DiffFPR: Diffusion Prior for Oversampled Fourier Phase Retrieval

Ji Li [*1] **Chao Wang** [*2]

Abstract

This paper tackled the challenging Fourier phase retrieval problem, the *absolute uniqueness* of which does not hold. The existence of *equivalent solution* (a.k.a. trivial solution ambiguity) hinders the successful recovery, especially for multi-channel color image. The traditional iterative engine, such as the Relaxed Averaged Alternating Reflections (RAAR), can be applied to reconstruct the image channel-wisely. However, due to the *relative uniqueness* of the solution, the restoration is not automatically aligned with the accurate orientation for each channel, resulting in a reconstructed image that deviates significantly from the true solution manifold. To address this issue, by penalizing the mismatch of the image channels, a diffusion model as the strong prior of the color image is integrated into the iterative engine. The combination of the traditional iterative engine and the diffusion model provides an effective solution to the oversampled Fourier phase retrieval. The formed algorithm, *DiffFPR*, is validated by experiments. The code is available at `https://github.com/Chilie/DiffFPR`.

1. Introduction

The phase retrieval problem arises from a broad spectrum of applications, including the astronomy (Gonsalves, 2014; Fienup & Dainty, 1987), optics (Shechtman et al., 2015), ptychography (Qian et al., 2014; Wen et al., 2012), as well as the coherent diffraction imaging (CDI) (Candès et al., 2015). For the CDI application, the specimen is imaged using X-ray and the propagation of the X-ray is specified by Fourier transform (Goodman, 2005), especially in the Fraunhofer regime (the "far field"). The specimen structure can be explored from the optical field after the reaction between the illumination optics and the specimen. There are two components of the optical field: the magnitude and the phase. Due to the limitation of the measurement instrument, only the magnitude of the optical field can be recorded. Phase retrieval means to reconstructing the unobserved phase information of the optical field. With the resolved optical phase, the specimen structure can be resolved using the back-propagation of the obtained full optical field.

Due to the missing of phase information, it is impossible to reconstruct the image without additional constraints. To this end, the oversampling in measurement compensates the missing information to some extent. It equals to warp the image with zero-padding boundary. In this paper, we focus on the three-channel color image $\boldsymbol{x}_0 \in \mathbb{R}^{h\times w\times 3}$. To facilitate the presentation, we define the related two operators: Zero-padding operator and its adjoint (effect-reverting) operator. The $\mathrm{zpad}_{r^2} : \mathbb{R}^{h\times w\times 3} \to \mathbb{R}^{rh\times rw\times 3}$ denotes padding the centered original image using zero at the boundary, and the $\mathrm{revzpad}_{r^2} : \mathbb{R}^{rh\times rw\times 3} \to \mathbb{R}^{h\times w\times 3}$ takes off the added padding region. In the notations, the subscripts denote the oversampling ratio, which is omitted if the oversampling ratio is predefined in the context. Then we applied a channel-wise two-dimensional Fourier transform to the padded image $\mathrm{zpad}_{r^2}(\boldsymbol{x}_0)$ to produce the measurement $\boldsymbol{y} = |\mathcal{F}(\mathrm{zpad}_{r^2}(\boldsymbol{x}_0))| \in \mathbb{R}^{rh\times rw\times 3}$. Here the $\mathcal{F}$ is the standard normalized discrete Fourier transform.

There have been a large body of works on the conditions of the oversampling ratio to ensure the *relative uniqueness* of the phase retrieval problem (Balan et al., 2006; Balan, 2015a;b; Hayes, 1982). The relative uniqueness means the solution is unique up to trivial shifting, twin image with rotation by $180°$, and the minus image with multiplication by -1. Though the solution ambiguity does not impact the image contents, it plays a negative impact on the numerical solution to phase retrieval.

The translation and the minus image solution ambiguity can be avoided by constraining the support and the nonnegativity of the solution. For gray-scale image, though phase retrieval is nonconvex, there exists iterative engine to obtain the solution up to trivial ambiguities. For one-channel gray-scale image, the ambiguity is a negligible issue, however, for three-channel color problem, the orientation ambiguity leads to erroneous reconstruction. See Figure 1 for the

[*]Equal contribution [1]Academy for Multidisciplinary Studies, Capital Normal University, Beijing, China [2]University of Kansas Medical Center, Kansas City, US. Correspondence to: Ji Li <matliji@163.com>, Chao Wang <wywwwnx@163.com>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

representative results of Relaxed Averaged Alternating Reflections (RAAR) (Li & Zhou, 2017; Luke, 2004), with the comparison to our approach. The main issue of RAAR is the mismatch of the image channels. In this case, one may manually align the reconstructed channels, which is image-specific. It is desired to achieve an automatic solving approach for color image.

Figure 1. **Illustration of the results of noiseless $4\times$ oversampled phase retrieval.** RAAR produces restoration with misaligned channel orientation, DPS (Chung et al., 2022a) generates deviated solution, while ours produces much better solution.

Among generative models, we choose the diffusion model as the prior provider for several reasons: its notable modeling performance, the iterative generation process that naturally integrates the previous traditional engine into the diffusion generation, and the advantage of not requiring retraining for different oversampling ratios. Consequently, in this paper, we propose using a pretrained diffusion model to penalize the deviation of the RAAR iteration from the underlying image manifold. The unconditional generation process can be interpreted as the transition from the noisy image manifold to the clean image manifold gradually. With this interpretation, we combined the traditional RAAR and the diffusion model together to solve multi-channel Fourier phase retrieval. The general flowchart of our approach is depicted in Figure 2. Specially, the main steps for such conditional generation process are as follows.

1. We use the Tweedie's formula to push the intermediate generation $\boldsymbol{x}_{t+1}$ back to the clear image manifold of the diffusion model and achieve $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_{t+1})$;
2. The pulled-back point is the initialization to drive the traditional RAAR engine and we achieve $\hat{\boldsymbol{x}}_0'(\boldsymbol{x}_{t+1})$;
3. The updated estimation $\hat{\boldsymbol{x}}_0'(\boldsymbol{x}_{t+1})$ and the unconditional generative point $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_{t+1})$ is weighted to produce the new point $\hat{\boldsymbol{x}}_0'(\boldsymbol{x}_{t+1})$;
4. Finally, the point $\hat{\boldsymbol{x}}_0'(\boldsymbol{x}_{t+1})$ is remapped back to $\boldsymbol{x}_t$, which is supposed lying on the next-level noisy image manifold.

In this manner, the diffusion model serves as the prior of the color image solution and the traditional iterative engine is the driving force to produce next estimation with the measurement guidance consistency. To the best of our knowledge, this is the first work to solve color-scale Fourier phase retrieval with the combination of diffusion model and the traditional iterative engine.

2. Background

In this section, we introduce the related preliminaries of our approach for phase retrieval. The traditional iterative engine for phase retrieval and the diffusion model as prior for inverse problems will be reviewed.

2.1. Iterative Engine for Phase Retrieval

For projection-based approach of phase retrieval, the zero-padding is viewed as the known support for the unknown image. In this case, we denote the unknown truth image by $\tilde{\boldsymbol{x}}_0 = \mathrm{zpad}(\boldsymbol{x}_0)$, and its zero-valued set is

$$S := \{i | i \text{ is the index of the zero-padding region.}\}.$$

With such view, the noiseless Fourier phase retrieval can be cast as a two-set feasibility problem (Bauschke et al., 2002):

$$\text{find} \quad \tilde{\boldsymbol{x}}_0 \in \mathcal{X} \cap \mathcal{Y}, \tag{1}$$

where the measurement constraint set $\mathcal{Y} := \{\boldsymbol{x} | |\mathcal{F}(\boldsymbol{x})| = \boldsymbol{y} \in \mathbb{R}^{hr\times wr\times 3}\}$, and the prior constraint set $\mathcal{X} := \{\boldsymbol{x} | \boldsymbol{x}[j] = 0, j \in S, \text{and } \boldsymbol{x} \geq 0\}$. Here the indices of the three-channel image are transformed into one dimension for simple notation. The projections onto the two sets have closed-form expressions:

$$\mathcal{P}_{\mathcal{Y}}(\boldsymbol{z}) := \mathcal{F}^{-1}\left(\boldsymbol{y} \circ \frac{\mathcal{F}(\boldsymbol{z})}{|\mathcal{F}(\boldsymbol{z})|}\right),$$
$$\mathcal{P}_{\mathcal{X}}(\boldsymbol{z}) := \begin{cases} 0 & \text{if } j \in S \\ \max(\boldsymbol{z}[j], 0) & \text{otherwise.} \end{cases}$$

With the two projections, the Douglas-Rachford splitting (DRS) (Themelis & Patrinos, 2020) can be developed to solve the phase retrieval. DRS scheme generates the sequences $\{\boldsymbol{x}_k\}$ using the iteration:

$$\begin{aligned} \boldsymbol{x}_{k+1} = \mathcal{T}(\boldsymbol{x}_k) &:= \frac{1}{2}(\mathcal{R}_{\mathcal{X}}\mathcal{R}_{\mathcal{Y}} + I)(\boldsymbol{x}_k) \\ &= [I - \mathcal{P}_{\mathcal{X}} + \mathcal{P}_{\mathcal{X}}(2\mathcal{P}_{\mathcal{Y}} - I)](\boldsymbol{x}_k), \end{aligned} \tag{2}$$

where $\mathcal{R}_i = 2\mathcal{P}_i - I$ are the reflection operators.

There exists solution for noiseless case of (1). However, for noisy measurement, the DRS iterations exhibit oscillation phenomenon. To stabilize the iteration, RAAR iteration is proposed to solve noisy phase retrieval. The iteration reads as

$$\boldsymbol{x}_{k+1} = \beta\mathcal{T}(\boldsymbol{x}_k) + (1-\beta)\mathcal{P}_{\mathcal{Y}}(\boldsymbol{x}_k), \tag{3}$$

where β is the hyperparameter.

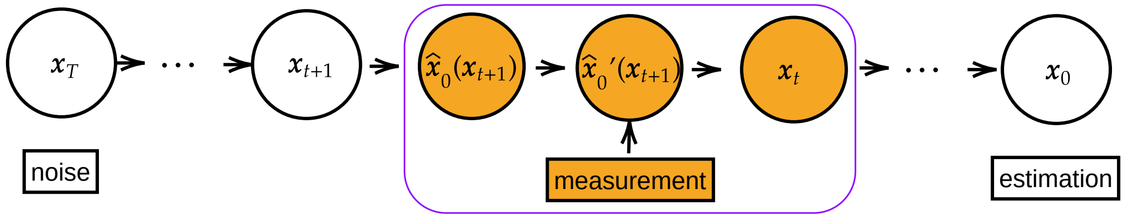


Figure 2. **The overview of our approach.** At each sampling step (in purple box), the Tweedie's formula is first executed and obtain $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_{t+1})$, then the measurement guidance is leveraged to achieve the updated point $\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_{t+1})$ with damping strategy. The update of the prediction $\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_{t+1})$ is implemented using the traditional iterative engine, such as RAAR (3).

2.2. Diffusion Model

Diffusion model is a powerful approach to generate a clean image from the underlying data distribution of a given dataset. The diffusion model comprises of two processes: *forward process* and *generative process*. The procedure of the forward process is to gradually perturb the clean data $\boldsymbol{x}_0$ with a series of pre-configured noise scales. From the continuous modeling perspective, the process is described using the stochastic differential equation (SDE):

$$d\boldsymbol{x}_t = f(t)\boldsymbol{x}_t dt + g(t)d\boldsymbol{w}_t, \tag{4}$$

where $\boldsymbol{w}_t$ is the standard Brownian process. Hence for a given clean image point $\boldsymbol{x}_0$, equation (4) generates a series of transition conditional distribution $p(\boldsymbol{x}_t|\boldsymbol{x}_0)$, which are a series of Gaussian distributions.

Ho et al. (2020) proposed a discrete scheme to generate these Gaussian transition conditional distributions $p(\boldsymbol{x}_t|\boldsymbol{x}_0)$. Ho et al. (2020) considered a sequence of positive noise scales $\{\beta_i\}_{i=1}^T$ and $0 < \beta_i < 1$. For each training data point $\boldsymbol{x}_0$, a Markov chain noisy degradation path $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ can be constructed such that $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t \boldsymbol{I})$. By induction, we have the degradation model $p(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_0, (1-\alpha_t)\boldsymbol{I})$, where $\alpha_t := \Pi_{j=1}^t(1-\beta_j)$.

The generative process is to revert the forward process, it obtains a clean image gradually from a Gaussian noise image $\boldsymbol{x}_T$. The reverse process is governed by the reverse SDE:

$$d\boldsymbol{x}_t = \left[f(t)\boldsymbol{x}_t - g^2(t)\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)\right]dt + g(t)d\overline{\boldsymbol{w}}_t, \tag{5}$$

where $d\overline{\boldsymbol{w}}_t$ corresponds the Brownian process running in backward time.

For the outstanding denoising diffusion probabilistic model (DDPM) (Ho et al., 2020), it uses a neural network $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)$ to estimate the scaled score function, *i.e.*, $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t) \simeq -\sqrt{1-\alpha_t}\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)$. There are two samplers, called DDPM and DDIM in the literature. As an efficient sampling, the DDIM is a non-Markov process to accelerate the sampling process (Song et al., 2021). DDIM is given by

$$\begin{aligned}\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) &= \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right)\\ \boldsymbol{x}_{t-1} &= \sqrt{\alpha_{t-1}}\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) + \sqrt{1-\alpha_{t-1}-\delta_t^2}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t) + \delta_t \boldsymbol{z},\end{aligned} \tag{6}$$

where

$$\delta_t = \eta\tilde{\beta}_t := \eta\sqrt{(1-\alpha_{t-1})/(1-\alpha_t)}\sqrt{(1-\alpha_t/\alpha_{t-1})}. \tag{7}$$

When set $\eta = 1$, it restores the DDPM, when set $\eta = 0$, it is DDIM without randomness in the iteration. Note that $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ is the predicted clean image $\boldsymbol{x}_0$ from $\boldsymbol{x}_t$, which is the expectation $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$. It is actually an application of the Tweedie's formula (Chung et al., 2022a).

2.3. Diffusion Model for Inverse Problem

Given a general inverse problem corrupted with Gaussian white noise, $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}_0) + \boldsymbol{n}$. Leveraging diffusion model as a prior, the natural way is to solve the conditional reverse SDE:

$$\begin{aligned}d\boldsymbol{x}_t = &\left[f(t)\boldsymbol{x}_t - g^2(t)(\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{y}|\boldsymbol{x}_t))\right]dt\\ &+ g(t)d\overline{\boldsymbol{w}}_t,\end{aligned} \tag{8}$$

where the Bayes' rule $\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t|\boldsymbol{y}) = \nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{y}|\boldsymbol{x}_t)$ is used. Based on the reverse SDE, the numerical conditional sampling method can be developed. The key issue is that there does not exist an analytical formulation for the likelihood term $p_t(\boldsymbol{y}|\boldsymbol{x}_t)$. To address the issue, there are two categories of works: 1) to adopt alternating projections onto the measurement subspace to guide the generation, such as DDRM (Kawar et al., 2022), DDNM (Wang et al., 2022), MCG (Chung et al., 2022b) and ILVR (Choi et al., 2021). The issue is that the projection requires calculation of the singular value decomposition (SVD), which may not feasible for a general operator. 2) to estimate the noisy likelihood with heuristic assumptions, such as (Chung et al., 2022a; Song et al., 2022), *etc*. There are two consecutive steps to accomplish the sampling. The unconditional sampling step is expressed as

$$x'_{t-1} = f(x_t, \epsilon_\theta(x_t, t)), \tag{9}$$

where f can be the DDPM or DDIM. The two kinds of guidance differ from each other the post-processing: 1) Update x'_{t-1} to match the measurement guidance by projection, such as (Chung et al., 2022b; Choi et al., 2021); 2) Perform gradient descent with data fidelity

$$x_{t-1} = x'_{t-1} - \eta_t \nabla_{x_t} \left\| y - A(\hat{x}_0(x_t)) \right\|_2^2, \tag{10}$$

where η_t is the tunable step-size for each iteration (Chung et al., 2022a).

3. Related Works

In this section, we introduce the Gaussian phase retrieval, as an extension to the Fourier phase retrieval, and the solving methods for phase retrieval.

3.1. Gaussian Phase Retrieval

As discussed in the introduction, the relative uniqueness of the Fourier phase retrieval results from the property of Fourier transform. The existence of equivalent solution leads to the difficulty of developing numerical methods. To diminish the set of equivalent solutions, Gaussian phase retrieval is proposed as a mathematical proxy to study the nonconvex phase retrieval problem. With this setting, the equivalent solution is only up to a constant phase shift (or a minus symbol difference for real image) (Candes et al., 2015; Candès et al., 2012). The condition on the complexity of measurement to ensure the uniqueness up to the constant phase shift is well studied, such as (Candes et al., 2015; Candès et al., 2012).

In terms of the numerical algorithms, optimizations built on different objective loss have been proposed (Candes et al., 2015; Chen & Candès, 2017; Wang et al., 2017). The global convergence is assured when a good enough initialization is provided (Candes et al., 2015; Ma et al., 2019). Fortunately, a close-to-the-truth initialization is provided by some initialization routines with enough measurement complexity. Similar to Gaussian phase retrieval, Fourier transform with randomized mask (Coded diffraction pattern) is proposed to diminish the gap between the Gaussian phase retrieval and the practical Fourier phase retrieval. However, due to the limitation on the resolution of the mask, coded diffraction pattern is only a proof-of-concept. Though its limited application, Gaussian phase retrieval is of high interests among the mathematicians (Luke, 2017). It is now a canonical problem to study the nonconvex continuous optimization.

3.2. Iterative Phase Retrieval

The Fourier phase retrieval is generally reformulated as a two-set feasibility problem. When the two sets are convex, the projection methods has been well understood in the literature (Bauschke et al., 2002). Several convergent methods have been proposed to solve the feasibility problem. These methods can be directly applied to phase retrieval without any modification, while losing the global convergence guarantee. Among the methods, DRS (2) is the outstanding solver, and the local convergence is assured. Actually, DRS is equivalent to ADMM iteration on the reformulation of the feasibility problem using indicator functions:

$$\begin{aligned} \min \quad & \mathbb{1}_{\mathcal{Y}}(z) + \mathbb{1}_{\mathcal{X}}(x) \\ \text{s.t.} \quad & z = x. \end{aligned}$$

Here the indicator function $\mathbb{1}_{\mathcal{X}}(x) = 0, \text{if } x \in \mathcal{X}$, otherwise, $\mathbb{1}_{\mathcal{X}}(x) = +\infty$. The DRS is also equivalent to the classical hybrid input-output (HIO) method with special hyperparameter configuration (Bauschke et al., 2002).

The feasibility problem yields empty intersection when the measurement is corrupted by noise. The direct application of DRS will lead to oscillation. To circumvent the oscillation, efforts have been made to diminish the effect of noise on the iterative methods. RAAR (3) introduced a weighted iteration to push the iteration close to the measurement constraints. Besides, there are other projection-based solvers, such as difference map (Fannjiang, 2014), graph projection splitting (Li & Zhao, 2020).

3.3. Deep Learning for Phase Retrieval

The supervised learning for image restoration has shown outstanding performance for certain degradations. The supervised learning is to learn an end-to-end mapping from the measurement to the clean image. Using a bulk of paired training data, the trained model achieved significant improvements over traditional methods in areas such as denoising, image deblurring, and general inverse problem, such as CT and MRI. The supervised learning for Gaussian phase retrieval has been developed (Yang et al., 2023; Chen et al., 2022; Kazemi et al., 2022; BaoShun & QiuSheng, 2022). The recovery improvement over traditional methods is promising, as the inference is fast. Manekar et al. (2020) proposed a passive loss, which is invariant to symmetries, for the supervised training to diminish the effect of trivial ambiguity. Cha et al. (2021) intimated the PhaseCut algorithm (Waldspurger et al., 2015) and formulated a novel loss function to train the networks. The trained model works for gray-scale images of small size, its application to color image has not explored yet.

The prediction mapping of the supervised learning is assumed one-to-one and well posed. However, such assumption is problematic for the Fourier phase retrieval. Even we provide the precise knowledge of the image support and the nonnegativity of the image, the existence of rotation ambiguity destroyed the assumptions. In this case, it is hard to learn en effective mapping for the Fourier phase retrieval in supervised manner.

3.4. Prior by Denoiser

Instead of employing the supervised learning for Fourier phase retrieval, the deep denoiser is adopted to improve the traditional optimization methods for phase retrieval:

$$\min \quad \mathcal{L} := l(\boldsymbol{x}; \boldsymbol{b}) + \mathcal{R}(\boldsymbol{x})$$

where l is the data fidelity term and $\mathcal{R}$ is a regularizer to promote the image statistics.

The general step to solve the optimization is to alternatively run the gradient descent using l, and then run the denoising step involving the regularizer $\mathcal{R}$. The denoising step is to solve a proximal problem:

$$D(\boldsymbol{z}) := \min_{\boldsymbol{x}} \quad \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \mathcal{R}(\boldsymbol{x}).$$

Metzler et al. (2018) proposed the Gaussian denoiser D to replace the proximal step.

Another choice of the regularizer is the Regularization-by-denoising (RED):

$$\mathcal{R}(\boldsymbol{x}) = \frac{\lambda}{2} \langle \boldsymbol{x}, \boldsymbol{x} - D(\boldsymbol{x}) \rangle.$$

It has been shown that if the denoiser D has the properties of local homogeneity and Jacobian symmetry, then the evaluation of the proximal operator at $\boldsymbol{x}$ required to solving

$$\boldsymbol{z} - \boldsymbol{x} + \lambda(\boldsymbol{z} - D(\boldsymbol{z})) = 0.$$

The properties rarely hold for common denoiser, such as the deep denoiser. Recent applications of RED to Fourier phase retrieval have validated the the performance improvement over the traditional iterative methods, such as (Metzler et al., 2016; 2018; Wang et al., 2020).

3.5. Generative Prior for Phase Retrieval

GAN-based prior. The GAN-based deep generative priors have be explored for the Gaussian phase retrieval problems. In these applications, the range constraint of the generative manifold is implicitly leveraged in the optimization. Such works include the general deep generative prior (Hand et al., 2018), the untrained neural network prior (Jagatap & Hegde, 2019; Li & Wang, 2022).

Diffusion model as a prior. The diffusion model can serve as a strong prior to solve inverse problem. The key idea is using the data fidelity to guide the unconditional generation process. The conditional generation process runs

$$\boldsymbol{x}_{t-1} = \underbrace{f(\boldsymbol{x}_t; \boldsymbol{\epsilon}_\theta, \boldsymbol{z})}_{\text{unconditional generation}} + \eta_t \underbrace{\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t)}_{\text{measurement guide}}, \tag{11}$$

where the unconditional generation can be any of sampling scheme, such as DDPM (Ho et al., 2020), DDIM (Song et al., 2021), *etc*. To run the iteration (11), one is required to derive the intractable $\log p(\boldsymbol{y}|\boldsymbol{x}_t)$. Using the decomposition relationship $p(\boldsymbol{y}|\boldsymbol{x}_t) = \int p(\boldsymbol{y}|\boldsymbol{x}_0) p(\boldsymbol{x}_0|\boldsymbol{x}_t) d\boldsymbol{x}_0$, Chung et al. (2022a) approximated

$$p(\boldsymbol{y}|\boldsymbol{x}_t) \simeq p(\boldsymbol{y}|\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]), \tag{12}$$

as only $p(\boldsymbol{y}|\boldsymbol{x}_0)$ is the known likelihood. The expectation $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$ can be estimated using the trained score network $\boldsymbol{\epsilon}_\theta$. The so-called diffusion posterior sampling (DPS) is applied to oversampled Fourier phase retrieval. However, it generally failed as illustrated in Figure 1. The existence of the trivial ambiguities and the nonconvexity of the function $p(\boldsymbol{y}|\boldsymbol{x}_0)$ account for the failure. Besides, the calculation of the gradient leads to back-propagation through the score network, which is computationally demanding.

4. Methodology

To address the difficulty of DPS for phase retrieval, we propose combining the diffusion prior and the traditional iterative engine for phase retrieval. The iterative engine can reconstruct each channel of the color image. The existence of the trivial rotation ambiguities lead to orientation misalignment, which leads to erroneous restoration. The combination approach leveraged the strong diffusion prior to rectify the orientation in the intermediate generation.

4.1. Geometry of the Diffusion and Generation Process

We first review the diffusion process, a series of noisy images are generated, in which the corrupted Gaussian noise level becomes larger gradually.

Geometric interpretation of the diffusion process (Chung et al., 2022b). Suppose that $\mathcal{M}_0 \subset \mathbb{R}^n$ is the data manifold, which denotes the set of all data points. Then the distribution of noisy data $p_t(\boldsymbol{x}_t) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_0) p(\boldsymbol{x}_0) d\boldsymbol{x}_0, p(\boldsymbol{x}_t|\boldsymbol{x}_0) \sim \mathcal{N}(\sqrt{\alpha_t}\boldsymbol{x}_0, (1-\alpha_t)\boldsymbol{I})$. Hence $p_t(\boldsymbol{x}_t)$ is concentrated on a manifold $\mathcal{M}_t := \{\boldsymbol{y} \in \mathbb{R}^n : d(\boldsymbol{y}, \sqrt{\alpha_t}\mathcal{M}_0) = r_t := \sqrt{1-\alpha_t}\sqrt{n-l}\}$, where l is the dimensionality of the manifold $\mathcal{M}_0$. The degradation path $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ can be interpreted as the transition from clean image manifold to noisy image manifold.

Geometric interpretation of the generation process. The generation process is to transit form the noisy image manifold $\mathcal{M}_T$ to the clean image manifold $\mathcal{M}_0$. We recall the

Algorithm 1 RAAR for Fourier phase retrieval.

Input: Iteration number T, parameter β, initialization $\boldsymbol{x}_0$ and measurement $\boldsymbol{y}$

Output: Estimated image $\boldsymbol{x}_0$.

1: Produce the zero-padded initialization $\text{zpad}(\boldsymbol{x}_0)$ and set $\boldsymbol{x}_T = \text{zpad}(\boldsymbol{x}_0)$

2: **for** $t = T : -1 : 1$ **do**

3: %% RAAR iteration:
$\boldsymbol{x}_{t-1} = [\beta(I + \mathcal{P}_\mathcal{X}(2\mathcal{P}_\mathcal{Y} - \boldsymbol{I})) - (2\beta - 1)\mathcal{P}_\mathcal{Y}]\,(\boldsymbol{x}_t)$

4: **end for**

5: Extract the centered image: $\boldsymbol{x}_0 = \text{revzpad}(\boldsymbol{x}_0)$

generation process of DDIM (DDPM is a special case of general DDIM). The DDIM consists of two steps:

1. Pull-back $\boldsymbol{x}_t \in \mathcal{M}_t$ to the clean image manifold $\mathcal{M}_0$ and get the predicted point $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$:
$$\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right)$$
2. Push point $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ to the less noisy manifold $\mathcal{M}_{t-1}$:
$$\boldsymbol{x}_{t-1} = \sqrt{\alpha_{t-1}}\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) + \sqrt{1-\alpha_{t-1}-\delta_t^2}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t) + \delta_t \boldsymbol{z}.$$

Hence the generation chain is
$$\boldsymbol{x}_T \to \cdots \to \boldsymbol{x}_{t+1} \to \hat{\boldsymbol{x}}_0(\boldsymbol{x}_{t+1}) \to \boldsymbol{x}_t \to \cdots \to \boldsymbol{x}_0.$$

4.2. Measurement Guidance

With such geometrical view, we propose leveraging the measurement condition to update the predicted point $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ and achieving a new point $\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t)$. To assure the closeness of the update to the measurement, we run traditional iterative engine to update the point $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ with a fixed iteration number T_{RAAR}. The remapping back to the manifold $\mathcal{M}_{t-1}$ is not changed. For better sampling efficiency, we adopted the DDIM iteration for the unconditional generation and we chose the RAAR method (Algorithm 1) to perform the measurement guidance for its better processing of noisy measurement than HIO. The overall algorithm is listed in Algorithm 2. Following such framework, for our approach, there are alternative options for the unconditional generative method and the traditional iterative engine.

Damping the update. The update scheme for unconditional $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ to $\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t)$ can be written as:
$$\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t) = \hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) + \underbrace{(\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t) - \hat{\boldsymbol{x}}_0(\boldsymbol{x}_t))}_{\text{ascent grad}}. \tag{13}$$

It is observed that directly using the updated point $\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t)$ increases the failure of the reconstruction. It indicts that the update shall be damped. Hence we propose the adaptive damped update
$$\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t) = \hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) + \sqrt{\alpha_t}(\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t) - \hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)). \tag{14}$$

Algorithm 2 Measurement-guided diffusion model for Fourier phase retrieval.

Input: Sampling iterations T, pretrained diffusion model ϵ_θ, hyperparameter β and T_{RAAR} of RAAR

Output: Estimated image $\boldsymbol{x}_0$.

1: Set $\boldsymbol{x}_T = \boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I})$

2: **for** $t = T : -1 : 1$ **do**

3: %% Prediction of clean image:
$\hat{\boldsymbol{x}}_0 = \frac{\boldsymbol{x}_t - \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)}{\sqrt{\alpha_t}}$

4: %% Run Algorithm 1 with T_{RAAR} iteration:
$\hat{\boldsymbol{x}}'_0 = \text{RAAR}(\hat{\boldsymbol{x}}_0, \beta, \boldsymbol{y}, T_{\text{RAAR}})$

5: %% Damping the update:
$\hat{\boldsymbol{x}}'_0 = (1 - \sqrt{\alpha_t})\hat{\boldsymbol{x}}_0 + \sqrt{\alpha_t}\hat{\boldsymbol{x}}'_0$

6: %% Push to $\mathcal{M}_{t-1}$ using DDPM or DDIM iteration:
$\boldsymbol{x}_{t-1} = \sqrt{\alpha_{t-1}}\hat{\boldsymbol{x}}'_0 + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t) + \sigma_t \boldsymbol{z}$

7: **end for**

Note that with t approaching to 0, α_t approaches to unity, and $\alpha_T \to 0$. Hence the gradient ascent contributes more and more when the generation process evolves.

Justification of the damped update. There are two reasons to showcase the possible failure of the undamped update. Regarding the unknown signal $\boldsymbol{x}$ is a random variable, the bias of the prediction $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ to solution $\boldsymbol{x}_0$ (up to possible ambiguities) decreases gradually. Though we can calculate the expectation of the distribution by Tweedie's formula, the distribution $p(\boldsymbol{x}_0|\boldsymbol{x}_t)$ is complicated and intractable. Its variance is expected increasing as time increases. Song et al. (2022) proposed the Gaussian approximation $p(\boldsymbol{x}_0|\boldsymbol{x}_t) \simeq \mathcal{N}(\hat{\boldsymbol{x}}_0, (1-\alpha_t)\boldsymbol{I})$. With this approximation, at time step t, the truth $\boldsymbol{x}_0$ should be approximated by
$$\boldsymbol{x}_0 = \hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) + \sqrt{1-\alpha_t}\boldsymbol{z}. \tag{15}$$

Therefore, we have the relationship
$$\begin{aligned}\boldsymbol{y} &= \mathcal{F}(\text{zpad}(\boldsymbol{x}_0)) + \boldsymbol{n} \\ &\simeq \mathcal{F}(\text{zpad}(\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) + \sqrt{1-\alpha_t}\boldsymbol{z})) + \boldsymbol{n}.\end{aligned} \tag{16}$$

From (16), the variance of $p(\boldsymbol{y}|\boldsymbol{x}_t)$ is larger at the beginning of the generation process. And it decreases gradually as the generation executes. In other word, the gradient ascent resulted from the measurement information should be damped to match the high variance at the beginning of the generation. Another reason comes from the *relative uniqueness* of phase retrieval problem. The erroneous alignment of $\hat{\boldsymbol{x}}'_0(\boldsymbol{x}_t)$ from iterative engine will lead itself off from the clean image manifold. To account for the possible mismatch, the damped gradient is adopted to impose strong diffusion prior at the beginning of the generation. Using the damped update, the failure of our combination approach is circumvented.

4.3. Discussion

Manifold projection does not contribute much. As discussed before, to some extent, the damped gradient ascent can avoid the erroneous alignment of the restoration at the beginning of the generation. If the orientation mismatch happens, the gradient is not tangent to the manifold $\mathcal{M}_0$. If we adopted the manifold constraint, we can compute the projected gradient on the tangent plane of $\mathcal{M}_0$. Motivated by (Chung et al., 2022b), the projection can be computed using the following PyTorch code.

```
x0_hat = (xt - np.sqrt(1-alpha[t])\
         model(xt,t))/np.sqrt(alpha[t])
# x0_hat_p is the return of RAAR
grad = (x0_hat_p - x0_hat).detach()
loss = (grad*x0_hat).sum()
grad_p = torch.autograd.grad(outputs=\
        loss,inputs=xt)[0]
# grad_p is the projection of the grad
# onto the tangent plane of the manifold
```

The projection will intrigue the back-propagation through the pretrained model, which is computationally demanding. The projection on the manifold does not improve the reconstruction significantly, see Appendix B.2. To trade off the computation cost and the performance, we did not consider the manifold constraint in the experiment.

Post-projection option underperforms Algorithm 2. Instead of performing the measurement guidance in $\mathcal{M}_0$, we also consider performing Algorithm 1 on the unconditional sampling $\boldsymbol{x}'_{t-1}$. Due to the execution order, this option is called *post-projection*. It is not reasonable to let the sampling $\boldsymbol{x}'_{t-1}$ match the measurement, as the sampling $\boldsymbol{x}'_{t-1}$ is farther away from $\boldsymbol{x}_0$ than that of the prediction $\hat{\boldsymbol{x}}_0(\boldsymbol{x}'_{t-1})$, especially at the beginning of the generation. See Section 5.2 for the ablation study.

5. Experiments

In this section, we conduct experiments to evaluate the performance of our approach on Fourier phase retrieval. Two datasets, including Flickr Faces High Quality (FFHQ) 256×256 and ImageNet 256×256 are considered. The pretrained diffusion models are directly downloaded from the open-source library without any refinement. Each test dataset contains 1K images. To quantitatively compare the performance, we report the reference-based PSNR/SSIM metric to measure the closeness to the original image, and the LPIPS metric to measure the perception quality of the restoration. There is no available deep learning approach tailed for multi-channel Fourier phase retrieval. The works, including prDeep (Metzler et al., 2018) and DeepITA (Wang et al., 2020), can not produce decent restoration with the color image deep denoiser, as the orientation misalignment is not addressed yet. The works based on DIP (Ulyanov et al., 2018) targeted vanishing losses, which does not imply the successful restoration due to the nonconvexity. Hence we compare our method to RAAR.

We first provide the quantitative results of our approach for the two datasets for noiseless and noisy measurement. Then we provide ablation studies to explore the effects on the performance with different options.

5.1. Quantitative Results

Suppose that $\boldsymbol{y} = |\mathcal{F}(\mathrm{zpad}(\boldsymbol{x}_0))| + \boldsymbol{n}$, where the noise $\boldsymbol{n} \sim \mathcal{N}(0, \sigma_y^2 \boldsymbol{I})$. Hence $\sigma_y = 0.00$ means noiseless measurement. In the default configuration, we select 1000 diffusion-step DDIM as the unconditional sampling, and the one-step RAAR as the inner iterative engine. The hyperparameter β is set to unity and 0.75 for noiseless and noisy cases respectively. We compare our method to DPS and RAAR. Due to the possible failure of all the three methods, for each test image, we run each method three times and take the best result as the final restoration. When calculating the quantitative metrics, the possible 180° rotation ambiguity is considered.

Table 1. Results of methods on the resized FFHQ.

Method	PSNR	SSIM	LPIPS
HIO	32.71	0.832	0.212
DIP	16.92	0.474	0.578
SIREN	15.54	0.292	0.610
E2E	14.25	0.273	0.506
CDM	13.11	0.300	0.470
Ours	**37.59**	**0.840**	**0.141**

Evaluation on $128\times$ 128 images. We initially compare DiffFPR with other methods using 100 noiseless images from the FFHQ dataset. The methods compared include the traditional HIO (hybrid input-output) method, neural network representation-based methods, and the end-to-end supervised method. In comparison to RAAR, the HIO method performed worse under noisy conditions. For neural network representation-based methods, we evaluate the deep image prior (DIP) (Ulyanov et al., 2018) and the SIREN network (Sitzmann et al., 2020), with the training loss taken from (Li et al., 2024). In these two methods, the network architecture acts as an implicit prior for the image. For end-to-end (E2E) supervised learning, the Unet architecture is employed. Additionally, we train a conditional diffusion model (CDM) similar to (Whang et al., 2022; Saharia et al., 2022) to address phase retrieval in a supervised learning context. As a simple demonstration, we resize the $256\times$ 256 images from FFHQ to $128\times$ 128, and the oversampling ratio 4 is considered. In this way, the measurement is of size 256×256. For the E2E and conditional diffusion model, we learn the mapping from the measurement to the zero-padded image. For supervised learning, the paired training dataset contains 1000 images. Table 1 shows the significant advantages of our method. The ambiguity solution still exists for SIREN and DIP, which accounts for the unreal restoration.

Evaluation on the 1K dataset. The high likelihood of failure for DPS happens for phase retrieval, as shown in Appendix B.1. Out of the 1K images in FFHQ and ImageNet, only 14.5% and 4.0% resulted in restoration with a PSNR exceeding 20dB, respectively. In contrast, RAAR achieved success rates of 52.4% and 48.2% for FFHQ and ImageNet, while Our method achieved a 100% success rate, highlighting its significant advantage over other methods. To make more fair comparison, DPS* is the average results of the selected image subset. See Table 2 for the results of phase retrieval (with oversampling ratio $r^2 = 4.0$) from noiseless and noisy measurement. The best performer is indicated in **bold** text.

Table 2. Quantitative results on the FFHQ and ImageNet dataset.

σ_y	Method	FFHQ			ImageNet		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
0.00	RAAR	22.50	0.601	0.508	21.80	0.546	0.509
	DPS*	24.85	0.704	0.237	22.89	0.676	0.343
	Ours	**37.24**	**0.902**	**0.087**	**31.76**	**0.751**	**0.204**
0.01	RAAR	12.74	0.250	0.731	12.80	0.220	0.705
	DPS*	25.89	0.736	0.228	21.52	0.653	0.358
	Ours	**29.17**	**0.774**	**0.197**	**25.94**	**0.663**	**0.277**
0.05	RAAR	12.51	0.198	0.752	12.66	0.173	0.722
	Ours	**22.15**	**0.429**	**0.458**	**19.81**	**0.343**	**0.504**

Overall the restoration performance almost decreases when noise level increases. It shows that the diffusion model is crucial and improve the RAAR significantly. The strong prior brought by diffusion model addresses the orientation mismatch problem of RAAR. See Figure 3 for the visualization comparison of the results. Our method produces very decent restoration. See Appendix B.3 for more visualization results. The results for oversampling ratio $r^2 = 2.0$ is provided in the Appendix B.2. For the more challenging scenario, the performance of ours is very promising.

Running time. Compared to the RAAR, our method requires extra computation to perform the diffusion-related step. The additional processing time varies based on the complexity of the diffusion model employed. See Table 3 for a comparison of the processing times required to handle a single image. Despite the extra computational overhead, this extra cost yields a notable enhancement in performance.

Table 3. Running time in (s).

Method	FFHQ	ImageNet
RAAR	0.075	0.075
Ours	1.089	4.482

Comparison of RAAR and HIO. While HIO was initially groundbreaking in addressing Fourier phase retrieval, the RAAR method has proven to be more effective in handling noisy scenarios. See Table 4 for a comprehensive comparison between RAAR and HIO within our framework. The table highlights RAAR's superiority over HIO, particularly in dealing with noise levels.

Table 4. Comparison of RAAR and HIO.

σ_y	Method	FFHQ			ImageNet		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
0.00	Ours-HIO	36.28	0.893	0.116	30.31	0.705	0.218
	Ours-RAAR	**37.24**	**0.902**	**0.087**	**31.76**	**0.751**	**0.204**
0.01	Ours-HIO	23.01	0.549	0.414	21.06	0.482	0.448
	Ours-RAAR	**29.17**	**0.774**	**0.197**	**25.94**	**0.663**	**0.277**
0.05	Ours-HIO	14.08	0.197	0.722	14.03	0.153	0.689
	Ours-RAAR	**22.15**	**0.429**	**0.458**	**19.81**	**0.343**	**0.504**

5.2. Ablation Study

In this section, we provide ablation studies on FFHQ dataset to explore the effects on the performance with different experimental configuration. First we illustrate the results when i) the post-projection option is adopted. Then we report the quantitative results to compare the performance of our approach with alternative iteration configuration,*i.e.*, ii) the iteration number of the outer sampling and the inner RAAR varies; For all experiments, the oversampling ratio is set to 4.

Table 5. Results of the ablation study on the FFHQ.

σ_y	Method	PSNR	SSIM	LPIPS
0.00	500-1	32.29	0.868	0.142
	500-2	44.44	0.948	0.056
	500-4	**49.98**	**0.956**	**0.043**
	1000-1	37.24	0.902	0.087
	1000-2	44.48	0.952	0.051
0.01	500-1	26.50	0.748	0.240
	500-2	28.81	**0.766**	**0.207**
	500-4	**28.82**	0.747	0.218
0.05	500-1	20.47	0.406	0.492
	500-2	**22.21**	**0.430**	**0.473**
	500-4	22.21	0.421	0.481

Result of ablation study (i). We only consider the noiseless case for this ablation study. To this end, with other part of our approach being fixed, we exchange the order of the measurement guidance and the unconditional sampling. We compare the *pre-projection* used in our approach to its alternative *post-projection*. Compared to ours, the PSNR/SSIM and LPIPS metrics of the post-projection degraded by 15.46dB/0.213 and 0.299 respectively. See Figure 4 for the results from the two options for four representative images. The orientation misalignment still exists for the post-projection, which leads to restoration failure. In contrast, ours produces decent restoration.

Result of ablation study (ii). The quantitative result is evaluated on the FFHQ dataset. The study is to investigate and compare their importance of the two iteration number, *i.e.*, the total diffusion steps T and the inner RAAR step T_{RAAR}, in our approach. See Table 5 for the comparison of our method with various iteration number configurations. The naming convention $m - n$ stands for performing Algorithm 2 with $T = m$ and $T_{\text{RAAR}} = n$. For noiseless case, fixing diffusion steps T, increasing T_{RAAR} from 1 to 4

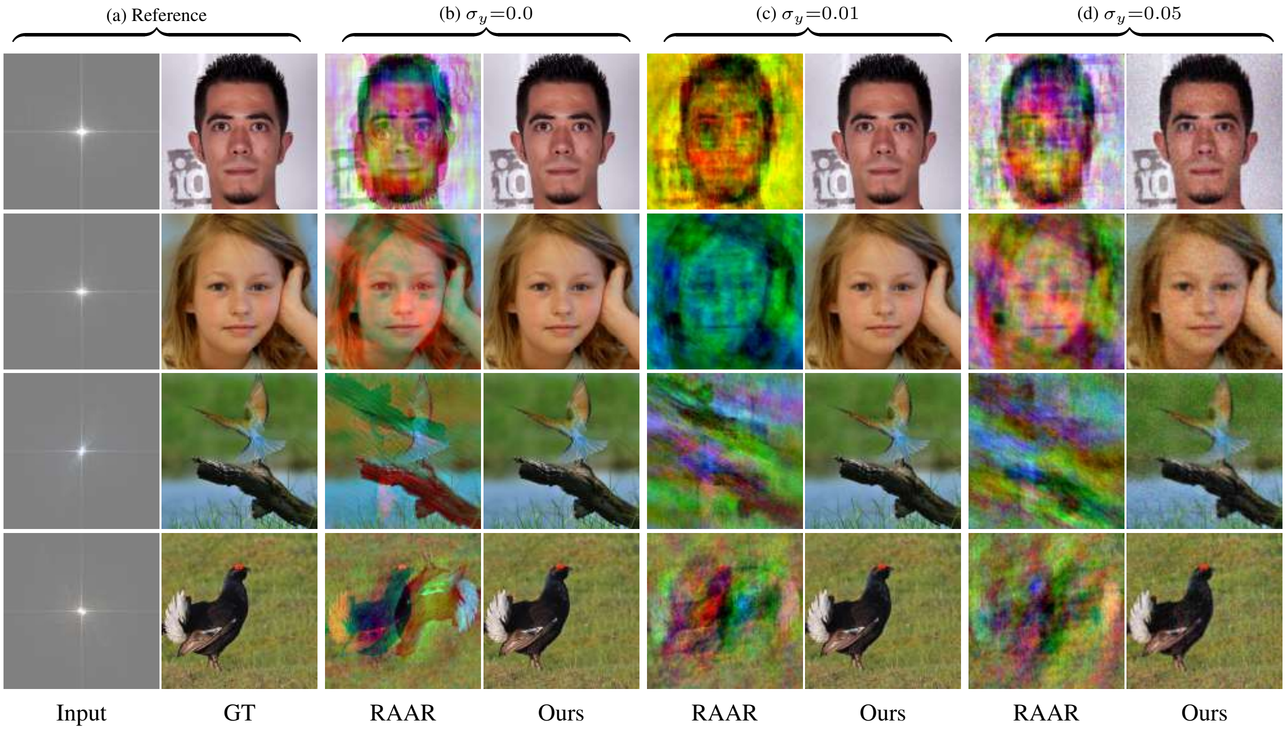


Figure 3. Visualization of the results from RAAR and Ours for FFHQ and ImageNet datasets with Gaussian noise (indicated as the top).

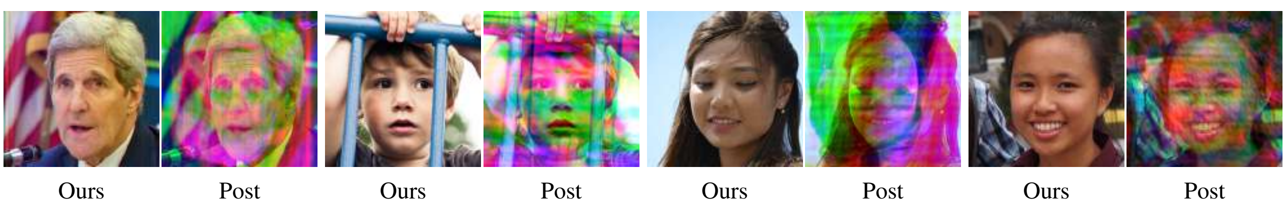


Figure 4. Comparison of the results from pre-/post-projection approaches for noiseless phase retrieval.

will gradually achieve performance improvement. For noisy case, the performance improvement achieve saturation for $T_{\text{RAAR}} = 2$. In contrast, fixing T_{RAAR}, the performance brought by more diffusion steps is slight. For noisy case, the configurations 500-2 and 1000-1 (see Table. 2) yield comparable results.

Limitation of our approach. The experiments have shown that our combined approach is less effective in noisy scenarios. As a result, the accumulated errors from the traditional algorithm may lead to the introduction of noise artifacts in the final restoration. To enhance the restoration process, it is crucial to address how to reduce noise in the final outcome. Another drawback of our study is its exclusive focus on synthetic data. Our future research will involve evaluating the performance of our method in real-world phase retrieval problems.

6. Conclusion

We proposed *DiffFPR*, a method for solving oversampled Fourier phase retrieval of color image by integrating the diffusion model and the traditional iterative engine. Motivated by the geometrical interpretation of the unconditional generation process, we consider exploiting the measurement guidance in the clean image manifold. To account for the bias of the prediction at the beginning of the sampling, we performed the damped gradient ascent. The proposed method avoided the orientation misalignment issue of the iterative engine, hence empowered the iterative engine significantly. As a comparison, the gradient descent manner of DPS failed for such highly nonconvex problem. Our approach illustrates the promising solving framework for other nonconvex problem with diffusion model.

Acknowledgements

We deeply appreciate the valuable and insightful feedback provided by the anonymous reviewers, which significantly enhanced our paper. We also extend our gratitude to the Academy for Multidisciplinary Studies of Capital Normal University for funding the computational resources; without their support, this paper would not have been possible.

Impact Statement

This paper presents an innovative approach to solving multi-channel Fourier phase retrieval by employing a diffusion model as the image prior. The effectiveness of our method demonstrates that integrating traditional optimization with diffusion generation iterations is successful for challenging ill-posed nonlinear issues. This approach exhibits considerable promise for other ill-posed challenges.

References

Balan, R. Reconstruction of signals from magnitudes of redundant representations: The complex case. *Foundations of Computational Mathematics*, pp. 1–45, apr 2015a.

Balan, R. The fisher information matrix and the CRLB in a non-Awgn model for the phase retrieval problem. In *2015 International Conference on Sampling Theory and Applications (SampTA)*. Institute of Electrical & Electronics Engineers (IEEE), may 2015b.

Balan, R., Casazza, P., and Edidin, D. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, may 2006.

BaoShun, S. and QiuSheng, L. Dualprnet: Deep shrinkage dual frame network for deep unrolled phase retrieval. *IEEE Signal Processing Letters*, 29:1177–1181, 2022.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. Phase retrieval, error reduction algorithm, and fienup variants: A view from convex optimization. *Journal of the Optical Society of America A*, 19(7):1334–1345, 2002. doi: 10.1364/josaa.19.001334. URL http://dx.doi.org/10.1364/josaa.19.001334.

Candès, E. J., Strohmer, T., and Voroninski, V. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, nov 2012. doi: 10.1002/cpa.21432. URL http://dx.doi.org/10.1002/cpa.21432.

Candès, E. J., Li, X., and Soltanolkotabi, M. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, oct 2015.

Candes, E. J., Li, X., and Soltanolkotabi, M. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

Cha, E., Lee, C., Jang, M., and Ye, J. C. Deepphasecut: Deep relaxation in phase for unsupervised fourier phase retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9931–9943, 2021.

Chen, H., Huang, L., Liu, T., and Ozcan, A. Fourier imager network (fin): A deep neural network for hologram reconstruction with superior external generalization. *Light: Science & Applications*, 11(1):254, 2022.

Chen, Y. and Candès, E. J. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Communications on pure and applied mathematics*, 70(5):822–883, 2017.

Choi, J., Kim, S., Jeong, Y., Gwon, Y., and Yoon, S. Ilvr: Conditioning method for denoising diffusion probabilistic models. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14347–14356. IEEE, 2021.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2022a.

Chung, H., Sim, B., Ryu, D., and Ye, J. C. Improving diffusion models for inverse problems using manifold constraints. volume 35, pp. 25683–25696, 2022b.

Fannjiang, A. Difference map for coded-aperture phasing has unique fixed point. Jun 2014. URL http://arxiv.org/abs/1406.2742v1.

Fienup, C. and Dainty, J. Phase retrieval and image reconstruction for astronomy. In Stark, H. (ed.), *Image Recovery: Theory and Application*. Academic Press, 1987.

Gonsalves, R. A. Perspectives on phase rertrieval and phase diversity in astronomy. In Marchetti, E., Close, L. M., and Véran, J.-P. (eds.), *Adaptive Optics Systems IV*, pp. 91482–91482. International Society for Optics and Photonics, SPIE-Intl Soc Optical Eng, jul 2014.

Goodman, J. W. *Introduction to Fourier Optics*. Roberts & Company Publishers, jan 2005.

Hand, P., Leong, O., and Voroninski, V. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems 31*, volume 31. NIPS Proceedings, 2018.

Hayes, M. H. The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(2):140–154, apr 1982. doi: 10.1109/tassp.1982.1163863. URL http://dx.doi.org/10.1109/tassp.1982.1163863.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. volume 33, pp. 6840–6851, 2020.

Jagatap, G. and Hegde, C. Algorithmic guarantees for inverse imaging with untrained network priors. *Advances in neural information processing systems*, 32, 2019.

Kawar, B., Elad, M., Ermon, S., and Song, J. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, 2022.

Kazemi, S., Yonel, B., and Yazici, B. Unrolled wirtinger flow with deep decoding priors for phaseless imaging. *IEEE Transactions on Computational Imaging*, 8:609–625, 2022.

Li, J. and Wang, C. Over-parameterized network solves phase retrieval effectively. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4133–4137. IEEE, 2022.

Li, J. and Zhao, H. Solving phase retrieval via graph projection splitting. *Inverse Problems*, 36(5):055003, 2020.

Li, J. and Zhou, T. On relaxed averaged alternating reflections (raar) algorithm for phase retrieval with structured illumination. *Inverse Problems*, 33(2):025012, 2017.

Li, T., Xu, Z., Chu, Y. S., Huang, X., and Li, J. Coordinate-based neural network for fourier phase retrieval. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2585–2589. IEEE, 2024.

Luke, D. R. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21(1):37–50, nov 2004.

Luke, D. R. Phase retrieval, what's new. *SIAG/OPT Views and News*, 25(1):1–5, 2017.

Ma, J., Xu, J., and Maleki, A. Optimization-based amp for phase retrieval: The impact of initialization and ℓ_2 regularization. *IEEE Transactions on Information Theory*, 65(6):3600–3629, 2019.

Manekar, R., Zhuang, Z., Tayal, K., Kumar, V., and Sun, J. Deep learning initialized phase retrieval. In *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*, 2020.

Metzler, C., Schniter, P., Veeraraghavan, A., et al. prdeep: robust phase retrieval with a flexible deep network. In *International Conference on Machine Learning*, pp. 3501–3510. PMLR, 2018.

Metzler, C. A., Maleki, A., and Baraniuk, R. G. Bm3d-prgamp: Compressive phase retrieval based on bm3d denoising. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2504–2508. IEEE, 2016.

Qian, J., Yang, C., Schirotzek, A., Maia, F., and Marchesini, S. Efficient algorithms for ptychographic phase retrieval. In *Contemporary Mathematics*. American Mathematical Society (AMS), 2014.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4713–4726, 2022.

Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J., and Segev, M. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, may 2015.

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *ICLR*, 2021.

Song, J., Vahdat, A., Mardani, M., and Kautz, J. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2022.

Themelis, A. and Patrinos, P. Douglas–rachford splitting and admm for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.

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.

Waldspurger, I., d'Aspremont, A., and Mallat, S. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1–2):47–81, dec 2015. doi: 10.1007/s10107-013-0738-9. URL `http://dx.doi.org/10.1007/s10107-013-0738-9`.

Wang, G., Giannakis, G. B., Saad, Y., and Chen, J. Solving most systems of random quadratic equations. *Advances in Neural Information Processing Systems*, 2017:1868–1878, 2017.

Wang, Y., Sun, X., and Fleischer, J. When deep denoising meets iterative phase retrieval. In *International Conference on Machine Learning*, pp. 10007–10017. PMLR, 2020.

Wang, Y., Yu, J., and Zhang, J. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2022.

Wen, Z., Yang, C., Liu, X., and Marchesini, S. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, 2012.

Whang, J., Delbracio, M., Talebi, H., Saharia, C., Dimakis, A. G., and Milanfar, P. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16293–16303, 2022.

Yang, Y., Lian, Q., Zhang, X., Zhang, D., and Zhang, H. Hionet: deep priors based deep unfolded network for phase retrieval. *Digital Signal Processing*, 132:103797, 2023.

A. Detailed Statistics of Restoration

It has been observed that phase retrieval solutions, such as RAAR and our proposed DiffFPR, may produce unrealistic results from a random initialization. We assess the stability, measured by the ratio where the reconstruction's PSNR exceeds 20 dB, of RAAR and DiffFPR. For each method, we run 20 runs with different initializations. The assessment is carried out on the first image of the 1K datasets of FFHQ and ImageNet for noiseless measurement. See Table 6 for the comparison. Our approach demonstrates notably superior performance compared to RAAR.

Table 6. Statistics of restoration from 20 runs.

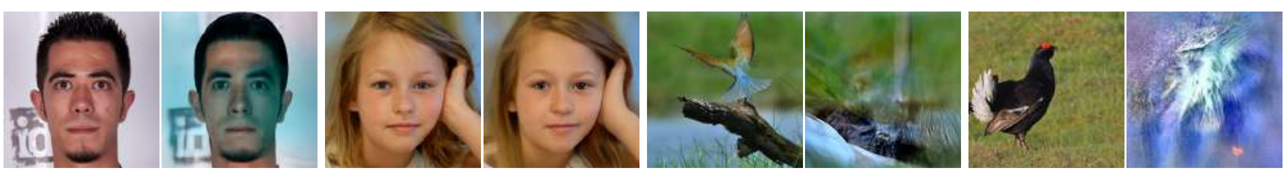
Method	Statistics	**FFHQ**			**ImageNet**		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
RAAR	mean	14.70	0.349	0.643	15.20	0.228	0.698
	best	17.15	0.398	0.612	16.53	0.250	0.675
	worst	11.58	0.303	0.671	14.32	0.215	0.712
	std	1.32	0.026	0.025	0.54	0.007	0.009
	# PSNR>=20dB		0			0	
Ours	mean	25.13	0.817	0.189	23.70	0.670	0.266
	best	44.31	0.990	0.006	26.32	0.862	0.063
	worst	13.44	0.436	0.540	22.11	0.544	0.373
	std	10.72	0.165	0.163	1.43	0.106	0.106
	# PSNR>=20dB		14			20	

B. More Visualization Results

In this section, we provide more visualization results to illustrate the performance of our method.

B.1. The Representative Results of DPS

First, we demonstrate the high failure probability of DPS. See Figure 5 for the reconstruction of the images for noiseless measurement. Note that the presented images are the best from three runs.



GT DPS GT DPS GT DPS GT DPS

Figure 5. Visualization of the results from DPS for noiseless phase retrieval.

B.2. Visualization Results for the Case of Oversampling Ratio 2.0

In this section, we provide the result of Fourier phase retrieval with oversampling ratio $r^2 = 2.0$. From the condition on the *relative uniqueness* of Fourier phase retrieval (Hayes, 1982), performing successful reconstruction requires that oversampling ratio is greater than 4.0. By this theorem, for the scenario of oversampling ratio being 2.0, the phase retrieval problem is very challenging. Hence we consider the noiseless measurement. See Table 7 for the result of our approach with the comparison to the RAAR. We also compare our baseline approach to its variant, including the different configuration of the iteration steps and the manifold constraint version of our approach. The notation "Ours-m-n" means that we run m-step diffusion steps and the inner RAAR is run with n steps. It shows that with the help of strong prior provided by the diffusion model, our approach outperformed RAAR by a large margin. Though Ours-Manifold yields the comparable result of Ours-1000-1, the manifold constraint requires the back-propagation through the network, it is computational demanding and memory intensive. See Figure 6 for the representative restoration results of ours and RAAR in the noiseless case.

Table 7. Results of noiseless phase retrieval with oversampling ratio 2.0 on the FFHQ.

Method	PSNR	SSIM	LPIPS
RAAR	11.48	0.18	0.778
Ours-500-1	19.98	0.583	0.397
Ours-1000-1	**24.26**	0.683	0.274
Ours-Manifold	24.18	**0.774**	**0.217**

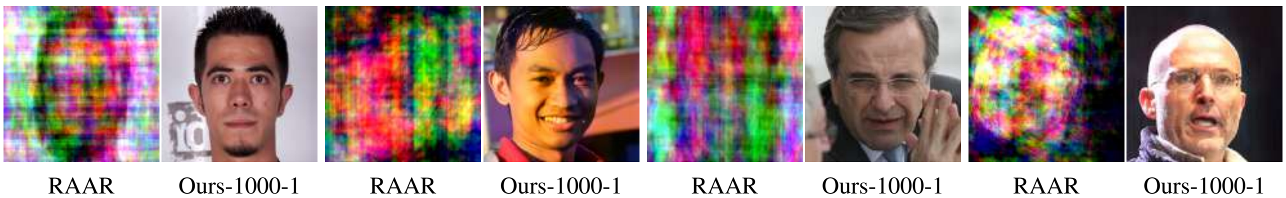


Figure 6. Visualization of the results from Ours for noiseless phase retrieval.

B.3. More Results of Our Methods on FFHQ and ImageNet

In this section, we illustrate more visualization of our methods, with comparison to the RAAR. See Figure 7 and 8 for the results on FFHQ and ImageNet datasets respectively. Our methods produce decent restoration, while RAAR produces the mis-aligned restoration without the help of the diffusion model.

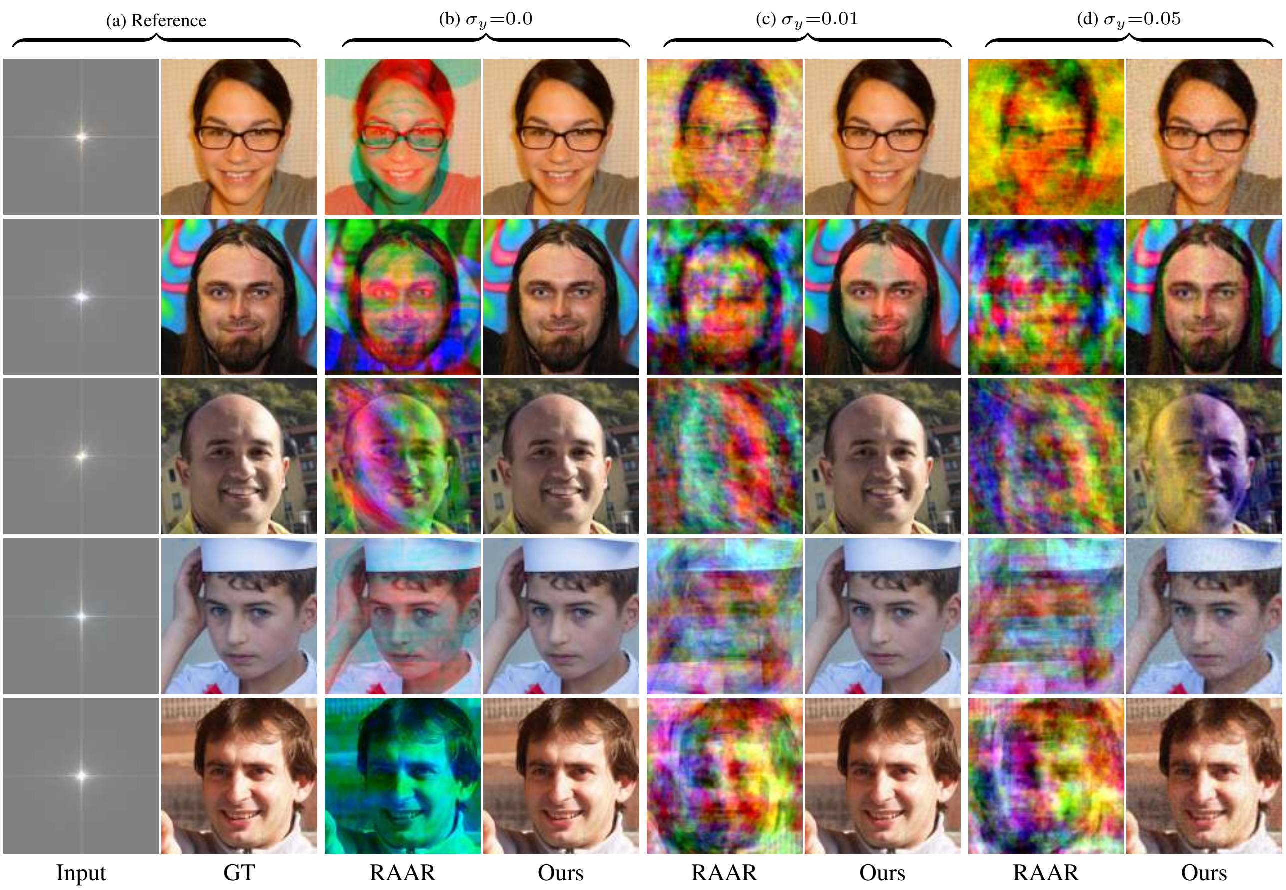


Figure 7. Visualization of the results from RAAR and Ours for FFHQ with Gaussian noise (indicated as the top).

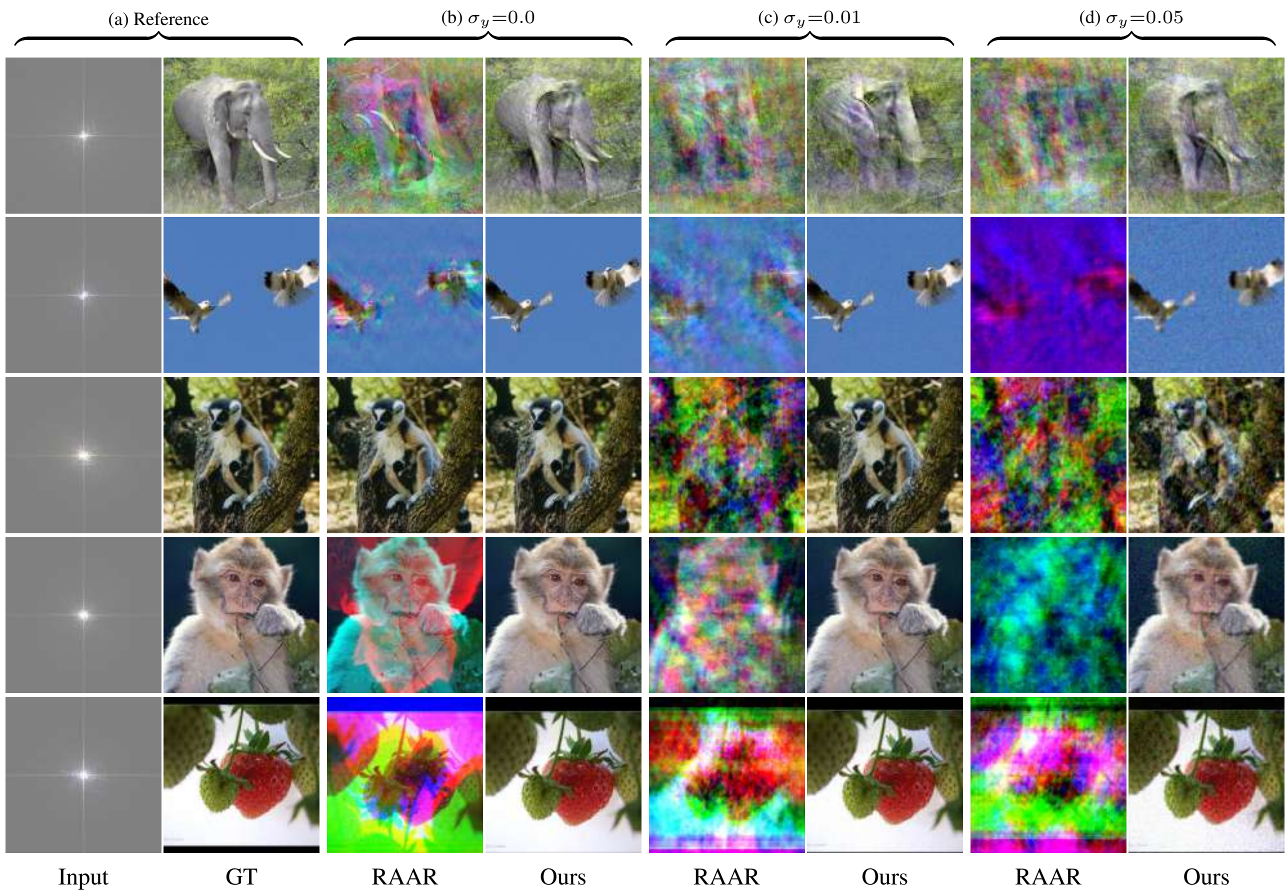


Figure 8. Visualization of the results from RAAR and Ours for ImageNet dataset with Gaussian noise (indicated as the top).