Proceedings of Machine Learning Research vol 107:1–31, 2021

# Phase Retrieval with Holography and Untrained Priors: Tackling the Challenges of Low-Photon Nanoscale Imaging

**Hannah Lawrence**[*1] HANLAW@MIT.EDU
**David A. Barmherzig**[*2] DBARMHERZIG@FLATIRONINSTITUTE.ORG
**Henry Li**[3] HENRY.LI@YALE.EDU
**Michael Eickenberg**[2] MEICKENBERG@FLATIRONINSTITUTE.ORG
**Marylou Gabrié**[†2,4] MGABRIE@FLATIRONINSTITUTE.ORG

*1. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*
*2. Center for Computational Mathematics, Flatiron Institute, New York, New York, USA*
*3. Program in Applied Mathematics, Yale University, Connecticut, USA*
*4. Center for Data Science, New York University, New York, New York, USA*
† *Corresponding author,* * *These authors contributed equally*

## Abstract

Phase retrieval is the inverse problem of recovering a signal from magnitude-only Fourier measurements, and underlies numerous imaging modalities, such as Coherent Diffraction Imaging (CDI). A variant of this setup, known as holography, includes a reference object that is placed adjacent to the specimen of interest before measurements are collected. The resulting inverse problem, known as holographic phase retrieval, is well-known to have improved problem conditioning relative to the original. This innovation, i.e. Holographic CDI, becomes crucial at the nanoscale, where imaging specimens such as viruses, proteins, and crystals require low-photon measurements. This data is highly corrupted by Poisson shot noise, and often lacks low-frequency content as well. In this work, we introduce a dataset-free deep learning framework for holographic phase retrieval adapted to these challenges. The key ingredients of our approach are the explicit and flexible incorporation of the physical forward model into an automatic differentiation procedure, the Poisson log-likelihood objective function, and an optional untrained deep image prior. We perform extensive evaluation under realistic conditions. Compared to competing classical methods, our method recovers signal from higher noise levels and is more resilient to suboptimal reference design, as well as to large missing regions of low frequencies in the observations. Finally, we show that these properties carry over to experimental data acquired on optical wavelengths. To the best of our knowledge, this is the first work to consider a dataset-free machine learning approach for holographic phase retrieval.

## 1. Introduction

Phase retrieval is a nonlinear inverse problem that arises ubiquitously in imaging sciences, and has gained much recent attention (Shechtman et al., 2015). In this work we focus on a practical instance of the problem that arises in Coherent Diffraction Imaging (CDI). Here, *holographic phase retrieval* consists of recovering an image $\boldsymbol{X}_0 \in \mathbb{R}^{m \times n}$ from a set of squared Fourier transform magnitudes

$$\boldsymbol{Y} = |F(\boldsymbol{X}_0 + \boldsymbol{R}_0)|^2, \tag{1}$$

where $F$ denotes an oversampled Fourier transform operator and $\boldsymbol{R}_0 \in \mathbb{R}^{m \times n}$ is a known reference image whose support does not intersect the support of $\boldsymbol{X}_0$. The known reference image $\boldsymbol{R}_0$ distin-

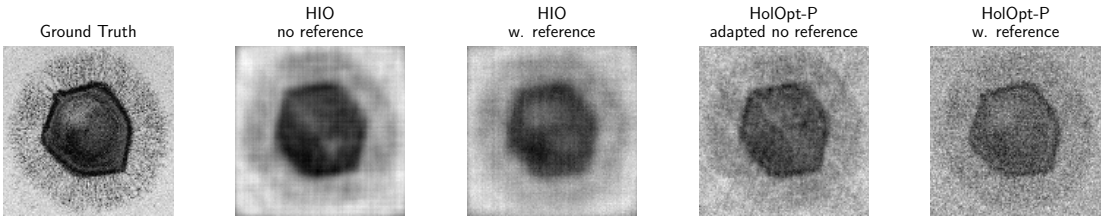



Figure 1: The advantage of using a reference for phase retrieval at $N_p = 1$ photon/pixel. Two algorithms, HIO and our HolOpt-P, are applied to reconstruct from Fourier magnitude data of image alone (classical phase retrieval), and image with reference (holographic phase retrieval). The setting and reference are given on Figure 2(*b*). Comparing between algorithms, observe the higher-quality reconstruction when the reference is present. (VIRUS image courtesy of Ghigo et al. (2008))

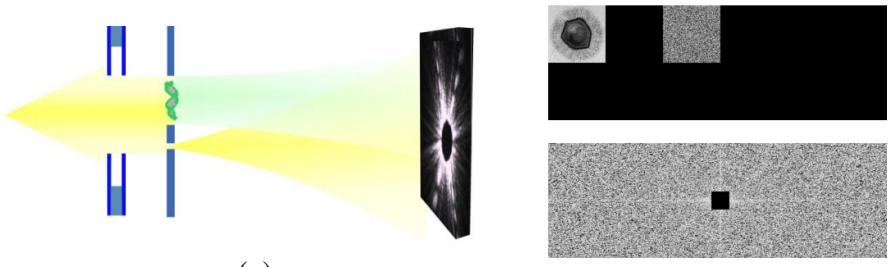
guishes holographic phase retrieval from the (classical) phase retrieval setting, where the goal is to recover $\boldsymbol{X}_0$ from $|F(\boldsymbol{X}_0)|^2$ alone. We focus entirely on the holographic version of the problem in realistic conditions: with high noise levels and missing low-frequency data. The advantage provided by the holographic reference is briefly illustrated on Figure 1.

In the remainder of this introduction we will situate this problem in the context of Coherent Diffraction Imaging, review related works, and list our contributions. Section 2 describes our setup in detail, and Section 3 our reconstruction strategy. In Section 4, we describe extensive experiments and compare to several baseline methods. Section 5 presents a validation in an optical laser CDI experiment. Section 6 concludes with a general discussion.

### 1.1. Holographic Coherent Diffraction Imaging and phase retrieval

Coherent Diffraction Imaging (CDI) is a scientific imaging technique used for resolving nanoscale scientific specimens, such as viruses, proteins, and crystals (Miao et al., 1999). In CDI, an image is sought to be reconstructed from X-ray diffraction measurements recorded on a CCD detector plane. By the *far-field approximation* of optical theory, these measurements are approximately proportional to the squared magnitude values of the Fourier transform of the electric field within the diffraction area. Thus, the specimen structure (e.g., its electron density) can be determined, in principle, by solving the phase retrieval problem. Holographic CDI is a popular setup to perform CDI experiments in which the object undergoing diffraction physically consists of a specimen together with a "reference", i.e. a portion of the object *a priori* known. This setup is illustrated in Figure 2. The inclusion of a reference in the CDI setup both enhances the quality of image reconstruction, and greatly simplifies the analysis and solution of the corresponding phase retrieval problem (Marchesini et al., 2008; Guizar-Sicairos and Fienup, 2007; Barmherzig et al., 2019).

Nevertheless, holographic CDI remains challenging in practice. Due to the quantum mechanical nature of photon emission and absorption, CDI measurements are inherently corrupted by *Poisson shot noise*. The severity of this noise corruption is inversely proportional to the strength of the X-ray source in use, which is in turn quantified via $N_p$, the number of photons per pixel reaching the detector plane. Nanoscale applications of CDI often necessitate imaging in the *low-photon regime*, where measurements are highly corrupted by noise. CDI measurements are also typically lacking low-frequency data, due to the presence of a *beamstop* apparatus which occludes direct measurement of these values (He et al., 2015; Latychevskaia, 2019; Barmherzig et al., 2020).



(*a*) (*b*)

Figure 2: Holographic CDI schematic. The upper portion of the diffraction area contains the specimen of interest $\boldsymbol{X}_0$, and the adjacent portion consists of a known "reference" $\boldsymbol{R}_0$. The recorded data $\boldsymbol{Y}$ has its low frequencies occluded by a beamstop. (Image courtesy of Saliba et al. (2012).)

The holographic phase retrieval problem is commonly solved by *inverse filtering* (Gabor, 1948; Kikuta et al., 1972), which amounts to solving a structured system of linear equations. While straightforward, this method is not well-suited for noisy data. *Wiener filtering* (Gorkhover et al., 2018) is a variant on this method with some denoising ability. Yet Wiener filtering is derived to account for an *additive* noise model — an assumption which is not true for Poisson shot noise at low photon counts, and only holds as an approximation at high photon count levels (Salditt et al., 2020). Moreover, these methods do not account explicitly for missing low-frequency data, and require a minimum separation between the specimen and the reference objects, the *holographic separation condition* (see Section 4.3). The most popular algorithm for the classical phase retrieval problem is the Hybrid Input-Output (HIO) algorithm (Fienup, 1978), which is based on an alternating projection scheme. This method can be modified to the holographic setting by adding an additional projection step to enforce the reference constraints, which greatly improves the algorithm's performance Barmherzig et al. (2019).

### 1.2. Related work

**Machine learning for inverse problems** Increasing research effort has been devoted to addressing inverse problems, even beyond phase retrieval, with deep learning approaches (see Ongie et al. (2020) for a recent review). Supervised strategies can be broadly divided into four main categories: end-to-end methods, (e.g. McCann et al. (2017)), "unrolling" algorithms (e.g. Meinhardt et al. (2017)), pretrained image denoisers (e.g. Romano et al. (2017)), and learned generative models as highly informative priors (e.g. Tramel et al. (2016)). All of these approaches require a training set, containing either matched signal-observation pairs or simply typical signals, with a large number of data points. The reconstruction improves drastically when this information is available. However, it is often unrealistic to assume such prior knowledge on the measured signal. Recently, for the reconstruction of images, it was found that untrained generative neural networks with appropriate architectures can still be efficient priors (Lempitsky et al., 2018; Heckel and Hand, 2019). By adjusting their parameters to fit a single output observation, they do not require a training set, and instead encourage naturalistic images due to architecture alone. In this paper we focus on the last approach, as it is widely applicable to image data and satisfies our requirement of applicability in a realistic setting.

**Machine learning for phase retrieval** More specifically, several variants of the phase retrieval problem have received attention in the context of machine learning for inverse problems. Compressive Gaussian phase retrieval, where one observes the amplitude of random complex Gaussian projections of the signal, is a popular setting in the machine learning community. It is easier than the Fourier phase retrieval problem and often more amenable to theoretical analysis (see e.g. Aubin et al. (2020)). For this version of the problem, trained generative models such as Generative Adversarial Networks (Shamshad and Ahmed, 2018; Hand et al., 2018), as well as untrained priors (Jagatap and Hegde, 2019), were found to be very effective on machine learning toy datasets. An increasing number of works now consider the more realistic problem of Fourier phase retrieval. Using pre-trained Gaussian denoisers and iterative algorithms, Deep-prior-based sparse representation (Shi et al., 2020), prDeep (Metzler et al., 2018) and Deep-ITA (Wang et al., 2020b), are solutions robust to noise in the case where the corruption is small enough to be approximately Gaussian. The end-to-end solution investigated by Uelwer et al. (2019) features some robustness to Poisson shot noise, but struggles to generalize to complicated datasets. Meanwhile, the "physics-informed" architecture of Goy et al. (2018), which includes information about the data generating process, is shown to perform well on realistic signals at very low photon counts, but requires a few thousand training examples. Closer to our work, Wang et al. (2020a) proposed a U-net and Bostan et al. (2020) tested the deep decoder, both untrained neural networks, for Fourier phase retrieval, but did not consider the holographic setting. To the best of our knowledge, Rivenson et al. (2018) is the only proposition considering holography, showing that deep neural networks trained end-to-end on a dataset of a few hundred images lead to state-of-the-art performance.

**Other optimization approaches** The use of auto-differentiation for Fourier phase retrieval was initiated by Jurling and Fienup (2014), while the convenience of deep learning packages was exploited later (Nashed et al., 2017; Kandel et al., 2019). Thibault and Guizar-Sicairos (2012) derived conjugate gradients to optimize the likelihood for classical (non-holographic) CDI, following either the Poisson or the Euclidean metric. Recently, Barmherzig and Sun (2020) pointed at the potential of likelihood optimization for holographic phase retrieval.

### 1.3. Our contributions

We address the holographic phase retrieval problem in the low-photon regime using a Poisson maximum likelihood framework and recent insights in machine learning for inverse problems. Our strategy combines three key ideas: (i) a realistic physical noise model for CDI, (ii) auto-differentiation and efficient optimization readily available in a package like PyTorch (Paszke et al., 2019), and lastly (iii) the option to add a neural network prior. We (a) compare these methods to baselines, exploring several experiment challenges; (b) demonstrate significant improvements at different noise levels; (c) investigate the impact of missing low-frequency data on our methods, and show that ours are more robust than baseline methods; (d) investigate the impact of the distance between object image and reference image on reconstruction quality, showing that our proposed method can easily deal with distances below the *holographic separation condition*; (e) vary the oversampling rate of observation, observing a graceful degradation of reconstruction with decreasing samples; and (f) perform a comparison of our different methods on an experimental data set in the optical range. Finally, we (g) provide a Python package[1] to run our implementation.

1. https://github.com/marylou-gabrie/deep-phase-retrieval-holo

## 2. Holographic CDI Setup

The data generation process mimics the key components of a holographic diffraction experiment as realistically as possible, namely by including two crucial ingredients: the Poisson shot noise model and the beamstop occluding low-frequency measurements.

**Coherent diffraction imaging** Let $\boldsymbol{Z} \in \mathbb{R}^{m \times n}$ represent a real $m \times n$-pixel image. As explained in Section 1.1, the recorded CDI measurements can be approximated by the square of the oversampled Fourier transform magnitude of the object image. Here, we assume an oversampling factor of two, which is the minimum oversampling factor theoretically required for perfect reconstruction in the noiseless setting (Hayes, 1982). Let $F : \mathbb{R}^{m \times n} \to \mathbb{C}^{2m \times 2n}$ be the doubly oversampled discrete Fourier transform operator. $F(\boldsymbol{Z})$ can be implemented as the discrete Fourier transform of a zero-padded version of $\boldsymbol{Z}$. Let $\tilde{\boldsymbol{Z}} = \left(\begin{array}{c|c} \boldsymbol{Z} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{0} \end{array}\right) \in \mathbb{R}^{2m \times 2n}$. Then let $F(\boldsymbol{Z}) = \mathcal{F}(\tilde{\boldsymbol{Z}})$, where $\mathcal{F}$ is the discrete Fourier transform operator. The intensity distribution at the detector is defined as $I(\boldsymbol{Z}) = |F(\boldsymbol{Z})|^2$ (where the absolute value here is understood in the pointwise sense).

**Beam stopping** A beam stopping mask, or "beamstop", is defined as $\boldsymbol{B} \in \{0, 1\}^{2m \times 2n}$ such that it takes the value 0 in a region of low frequency and 1 everywhere else. The beam-stopped intensity image can then be written as $I(\boldsymbol{Z}) \odot \boldsymbol{B}$, where $\odot$ represents pointwise multiplication.

**Measurement process** Let $N_p > 0$ represent the expected mean number of photons incident per detector pixel. Then $(2m \times 2n)N_p$ is the expected total number of photons incident on the detector. The measurement data vector $\boldsymbol{Y} \in \mathbb{R}^{2m \times 2n}$ is set to

$$\boldsymbol{Y} \sim \frac{C}{N_p} \text{Poisson}\left(\frac{N_p}{C} I(\boldsymbol{Z}) \odot \boldsymbol{B}\right). \tag{2}$$

The constant $C$ is equal to the sum of all square Fourier magnitudes over the detector. The inner normalization constant $N_p/C$ ensures that the simulated setting corresponds on average to the measurement of $N_p$ expected photons per pixel. The outer normalization constant is applied to make $\boldsymbol{Y}$ be of the same order of magnitude as $I(\boldsymbol{Z}) \odot \boldsymbol{B}$.

**Holography setup** We structure $\boldsymbol{Z}$ into an unknown object $\boldsymbol{X}$, and a known reference $\boldsymbol{R}$. The setting we will use throughout this paper is as follows (see also Figure 2). Let $\boldsymbol{X}, \boldsymbol{R} \in \mathbb{R}^{m \times m}$, and set $\boldsymbol{X}_0 = (\boldsymbol{X}|\boldsymbol{0}_{m \times m}|\boldsymbol{0}_{m \times m}), \boldsymbol{R}_0 = (\boldsymbol{0}_{m \times m}|\boldsymbol{0}_{m \times m}|\boldsymbol{R})$ with $\boldsymbol{X}_0, \boldsymbol{R}_0 \in \mathbb{R}^{m \times 3m}$. Then $\boldsymbol{Z} = \boldsymbol{X}_0 + \boldsymbol{R}_0 = (\boldsymbol{X}|\boldsymbol{0}_{m \times m}|\boldsymbol{R})$. The region of zeros $\boldsymbol{0}_{m \times m}$ separating object and reference represents the *holographic separation condition*. It is not necessary for our proposed methods (see Section 4.3), but required for several baseline methods. Thus to ensure a fair comparison, the separation setting will be our standard setting.

## 3. A reconstruction strategy adapted to low-photon CDI

### 3.1. HolOpt-P and HolOpt-P-DD

We propose to maximize the likelihood of the measurements $\boldsymbol{Y}$ given the underlying image $\boldsymbol{X}$ and the CDI model above. This objective involves the likelihood of the Poisson-distributed measurements, accounting for the nature of noise in the low-photon regime, as well as the full forward model (including reference and beamstop): $\hat{\boldsymbol{X}} = \arg\max_{\boldsymbol{X}} \log p(\boldsymbol{Y}|\boldsymbol{X}, N_p)$, where the distribution of $\boldsymbol{Y}$ conditional on $\boldsymbol{X}$ is given by Equation 2. Replacing the expression of the Poisson distribution

and dropping constants yields

$$\hat{\boldsymbol{X}} = \arg\max_{\boldsymbol{X}} \sum_{ij|(\boldsymbol{B})_{ij}=1} Y_{ij} \log I\left((\boldsymbol{X}|\boldsymbol{0}_{m\times m}|\boldsymbol{R})\right)_{ij} - I((\boldsymbol{X}|\boldsymbol{0}_{m\times m}|\boldsymbol{R}))_{ij} \tag{3}$$

where the sum is taken only over non-zero entries of the beamstop mask. The optimization of this objective is performed directly using gradient ascent in PyTorch.

We investigate two strategies. The optimization is done either directly on the pixels of $\boldsymbol{X}$, as in Equation 3, or on the parameters of a deep decoder neural network prior encoding $\boldsymbol{X}$ (Heckel and Hand, 2019). We refer to these two variants as HolOpt-P for holographic Poisson likelihood optimization, and HolOpt-P-DD for holographic Poisson likelihood optimization with a deep decoder.

### 3.2. The deep decoder

The deep decoder belongs to the class of untrained image priors: neural networks with image-shaped outputs trained by gradient descent to output one single image. The architecture of the network imposes an inductive bias favoring natural image statistics. In a sense, the architecture itself has been trained by decades of engineering in image processing.

**Deep decoder architecture** The deep decoder essentially consists of an alternation of two operations — convolutions with filter size of $1 \times 1$ pixels, and upsampling by a factor of 2 using bilinear interpolation. The input is a randomly initialized image of smaller size. In order to end up with a specific output image size, either the input image size or the number of layers $d$ is adjusted. It should be noted that pixels are spatially coupled only through the upsampling layers, while the $1 \times 1$ convolutions are pixel-wise linear transforms shared among all pixels. This weight-sharing allows to reduce the number of free parameter to mitigate overfitting (Heckel and Hand, 2019).

Let $c_i$ represent the number of channels at layer $i$ and let $\vartheta_i \in \mathbb{R}^{c_{i+1}\times c_i \times 1 \times 1}$ represent the convolution kernels at layer $i$. Denote by $\mathrm{conv}_{\vartheta_i}$ the typical deep learning convolution with $\vartheta_i$, by $\mathrm{up}_2$ the bilinear upsampling operation, and by $\mathrm{relu}(x) = x\mathbf{1}_{x>0}(x)$. Then we can define one component as

$$\mathrm{block}_i := \mathrm{up}_2 \circ \mathrm{relu} \circ \mathrm{conv}_{\vartheta_i}$$

and the full network as

$$\mathrm{net} := \mathrm{block}_d \circ \cdots \circ \mathrm{block}_1.$$

For an input image $z \in \mathbb{R}^{1\times c_1 \times \kappa \times \lambda}$, let

$$\sigma(\vartheta, z) = \sigma(\vartheta_d, \ldots, \vartheta_1, z) = \mathrm{net}(z),$$

where $\vartheta = (\vartheta_d, \ldots, \vartheta_1)$ collects all the convolution parameters. For HolOpt-P-DD, we set $\boldsymbol{X} = \sigma(\vartheta, z)$ and train all $\vartheta_i$ in $\vartheta$. Here the objective can be re-written as

$$\hat{\vartheta} = \arg\max_{\vartheta} \sum_{ij/(\boldsymbol{B})_{ij}=1} Y_{ij} \log I((\sigma(\vartheta, z)|\boldsymbol{0}_{m\times m}|\boldsymbol{R}))_{ij} - I((\sigma(\vartheta, z)|\boldsymbol{0}_{m\times m}|\boldsymbol{R}))_{ij}. \tag{4}$$

The reconstructed image is then the output of the deep decoder $\hat{\boldsymbol{X}} = \sigma(\hat{\vartheta}, z)$ after training on a single magnitude image $\boldsymbol{Y}$.

**Deep decoder depth** In our reconstruction experiments, we observe the number of channels of the convolutional filters to only marginally change the outcome of the reconstruction, while varying

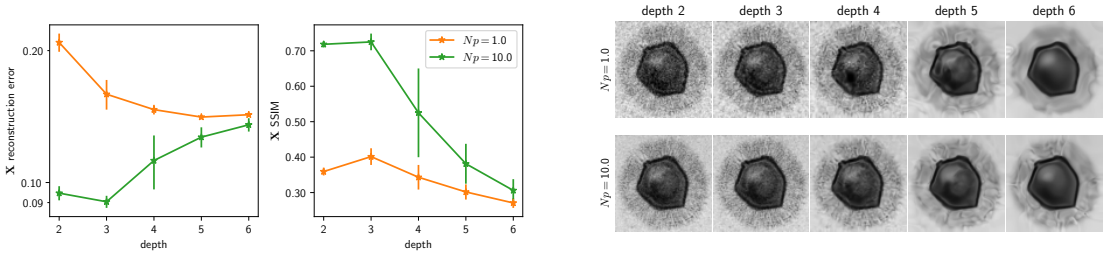



Figure 3: Impact of depth of a deep decoder prior on the reconstruction illustrated on VIRUS. Left: Relative reconstruction mean squared errors and structure similarity indices with ground truth as a function of depth. Right: Best reconstruction out of 4 runs.

the depth of the prior trades off precision of the reconstructed edges and finer details (shallower) with spatial regularity (deeper). To fine-tune a specific reconstruction it can be useful to adjust this parameter, e.g. to limit the fitting power of the model at high noise levels, as pointed out by Heckel and Hand (2019), yet its impact is much more subtle than that of depth. We observe that the scaling of the distribution of the random input vector $z$ does not have a significant impact on the reconstructions. Following the heuristic of Heckel and Hand (2019), $z$ was drawn from a uniform distribution between 0 and 0.1. Using different scalings seem to be compensated for by the training of convolutional layer weights, which were initialized with the PyTorch default.

In Figure 3, we illustrate the impact of depth by reporting errors and reconstructed images of the VIRUS as a function of depth for different noise levels in the setting of the experiment presented in Section 4.1 below. Visually, deeper decoders render smoother images. This is a direct consequence of the upsampling layers which correlates the neighboring pixels. The deeper the decoder, the smaller the latent representation $z$ and the less independent the output pixels. At high noise levels ($N_p = 1$) the smoothing reduces the reconstruction error estimated using Euclidean distance. Perceptually however, the smoothing is only beneficial to some extent, and a practitioner would likely prefer a shallow network. At lower noise levels ($N_p = 10$), the excess of depth can be spotted directly in the reconstruction error as we observe a dip in the curve.

## 4. Experiments

**Data** Our strategy is demonstrated on the following datasets. SET12 is a dataset consisting of 12 images used as a traditional benchmark in image processing, here resized to $128 \times 128$ pixels, while BIO10 , resized to $256 \times 256$ pixels, contains 10 more realistic biological samples (these two datasets are available with the paper code). We also consider the COIL100 dataset (Nene et al., 1996) which contains 100 objects on a black background with $128 \times 128$ pixels. We explicitly zero-out the background such that the support of the objects is not perfectly known. In contrast to non-holographic phase retrieval, a good reference should disambiguate the position of the sample within the frame. Hence, COIL100 allows us to test the robustness of the different algorithms to reference design. All images are converted to gray scale, and examples are presented in Figure 15 of the Appendix.

**Benchmark setup** The strategy proposed in this work is compared against three algorithms for holographic CDI: inverse filtering, Wiener filtering and Hybrid Input-Output modified for holographic phase retrieval, here referred to as HIO-Holo. We further augment HIO-Holo by selecting

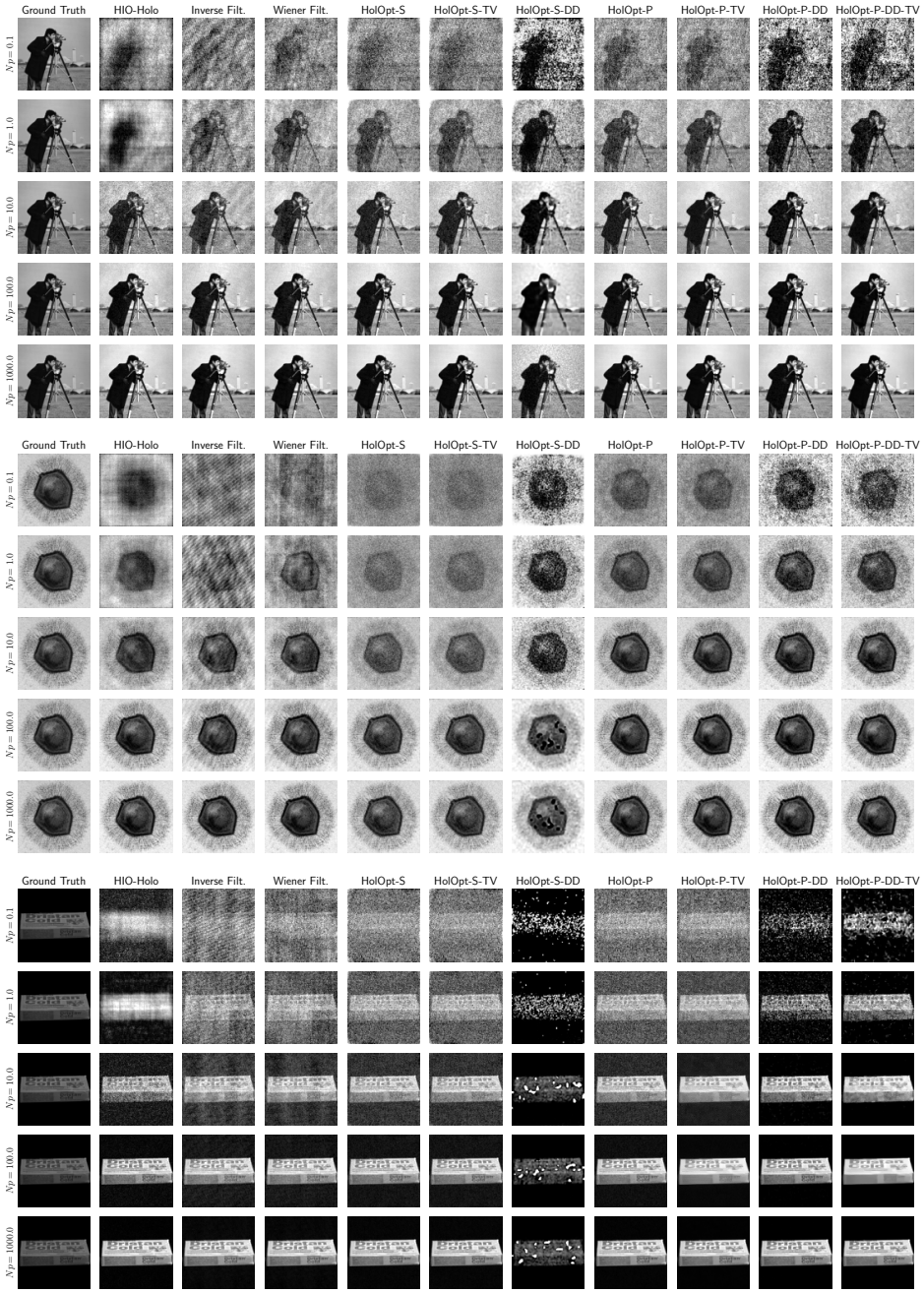



Figure 4: Comparing reconstructions across algorithms and noise levels on a sample image from SET12 (CAMERA, top), a sample image from BIO10 (VIRUS, middle) and a sample image from COIL100 (bottom) with a binary random reference and without beamstop. Figure 16 presents 3 more images. Corresponding SSIM scores can be found in Figure 5. To improve contrast black and white are set respectively to first and last (99th) percentile of all pixels within each image. This convention is adopted for all visuals in the paper. Revision: Results with TV regularization where added.

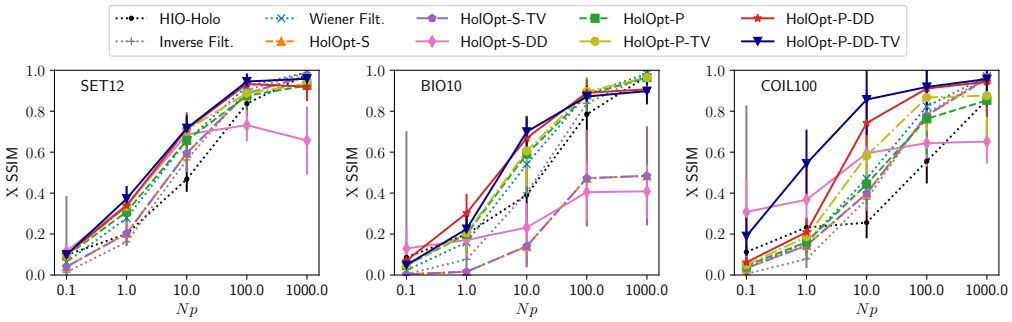



Figure 5: Reconstruction scores (SSIM) for SET12 (left), BIO10 (center), and COIL100 (right) as a function of the photon count $N_p$ with a binary random reference and without beamstop. Corresponding reconstruction MSEs and residual observations MSEs are given in Figure 17 of the Appendix. Results with TV regularization where added.

the best residual reconstruction over all iterations, to give it the fairest possible chance in our comparison. As discussed in Section 1.2, there is no comparable machine learning method that can be used here as a benchmark. However, we test variants of HolOpt-P and HolOpt-P-DD, termed HolOpt-S and HolOpt-S-DD respectively, in which mean squared error (MSE) is optimized instead of the Poisson likelihood. At high photon counts (low noise), Poisson noise is well-approximated by Gaussian noise and the two objectives are expected to perform similarly. However, their difference becomes significant at lower photon counts (high noise). The benefits of taking into account the Poisson nature of the noise are shown in experiments below. We also test variants with total variation (TV) regularization, a common strategy in image reconstruction.

**Hyperparameters** Without access to a ground truth, the selection of hyperparameters of any of the compared algorithms inevitably relies on heuristics. The heuristic adopted here is to tune the hyperparameters by visual inspection on one of the specimens of each dataset at each noise level, and leave these parameters fixed for all subsequent specimens. This would correspond to determining parameters once for each specific experimental setup. Practitioners are encouraged to proceed similarly choosing preferably the specimen for which they have the strongest prior knowledge for the hyperparameter selection. In Section 4.5 we propose a direction to formalize this procedure based on statistics of collection of images.

For gradient descent, we use the Adam optimizer (Kingma and Ba, 2015) and learning rates varying between 0.01 and 0.1, depending on the loss and prior. Maximum number of iterations is also fixed depending on noise level and the reconstruction at the best magnitude residual value along the iterations is retained as the final reconstruction. Deep decoder parameters are gathered in Table 1 which also includes number of steps in the gradient descent. Some early stopping was found beneficial in order to avoid overfitting at high noise levels. As a result we adapt the number of iterations to the photon count. TV regularization coefficients are chosen at the highest possible value that yields good visual reconstruction and does not deteriorate the final residual error. Selected values ranged between $10^2$ and $10^3$.

**Evaluation** Algorithms are compared in terms of Structural Similarity Index (SSIM) Wang et al. (2004). Comparisons in terms of relative reconstruction error — Euclidean distance between the reconstructed $\hat{\boldsymbol{X}}$ and ground truth specimen $\boldsymbol{X}$ normalized by the $\ell_2$-norm of $\boldsymbol{X}$ — and relative residual error — understood as the Euclidean distance between the observations $\boldsymbol{Y}$ and the noiseless

output of the forward model for a specimen $\hat{\boldsymbol{X}}$ normalized by the $\ell_2$-norm of $\boldsymbol{Y}$ — are also reported in the Appendix. Errors are averaged over images of each dataset as well as a few different random seeds for BIO10 and SET12 to increase the statistics of the small datasets. Error bars correspond to standard deviations.

### 4.1. Noisy reconstruction with and without the deep decoder

In a first series of experiments, we examine the robustness to noise of HolOpt-P and HolOpt-P-DD and their TV variants. We set the reference $\boldsymbol{R}$ to a $m \times m$ binary array with entries 0 or 1 sampled uniformly and independently, a reference design generally very favorable to the reconstruction (Candès et al., 2015; Marchesini et al., 2008). No beamstop mask is included.

Figure 4 shows a clear qualitative improvement of Poisson likelihood optimization methods over the baselines as noise increases (observed consistently over the different datasets, see also Figure 16 in the Appendix). In Figure 5, HolOpt-P-DD consistently reaches higher SSIM scores in the noise range $Np = 1$ to $Np = 100$ across datasets. For SET12 and COIL, which both feature images with large constant regions, the TV regularized version HolOpt-P-DD-TV yields the best SSIMs. At very low noise $Np = 1000$, the ordering of the methods varies, yet visual inspection confirms that all algorithms perform similarly and well, except for HolOpt-S-DD combining squared loss and deep decoder (see discussion below). On the other hand, at $Np = 0.1$, little information is left for the algorithms to retrieve: all methods reconstruct images with SSIM scores close to $0.1$, visually failing in different ways.

Among the variants of HolOpt, we observe the following trends. The reconstruction loss and visual quality of the samples are in almost all cases better with HolOpt-P than with HolOpt-S, validating our adoption of the most realistic noise model and Poisson likelihood objective. The difference between MSE and Poisson likelihood objectives is most drastic when including the deep decoder prior. In particular, the MSE loss sometimes leads to artifacts in the images reconstructed by HolOpt-S-DD (see Figures 4 and 16 in Appendix B).

Thus, we focus only on the Poisson likelihood objective going forward. Regarding the use of a deep decoder, we distinguish several performance regimes for our method. At low noise ($N_p \simeq 1000$ to $100$), there is no need for regularization by a deep decoder, and HolOpt-P-DD achieves reconstructions of similar quality to HolOpt-P. Here, the SSIM is typically slightly worse with a deep decoder than without (except for COIL images). Including a prior is not harmful, but often unnecessary. At higher levels of noise, the denoising power of the deep decoder is beneficial. These observations are in accordance with the intuition that a prior helps when information is scarce, but is less helpful when observations of high quality are available.

Focusing on visuals, Figures 4 and 16, we observe that the relative performances of HolOpt-P-TV, HolOpt-P-DD and HolOpt-P-DD-TV are image dependent. The deep decoder usually allows better contrast. TV regularization smooths constant backgrounds, but sometimes also erases details (see e.g. the writing of the COIL box). Adding TV regularization to the deep decoder can hurt (COIL box, $Np = 10$ ) or help (COIL box, $Np = 1$). In the next experiments we focus on the behavior of HolOpt-P and HolOpt-P-DD, noting that a practitioner should also check for improvements with a TV regularization tuned to their setup.

Finally, we note that training the deep decoder incurs an additional computational cost. For our implementations, HoloOpt-P-DD is about 3 times slower than the classical HIO-Holo, an acceptable slow-down in scientific imaging applications, if traded with reconstruction improvement.[2]

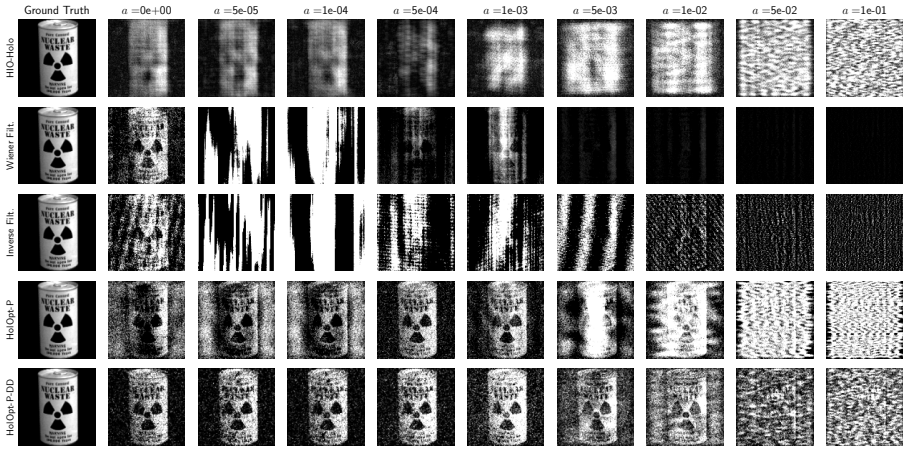



Figure 6: An example reconstructed image from COIL dataset as a function of beamstop area fraction $a$ for fixed photon count $N_p = 1$. See also Figure 26 in the Appendix for a differently scaled visualization.

### 4.2. Reconstructing with missing low frequencies

As discussed in Section 2, a universal feature of CDI experiments is a beamstop which obscures low-frequency magnitudes. In Wiener and inverse filtering, one simply sets the missing magnitudes to zero (Guizar-Sicairos and Fienup, 2007), whereas HIO-Holo can be made agnostic to the missing magnitudes.We show here that our optimization-based methods can effectively incorporate an arbitrary beamstop in the forward model, as defined in Equation 3. Moreover, we expect and indeed observe that the deep decoder prior can be particularly useful in compensating for the missing magnitudes.

We evaluate our methods HolOpt-P and HolOpt-P-DD at several noise levels and beamstop sizes, example reconstructions and SSIMs plots for COIL100 are in Figures 25 and 7 respectively. Supplemental plots for all datasets and noise levels are available in Figures 22-35 of the Appendix, including mean-squared error, which corroborate the trends observed here. We consider square beamstop masks centered at the 0 frequency, identified by their area fraction $a$: the fraction of the total measured magnitudes which are lost (visualized in Figure 22 of the Appendix).

We find that both HolOpt-P and HolOpt-P-DD vastly outperform HIO-Holo, Wiener and inverse filtering at near-all noise levels, test images, and beamstop area fractions. The advantage in terms of SSIM of the deep decoder prior is image-dependent: HolOpt-P-DD yields most significantly improved SSIM relative to HolOpt-P on BIO10 at low photon counts, and at all noise levels for COIL100, but makes little difference in SSIM on the SET12 dataset (see Figure 7). Yet visually,

2. We report computational costs for runs on an NVIDIA Tesla V100SXM2 GPU reconstructing a $128 \times 128$ pixel image: HIO-Holo (1000 iterations) $\sim$27s, HoloOpt (2500 iterations) $\sim$3s and HoloOpt-P-DD (2-layer) (2500 iterations) $\sim$92s. Inverse filtering and Wiener filtering running times are negligible. Running times have not been optimized.

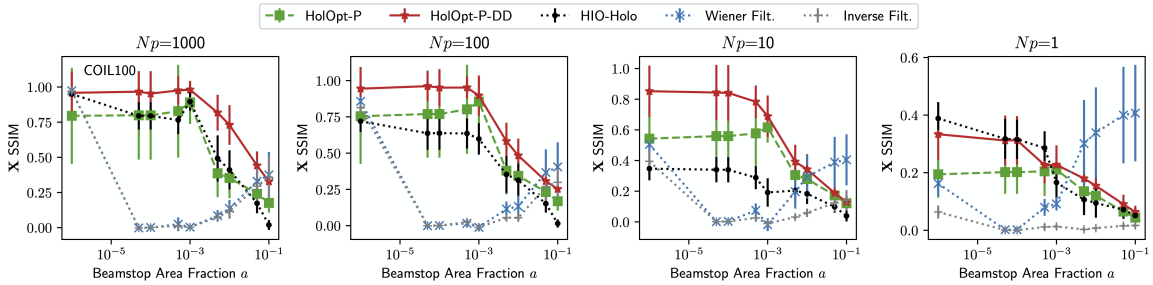



Figure 7: Reconstruction SSIM as a function of beamstop area fraction. Baseline methods are run for 5 trials per image while for tractability our methods are run 1 trial each on COIL100. Average SSIM and one standard deviation error bars are shown. The leftmost datapoints correspond to no missing data, i.e. formally $a = 0$ at the leftmost points. See Figure 23 for a comparision over all datasets.

as in Figure 25, the reconstructions via HolOpt-P-DD on all datasets have enhanced contrast and details relative to HolOpt-P; furthermore, both HolOpt-P and HolOpt-P-DD visually outperform baselines. We note that even at settings where baselines achieve higher SSIM, notably with largest beamstop size and low photon counts on the COIL100 dataset, visual inspection of Figures 25 and 30 illustrate that no method performs recognizable reconstruction at this beamstop size; the supposed improvement thus appears to be an idiosyncrasy of the SSIM[3]. In sum, the performances of our methods smoothly degrade with increasing fraction of missing magnitudes $a$, and enable reconstructions with lower error (Figure 7) and visually improved features (Figure 25) relative to baselines. This provides powerful evidence that our method can enable refined reconstructions given even large fractions of lost magnitude data at the highest noise levels.

### 4.3. Robustness to reference separation

The separation distance between the specimen and the reference limits the smallest resolution that can be achieved (Salditt et al., 2020), and thus would ideally be minimized. For a specimen of size $m \times m$ pixels, the *holographic separation condition* dictates that inverse filtering and Wiener filtering require a full separation $(\boldsymbol{X}|\mathbf{0}_{m\times m}|\boldsymbol{R})$. In contrast, our approach only requires that the forward model be differentiable, which is the case for any reference placement.

Here, we explore signal-reference association of the form $(\boldsymbol{X}|\boldsymbol{R}|\mathbf{0}_{m\times m})$, where $\boldsymbol{R}$ is a random binary block of small size: it is non-zero only on a box of size $0.1m \times 0.1m$ with uniformly and independently chosen entries in $\{0, 1\}$, and the box position is varied between experiments. No beamstop is included. Figure 8 displays example reconstructions for $N_p = 10$ and $N_p = 1$ and Figure 9 reports achieved SSIM scores as a function of the relative specimen-reference separation for the three datasets and multiple level of noises. The only applicable baseline is HIO-Holo, illustrating the challenging nature of this setting.

For the three compared methods, the quality of the reconstruction depends slightly on the separation, only HIO-Holo at 0 separation generates clearly poorer images. Visually, HIO-Holo yields

3. We note that the performance of Wiener and inverse filtering suffers at small beamstop sizes due to the "ringing" effect of low frequencies, an effect which occurs regardless of the method for filling the low magnitudes. Evident as a striping effect in Figure 25, this explains the occasional increase in SSIM with increasing beamstop size evident in Figure 7.

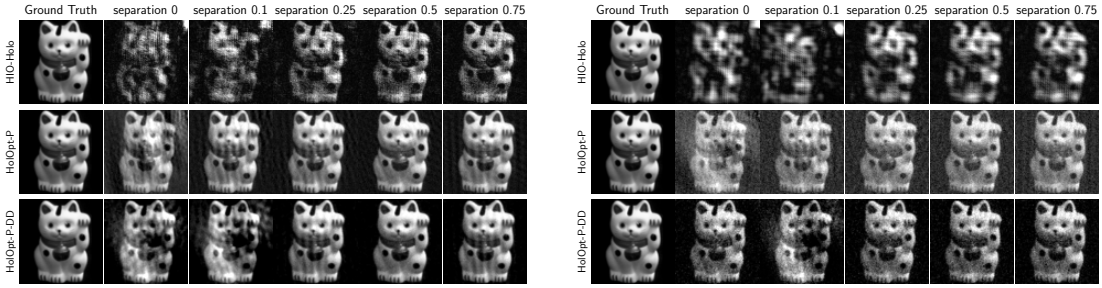



Figure 8: Reconstructions for photon counts $N_p = 10$ (left) and $N_p = 1$ (right) with a $0.1m \times 0.1m$ binary random reference as a function of the relative separation. A separation of $0.5$ implies that the left-most non-zero pixel of the reference is $0.5n$ pixels away form the image. See Figure 19-18 in Appendix for $N_p = 100$ and $N_p = 0.1$ and different samples.

the less sharp images and HolOpt-P-DD is almost always the best algorithm. Without the deep decoder, HolOpt-P reconstructions are also of good quality but sometimes show less contrast. The neural network prior makes a difference in particular at high levels of noise and on COIL100 images, for which the deep decoder captures well the solid black background. These observations are consistent with the average SSIM, but we note that the variance across images in the dataset can be significant. At very high noise $Np = 0.1$ photon/pixel, HIO-Holo has better SSIM scores than HolOpt-P methods, yet this is arguably the limit of reconstruction possibilities for any of the methods (SSIM $< 0.2$ and visuals in Figure 18).

Overall, this more challenging experimental setting confirms the efficiency of our proposed method, allowing reconstruction even if the holographic separation condition is broken. Moreover, a deep decoder prior provides clear benefits even over direct optimization in certain cases.

### 4.4. Varying the oversampling rate

In a last experiment, we consider yet another challenge for reconstruction by lowering the oversampling rate of the observation. For oversampling factors larger or equal to 2, the inverse problem can be solved exactly in the absence of noise (Hayes, 1982). In the presence of noise, increased oversampling is all the more beneficial, as collecting more information can compensate for the noise corruption. We test how our methods perform relative to HIO-Holo for oversampling ratios around the critical value of 2. Note that inverse and Wiener filtering are not well-defined for oversampling rates smaller than 2. Figures 10 and 11 display results for a well-separated binary reference and without beamstop at $N_p = 10$ photon/pixel (see Appendix B.4 for identical figures at $N_p = 1$). All images reconstructed by HolOpt-P-DD are visually superior to images reconstructed by HIO-Holo at oversampling factors both smaller and larger than 2. HolOpt-P-DD also produces sharper images than HIO-Holo, although sometimes with less contrast. This phenomenology is largely captured by the SSIM.

### 4.5. Selecting hyperparameters using group statistics

In order for systematic hyperparameter selection to be practically useful, it is required that it can be performed without access to the ground truth image sought to be reconstructed from measurements.

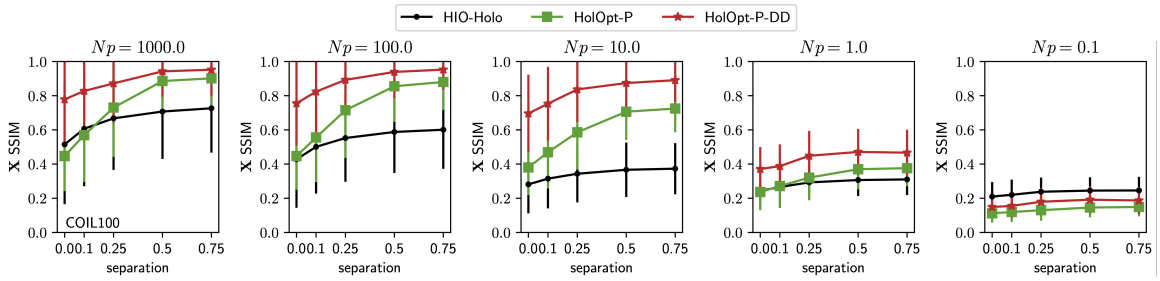



Figure 9: Reconstruction SSIM for decreasing photon counts $N_p$ as a function of the relative separation (see caption of Figure 8). Similar plot for all datasets is given in Figure 20. Corresponding MSEs are plotted in Figure 21 of the Appendix.

For a reconstructed image or a group of reconstructed images with unknown ground truth stemming from a specific hyperparameter setting, an idea is to quantify the extent to which the reconstructions look *in-distribution* or expected with respect to the general class of images under investigation. We propose to do this using the Fréchet Inception Distance (FID, Heusel et al. (2017)). Since the distribution of natural images is extremely high-dimensional and complicated, it needs to be approximated. This can be done by selecting a collection of images, extracting features, and using the statistic of the features on the collection to model a Gaussian distribution. The feature extraction should ideally be Gaussianizing, contracting for images within the collection and mapping non-images, *out-of-distribution* samples, to outlying points.

As a proof of concept, we study parameter selection on the COIL100 data set using features from a pretrained VGG16 convolutional neural network (Simonyan and Zisserman, 2015). Let $f_{\text{VGG}}(x)$ denote the function returning last-layer logits from a pretrained VGG16 network[4]. In order to create a suitable image distribution, we divide the COIL100 dataset into two halves. The first 50 images will serve as an image prior, and the second half will be used to perform hyperparameter selection without ground truth. We compute

$$\mu_{\text{prior}} = \frac{1}{50}\sum_{i=1}^{50} f_{\text{VGG}}(x_i), \qquad \Sigma_{\text{prior}} = \frac{1}{49}\sum_{i=1}^{50}(f_{\text{VGG}}(x_i) - \mu_{\text{prior}})(f_{\text{VGG}}(x_i) - \mu_{\text{prior}})^T. \tag{5}$$

For a given hyperparameter setting, let $\hat{x}_i, i = 51, \ldots, 100$ be the reconstructed images. Analogously to the computation of the prior distribution parameters, we compute

$$\mu_{\text{recon}} = \frac{1}{50}\sum_{i=51}^{100} f_{\text{VGG}}(\hat{x}_i), \qquad \Sigma_{\text{recon}} = \frac{1}{49}\sum_{i=51}^{100}(f_{\text{VGG}}(\hat{x}_i) - \mu_{\text{recon}})(f_{\text{VGG}}(\hat{x}_i) - \mu_{\text{recon}})^T. \tag{6}$$

The Gaussians defined by $(\mu_{\text{prior}}, \Sigma_{\text{prior}})$ and $(\mu_{\text{recon}}, \Sigma_{\text{recon}})$ can now be compared using the Wasserstein 2-distance between distributions, which reduces to

$$d_{\text{prior, recon}} = \|\mu_{\text{prior}} - \mu_{\text{recon}}\|^2 + \text{trace}(\Sigma_{\text{prior}} + \Sigma_{\text{recon}} - 2\sqrt{\Sigma_{\text{prior}}\Sigma_{\text{recon}}}). \tag{7}$$

Using this distance, we can evaluate reconstructions with different hyperparameter settings and select the one with the smallest distance to the collection of prior images.

4. We use the one available in the `torchvision` python package

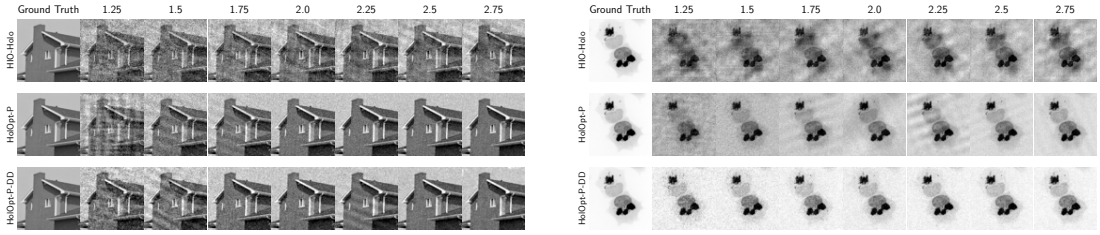



Figure 10: Reconstructed images for samples from SET12, BIO10 datasets with varying oversampling factors (numbers above each column) at $N_p = 10$ photon/pixel. See Figure for a figure incuding COIL images.

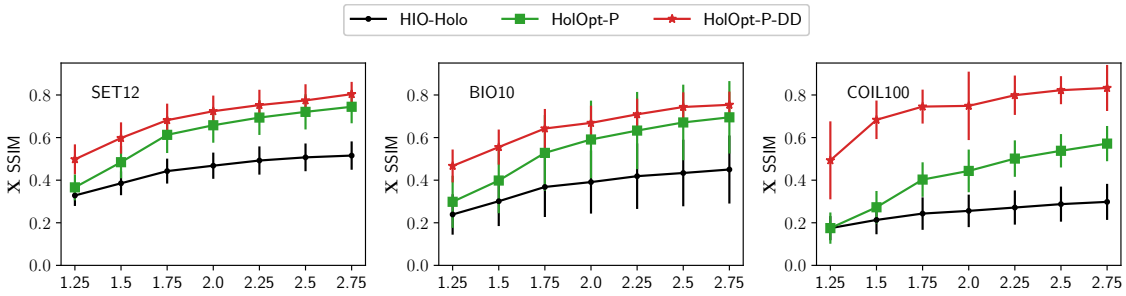



Figure 11: Average SSIM scores for SET12, BIO10 and COIL100 with varying oversampling factor at $N_p = 10$ photon/pixel. Errors bars represent standard deviations.

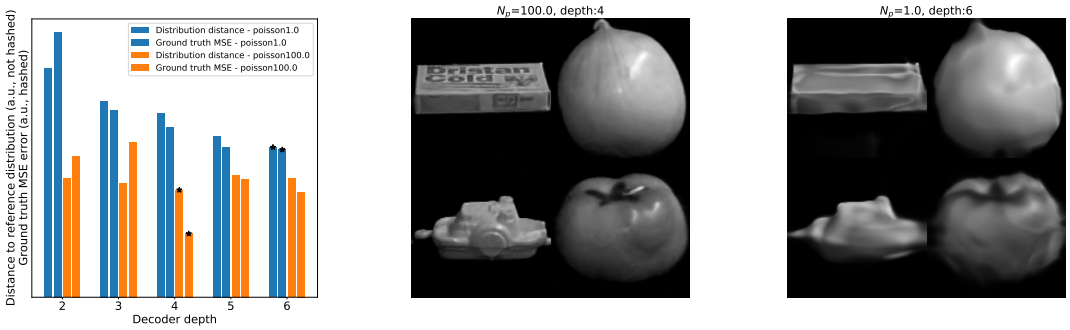



Figure 12: Evaluation of deep decoder depth selection using Fréchet Inception Distance (FID) to a natural image prior. Different colors represent different noise levels. The unhashed bars represent Fréchet Inception Distance, and the hashed bars represent MSE error compared to ground truth. Asterisks are placed at the decoder depth representing minimal FID and minimal MSE error. They correspond for all noise levels. Reconstruction examples at the selected decoder depths are shown on the right. While $N_p = 100$ shows sharp reconstructions at depth 4, the noisy $N_p = 1.0$ leads to selection of depth 6, which blurrs the finer features but preserves the shape.

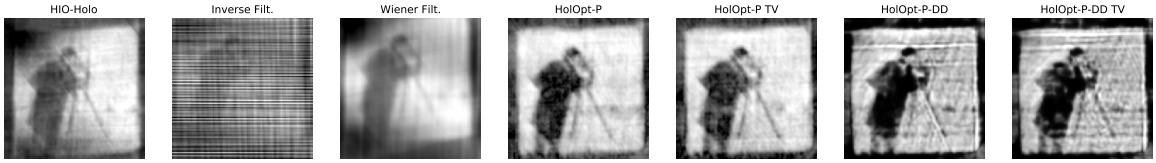



Figure 13: Reconstructions from experimental data. Further details on the experimental setup and methods can be found in Appendix A.

We exemplify this strategy for setting depth of deep decoders. Figure 12 shows FIDs to the natural image prior for different depths at two different noise levels. For a given noise level, the FID is minimal at the same decoder depth as the MSE with ground truth. We have thus created a model selection heuristic that can select the decoder depth leading to minimal squared reconstruction error. While these statitics offer a promising direction of systematic hyperparameter selection, the FID evaluated here exhibits the same selection biases as MSE (e.g. privileging low-frequency accuracy over high frequencies). A practitioner may prefer shallower networks, as in the experiments from the sections above.

## 5. Optical Laser Coherent Imaging Experimental Data

In this section, we evaluate HolOpt-P and HolOpt-P-DD, with and without TV regularization, on the experimental data used in Guizar-Sicairos and Fienup (2008), obtained with permission from the authors. In this experimental setup, a slide printed with the CAMERA image and an adjacent triangular reference object were exposed to a collimated wave from a He-Ne laser (632.8 nm wavelength), and the resultant Fourier intensity pattern was measured. As this was not an X-ray experiment, no beamstop was necessary. We refer to Guizar-Sicairos and Fienup (2008) for precise experimental details and to Appendix A for detail of our methods. We compare variants of HolOpt-P to baselines of Wiener filtering, inverse filtering, and HIO-Holo.[5] The results are shown in Figure 13. HolOpt-P produces superior reconstructions, with the qualitative best by either HolOpt-P or HolOpt-P-DD TV. Those produced by the Wiener and inverse filtering baselines have important artifacts; that of HIO-Holo looks better than the former two, but still suffers from significant shadowing artifacts in the upper and lower left corners. We also observe that TV-regularization alone noticeably smooths out the horizon line, effectively removing it, while the addition of a deep decoder prevents this phenomenon.

## 6. Discussion and Conclusion

In this paper, we have shown that recent progress at the intersection of machine learning and inverse problems can yield highly successful algorithms which also account for realistic experimental challenges. Our novel optimization framework for holographic phase retrieval improves on state-of-the-art reconstruction, even in the most difficult experimental settings and without external training data. Untrained image priors are confirmed to be powerful tools, especially when a significant

5. While the HERALDO method can yield sharper reconstructions, it is not directly comparable to the techniques presented here. In particular, HERALDO requires specific assumptions on the reference geometry and uses a tailored reconstruction operator for each new reference shape, whereas our method is one-size-fits-all (Guizar-Sicairos and Fienup, 2007).

amount of information is missing from the measured data due to low photon counts, beamstop-obscured frequencies and small oversampling. Due to its practicality and flexibility, we believe our methodology should prove valuable for practitioners. We confirm the success of our approach on the experimental data in the optical range. For reasons of data availability, we leave an evaluation of the method on X-ray holographic CDI to future work. Finally, our framework is easily adaptable to different variants of the problem — even the mathematically distinct non-holographic setting — and should enable similarly improved reconstruction with other imaging modalities than Holographic CDI, such as optical holography (He et al., 2015), magnetic holographic imaging (Hu et al., 2019), and ptychography (Wen et al., 2012).

ACKNOWLEDGMENTS

We are grateful to Charles Epstein, Leslie Greengard and Alex Barnett for insightful discussions. We also thank Manuel Guizar-Sicairos and Jim Fienup for sharing Heraldo experimental data, and Ju Sun for suggesting SSIM as an image metric and other helpful comments.

## References

Benjamin Aubin, Bruno Loureiro, Antoine Baker, Florent Krzakala, and Lenka Zdeborova. Precise asymptotics for phase retrieval and compressed sensing with random generative priors. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference, PMLR*, volume 107, pages 55–73, 2020.

David A. Barmherzig and Ju Sun. Low-photon holographic phase retrieval. In *Imaging and Applied Optics Congress*, page JTu4A.6. Optical Society of America, 2020. URL http://www.osapublishing.org/abstract.cfm?URI=COSI-2020-JTu4A.6.

David A Barmherzig, Ju Sun, Po-Nan Li, T J Lane, and Emmanuel J Candès. Holographic phase retrieval and reference design. *Inverse Problems*, 35(9):094001, Aug 2019. doi: 10.1088/1361-6420/ab23d1. URL https://doi.org/10.1088%2F1361-6420%2Fab23d1.

David A. Barmherzig, Alex H. Barnett, Charles L. Epstein, Leslie F. Greengard, Jeremy F. Magland, and Manas Rachh. Recovering missing data in coherent diffraction imaging, 2020.

Emrah Bostan, Reinhard Heckel, Michael Chen, Michael Kellman, and Laura Waller. Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network. *Optica*, 7 (6):559, jun 2020. ISSN 2334-2536. doi: 10.1364/OPTICA.389314. URL https://www.osapublishing.org/abstract.cfm?URI=optica-7-6-559.

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

J. R. Fienup. Reconstruction of an object from the modulus of its fourier transform. *Opt. Lett.*, 3 (1):27–29, Jul 1978. doi: 10.1364/OL.3.000027. URL http://ol.osa.org/abstract.cfm?URI=ol-3-1-27.

D. Gabor. A new microscopic principle. *Nature*, 161:777–778, 1948.

Eric Ghigo, Jürgen Kartenbeck, Pham Lien, Lucas Pelkmans, Christian Capo, Jean-Louis Mege, and Didier Raoult. Ameobal Pathogen Mimivirus Infects Macrophages through Phagocytosis. *PLOS Pathogens*, 4(6):1–17, 2008. doi: 10.1371/journal.ppat.1000087. URL https://doi.org/10.1371/journal.ppat.1000087.

Tais Gorkhover, Anatoli Ulmer, Ken Ferguson, Max Bucher, Filipe R. N. C. Maia, Johan Bielecki, Tomas Ekeberg, Max F. Hantke, Benedikt J. Daurer, Carl Nettelblad, Jakob Andreasson, Anton Barty, Petr Bruza, Sebastian Carron, Dirk Hasse, Jacek Krzywinski, Daniel S. D. Larsson, Andrew Morgan, Kerstin Mühlig, Maria Müller, Kenta Okamoto, Alberto Pietrini, Daniela Rupp, Mario Sauppe, Gijs van der Schot, Marvin Seibert, Jonas A. Sellberg, Martin Svenda, Michelle Swiggers, Nicusor Timneanu, Daniel Westphal, Garth Williams, Alessandro Zani, Henry N. Chapman, Gyula Faigel, Thomas Möller, Janos Hajdu, and Christoph Bostedt. Femtosecond x-ray fourier holography imaging of free-flying nanoparticles. *Nature Photonics*, 12(3):150–153, Mar 2018. ISSN 1749-4893. doi: 10.1038/s41566-018-0110-y. URL https://doi.org/10.1038/s41566-018-0110-y.

Alexandre Goy, Kwabena Arthur, Shuai Li, and George Barbastathis. Low Photon Count Phase Retrieval Using Deep Learning. *Physical Review Letters*, 121(24):1–8, 2018. ISSN 10797114. doi: 10.1103/PhysRevLett.121.243902.

Manuel Guizar-Sicairos and James R. Fienup. Holography with extended reference by autocorrelation linear differential operation. *Opt. Express*, 15(26):17592–17612, Dec 2007. doi: 10.1364/OE.15.017592. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-15-26-17592.

Manuel Guizar-Sicairos and James R. Fienup. Direct image reconstruction from a fourier intensity pattern using heraldo. *Opt. Lett.*, 15(33(22)):2668–70, Nov 2008. doi: 10.1364/ol.33.002668. URL https://pubmed.ncbi.nlm.nih.gov/19015703/.

Paul Hand, Oscar Leong, and Vladislav Voroninski. Phase Retrieval Under a Generative Prior. In *Neural Information Processing Systems 2018*, number NeurIPS, 2018. URL http://arxiv.org/abs/1807.04261.

M. Hayes. 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, 1982.

Kuan He, Manoj Kumar Sharma, and Oliver Cossairt. High dynamic range coherent imaging using compressed sensing. *Opt. Express*, 23(24):30904–30916, Nov 2015. doi: 10.1364/OE.23.030904. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-23-24-30904.

Reinhard Heckel and Paul Hand. Deep Decoder: Concise Image Representations from Untrained Non-convolutional Networks. *International Conference on Learning Representations*, 2019. URL http://arxiv.org/abs/1810.03982.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In

I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf.

Wen Hu, Claudio Mazzoli, and Stuart Wilkins. New advances at CSX beamline in magnetic imaging (Conference Presentation). In Barry Lai and Andrea Somogyi, editors, *X-Ray Nanoimaging: Instruments and Methods IV*, volume 11112. International Society for Optics and Photonics, SPIE, 2019. doi: 10.1117/12.2528058. URL https://doi.org/10.1117/12.2528058.

Gauri Jagatap and Chinmay Hegde. Phase Retrieval using Untrained Neural Network Priors. *NeurIPS workshop on Inverse Problems*, 2019.

Alden S. Jurling and James R. Fienup. Applications of algorithmic differentiation to phase retrieval algorithms. *Journal of the Optical Society of America A*, 31(7):1348, 2014. ISSN 1084-7529. doi: 10.1364/josaa.31.001348.

Saugat Kandel, S. Maddali, Marc Allain, Stephan O. Hruszkewycz, Chris Jacobsen, and Youssef S. G. Nashed. Using automatic differentiation as a general framework for ptychographic reconstruction. *Optics Express*, 27(13):18653, 2019. ISSN 1094-4087. doi: 10.1364/oe.27.018653.

S. Kikuta, S. Aoki, S. Kosaki, and K. Kohra. X-ray holography of lensless fourier-transform type. *Optics Communications*, 5(2):86 – 89, 1972. ISSN 0030-4018. doi: https://doi.org/10.1016/0030-4018(72)90005-3. URL http://www.sciencedirect.com/science/article/pii/0030401872900053.

Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

Tatiana Latychevskaia. Reconstruction of missing information in diffraction patterns and holograms by iterative phase retrieval. *Optics Communications*, 452:56 – 67, 2019. ISSN 0030-4018. doi: https://doi.org/10.1016/j.optcom.2019.07.021. URL http://www.sciencedirect.com/science/article/pii/S0030401819306054.

Victor Lempitsky, Andrea Vedaldi, and Dmitry Ulyanov. Deep Image Prior. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9446–9454. IEEE, jun 2018. ISBN 978-1-5386-6420-9. doi: 10.1109/CVPR.2018.00984. URL https://box.skoltech.ru/index.php/s/ib52BOoV58ztuPM{#}pdfviewerhttps://ieeexplore.ieee.org/document/8579082/.

Stefano Marchesini, Sébastien Boutet, Anne E. Sakdinawat, Michael J. Bogan, Săa Bajt, Anton Barty, Henry N. Chapman, Matthias Frank, Stefan P. Hau-Riege, Abraham Szöke, Congwu Cui, David A. Shapiro, Malcolm R. Howells, John Spence, Joshua W. Shaevitz, Joanna Y. Lee, Janos Hajdu, and Marvin M. Seibert. Massively parallel x-ray holography. *Nature Photonics*, 2(9): 560–563, September 2008. ISSN 1749-4885. doi: 10.1038/nphoton.2008.154.

Michael T. McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017. ISSN 10535888. doi: 10.1109/MSP.2017.2739299.

Tim Meinhardt, Michael Moeller, Caner Hazirbas, and Daniel Cremers. Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems. In *2017 IEEE International Conference on Computer Vision (ICCV)*, number 12, pages 1799–1808. IEEE, oct 2017. ISBN 978-1-5386-1032-9. doi: 10.1109/ICCV.2017.198. URL http://ieeexplore.ieee.org/document/8237460/.

Christopher A. Metzler, Philip Schniter, Ashok Veeraraghavan, and Richard G. Baraniuk. prDeep: Robust phase retrieval with a flexible deep network. *35th International Conference on Machine Learning, ICML 2018*, 8:5654–5663, 2018.

Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400:342–344, jul 1999. doi: 10.1038/22498.

Youssef S.G. Nashed, Tom Peterka, Junjing Deng, and Chris Jacobsen. Distributed Automatic Differentiation for Ptychography. *Procedia Computer Science*, 108:404–414, 2017. ISSN 18770509. doi: 10.1016/j.procs.2017.05.101. URL http://dx.doi.org/10.1016/j.procs.2017.05.101.

S. A. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library. *Technical Report CUCS-006-96*, February, 1996.

Gregory Ongie, Ajil Jalal, Christopher A. Metzler, Richard G. Baraniuk, Alexandros G. Dimakis, and Rebecca Willett. Deep Learning Techniques for Inverse Problems in Imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, may 2020. ISSN 2641-8770. doi: 10.1109/JSAIT.2020.2991563. URL https://ieeexplore.ieee.org/document/9084378/.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dÁlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Yair Rivenson, Yibo Zhang, Harun Günaydın, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science and Applications*, 7(2):17141, 2018. ISSN 20477538. doi: 10.1038/lsa.2017.141.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by Denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. ISSN 19364954. doi: 10.1137/16M1102884.

Tim Salditt, Alexander Egner, and D. Russell Luke, editors. *Nanoscale Photonic Imaging*. Springer International Publishing, 2020. doi: 10.1007/978-3-030-34413-9. URL https://doi.org/10.1007/978-3-030-34413-9.

M Saliba, T Latychevskaia, J Longchamp, and H Fink. Fourier Transform Holography: A Lensless Non-Destructive Imaging Technique. *Microscopy and Microanalysis*, 18(S2):564–565, 2012. doi: 10.1017/S1431927612004679.

Fahad Shamshad and Ali Ahmed. Robust Compressive Phase Retrieval via Deep Generative Priors. *arXiv preprint*, 1808.05854:1–19, 2018. URL http://arxiv.org/abs/1808.05854.

Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015. doi: 10.1109/MSP.2014.2352673.

Baoshun Shi, Qiusheng Lian, and Huibin Chang. Deep prior-based sparse representation model for diffraction imaging: A plug-and-play method. *Signal Processing*, 168:107350, 2020. ISSN 01651684. doi: 10.1016/j.sigpro.2019.107350. URL https://doi.org/10.1016/j.sigpro.2019.107350.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

P. Thibault and M. Guizar-Sicairos. Maximum-likelihood refinement for coherent diffractive imaging. *New Journal of Physics*, 14, 2012. ISSN 13672630. doi: 10.1088/1367-2630/14/6/063004.

Eric W. Tramel, Andre Manoel, Francesco Caltagirone, Marylou Gabrié, and Florent Krzakala. Inferring sparsity: Compressed sensing using generalized restricted Boltzmann machines. In *2016 IEEE Information Theory Workshop (ITW)*, pages 265–269. IEEE, sep 2016. ISBN 978-1-5090-1090-5. doi: 10.1109/ITW.2016.7606837. URL http://ieeexplore.ieee.org/document/7606837/.

Tobias Uelwer, Alexander Oberstraß, and Stefan Harmeling. Phase Retrieval using Conditional Generative Adversarial Networks. *arXiv preprint*, 1912.04981, 2019. URL http://arxiv.org/abs/1912.04981.

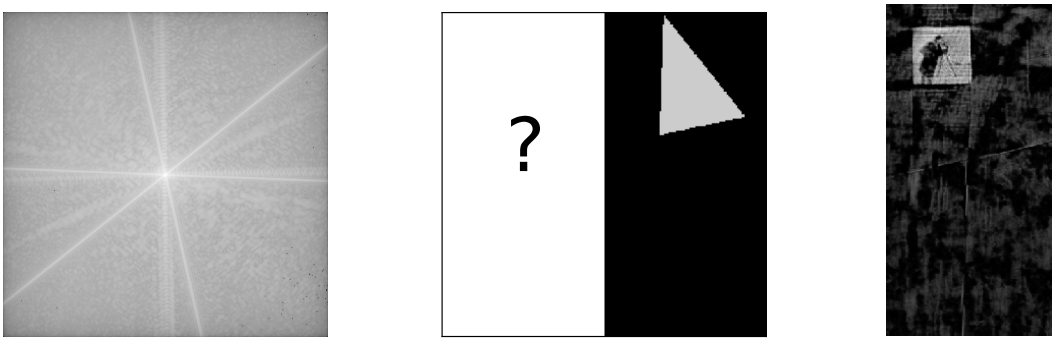
Fei Wang, Yaoming Bian, Haichao Wang, Meng Lyu, Giancarlo Pedrini, Wolfgang Osten, George Barbastathis, and Guohai Situ. Phase imaging with an untrained neural network. *Light: Science and Applications*, 9(1), 2020a. ISSN 20477538. doi: 10.1038/s41377-020-0302-3. URL http://dx.doi.org/10.1038/s41377-020-0302-3.

Yaotian Wang, Xiaohang Sun, and Jason W. Fleischer. When deep denoising meets iterative phase retrieval. In *International Conference on Machine Learning*, 2020b. URL http://arxiv.org/abs/2003.01792.

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*, 13(4):600–612, apr 2004. ISSN 1057-7149. doi: 10.1109/TIP.2003.819861. URL http://ieeexplore.ieee.org/document/1284395/.

Zaiwen Wen, Chao Yang, Xin Liu, and Stefano Marchesini. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, oct 2012. doi:

10.1088/0266-5611/28/11/115010. URL https://doi.org/10.1088%2F0266-5611%2F28%2F11%2F115010.




(*a*) (*b*) (*c*)

Figure 14: Reconstruction from experimental data. (a) Measurements ($1024 \times 1024$ pixels), (b) Reference positioning ($512 \times 512$ pixels), (c) Full reconstruction for HolOpt-P-DD-TV ($512 \times 256$ pixels)

| Noise | $N_p = 1000$ | $N_p = 100$ | $N_p = 10$ | $N_p = 1$ | $N_p = 0.1$ |
|---|---|---|---|---|---|
| SET 12 | d2 c128 | d3 c128 | d2 c128 | d1 c128 | d1 c128 |
| COIL100 | d2 c128 | d2 c128 | d2 c128 | d1 c128 | d1 c128 |
| BIO10 | d3 c128 | d3 c128 | d2 c128 | d2 c128 | d2 c128 |
| # iterations | 10000 | 10000 | 5000 | 2500 | 1250 |

Table 1: Depth, number of channels, and number of optimizer steps used for the deep decoder in all experiments.

## Appendix A. Optical Laser Coherent Imaging Experiment Details

We follow the same preprocessing steps on the raw data as in Guizar-Sicairos and Fienup (2008): namely, averaging over multiple exposures and frames, removing detector artifacts, and attenuating higher frequencies. The $1024 \times 1024$ processed magnitude measurements are displayed in Figure 14(*a*). Considering an oversampling ratio equal to 2, we assume half of a $512 \times 512$-pixel frame known, including the triangular reference, and reconstruct the remaining half. The deep decoder includes 3 layers and 256 channels. Figure 13 presents a zoom taken on the CAMERA image reconstructed close to the top left corner of the frame for each methodm (see Figure 14(*c*) for the full frame on the example of HolOpt-P-DD-TV).

## Appendix B. Additional information for numerical experiments

In this section we provide additional figures for experiments presented in the main text as well as hyperparameters settings. Figure 15 display sample images of the three datasets used for experiments. Table 1 gathers deep decoder hyperparameters selected.

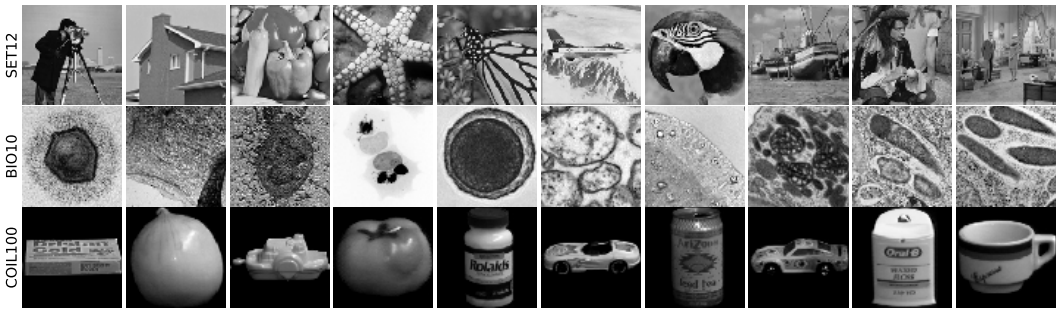



Figure 15: Example images from datsets used in the experiments of this paper: SET12 (128x128 pixels), BIO10 (256x256 pixels) and COIL100 (128x128pixels)

### B.1. Noisy reconstruction experiment

Figures 16 and 17 present results for the experiment described in 4.1 of the main text. The superior visual quality of our algorithms is confirmed on samples form BIO10 and COIL100. In Figure 17 we compare methods in terms of residual $\ell 2$ error minimization and MSE. These measure appear less informative of the perceived similarity of ground truth and reconstructed images than SSIM. In particular, it should be observed that the residual error reflects performance neither perceptually nor according to the other error metrics.

### B.2. Robustness to separation experiment

Figure 18, similar to 8, displays visual of the reconstruction as a function of the separation between sample and reference for two additional noise levels. in Figure 21 we report MSEs averaged over the datasets as a function of the separation. Observations on the ordering of the methods and dependence with separation are comparable with the observations drawn from SSIM plots in the main text.

### B.3. Missing low-frequencies (beamstop) experiment

We now visualize the beamstop area fractions tested in Section 4.2 (see Figure 22), as well as show analogous plots to Figures 25 for all three test images and for all remaining noise levels, $N_p = 10, 100, 1000$ (Figures27-35). Figure 24 also plots all settings using $\ell - 2$ errors.

In Figure 26, for completeness we replicate exactly the reconstructions of Figure 25 with an alternate image scaling technique. In particular, we map colors to pixel values according the absolute minimum and maximum value per reconstruction, instead of the more nuanced quantile approach of the main body (done to maintain fairness in intensity comparisons between methods). Hints of the underlying images are newly visible in the worst reconstructions, particularly in Wiener and inverse filtering, but the relative quality of the competing methods is wholly unchanged.

### B.4. Oversampling experiment

To complement results in the main text, Figure 36 displays reconstruction MSEs as a function of the oversampling rate as a function of the oversampling rate of the observations at $N_p = 10$. Figure

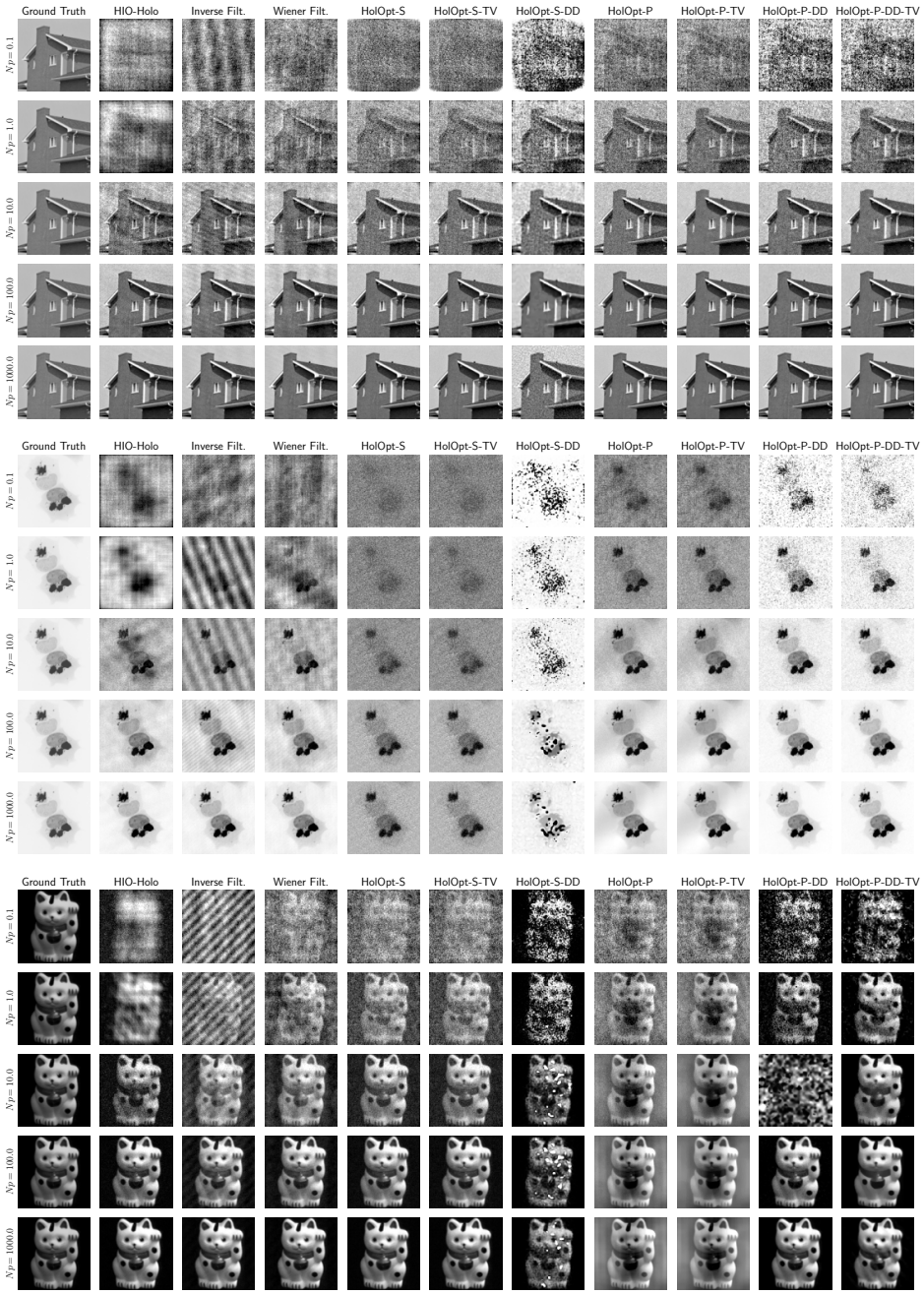



Figure 16: Comparing reconstructions across algorithms and noise levels on a sample of images from SET12, BIO10 and COIL100 with a binary random reference and without beamstop. Same as Figure 4. Results with TV regularization were added.

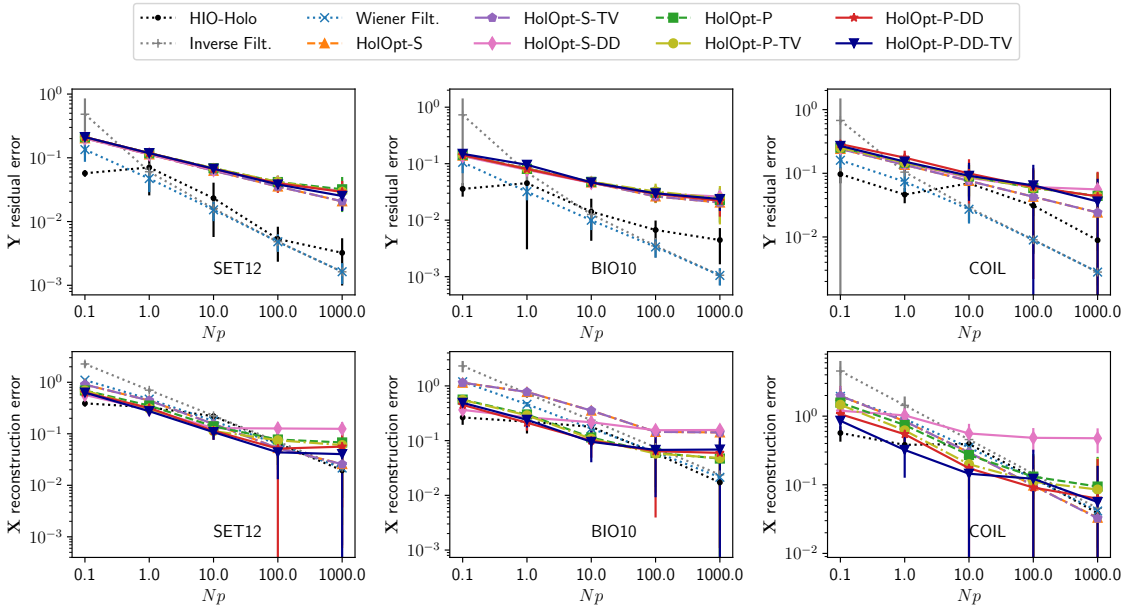
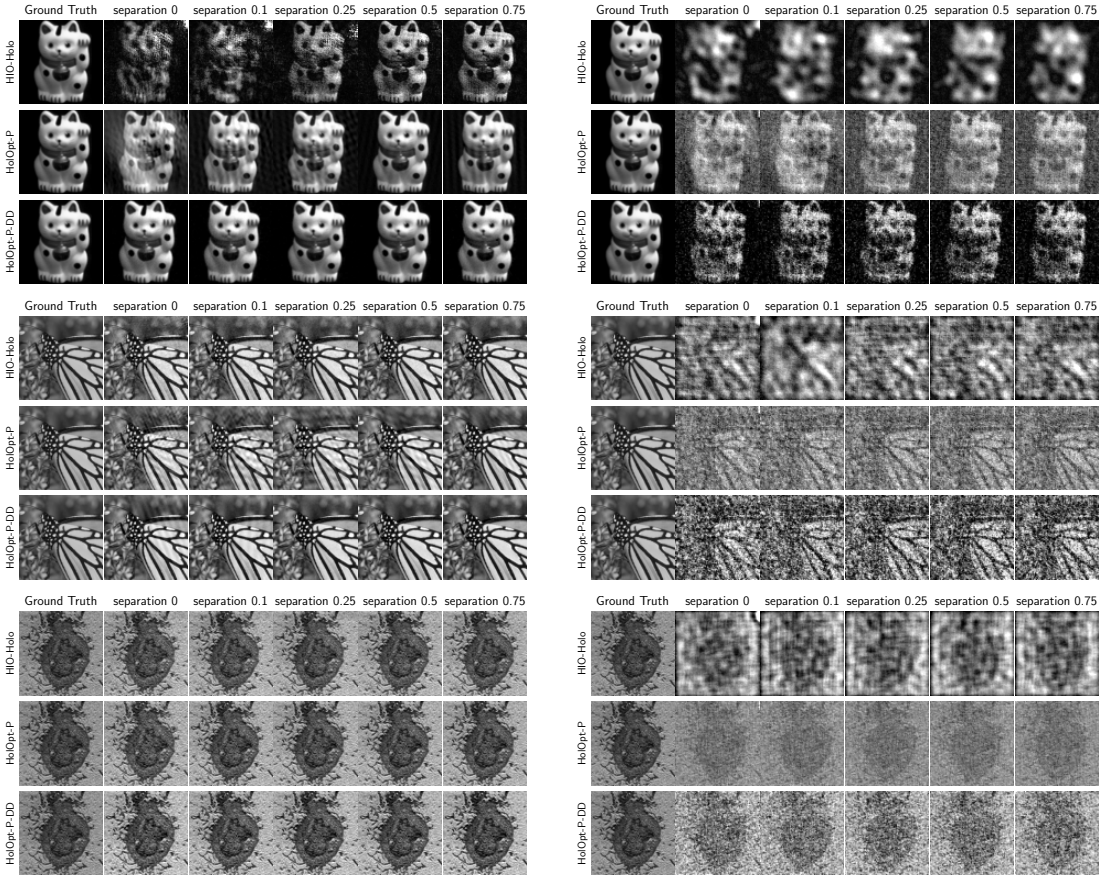
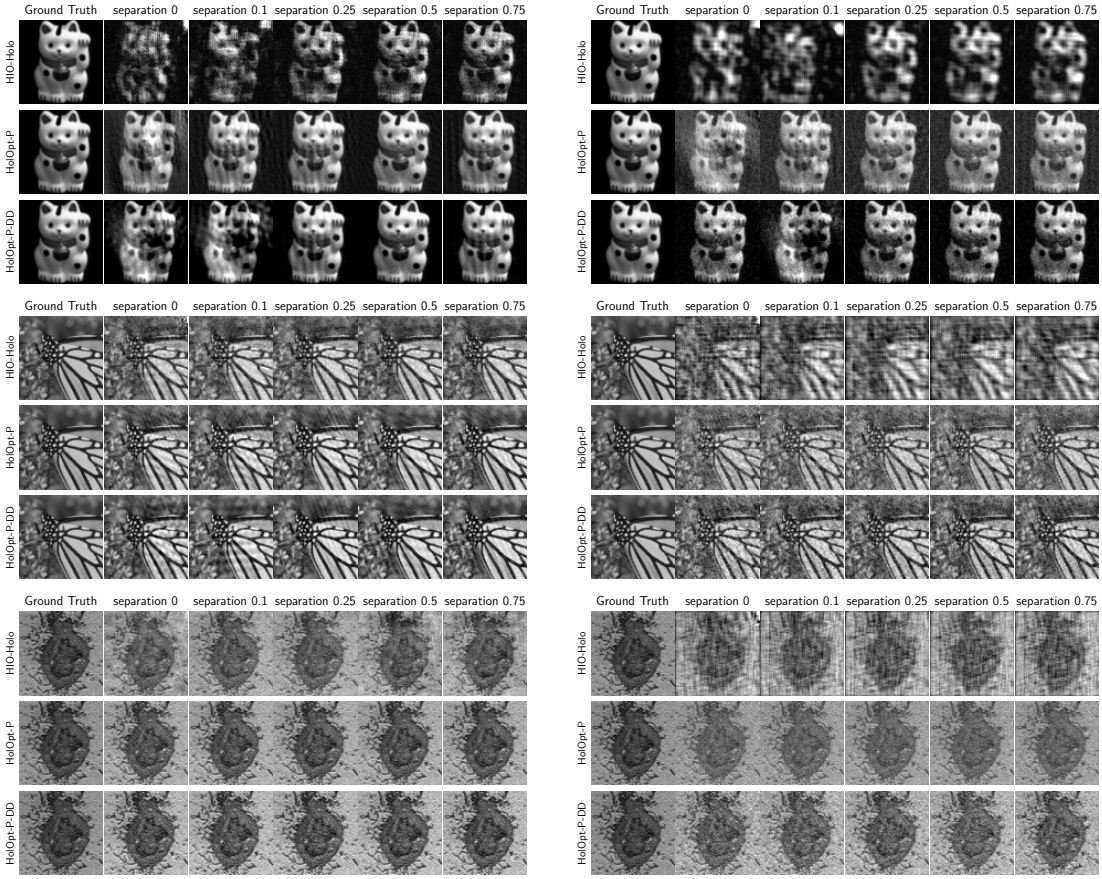



Figure 17: Relative mean squared errors on the reconstructions and residual mean squared error averaged over the datasets. These are different metrics from the SSIM scores presented in Figure 5 to analyze the results of the experiment presented of Section 4.1. Results with TV regularization where added.




(*a*) $N_p = 100$ photons / pixel (*b*) $N_p = 0.1$ photons / pixel

Figure 18: Reconstructed images as a function of the relative separation as described in caption of Figure 9. Images correspond to the best of 10 runs in terms of residual error.




(*a*) $N_p = 10$ photons / pixel (*b*) $N_p = 1$ photons / pixel

Figure 19: Reconstructions for photon counts $N_p = 10$ (left) and $N_p = 1$ (right) with a $0.1m \times 0.1m$ binary random reference as a function of the relative separation. A separation of $0.5$ implies that the left-most non-zero pixel of the reference is $0.5n$ pixels away form the image.

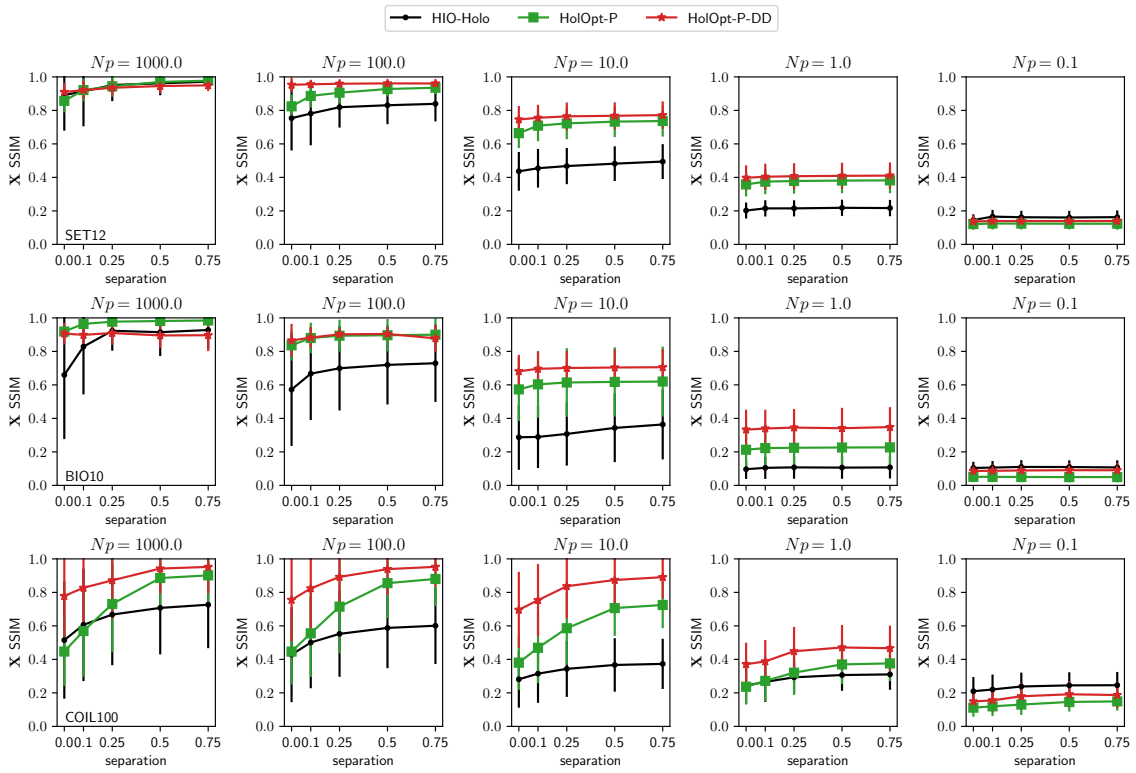



Figure 20: Reconstruction SSIM for decreasing photon counts $N_p$ as a function of the relative separation (see caption of Figure 19).

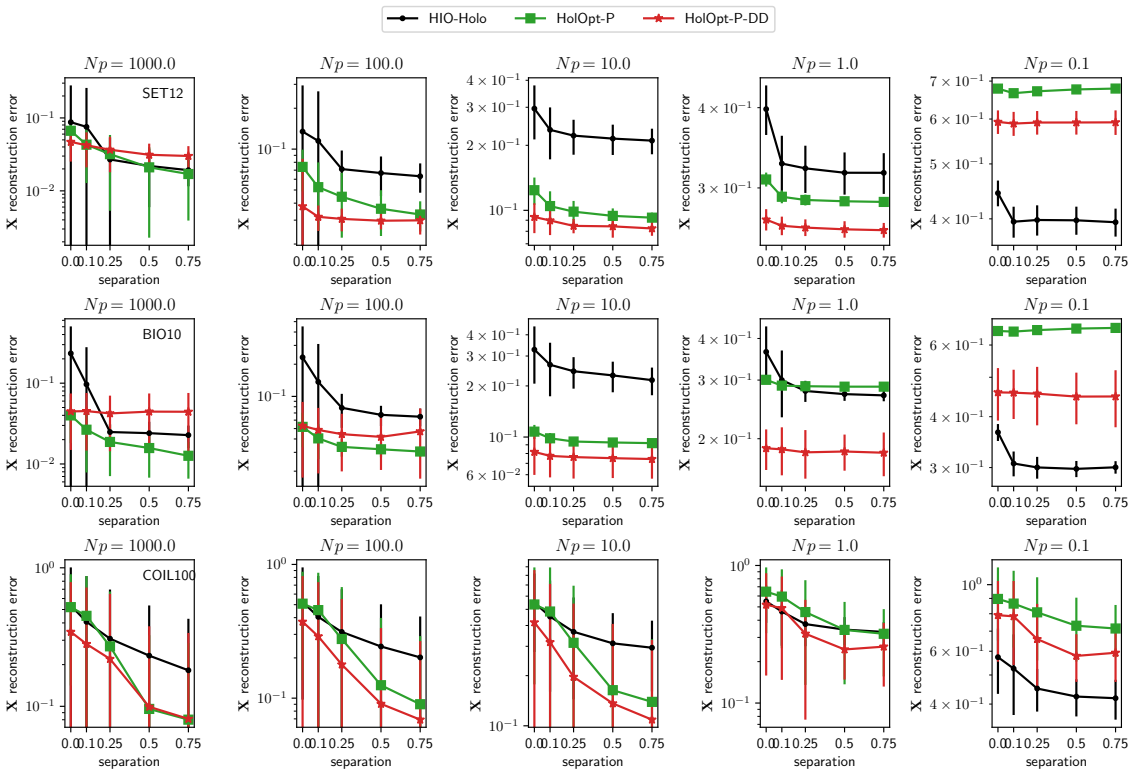



Figure 21: Reconstruction errors for decreasing photon counts $N_p$ with a $0.1m \times 0.1m$ binary random reference as a function of the relative separation. A separation of $0.5$ implies that the left-most non-zero pixel of the reference is $0.5n$ pixels away form the image. Dashed lines corresponds to best run out of the 10 runs in terms of residual error.

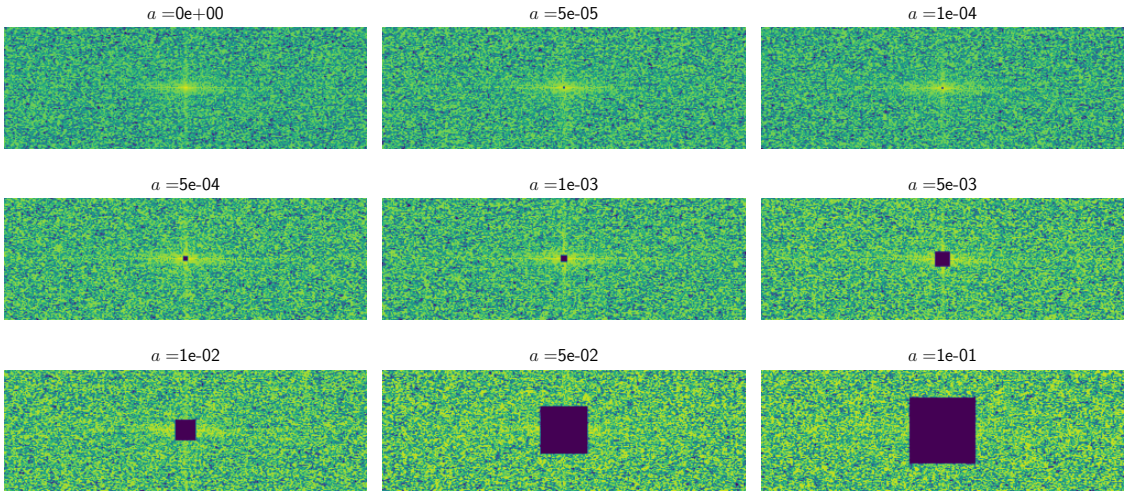



Figure 22: Examples of the measured magnitudes corresponding to varying beamstop area fractions for the first image of SET12, CAMERA, with $N_p = 10$.

38 and Figure 39 focus on the noise level $N_p = 1$ photon / pixel and are similar to Figures already given at $N_p = 10$.

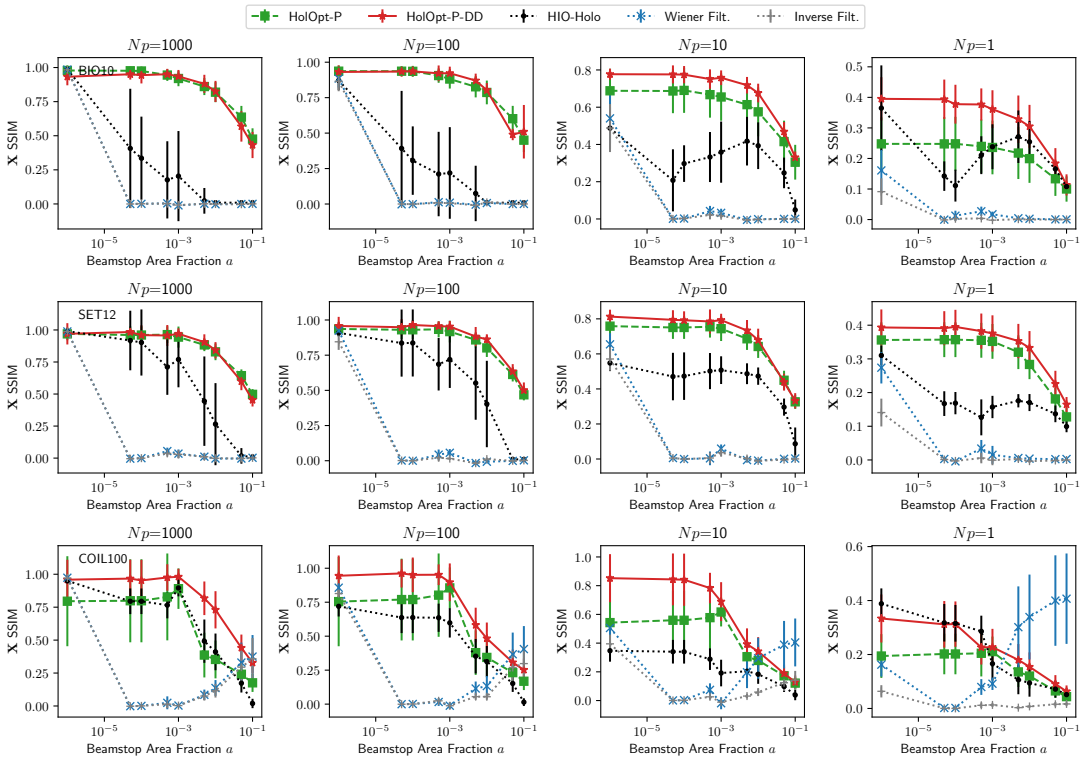



Figure 23: Reconstruction SSIM as a function of beamstop area fraction. Baseline methods are run for 5 trials per image across all datasets, while for tractability our methods are run for 2 trials each on BIO10 and SET12, and 1 trial each on COIL100. Average SSIM and one standard deviation error bars are shown. The leftmost datapoints correspond to no missing data, i.e. formally $a = 0$ at the leftmost points.

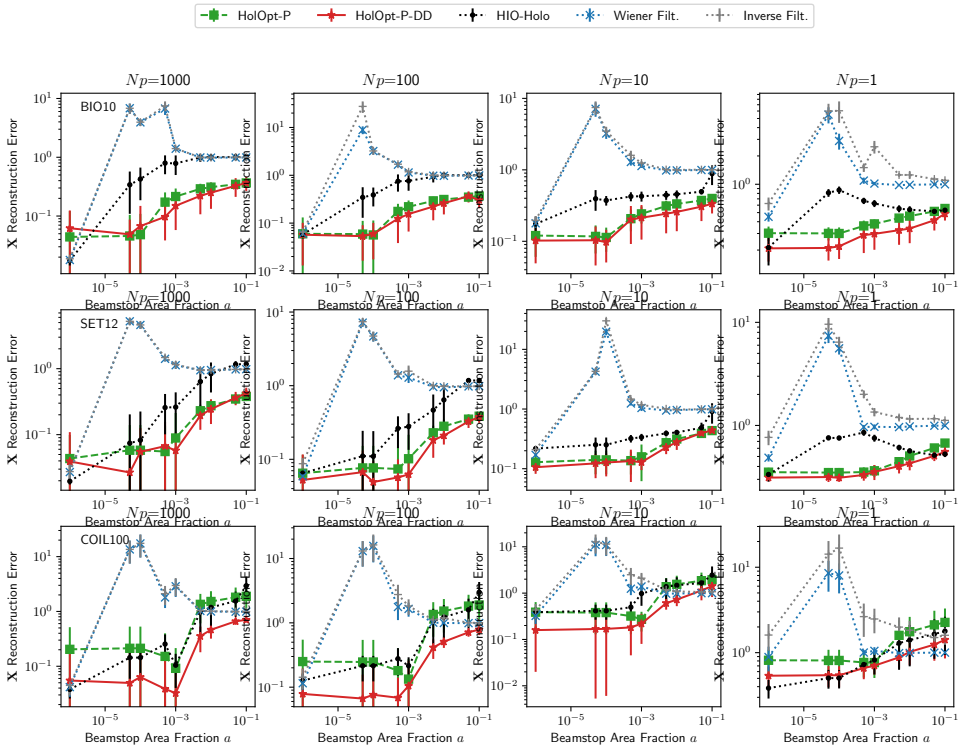



Figure 24: Reconstruction errors on all three datasets (one per row, as indicated by the leftmost row levels) as a function of beamstop area fraction. Errors and error bars are computed as in Figure 5. The leftmost datapoints at $a = 1\text{e}{-6}$ correspond to no missing data, i.e. $a = 0$ at the leftmost points.

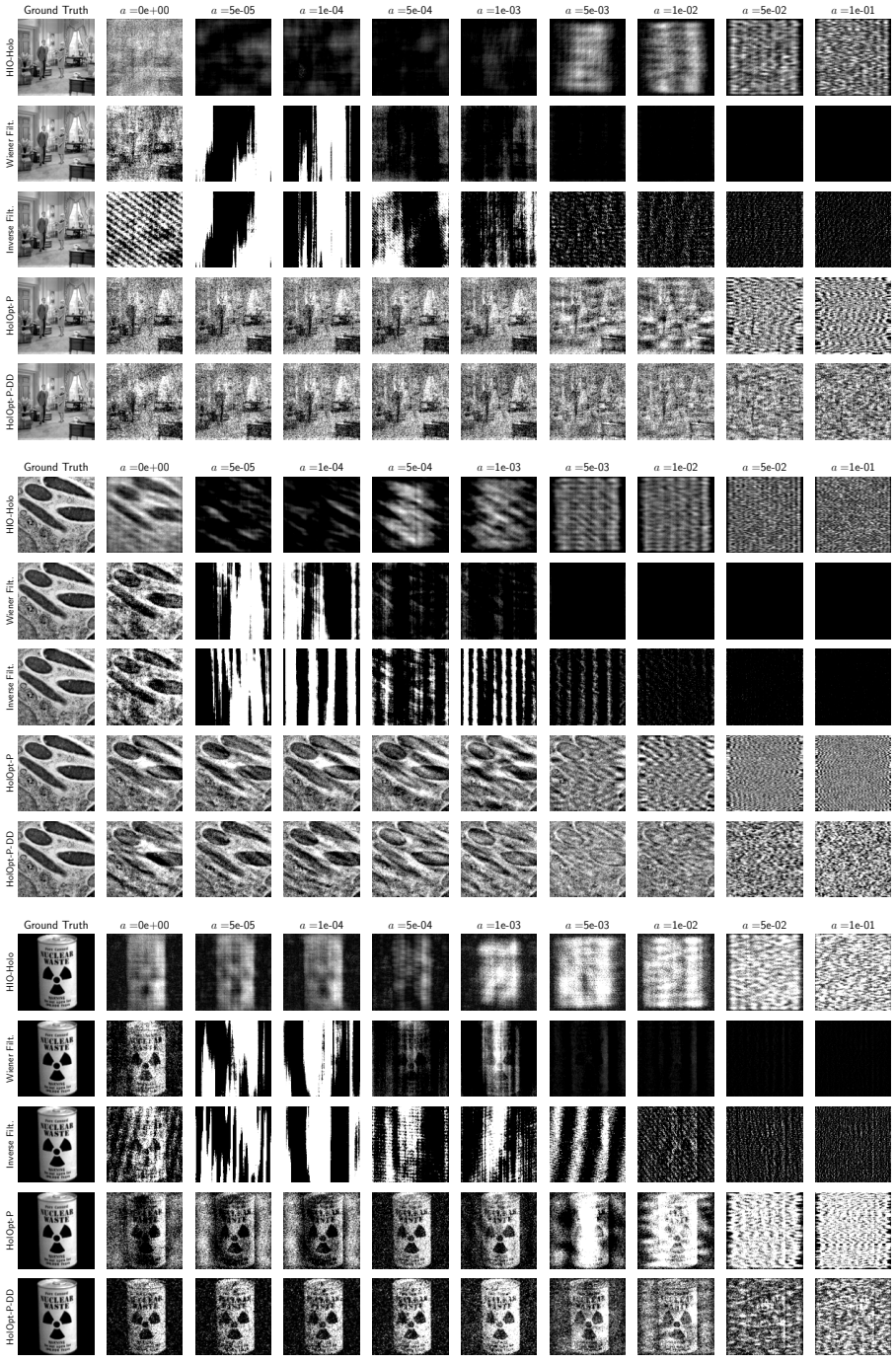



Figure 25: An example reconstructed image from each of the SET12, BIO10, and COIL datasets, respectively, as a function of beamstop area fraction $a$ for fixed photon count $N_p = 1$. See also Figure 26 for a differently scaled visualization.

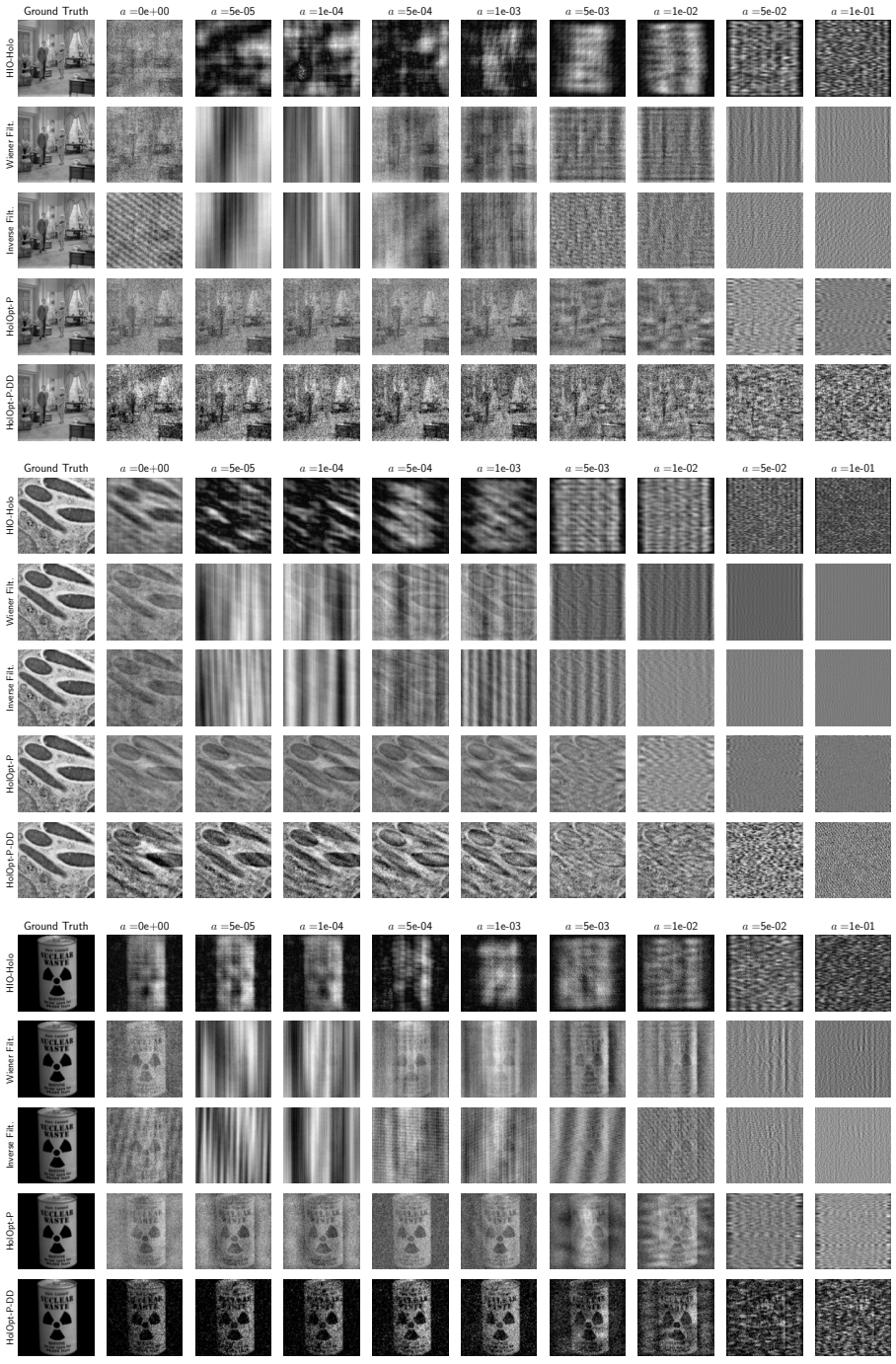



Figure 26: Figure 25 scaled without quantiles. An example reconstructed image from each of the SET12, BIO10, and COIL datasets, respectively, as a function of beamstop area fraction $a$ for fixed photon count $N_p = 1$.

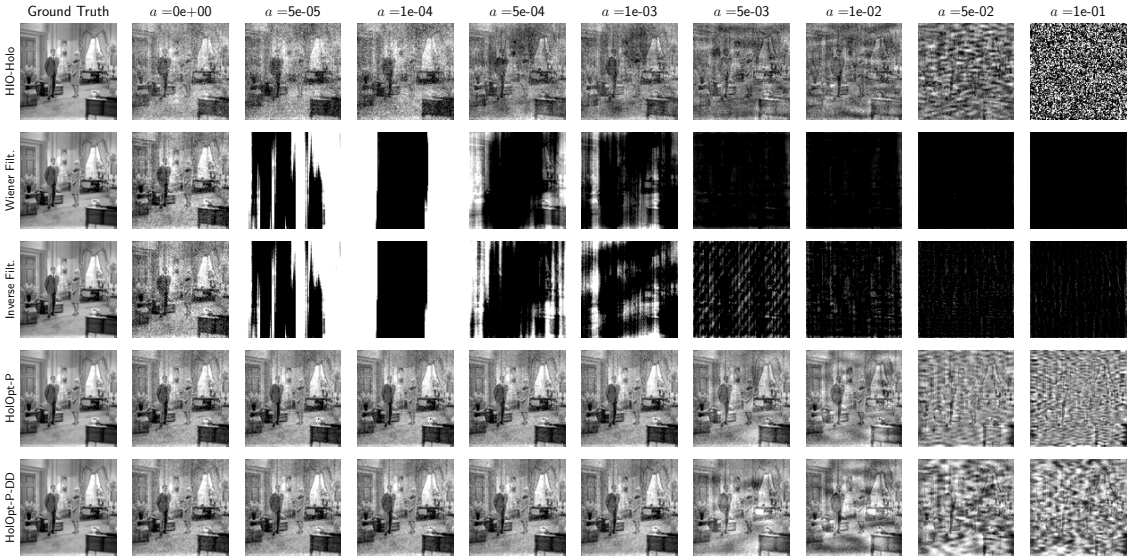



Figure 27: Reconstructed SET12 image with varying beamstop at $N_p = 10$ photons.

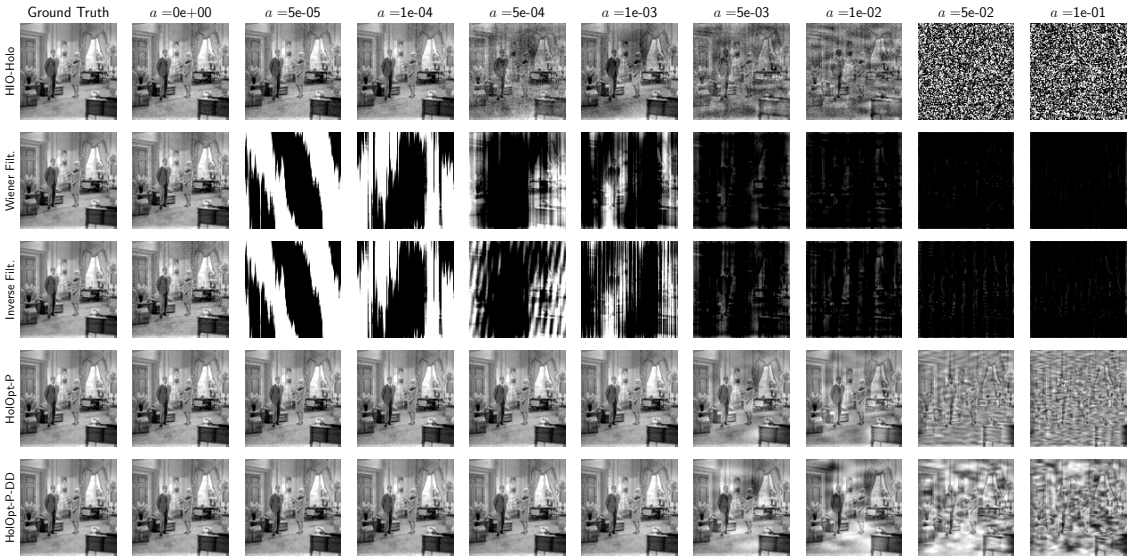



Figure 28: Reconstructed SET12 image with varying beamstop at $N_p = 100$ photons.

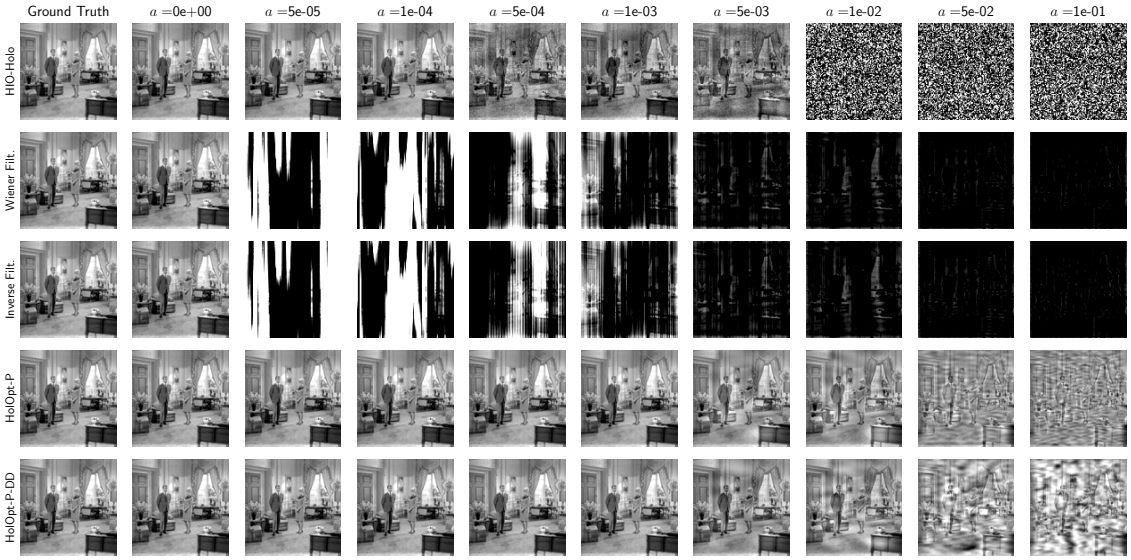



Figure 29: Reconstructed SET12 image with varying beamstop at $N_p = 1000$ photons.

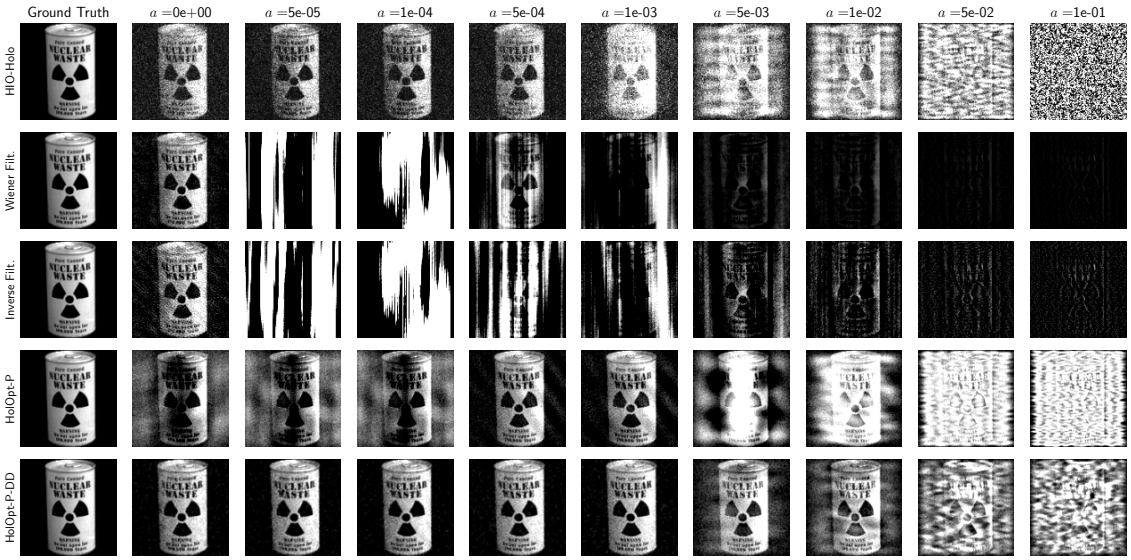



Figure 30: Reconstructed COIL image with varying beamstop at $N_p = 10$ photons.

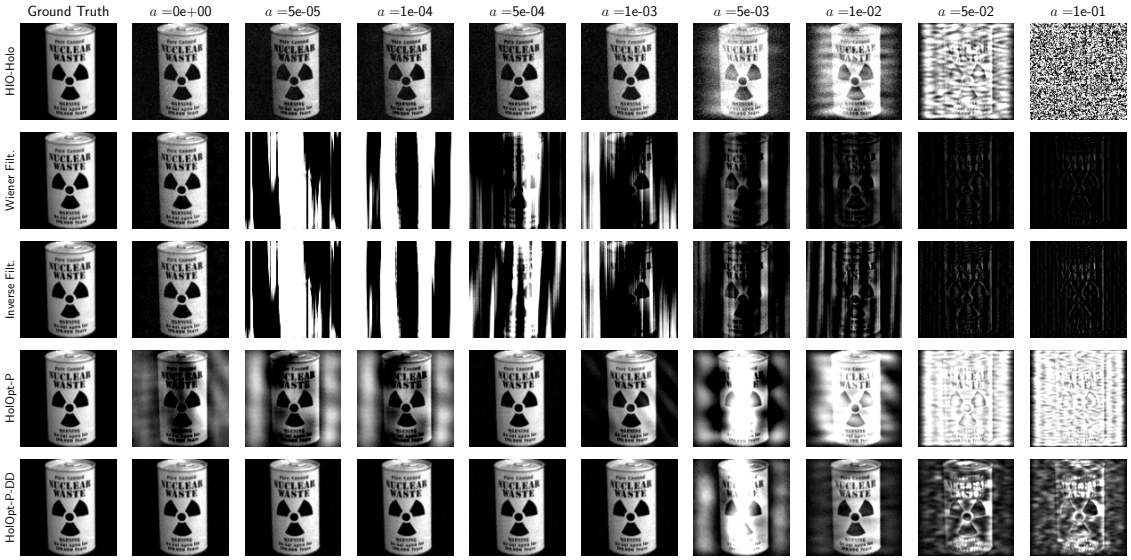



Figure 31: Reconstructed COIL image with varying beamstop at $N_p = 100$ photons.

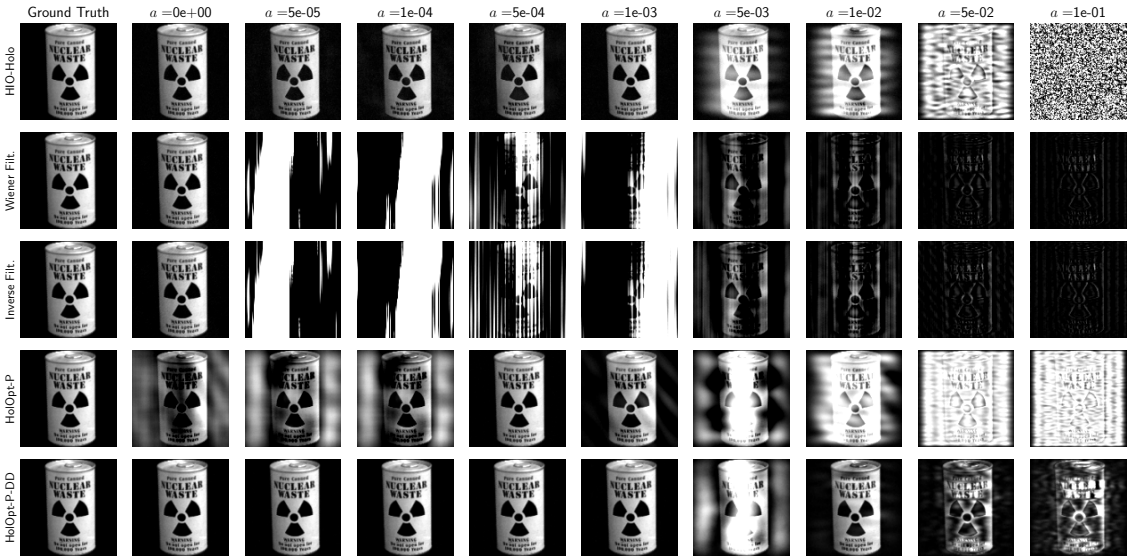



Figure 32: Reconstructed COIL image with varying beamstop at $N_p = 1000$ photons.

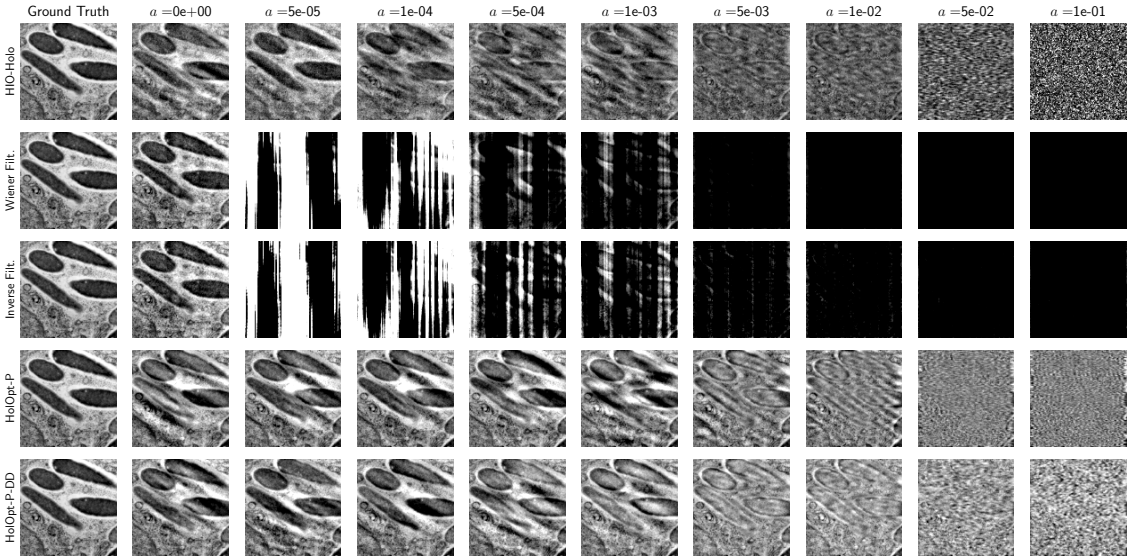



Figure 33: Reconstructed BIO10 image with varying beamstop at $N_p = 10$ photons.

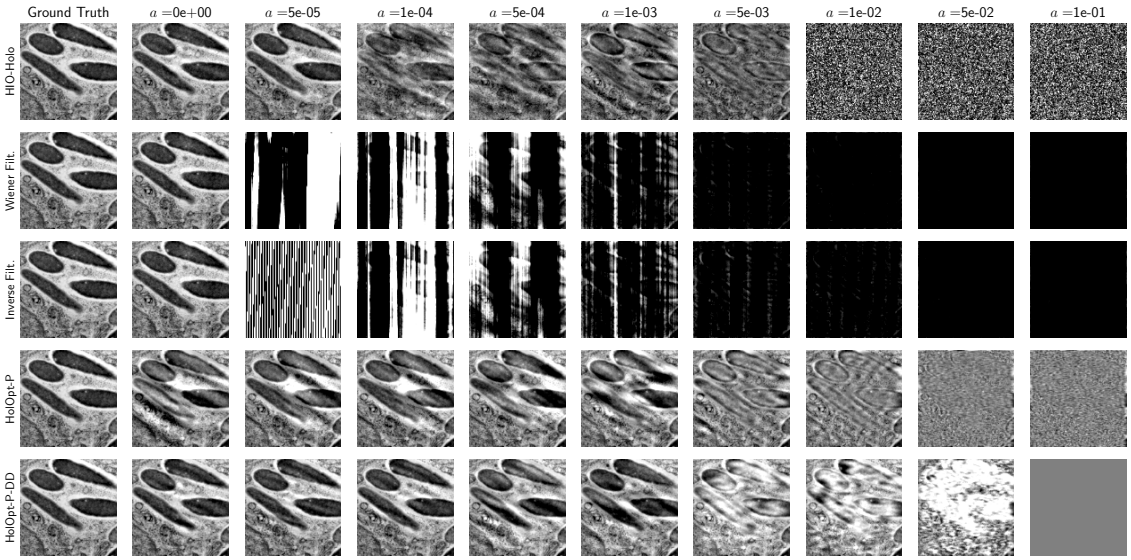



Figure 34: Reconstructed BIO10 image with varying beamstop at $N_p = 100$ photons.

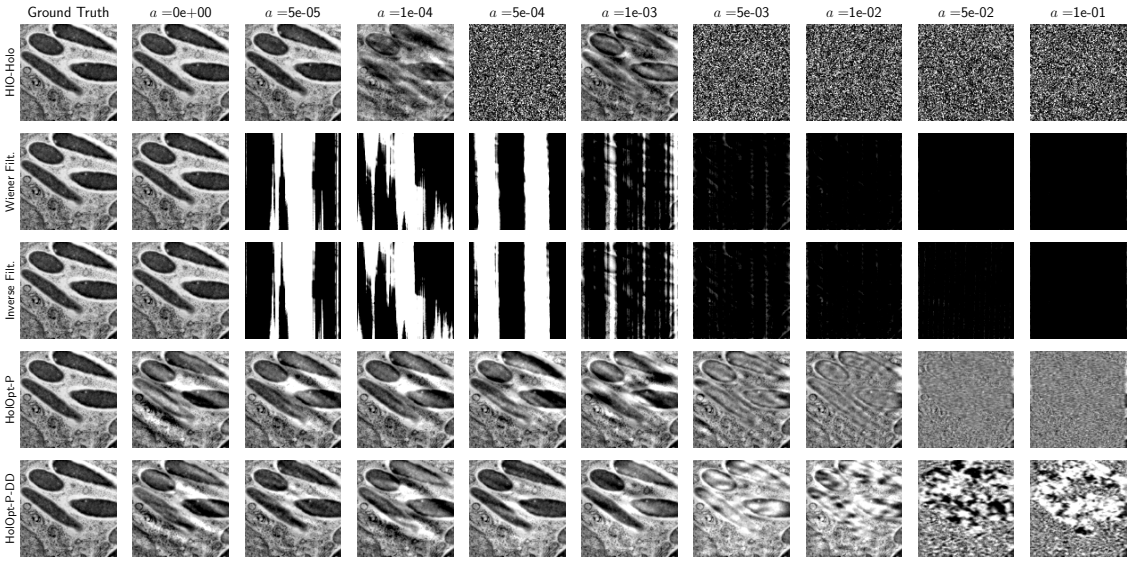



Figure 35: Reconstructed BIO10 image with varying beamstop at $N_p = 1000$ photons.

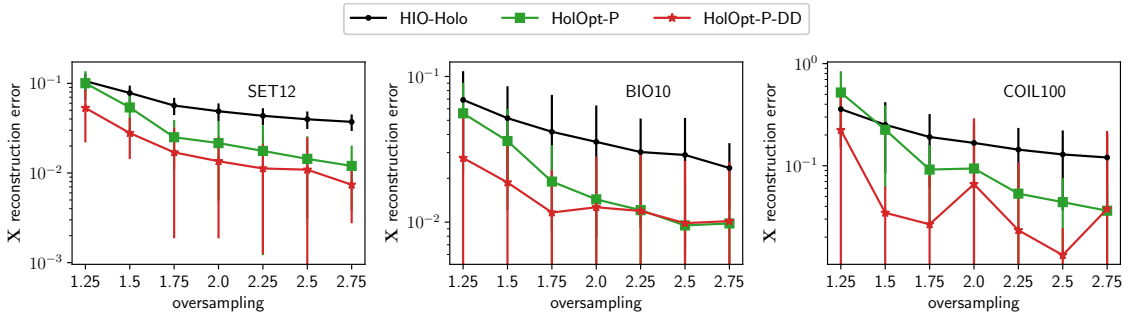



Figure 36: Reconstruction $\ell 2$-errors for SET12, BIO10 and COIL100 with varying oversampling factor at $N_p = 10$ photon/pixel. Corresponding SSIM and visuals are displayed in Figures 11 and 10.

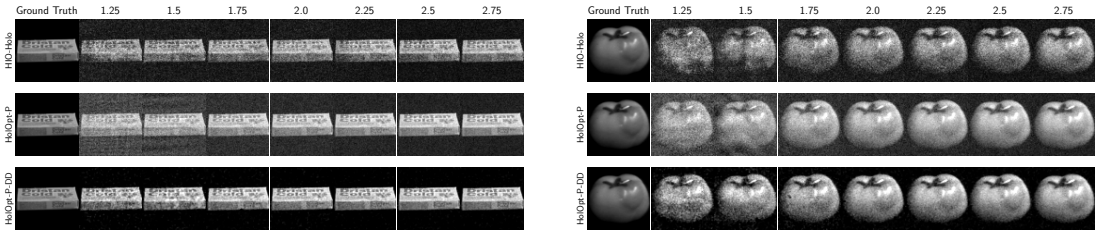



Figure 37: Reconstructed images for samples from COIL100 with varying oversampling factors (numbers above each column) at $N_p = 10$ photon/pixel. Same as Figure 10.

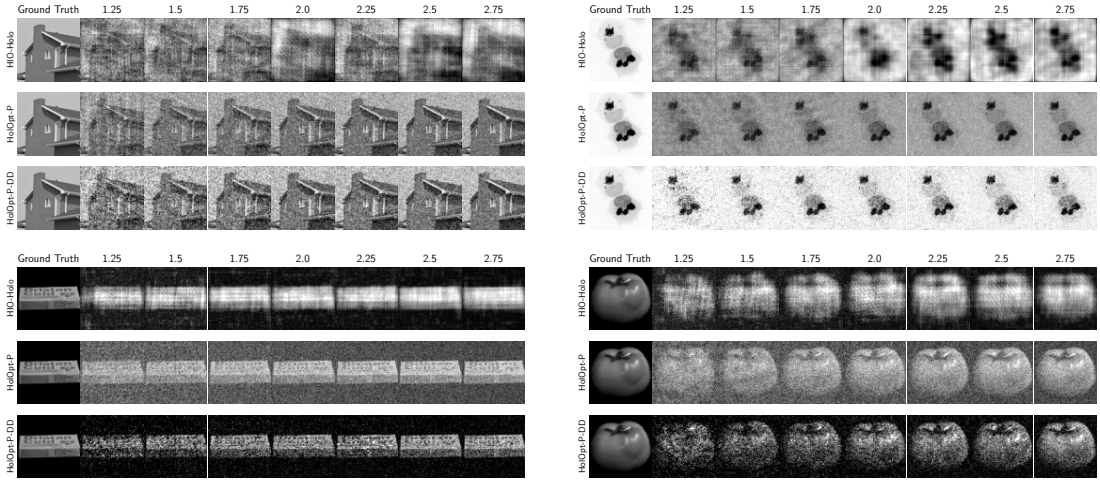



Figure 38: Reconstructed images for SET12, BIO10 and COIL100 with varying oversampling factors (numbers above each column) at $N_p = 1$ photon/pixel.

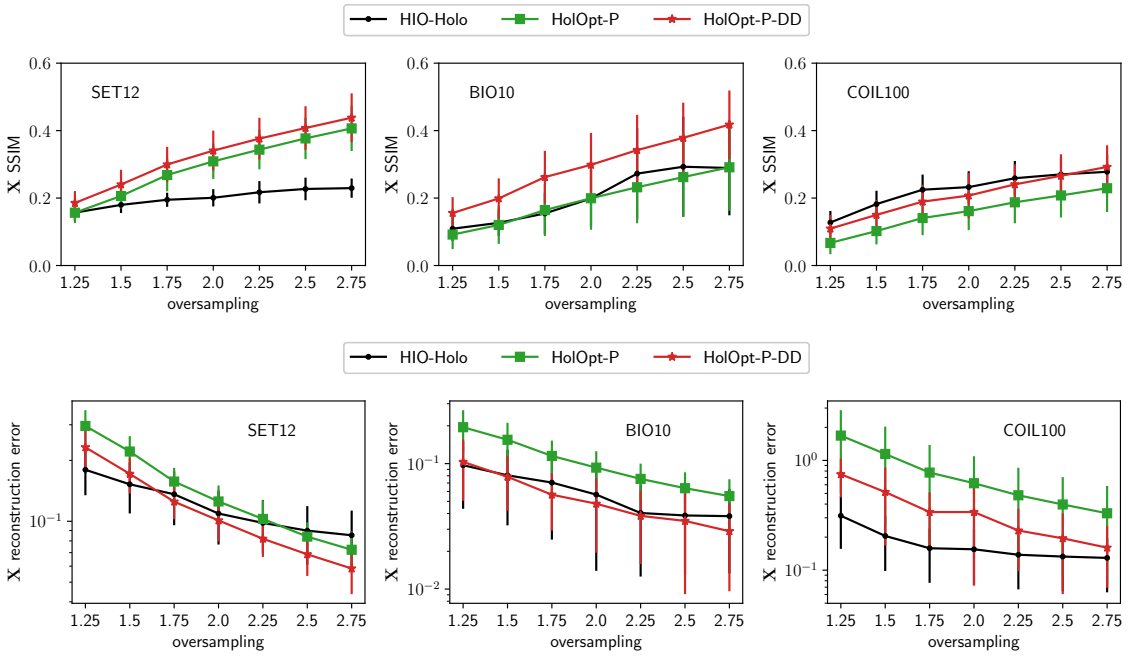



Figure 39: Reconstruction SSIM and $\ell 2$-errors for SET12, BIO10 and COIL100 with varying oversampling factor at $N_p = 1$ photon/pixel. Corresponding visuals are displayed in Figure 38.