I2I-PR: Data-Driven Phase Retrieval Using Image-to-Image Diffusion Models

Mehmet Onurcan Kaya[†] and Figen S. Oktem[‡]
[†]Dept. of Applied Math. and Computer Science, Technical University of Denmark, Lyngby, 2800, Denmark
[‡]Dept. of Electrical Eng., Middle East Technical University (METU), Ankara, 06800, Turkey
Email: monka@dtu.dk, figeno@metu.edu.tr

***Abstract*—Phase retrieval (PR) involves recovering a signal from intensity-only measurements, and arises in many fields such as imaging, holography, microscopy, optical computing, and crystallography. Although there are several well-known phase retrieval algorithms, including classical iterative solvers, the reconstruction performance often remains sensitive to measurement noise and initialization. Recently, image-to-image (I2I) diffusion models have gained popularity in various image reconstruction tasks, yielding significant theoretical insights and practical advances. In this work, we propose I2I-PR, a data-driven phase retrieval approach that unrolls classical solvers into a deep iterative refinement framework inspired by modern I2I pipelines. Our method begins with an enhanced initialization stage, combining Hybrid Input-Output and Error Reduction methods with a novel acceleration mechanism to obtain multiple crude estimates. The reconstruction is then iteratively refined using these multiple estimates with a tailored image-to-image diffusion pipeline, while simultaneously enforcing measurement consistency. We demonstrate that I2I-PR outperforms both classical and deep learning-based methods, highlighting its potential for robust and efficient phase retrieval in various applications.**

***Index Terms*—Phase retrieval, deep learning, image-to-image diffusion models, unrolling, computational imaging.**

I. Introduction

Phase retrieval (PR) is a fundamental inverse problem in many scientific and engineering disciplines, where the goal is to reconstruct a signal using only intensity measurements such as Fourier intensities. This problem is critical in applications such as microscopy, holography, crystallography, and coherent diffraction imaging [1]–[3]. Mathematically, the PR problem involves reconstructing an unknown signal $\mathbf{x} \in \mathbb{C}^n$ from its noisy intensity measurements:

$$\mathbf{y^2} = |\mathbf{A}\mathbf{x}|^{\mathbf{2}} + \mathbf{w} \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{m \times n}$ is a known measurement operator, $\mathbf{y^2} \in \mathbb{R}^m$ denotes intensity measurements, and $\mathbf{w}$ represents measurement noise, often modeled as Poisson-distributed but approximated as Gaussian in many practical cases [4]. An important special case is Fourier PR, where $\mathbf{A}$ corresponds to the Fourier transform.

The primary challenge in PR stems from the loss of phase information, making the problem highly non-linear and ill-posed. Classical iterative methods such as Hybrid Input-Output (HIO) and Error Reduction (ER) rely heavily on projections onto constraint sets, but their sensitivity to noise and initialization can lead to suboptimal solutions, particularly in noisy and large-dimensional settings [5].

In response to these challenges, deep learning has emerged as a powerful tool for solving inverse problems, including PR [6], [7]. Data-driven approaches based on deep neural networks (DNNs) have shown promise in directly reconstructing images from measurements or refining initial estimates from classical solvers. However, these methods face significant limitations, including sensitivity to domain shifts between training and test data and lack of interpretability. In addition, they require extensive hyperparameter tuning, making real-world use difficult [6], [7]. Furthermore, such deep-learning based PR methods tend to suffer from over-smoothing, causing loss of fine details in reconstructions [8].

Among deep learning-based approaches, physics-based unrolling methods are the most promising one for PR, as they mimic classical iterative solvers through trainable network layers. However, they often suffer from long training times, as well as computational and memory inefficiencies, due to their sequential nature and extensive parameters [7], [9]. Classical diffusion pipelines also suffer from similar disadvantages since they initiate reconstruction from random noise instead of using a warm-start with an initial crude estimate, underutilizing the denoiser's capacity and leading to slower convergence [10]–[12]. These inefficiencies make their practical application challenging and highlight the need for more effective initialization strategies and computationally efficient models.

In contrast, recent image-to-image (I2I) diffusion models introduce a warm-start approach by initializing the reconstruction with a plausible estimate rather than pure noise. This refinement-based process simplifies learning by focusing on improving an initial estimate rather than generating an image from scratch. The effectiveness of I2I pipelines has been demonstrated in various applications, including deblurring and super-resolution [11]–[14].

In this work, we propose image-to-image-PR (I2I-PR), a data-driven phase retrieval approach that unrolls classical solvers into a deep iterative refinement framework inspired by modern I2I pipelines. More specifically, our method extends the Inversion by Direct Iteration (InDI) [12] framework to phase retrieval by introducing novel strategies for denoising, initialization, and measurement consistency. Our approach begins with an enhanced initialization stage, combining Hybrid Input-Output and Error Reduction methods with a novel acceleration mechanism to generate multiple crude estimates. These initial estimates provide diverse candidate solutions with

different artifacts. Unlike traditional diffusion methods that start from random noise, I2I-PR leverages these initial estimates, allowing for more effective utilization of the denoiser's capacity and a significant reduction in training time. By iteratively refining the reconstruction with learned I2I priors and enforcing measurement consistency, I2I-PR enhances both robustness and reconstruction quality.

Our hybrid framework combines the strengths of physics-based classical solvers and data-driven diffusion models, enabling improved reconstruction quality with training efficiency. Unlike conventional diffusion models that start from random noise/initialization, our method refines a coarse estimate obtained via classical solvers. Similar to unrolling-based methods, I2I-PR iteratively refines this estimate by decomposing the phase retrieval problem into smaller, less ill-posed subproblems. We demonstrate that I2I-PR outperforms both classical and deep learning-based PR methods, highlighting its potential for broader applications beyond phase retrieval, including other linear and nonlinear inverse problems.

II. Related Work

A notable I2I framework is the Inversion by Direct Iteration (InDI) method, which formulates a structured diffusion process from an initial estimate [12]. In its stochastic version, InDI integrates denoising diffusion probabilistic models into an iterative refinement framework, progressively improving image quality while incorporating stochastic perturbations to enhance robustness against degradation and noise.

The InDI method effectively mitigates the "regression to the mean" effect observed in traditional single-step regression models, which tend to produce oversmoothed outputs [15]. Instead, it incrementally refines images, similar to generative denoising diffusion models, but without requiring explicit knowledge of the degradation process. By progressively refining images over multiple steps, InDI helps address the challenges associated with the ill-posed nature of phase retrieval. By decomposing the original PR problem into a sequence of subproblems—each of which is less ill-posed than the full reconstruction problem—InDI reduces the overall ill-posedness. Unlike single-step regression models, which often struggle to recover fine details, InDI avoids excessive smoothing through its iterative update process [12].

Training in the InDI framework follows a structured degradation schedule, simulating noise and degradation levels computationally efficiently. The degradation process is defined as [12]:

$$\mathbf{x}_t = (1-t)\mathbf{x} + t\mathbf{z} + t\sigma_t\boldsymbol{\epsilon}, \tag{2}$$

where $\mathbf{x}$ is the clean image, $\mathbf{z}$ is the degraded input, $t \in [0,1]$ represents the noise level, σ_t is a time-dependent noise standard deviation, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ introduces stochasticity. The training objective for the denoiser model is given by:

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \mathbb{E}_{\mathbf{x},\mathbf{z}\sim p(\mathbf{x},\mathbf{z})} \left[\mathbb{E}_{t\sim p(t)} \left[\|\text{Denoiser}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \mathbf{x}\|_2^2 \right] \right], \tag{3}$$

minimizing the mean squared error between the clean image and its denoised counterpart across noise levels.

Once trained, the denoiser refines images iteratively, following the recurrence relation:

$$\widehat{\mathbf{x}}_{t-\tau} = \frac{\tau}{t}\,\text{Denoiser}(\widehat{\mathbf{x}}_t, t) + \left(1 - \frac{\tau}{t}\right)\widehat{\mathbf{x}}_t + (t-\tau)\sqrt{\sigma_{t-\tau}^2 - \sigma_t^2}\,\boldsymbol{\epsilon}, \tag{4}$$

where τ is a small backward step, ensuring gradual refinement. The noise variance reduction term $\sigma_{t-\tau}^2 - \sigma_t^2$ enables controlled denoising, enhancing image quality. Starting from $t = 1$ with $\widehat{\mathbf{x}}_1 = \mathbf{z} + \sigma_1\mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the process iteratively improves $\widehat{\mathbf{x}}_t$ until convergence.

The function $\text{Denoiser}(\widehat{\mathbf{x}}_t, t)$ estimates the expected clean image given the noisy input, approximating $\mathbb{E}[\mathbf{x}_{t-1} \mid \mathbf{x}_t]$. This iterative refinement structure allows InDI to outperform traditional single-step models in preserving fine details and adapting to varying degradation levels, making it well-suited for inverse problems such as phase retrieval.

III. Developed Method

Our method builds upon the Inversion by Direct Iteration (InDI) framework. We employ a hybrid iterative initialization stage that combines the Hybrid Input-Output (HIO) and Error Reduction (ER) methods with a novel acceleration mechanism to generate multiple crude estimates. These initial estimates provide diverse candidate solutions with different artifacts. The reconstruction is then iteratively refined using these multiple estimates with an image-to-image (I2I) diffusion pipeline based on InDI, which progressively improves the reconstruction while enforcing data consistency with the measured Fourier intensities.

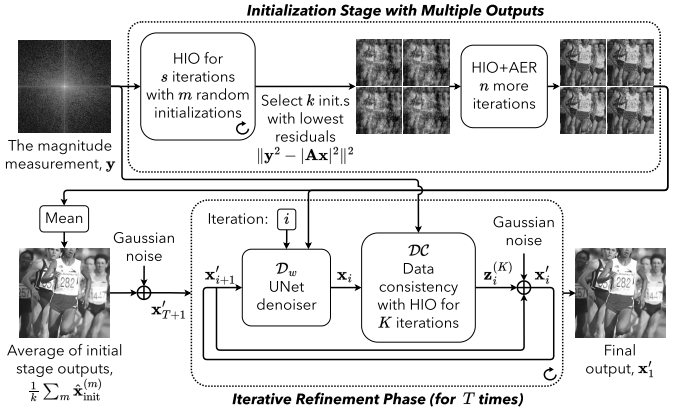

Fig. 1. The overall pipeline of I2I-PR.

A. Iterative Refinement Stage

In the iterative refinement stage, we learn and use an image-to-image translation model from the initialization output to the ground truth image using the InDI framework given in Eq. (4). However, we modify the standard InDI framework to better suit the phase retrieval problem. These modifications ensure a more effective refinement of the initialization estimates, leading to improved reconstruction quality and robustness.

In our approach, the initialization stage produces k distinct crude estimates $\{\hat{\mathbf{x}}_{\text{init}}^{(m)}\}_{m=1}^{k}$ from the same measurement $\mathbf{y}$.

These diverse reconstructions are critical because each estimate may correctly recover different regions of the image while also containing unique artifacts. Their mean,

$$\mathbf{z} = \frac{1}{k} \sum_{m=1}^{k} \hat{\mathbf{x}}_{\text{init}}^{(m)}, \tag{5}$$

serves as a preliminary, stable starting point for the subsequent refinement stage. Averaging multiple initial reconstructions, while beneficial for providing a stable starting point, inevitably leads to some information loss, as fine details and distinct features present in individual estimates may get blurred. To mitigate this, our denoiser is conditioned on the entire set of k initial estimates instead of solely relying on their average. In contrast to the denoiser component in the standard InDI formulation $\text{Denoiser}(\widehat{\mathbf{x}}_t, t)$ given in Eq. (4), our model uses

$$\text{Denoiser}(\widehat{\mathbf{x}}_t, t, \{\hat{\mathbf{x}}_{\text{init}}^{(m)}\}_{m=1}^{k}), \tag{6}$$

which effectively incorporates the unique contributions of each crude estimate into the refinement process.

At each refinement step, after the denoising operation, the model enforces measurement consistency through the HIO update. Formally, this update is performed as:

$$\text{HIO}(\text{Denoiser}(\widehat{\mathbf{x}}_t, t, \{\hat{\mathbf{x}}_{\text{init}}^{(m)}\}_{m=1}^{k}), \mathbf{y}), \tag{7}$$

which combines the Fourier magnitude of the denoised image with the measurement, ensuring that the refined image adheres to the physical constraints imposed by the data. Gaussian noise is then added according to the InDI formulation to simulate the reverse diffusion process. Simply applying a denoiser at each refinement step can lead to solutions that deviate from the measurements, especially in nonlinear inverse problems such as phase retrieval. To address this, we incorporate a data consistency enforcement block that ensures the refined image remains faithful to the observed measurements. By enforcing measurement consistency through the HIO update, we prevent divergence from the given Fourier magnitudes while still leveraging the denoiser for structured image priors.

The overall iterative update is expressed as:

$$\begin{aligned} \widehat{\mathbf{x}}_{t-\tau} = & \frac{\tau}{t} \text{ HIO} \left(\text{Denoiser}(\widehat{\mathbf{x}}_t, t, \{\hat{\mathbf{x}}_{\text{init}}^{(m)}\}_{m=1}^{k}), \mathbf{y} \right) \\ & + \left(1 - \frac{\tau}{t}\right) \widehat{\mathbf{x}}_t + (t-\tau)\sqrt{\sigma_{t-\tau}^2 - \sigma_t^2}\, \boldsymbol{\epsilon}, \end{aligned} \tag{8}$$

where τ is a small backward time-step, σ_t is the noise schedule, and $\boldsymbol{\epsilon}$ is sampled from a standard normal distribution. By discretizing the timestep interval $[0, 1]$ into T steps, we derive Algorithm 1. As shown in Fig. 1, this iterative refinement, performed over T discrete steps, gradually corrects errors and mitigates the ill-posedness inherent in phase retrieval.

A key advantage of this iterative refinement is its ability to decompose the overall PR challenge into a sequence of simpler denoising steps. By gradually reducing the noise and refining the estimate, the model can correct errors in a controlled manner and maintain stability even under severe noise conditions. It also achieves efficient use of the denoiser's model capacity and strikes a balance between denoising (regularization) and data consistency (fidelity) to produce high-quality results.

Our UNet denoiser is a central component of the refinement stage. It operates on $k+1$ input channels—one channel representing the current refined estimate and k channels corresponding to the crude initial estimates. This multi-channel input provides rich contextual information that guides the denoising process. Additionally, an auxiliary timestep input, encoded via positional encoding followed by a linear transformation, conditions the network on the current noise level. The network is designed to predict the residual—the difference between the noisy input and the desired clean image—enabling efficient, small-step corrections. To further enhance performance, our UNet incorporates attention mechanisms in both the downsampling and upsampling paths. These attention modules enable the network to focus on the most informative features, such as edges and textures, thereby preserving high-frequency details throughout the refinement process.

Algorithm 1 I2I-PR

Input: $\mathbf{y}$ (noisy magnitude measurements), $T, K, \beta, \{\sigma_i\}_{i=0}^{T}$ (fixed hyperparameters), $\boldsymbol{\lambda} \in \mathbb{R}^T$ (learnable vector, initialized with logarithmically increasing values)

Initialization:
$\{\hat{\mathbf{x}}_{\text{init}}^{(m)}\}_{m=1}^{k} \leftarrow$ Initialization stage($\mathbf{y}$)
$\mathbf{w} \leftarrow$ sample from $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$
$\mathbf{x}'_{T+1} \leftarrow \frac{1}{k}\sum_m \hat{\mathbf{x}}_{\text{init}}^{(m)} + \sigma_T \mathbf{w}$
Main loop:
for $i = T$ to 1 **do**
  $\mathbf{x}_i \leftarrow$ Denoiser($\mathbf{x}'_{i+1}$, i, $\{\hat{\mathbf{x}}_{\text{init}}^{(m)}\}_{m=1}^{k}$)
  $\mathbf{z}_i^{(0)} \leftarrow \mathbf{x}_i$
  $\mathbf{y}_i{}' \leftarrow \boldsymbol{\lambda}_i \mathbf{y} + (1 - \boldsymbol{\lambda}_i)|\mathbf{A}\mathbf{z}_i^{(0)}|$
  for $k = 1$ to K **do**
    $\mathbf{z}_i^{(k)'} \leftarrow \mathbf{A}^\dagger \left(\mathbf{y}_i{}' \odot \frac{\mathbf{A}\mathbf{z}_i^{(k-1)}}{|\mathbf{A}\mathbf{z}_i^{(k-1)}|} \right)$
    $\gamma \leftarrow$ indices where $\mathbf{z}_i^{(k)'}$ violates spatial constraints
    $\mathbf{z}_i^{(k)}[n] \leftarrow \begin{cases} \mathbf{z}_i^{(k)'}[n] & n \notin \gamma, \\ \mathbf{z}_i^{(k-1)}[n] - \beta\, \mathbf{z}_i^{(k)'}[n] & n \in \gamma. \end{cases}$
  $\boldsymbol{\epsilon} \leftarrow$ sample from $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$
  $\mathbf{x}'_i \leftarrow \frac{1}{i}\mathbf{z}_i^{(K)} + \left(1 - \frac{1}{i}\right) \mathbf{x}'_{i+1} + \frac{i-1}{T}\sqrt{\sigma_{i-1}^2 - \sigma_i^2}\, \boldsymbol{\epsilon}$
return $\mathbf{x}'_1$

The training of our model follows the InDI strategy given in Eq. (3), which relies on a carefully defined degradation/noising process. During training, the model is exposed to simulated noisy images generated by a fixed degradation schedule that gradually transitions from a clean image to a highly degraded one. This process ensures that the model learns to reverse the degradation over multiple timesteps. By aligning the network's behavior during both training and testing, this framework enhances performance and computational efficiency. Extensive experiments have demonstrated that careful tuning of hyperparameters—such as the number of timesteps T, the number of inner iterations K, the regularization parameter β, and the

noise schedule $\{\sigma_i\}_{i=0}^{T}$—is critical for balancing convergence speed and reconstruction accuracy.

B. Initialization Stage

Inherent nonlinearity of PR makes the reconstruction highly sensitive to initialization. To improve robustness, we adopt a hybrid strategy that combines HIO with ER and an acceleration mechanism, as described in [4]. Initially, m different random HIO initializations are run for s iterations to explore the solution space in parallel. The k estimates with the lowest residuals, $\|\mathbf{y} - |\mathbf{A}\mathbf{x}|\|_2^2$, are selected for further refinement.

These chosen estimates then undergo additional HIO and ER cycles for n iterations, alternating between applying the HIO constraint to enforce consistency with the measured Fourier magnitudes and the ER constraint to directly minimize the image-domain error. This cyclic alternation helps to prevent the reconstruction from getting trapped in local minima by balancing aggressive exploration (via HIO) with fine-tuned refinement (via ER). During the ER phases, an acceleration step is periodically applied to speed up convergence. Specifically, the accelerated ER (AER) method given in Algorithm 2 leverages a convex combination of the current and previous estimates, modulated by a scaling factor ζ. Here, $\mathcal{P}_F$ denotes the projection onto the Fourier magnitude constraint set, while $\mathcal{P}_S$ represents the projection onto the image-domain constraint set (e.g., enforcing spatial priors such as known object support). At each acceleration step, the algorithm computes the difference between two successive projections, $\mathbf{x}'_n$ and $\mathbf{x}''_n$, forms an intermediate point $\mathbf{c}_n$, and extracts a directional vector $\mathbf{a}$. The local progression distance is then used to update the estimate as follows:

$$\mathbf{x}_{n+1} = \mathbf{c}_n + \zeta r\, \mathbf{a} \quad \text{where} \quad r = \frac{1}{2}\|\mathbf{x}'_n - \mathbf{x}''_n\|. \tag{9}$$

This momentum-like adjustment not only helps the algorithm escape stagnation but also accelerates convergence, leading to lower residuals and higher-quality reconstructions.

Algorithm 2 Proposed accelerated ER (AER) algorithm

for $n = 1$ to K **do**
  $\mathbf{x}'_n \leftarrow \mathcal{P}_F \mathbf{x}_n$
  $\mathbf{x}''_n \leftarrow \mathcal{P}_S \mathbf{x}'_n$
  if $n \equiv -1 \pmod{t}$ **then**
    $\mathbf{c}_n \leftarrow \frac{1}{2}(\mathbf{x}'_n + \mathbf{x}''_n)$
    $\mathbf{a} \leftarrow \frac{\mathbf{c}_n - \mathbf{c}_{n-t}}{\|\mathbf{c}_n - \mathbf{c}_{n-t}\|}$
    $r \leftarrow \frac{1}{2}\|\mathbf{x}'_n - \mathbf{x}''_n\|$
    $\mathbf{x}_{n+1} \leftarrow \mathbf{c}_n + \zeta r\, \mathbf{a}$
  else
    $\mathbf{x}_{n+1} \leftarrow \mathbf{x}''_n$

IV. Experiments

The performance of our method for the Fourier PR problem is evaluated through numerical simulations using a large image dataset. We analyze generalization capacity and computational cost, comparing reconstruction performance against both classical and state-of-the-art phase retrieval methods. Noisy Fourier measurements are simulated using Eq. (1), with average SNR values presented in Table I (SNR $= 10\log(\||\mathbf{F}\mathbf{x}|^2\|_2 / \|\mathbf{y}^2 - |\mathbf{F}\mathbf{x}|^2\|_2)$). To ensure uniqueness (aside from trivial ambiguities), we use an oversampled DFT with an oversampling rate of $m = 4n$ [16].

Our method is trained exclusively on natural images, using a dataset of 44,000 images: 200 training and 100 validation images from the Berkeley Segmentation Dataset (BSD) [17], 41,400 images from ImageNet [18], and 2,300 images from the Waterloo Exploration Database [19]. We optimize the MSE loss using decoupled weight decay regularization [20], cosine annealing, and linear warming [21]. The model is implemented in PyTorch and trained on an NVIDIA A100 80GB PCIe GPU over 60 hours (27 epochs). Our training follows the InDI framework, where a structured degradation process progressively refines noisy images into clean reconstructions. Our method is trained and tested for a fixed noise level of $\alpha = 3$.

To assess generalization, we evaluate on both natural images and other type of images corresponding to the *unnatural* image dataset of prDeep. The test dataset, as used in [8], [22], contains 236 images (230 natural and 6 *unnatural*), including 200 images from BSD, 24 from Kodak [23], and additional images from [4]. All test images are 256×256 with pixel values in $[0, 255]$. Evaluating on both image types allows us to measure the model's ability to reconstruct unseen samples with varying characteristics.

In the initialization stage, HIO is run with $m = 100$ random initializations for $s = 50$ iterations. The $k = 10$ reconstructions with the lowest residuals, $\|\mathbf{y}^2 - |\mathbf{F}\mathbf{x}|^2\|_2$, are selected for further refinement using AER+HIO for $n = 1700$ iterations with a scaling factor $\zeta = 0.6$. In the iterative stage, hyperparameters are set as $K = 5$, $\beta = 0.9$, and $\sigma_i = 1$ for all i. The learnable vector $\boldsymbol{\lambda} \in \mathbb{R}^T$ is initialized with logarithmically increasing values from $10^{-0.64}$ to $10^{-0.10}$, and we set $T = 4$.

We evaluate reconstruction accuracy using PSNR and SSIM [24]. For comparison, results are obtained on the same test dataset using prDeep [4], HIO [25], DIR [8], and PnP HIO [22], [26]. Table I reports the average performance over 236 test images from 5 Monte Carlo runs. Our method consistently outperforms all baselines in both PSNR and SSIM, with significant gains over classical HIO. The results after the initialization stage further highlight the substantial improvement achieved through iterative refinement.

Table I also presents the reconstruction performance separately for natural and other images. Even though our network is trained exclusively on natural images, it achieves the highest PSNR for both image types, demonstrating strong generalization. However, SSIM scores for other (unnatural) images are occasionally lower, suggesting room for improvement in capturing specific textural features. This discrepancy can be attributed to phase retrieval ambiguities, particularly spatial circular shift ambiguity, which affects reconstructions of unnatural images such as "E. Coli" and "Yeast." While our model inherently mitigates trivial phase shift ambiguity

via realness and positivity assumptions, we do not use ground truth alignment to resolve spatial circular shifts. This choice ensures a more realistic evaluation setting, as such ambiguities remain unresolved in practical imaging applications. The lower SSIM observed in some cases highlights the challenge of reconstructing highly structured images where fine details play a significant role in perceived similarity.

The effectiveness of our approach is visually demonstrated in Fig. 2, where our method significantly reduces HIO artifacts and better preserves image details compared to prior methods. By explicitly considering the perception-distortion tradeoff, our approach minimizes the smoothing artifacts common in phase retrieval techniques [8], leading to sharper and more perceptually realistic reconstructions. Compared to purely optimization-based methods, our approach effectively balances distortion minimization with detail preservation, ensuring that reconstructions retain high-frequency textures while suppressing artifacts.

TABLE I
AVERAGE RECONSTRUCTION PERFORMANCES FOR 236 TEST IMAGES ACROSS 5 MONTE CARLO RUNS.

Method	**PSNR (dB)** ↑			**SSIM** ↑			**Runtime**
	Overall	**Natural**	**Other**	**Overall**	**Natural**	**Other**	**(s)** ↓
HIO [27]	18.92	18.89	20.34	0.43	0.43	0.43	0.27
prDeep [4]	22.06	22.09	20.91	0.59	0.59	0.54	59.41
DIR [8]	22.87	22.85	23.50	0.68	0.68	0.71	21.72
PnP HIO [22], [26]	23.92	23.92	23.98	0.70	0.70	0.69	24.35
Initialization stage	20.17	20.12	22.09	0.51	0.51	0.54	0.90
I2I-PR	26.78	26.85	24.18	0.73	0.73	0.61	1.11

Ground truth prDeep [4] DIR [8] Init. stage I2I-PR

Fig. 2. The outputs of various algorithms for the "Cameraman" test image subjected to $\alpha = 3$ noise (SNR=31.61dB).

V. CONCLUSION

In this paper, we introduced I2I-PR, a data-driven phase retrieval approach that unrolls classical solvers into a deep iterative refinement framework inspired by image-to-image pipelines. By incorporating advanced initialization strategies, measurement consistency constraints, and a tailored diffusion process, I2I-PR effectively refines multiple initial estimates, leading to improved reconstruction quality and computational efficiency. Unlike traditional diffusion-based methods that start from random noise, our approach leverages informative initial estimates to better utilize the denoiser's capacity and accelerate training. Experimental results demonstrate that I2I-PR outperforms both classical and deep learning-based phase retrieval methods. Beyond phase retrieval, our framework has the potential to generalize to broader inverse problems in computational imaging.

ACKNOWLEDGMENT

This study was funded by Scientific and Technological Research Council of Turkey (TUBITAK) under the Grant Number 120E505.

REFERENCES

[1] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.

[2] J. Dong, L. Valzania, A. Maillard, T.-a. Pham, S. Gigan, and M. Unser, "Phase retrieval: From computational imaging to machine learning," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 45–57, 2023.

[3] J. R. Fienup, "Phase retrieval algorithms: A personal tour," *Applied Optics*, vol. 52, no. 1, pp. 45–56, 2013.

[4] C. Metzler, P. Schniter, A. Veeraraghavan, and R. Baraniuk, "PrDeep: Robust phase retrieval with a flexible deep network," in *ICML*, 2018.

[5] S. Marchesini, "Invited article: A unified evaluation of iterative projection algorithms for phase retrieval," *Review of scientific instruments*, vol. 78, no. 1, 2007.

[6] S. López-Tapia, R. Molina, and A. Katsaggelos, "Deep learning approaches to inverse problems in imaging: Past, present and future," *Digital Signal Processing*, vol. 119, p. 103 285, 2021.

[7] K. Wang, L. Song, C. Wang, *et al.*, "On the use of deep learning for phase recovery," *Light: Science & Applications*, vol. 13, no. 1, 2024.

[8] Ç. Işıl, F. S. Oktem, and A. Koç, "Deep iterative reconstruction for phase retrieval," *Applied Optics*, vol. 58, no. 20, pp. 5422–5431, 2019.

[9] R. Heckel, M. Jacob, A. Chaudhari, O. Perlman, and E. Shimron, "Deep learning for accelerated and robust mri reconstruction," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 37, no. 3, pp. 335–368, 2024.

[10] G. Liu, H. Sun, J. Li, F. Yin, and Y. Yang, "Accelerating diffusion models for inverse problems through shortcut sampling," in *IJCAI*, 2024.

[11] A. Bansal, E. Borgnia, H.-M. Chu, *et al.*, "Cold diffusion: Inverting arbitrary image transforms without noise," in *NeurIPS*, 2023.

[12] M. Delbracio and P. Milanfar, "Inversion by direct iteration: An alternative to denoising diffusion for image restoration," *TMLR*, 2023.

[13] C. Saharia, W. Chan, H. Chang, *et al.*, "Palette: Image-to-image diffusion models," in *ACM SIGGRAPH*, 2022.

[14] J. Whang, M. Delbracio, H. Talebi, C. Saharia, A. G. Dimakis, and P. Milanfar, "Deblurring via stochastic refinement," *CVPR*, 2022.

[15] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," in *NeurIPS*, 2022.

[16] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 30, no. 2, pp. 140–154, 1982.

[17] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *ICCV*, 2001.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *CVPR*, 2009.

[19] K. Ma, Z. Duanmu, Q. Wu, *et al.*, "Waterloo Exploration Database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, 2017.

[20] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *ICLR*, 2017.

[21] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *ICLR*, 2017.

[22] Ç. Işıl and F. S. Oktem, "Model-based phase retrieval with deep denoiser prior," in *Imaging and Applied Optics Congress*, 2020.

[23] R. W. Franzen, "Kodak images," *http://r0k.us/graphics/kodak*, 2013.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[25] J. R. Fienup, "Reconstruction of an object from the modulus of its fourier transform," *Optics Letters*, vol. 3, no. 1, pp. 27–29, 1978.

[26] Ç. Işıl and F. S. Oktem, "Deep plug-and-play hio approach for phase retrieval," *Applied Optics*, vol. 64, no. 5, A84–A94, 2025.

[27] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.