20] Lokesh Boominathan, Mayug Maniparambil, Honey Gupta, Rahul Baburajan, and Kaushik Mitra. Phase retrieval for fourier ptychography under varying amount of measurements. *arXiv preprint arXiv:1805.03593*, 2018.

[21] Zhengyu Chen, Gauri Jagatap, Seyedehsara Nayer, Chinmay Hegde, and Namrata Vaswani. Low rank fourier ptychography. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6538–6542. IEEE, 2018.

[22] 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*, pages 537–546. JMLR. org, 2017.

[23] Fahad Shamshad and Ali Ahmed. Robust compressive phase retrieval via deep generative priors. *arXiv preprint arXiv:1808.05854*, 2018.

[24] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, pages 9136–9146, 2018.

[25] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Blind image deconvolution using deep generative priors. *arXiv preprint arXiv:1802.04073*, 2018.

[26] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, et al. Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging*, 37(6):1310–1321, 2017.

[27] Zhaoyu Zhang, Mengyan Li, and Jun Yu. On the convergence and mode collapse of gan. In *SIGGRAPH Asia 2018 Technical Briefs*, page 21. ACM, 2018.

[28] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[30] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[31] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011.

[32] Aron Yu and Kristen Grauman. Fine-grained visual comparisons with local learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 192–199, 2014.

[33] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

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