PRACTICAL PHASE RETRIEVAL: LOW-PHOTON HOLOGRAPHY WITH UNTRAINED PRIORS

Anonymous authors
Paper under double-blind review

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. 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 $\boldsymbol{R}_0$ distinguishes 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 in 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 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

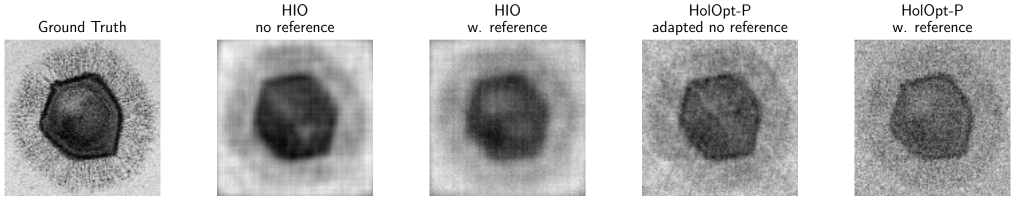


Figure 1: The advantage of using a reference for phase retrieval at $N_p = 1$ photon/pixel. Two algorithms, are applied to reconstruct from Fourier magnitude data of image alone and image+reference. Comparing between algorithms, HIO and our HolOpt-P, observe the higher-quality reconstruction when the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

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 (Barmherzig et al., 2019).

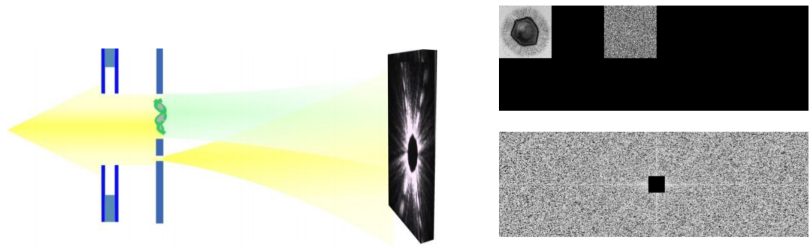
Nevertheless, holographic CDI remains challenging in practice. Due to the quantum mechanical nature of photon emission and absorption, CDI measurements are inherently corrupted by *Poisson shot noise*. The severity of this noise corruption is inversely proportional to the strength of the X-ray source 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).

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). 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 can as well be adapted to the holographic setting.

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 & 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. Com-



(a) Experimental setup.

(b) The zero-padded data and its measured Fourier magnitudes.

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).)

pressive 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 & Ahmed, 2018; Hand et al., 2018), as well as untrained priors (Jagatap & 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 & Fienup (2014), while the convenience of deep learning packages was exploited later (Nashed et al., 2017; Kandel et al., 2019). However, none of these works considered the Poisson likelihood objective, nor holography, nor neural network priors. Recently, Barmherzig & 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 and 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*. Finally, we (e) provide a Python package[1] to run our implementation.

[1]URL to be disclosed after anonymous review, please find the code in the Supplementary Material.

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 $\boldsymbol{Y} \in \mathbb{R}^{2m \times 2n}$ 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 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

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 over the pixels of $\boldsymbol{X}$, as in Equation 3, or over the parameters of a deep decoder neural network prior encoding $\boldsymbol{X}$ (Heckel & 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.

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.

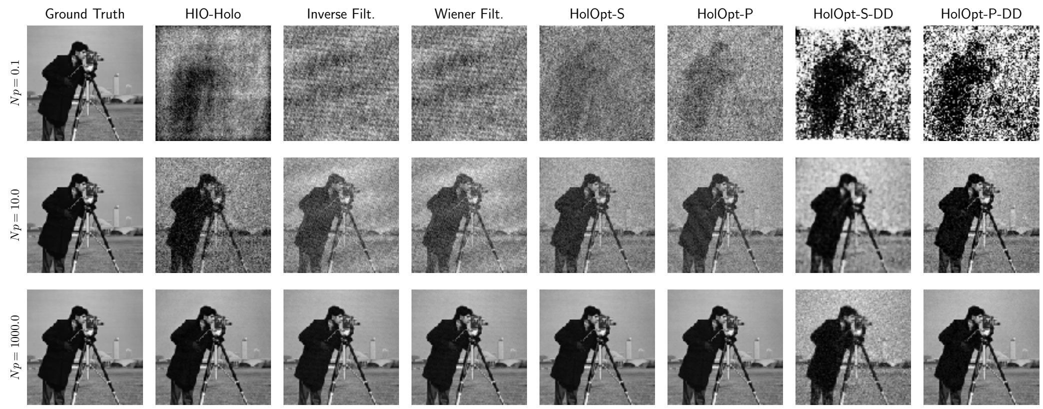


Figure 3: Reconstructed CAMERA by the various compared algorithms as a function of photon count N_p, with binary random reference and without beamstop. Images correspond to the best of 10 runs in terms of residual error. See Figures 11, 12, and 14 in Appendix B for VIRUS, COIL1, and COIL100 images.

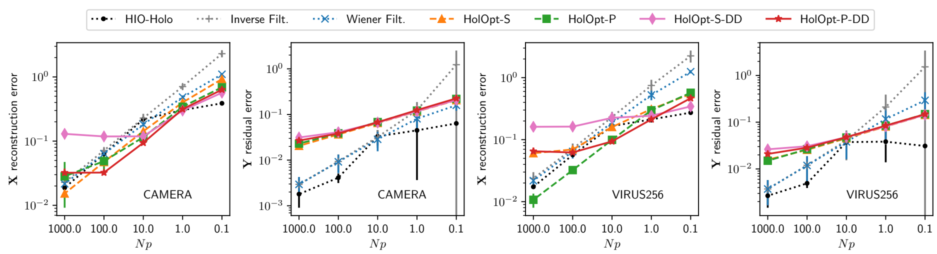


Figure 4: Reconstruction errors and residual errors for CAMERA (left two panels) and VIRUS (right two panels), as a function of the photon count N_p with a binary random reference and without beamstop. See Figures 13 in Appendix B for a similar figure for COIL1 and COIL100.

To formally describe HolOpt-P-DD, we denote by $\sigma(\vartheta, z)$ the output of the deep decoder, where z corresponds to a fixed latent image and $\vartheta = (\vartheta_1, \cdots, \vartheta_d)$ represents all the parameters ϑ_i of the convolutional layers of a deep decoder of depth d. We review the architecture of these networks in more detail in Appendix A. For HolOpt-P-DD, we set $\boldsymbol{X} = \sigma(\vartheta, z)$ and train all ϑ_i in ϑ. In this second case the objective can be re-written as

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

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 $\boldsymbol{Y}$.

4 Experiments

Data Our strategy is demonstrated on the following images. CAMERA is a traditional benchmark in signal processing, here resized to 128×128 pixels, while VIRUS (Ghigo et al., 2008), resized to 256×256 pixels, is a more realistic biological sample. We also consider the COIL100 dataset (Nene et al., 1996) which contains 100 objects on a black background with 128×128 pixels (COIL1 being the first image). 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.

Benchmark setup The strategy proposed in this work is compared against three algorithms for holographic CDI: inverse filtering, Wiener filtering and Hybrid Input-Output (HIO) modified for

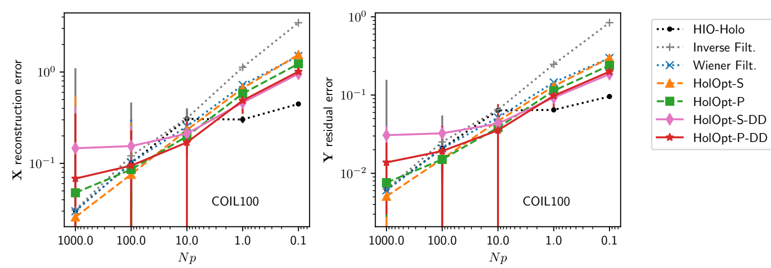


Figure 5: Our methods and baselines evaluated as in Figure 4 on the COIL100 dataset.

holographic phase retrieval, here referred to as HIO-Holo. We further augment HIO-Holo by selecting 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.

Hyperparameters We use the Adam optimizer (Kingma & Ba, 2015) and learning rates varying between 0.01 and 0.1, depending on the loss and prior. For deep decoders, we observe the number of channels to marginally change the outcome of the reconstruction, while varying the depth of the prior trades off precision of the reconstructed edges and finer details (shallower) with spatial regularity (deeper). Practitioners are encouraged to select it by visual inspection. Within image type and noise level, we observed a fixed architecture to yield consistent results. For optimal settings (Table 1) and a visualization of the effect of depth see Appendix A.2.

Evaluation Algorithms are compared in terms of relative reconstruction error, Euclidean distance between the reconstructed $\hat{\boldsymbol{X}}$ and ground truth specimen $\boldsymbol{X}$ normalized by the ℓ_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 ℓ_2-norm of $\boldsymbol{Y}$. Errors are averaged over 10 runs and error bars correspond to standard deviations. Comparisons in terms of Structural Similarity Index (SSIM) are also reported in the Appendix, Figures 15, 18, 19, 22, 24, and 36.

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. 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. In Figure 4, for all images and all but one noise level, our optimization procedure reaches lower reconstruction errors than the baselines for at least one of the variants. Interestingly, visual inspection of Figure 3 shows a clear qualitative improvement in even the highest noise level $N_p = 0.1$, where according to the error metric it appears to be outperformed by the HIO-Holo baseline. It should also be observed that the residual error reflects performance neither perceptually nor according to the error metric. An experiment on the full COIL100 dataset (Figures 5 and 14) confirms our conclusions on a large dataset.

We distinguish several performance regimes for our method. At low noise ($N_p = 1000, 100$), there is no need for regularization by a deep decoder, and HolOpt-P-DD and HolOpt-S-DD achieve higher error rates than HolOpt-P and HolOpt-S. Including a prior is not harmful, but unnecessary. At higher levels of noise ($N_p \leq 10$), the denoising power of the deep decoder is beneficial.

We note finally that the difference between MSE and Poisson likelihood objectives is most drastic when including the deep decoder prior. In particular, MSE sometimes creates artifacts in the images reconstructed by HolOpt-S-DD (see Figures 11, 12 and 14 in Appendix B). Lastly, the reconstruction loss and visual quality of the samples are in almost all cases better with HolOpt-P than with HolOpt-

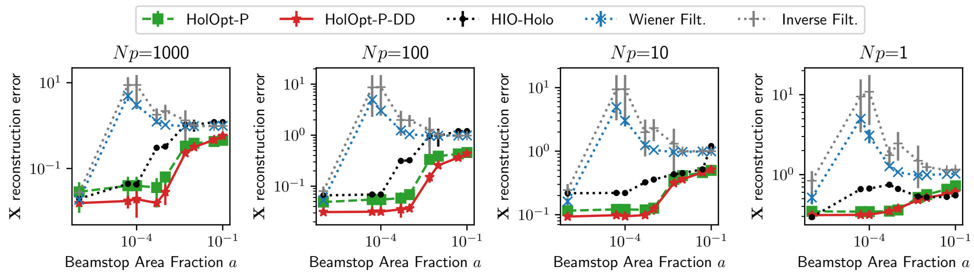


Figure 6: Reconstruction error on CAMERA as a function of beamstop area fraction. Errors and error bars are computed as in Figure 4. The leftmost datapoints at $a = 1\text{e}{-6}$ correspond to no missing data due to rounding, i.e. formally $a = 0$ at the leftmost points.

S, validating our adoption of the most realistic noise model and Poisson likelihood objective. Thus, we focus only on the Poisson likelihood objective going forward.

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 & 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 useful in compensating for the missing magnitudes.

We evaluate our methods HolOpt-P and HolOpt-P-DD at several noise levels and beamstop sizes, with error plots and example reconstructions for CAMERA in Figures 6 and 7 respectively. Supplemental plots for all images and noise levels are in Figures 21-35 of the Appendix, including errors in terms of SSIM and on the complete COIL dataset, which verify 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 20 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 of the deep decoder prior is image-dependent: HolOpt-P-DD yields improved reconstruction relative to HolOpt-P on CAMERA and COIL1, but not on VIRUS (see Figure 21). Even at $Np = 1$ and $a = 0.1$ in Figure 6, the *only* setting for CAMERA at which any baseline (HIO-Holo) attains marginally lower reconstruction error, visual inspection of Figure 7 shows that our methods correctly recover image motifs that HIO-Holo does not. 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 6) relative to baselines. This provides powerful evidence that our method can enable vastly refined reconstructions given even $\frac{1}{10}$ 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. In Figure 8, we report reconstruction errors as a function of the relative specimen-reference separation. Figure 9 displays reconstruction for $N_p = 1$. The only applicable baseline is HIO-Holo.

For VIRUS, the reconstruction appears equally good at all separations, and is comparable to the previous experiment with a full and well-separated random binary reference. At low photon counts,

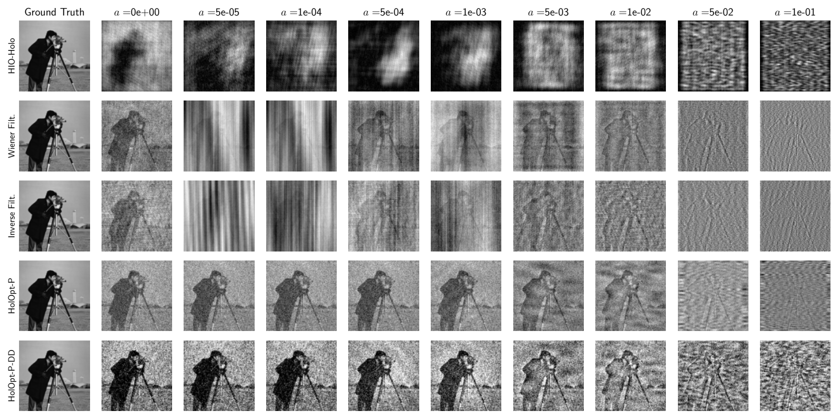


Figure 7: Reconstructed images as a function of beamstop area fraction a for fixed photon count $N_p = 1$. Images correspond to the best of 10 runs in terms of residual error.

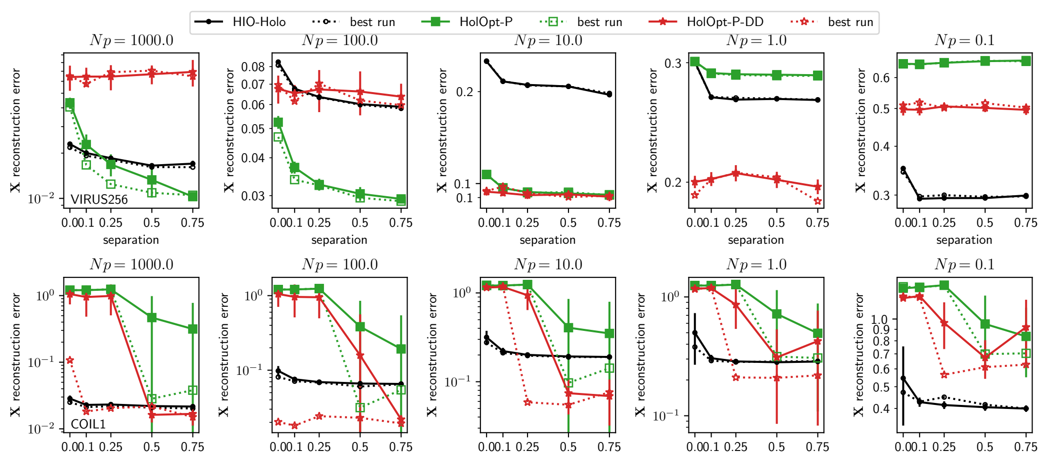


Figure 8: Reconstruction errors for decreasing photon counts N_p with a $0.1m \times 0.1m$ binary random reference as a function of the relative separation. A separation of 0.5 implies that the left-most non-zero pixel of the reference is $0.5n$ pixels away form the image. Dashed lines corresponds to best run out of the 10 runs in terms of residual error.

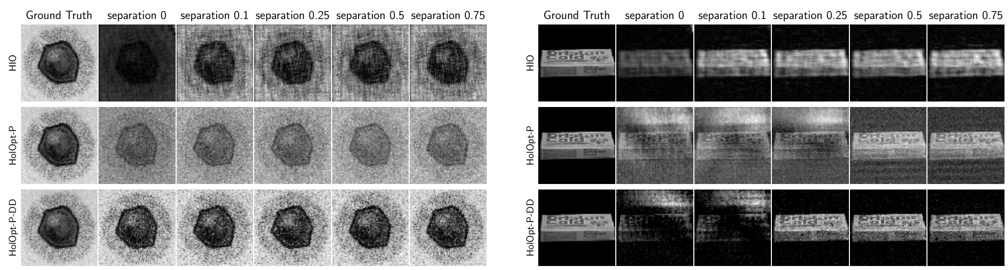


Figure 9: Reconstructed images as a function of the relative separation as described in caption of Figure 8. Images correspond to the best of 10 runs in terms of residual error.

HIO-Holo errors are typically smaller than HolOpt-P errors, and sometimes even smaller than HolOpt-P-DD errors. However, as in the previous experiments, our methods outperform HIO-Holo in perceptual quality. Note that reconstruction with a deep decoder is generally less sensitive to the separation than other methods. These observations are consistent with CAMERA (see Figures 16 and 17 in Appendix B). For COIL1, the impact of the separation distance is stronger. We observe a degradation in reconstruction quality at low separation distances. The most likely reason is that the zero background makes several support locations possible, because the Fourier interaction between the image and reduced reference is not strong enough to enforce the correct support. While HIO does not seem to suffer from this difficulty, HolOpt-P and HolOpt-P-DD sometimes show two shifted duplicates of COIL1 at small distances. Re-running the algorithm can yield reconstructions much better than HIO-Holo, consistent with the best-run errors reported in Figure 8. The two possible outcomes explain the large standard deviations. Interestingly, the threshold separation above which the optimization method can succeed is 0.5 for all noise levels for HolOpt-P, while the deep decoder helps achieve good reconstruction for smaller separations with a threshold that decreases as N_p increases for HolOpt-P-DD. Analogous results for the full COIL dataset can be found in Figure 19 of the Appendix.

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. An additional experiment, in Section C of the appendix, shows that for yet another challenge for reconstruction, namely small oversampling ratios, HoloOpt-P and in particular HoloOpt-P-DD enables better reconstructions than the baseline HIO-Holo.

5 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 amount of information is missing from the measured data due to low photon counts and beamstop-obscured frequencies. Due to its practicality and flexibility, we believe our methodology should prove quite valuable for practitioners. While this work demonstrated our approach's success on simulated observations, we expect to confirm its findings on experimental data in 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).

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, pp. 55–73, 2020.

David A. Barmherzig and Ju Sun. Low-photon holographic phase retrieval. In *Imaging and Applied Optics Congress*, pp. 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.

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.

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.

Wen Hu, Claudio Mazzoli, and Stuart Wilkins. New advances at CSX beamline in magnetic imaging (Conference Presentation). In Barry Lai and Andrea Somogyi (eds.), *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.

Victor Lempitsky, Andrea Vedaldi, and Dmitry Ulyanov. Deep Image Prior. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 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, pp. 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. dAlché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 32*, pp. 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 (eds.). *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.

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)*, pp. 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`.

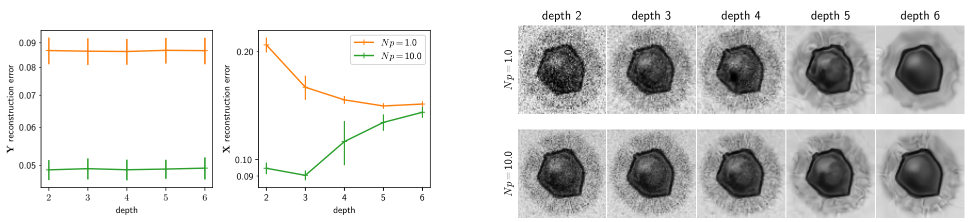


Figure 10: Impact of depth of a deep decoder prior on the reconstruction illustrated on VIRUS. Left: Relative residual errors and reconstruction errors as a function of depth. Right: Best reconstruction out of 10 runs.

A THE DEEP DECODER

In this section we provide further information regarding the deep decoder prior.

A.1 DEEP DECODER ARCHITECTURE

The deep decoder essentially consists of an alternation of two operations — convolutions with filter size of 1×1 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.

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 ϑ_i, by 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. In our experiments we set $\boldsymbol{X} = \sigma(\vartheta, z)$ and train the parameters ϑ_i.

A.2 SELECTION OF DEPTH AND NUMBER OF CHANNELS

When using a deep decoder to reconstruct one needs to specify the number of channels of its convolutional filters, as well as the number of layers. Across images and noise levels, we observed that the performance is marginally affected by the number of channels, except for the fact that it deteriorates for number of channels below a certain threshold (around 64 for our examples). 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 Hand et al. (2018), yet its impact is much more subtle than that of depth.

In Figure 10, we illustrate this fact by reporting errors and reconstructed images of VIRUS as a function of depth for different noise levels. 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. Note that at high noise levels ($N_p = 1.0$) the smoothing reduces the reconstruction error estimated using Euclidean distance. Perceptually however, the smoothing is only beneficial to some extent. Note that while the sum of errors across pixels is greater for shallow models, large mistakes are more evenly distributed whereas for the deep models they are concentrated on edges of the object in the picture. Our eyes are found to be excellent denoisers for speckles. At lower noise levels ($N_p = 10.0$), the excess of depth can be spotted directly in the reconstruction error as we observe a dip in the curve.

noise	$N_p = 1000$	$N_p = 100$	$N_p = 10$	$N_p = 1$	$N_p = 0.1$
CAMERA	d2 c128	d3 c128	d2 c128	d1 c128	d1 c128
COIL1	d2 c128	d2 c128	d2 c128	d1 c128	d1 c128
VIRUS	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 steps of optimizer used for the deep decoder in all experiments.

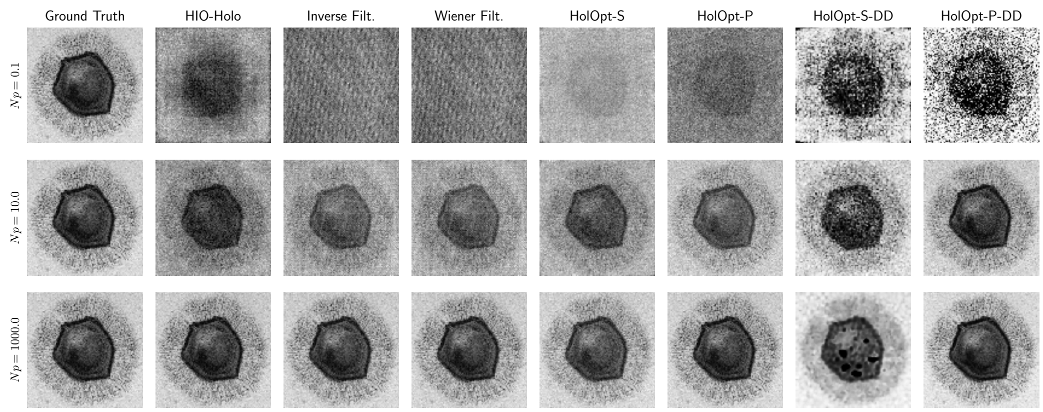


Figure 11: Same as Figure 3. Corresponding errors can be found in Figure 4

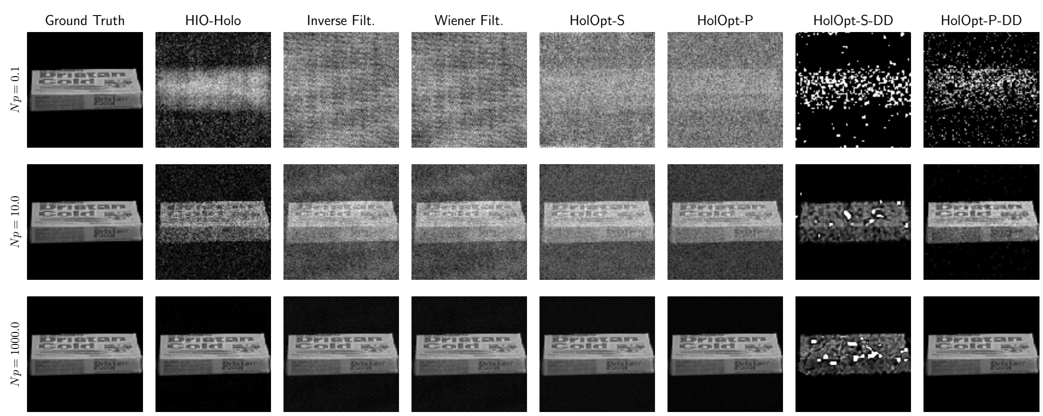


Figure 12: Same as Figure 3. Corresponding errors can be found in Figure 13

Following these observations, we select the parameters of the deep decoders used in our experiments according to the best visual outcomes. These 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.

B Additional figures

In this section we provide additional figures for experiments presented in the main text.

B.1 Noisy reconstruction experiment

Figures 11 through 15 present results for the experiment described in 4.1 of the main text. The superior visual quality of our algorithms is confirmed on VIRUS, COIL1 and COIL100. The trend in errors achieved by the different algorithms is also show to be similar for COIL1. One particularity

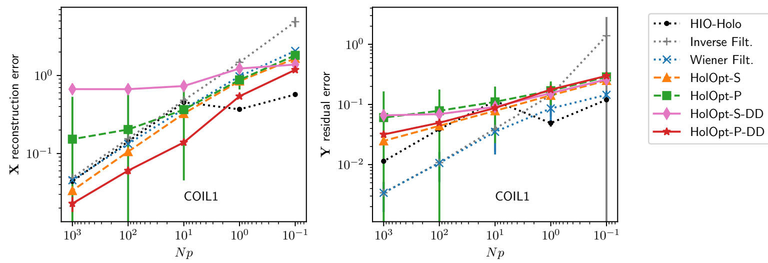


Figure 13: Same as Figure 4

of COIL1 is that HolOpt-P-DD provides the best reconstruction across noise levels on average, while HolOpt-P has high variance.

On Figure 15 we compare methods in terms of Structural Similarity Index (SSIM) Wang et al. (2004). This measures is more informative of the perceived similarity than MSE. At high photon counts, the methods without deep decoder are typically performing best, while the deep decoder proves useful at higher levels of noise. While more informative than the MSE, we note that this is still not a perfect measure, for example focusing on COIL1 at low-photon counts and comparing Figure 15 with Figure 12.

B.2 Robustness to separation experiment

Figures 16 and 17 present results of the experiment presented in Section 4.3 for CAMERA. For CAMERA the addition of a deep decoder prior is almost always beneficial in terms of reconstruction error (HolOpt-P-DD achieves lower errors than HolOpt-P), but there is a clear crossover for VIRUS between low and high noise (see Figure 8). One element of explanation for the difference can be that CAMERA carries larger solid patches, for which the deep decoder acts as a good regularizer. For VIRUS instead, the ground truth image itself presents many small speckles and the reconstruction at low noise is better without the deep decoder regularization. This is also consistent with our first experiment results presented in Figure 4. In Figure 19, we show analogous results to the previous plots for the full COIL dataset.

B.3 Missing low-frequencies (beamstop) experiment

We now visualize the beamstop area fractions tested in Section 4.2, as well as show analogous plots to Figures 6 and 7 for all three test images and for all noise levels, $N_p = 1, 10, 100, 1000$. Figure 22 also plots all settings using the SSIM.

C Oversampling experiment

We report an additional experiment varying the oversampling factor used to simulate observations. For oversampling factors larger or equal to 2, the inverse problem can be solved exactly in the absence of noise. We test how our methods perform compared to HIO-Holo for oversampling ratios around this critical value. Figures 36 and 37 display results for a well-separated binary reference and without beamstop for CAMERA, VIRUS and statistics over the COIL100 datasets. All images reconstructed by HoloOpt methods are visually superior to images reconstructed by HIO-Holo at oversampling factors both smaller and larger than 2, a phenomenon which is largely (although imperfectly) captured by the MSE and SSIM as well.

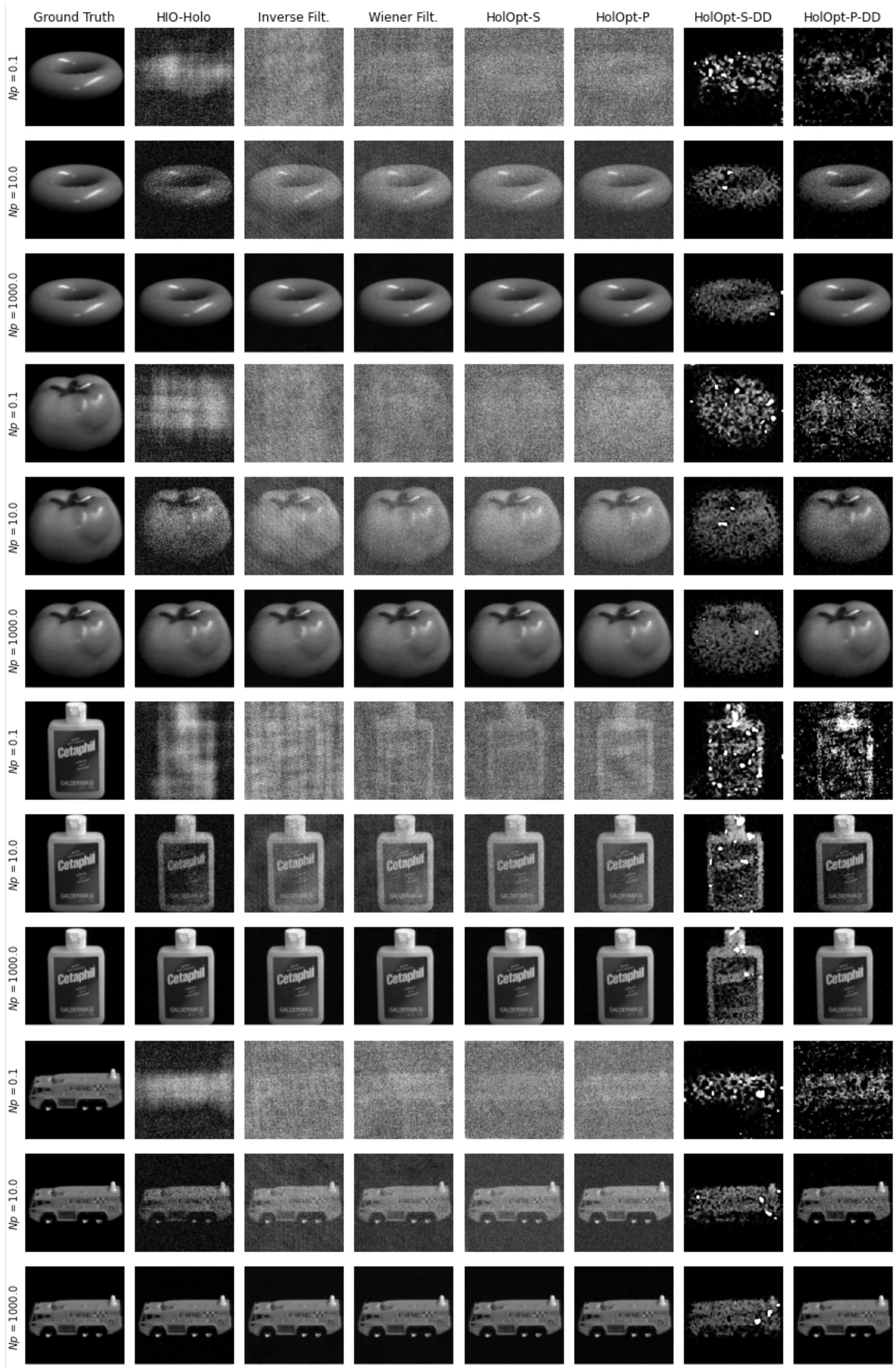


Figure 14: Representative reconstructed images from the COIL100 dataset. Same as Figure 3. Mean error over COIL100 dataset can be found in Figure 5.

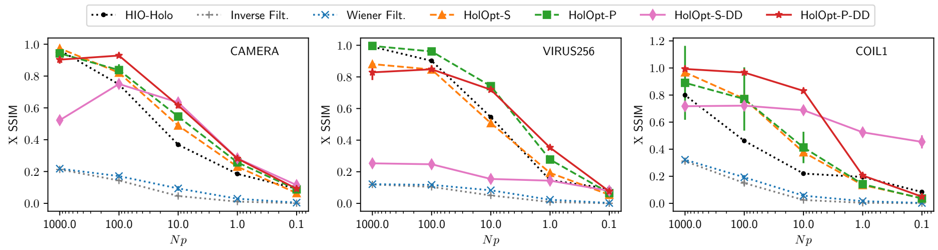


Figure 15: Reconstruction SSIM for CAMERA (left), VIRUS (center), and COIL (right) as a function of the photon count N_p with a binary random reference and without beamstop.

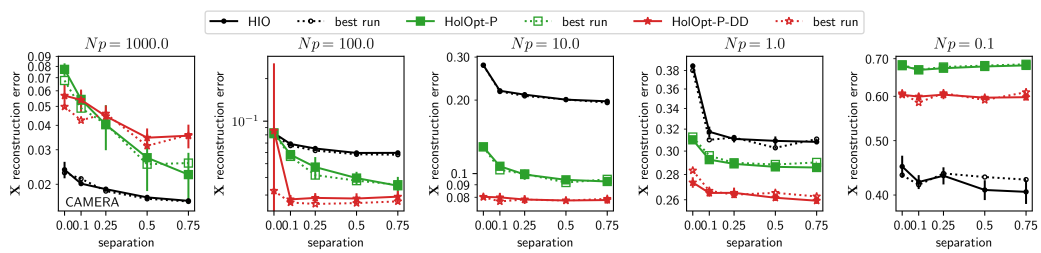


Figure 16: Same as Figure 8 for CAMERA.

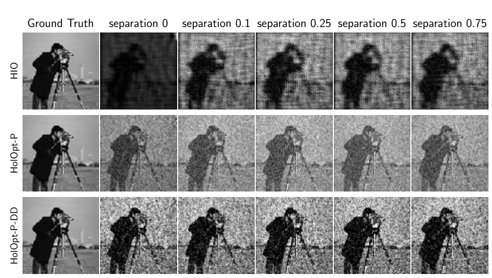


Figure 17: Same as Figure 9.

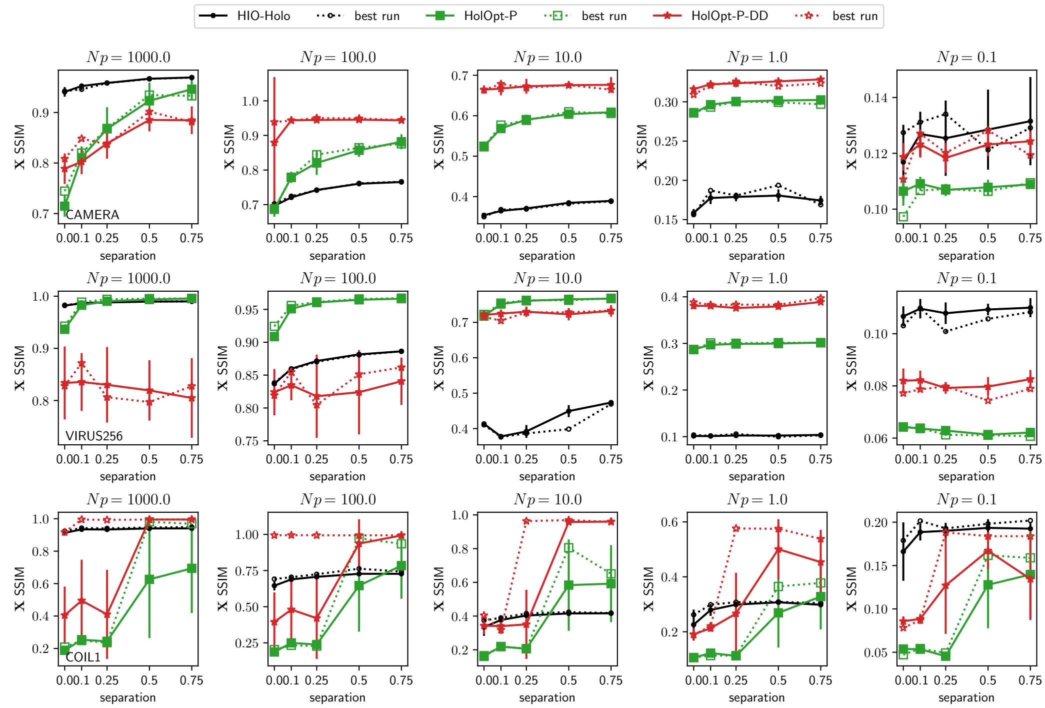


Figure 18: Same as Figure 16, but for all test images and with SSIM rather than MSE.

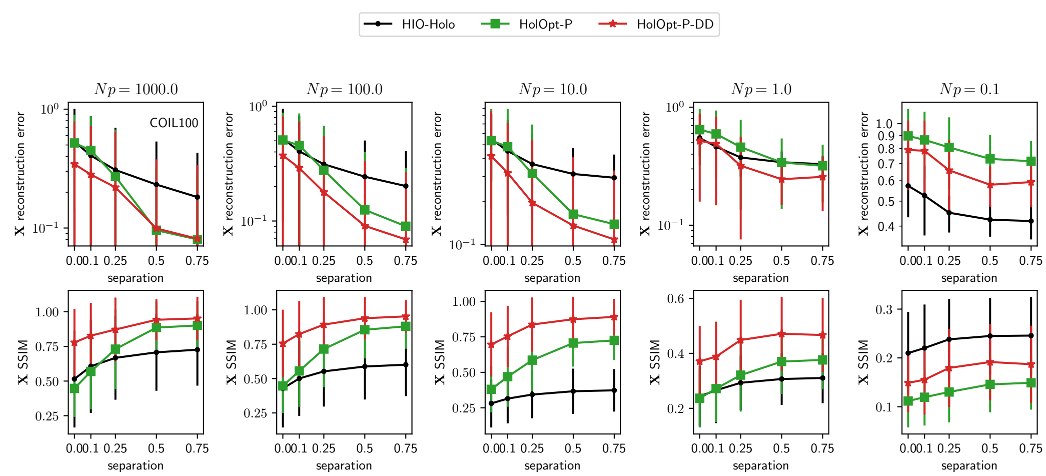


Figure 19: Same as Figure 16, but for all images in the COIL dataset and with both SSIM and MSE.

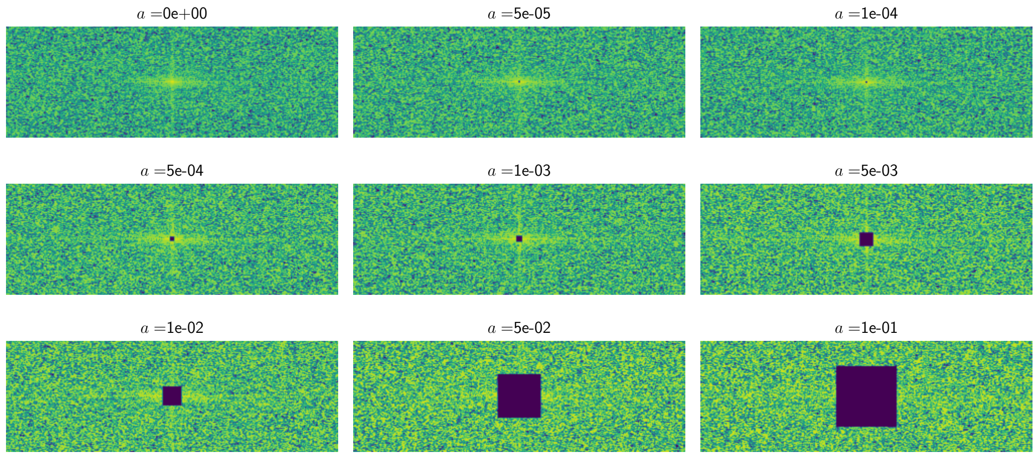


Figure 20: Examples of the measured magnitudes corresponding to varying beamstop area fractions for the CAMERA image with $N_p = 10$.

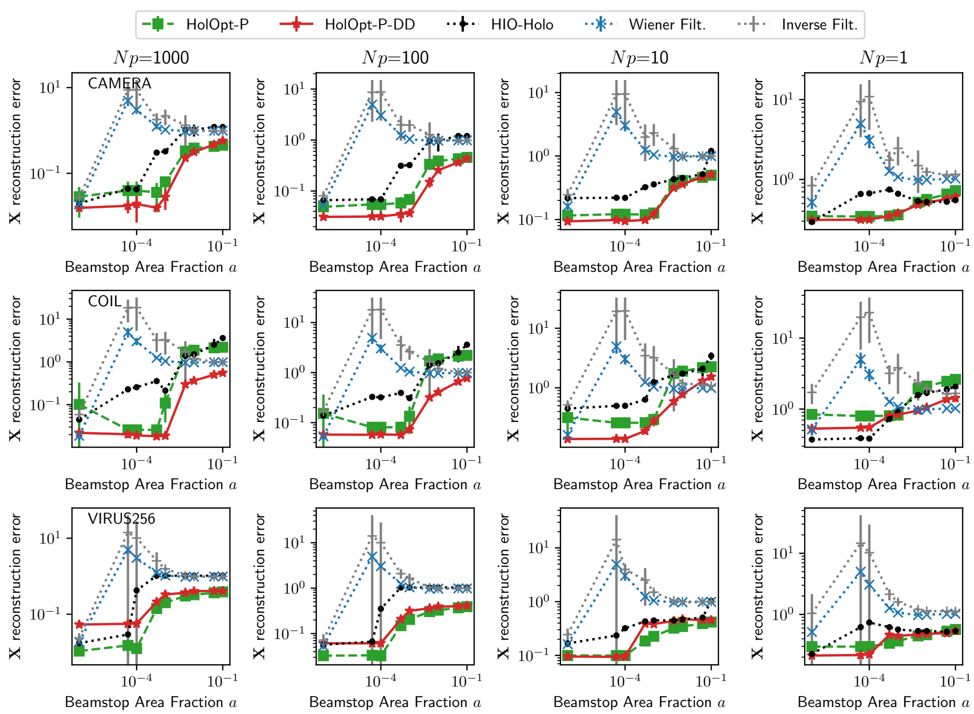


Figure 21: Reconstruction error on all three images (one per row, as indicated by the leftmost row levels) as a function of beamstop area fraction. Errors and error bars are computed as in Figure 4. The leftmost datapoints at $a = 1\mathrm{e}{-6}$ correspond to no missing data due to rounding, i.e. formally $a = 0$ at the leftmost points.

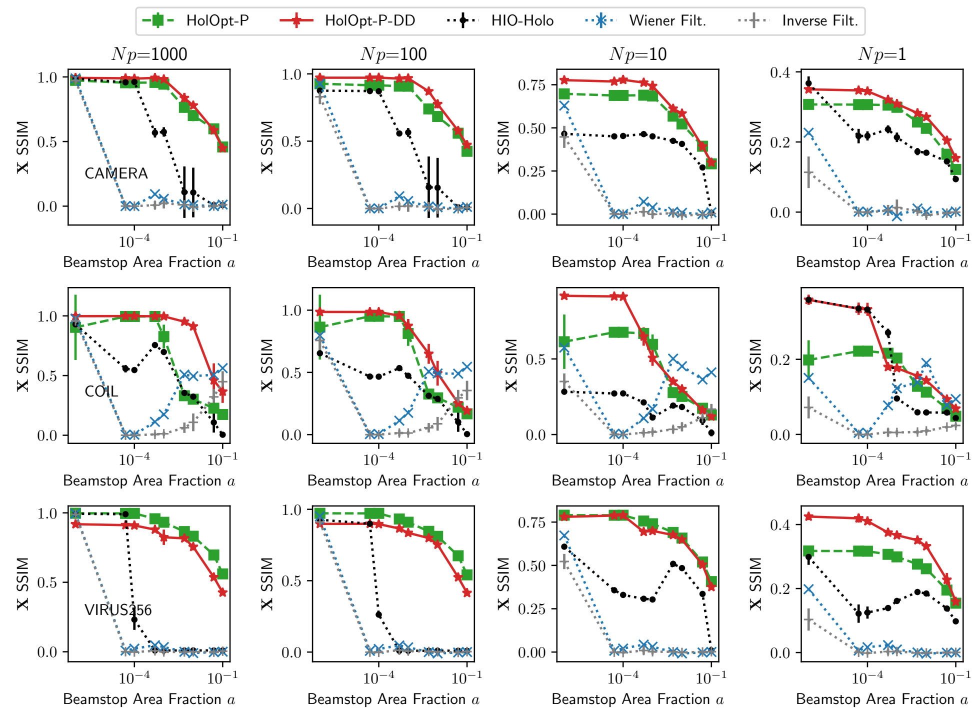


Figure 22: SSIM on all three images (one per row, as indicated by the leftmost row levels) as a function of beamstop area fraction. Errors and error bars are computed as in Figure 4. The leftmost datapoints at $a = 1\text{e}{-6}$ correspond to no missing data due to rounding, i.e. formally $a = 0$ at the leftmost points.

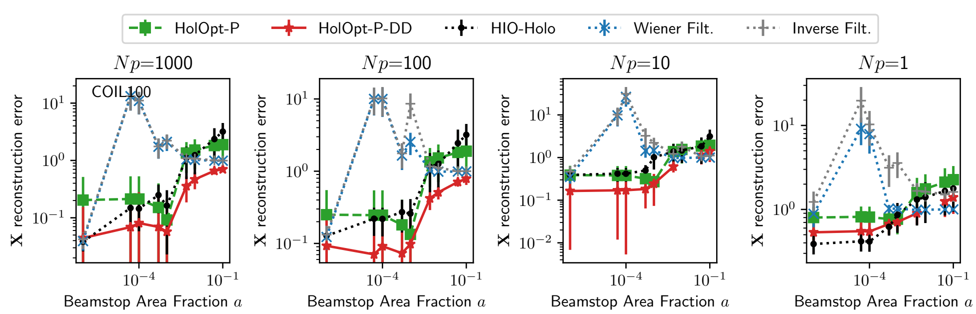


Figure 23: Reconstruction error on the full COIL dataset as a function of beamstop area fraction. Errors and error bars are computed over all 100 images, one trial each. The leftmost datapoints at $a = 1\text{e}{-6}$ correspond to no missing data due to rounding, i.e. formally $a = 0$ at the leftmost points.

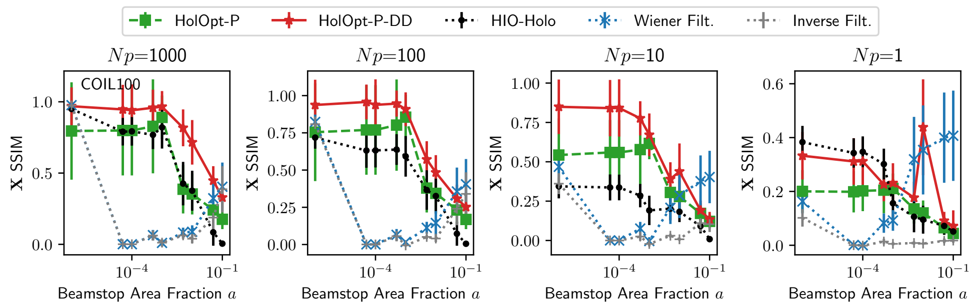


Figure 24: SSIM on the full COIL dataset as a function of beamstop area fraction. Errors and error bars are computed over all 100 images, one trial each. The leftmost datapoints at $a = 1\text{e}{-6}$ correspond to no missing data due to rounding, i.e. formally $a = 0$ at the leftmost points.

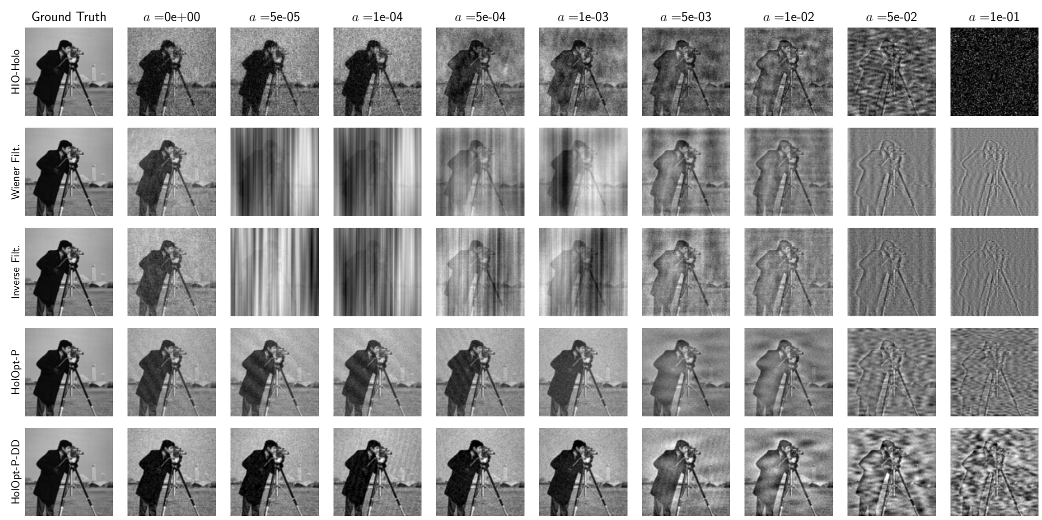


Figure 25: Reconstructed CAMERA image with varying beamstop at $N_p = 10$ photons.

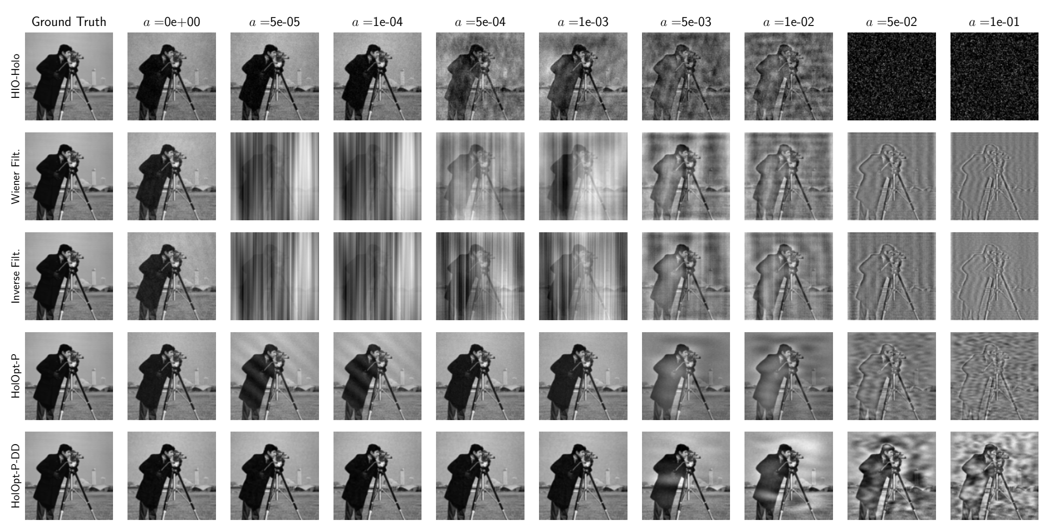


Figure 26: Reconstructed CAMERA image with varying beamstop at $N_p = 100$ photons.

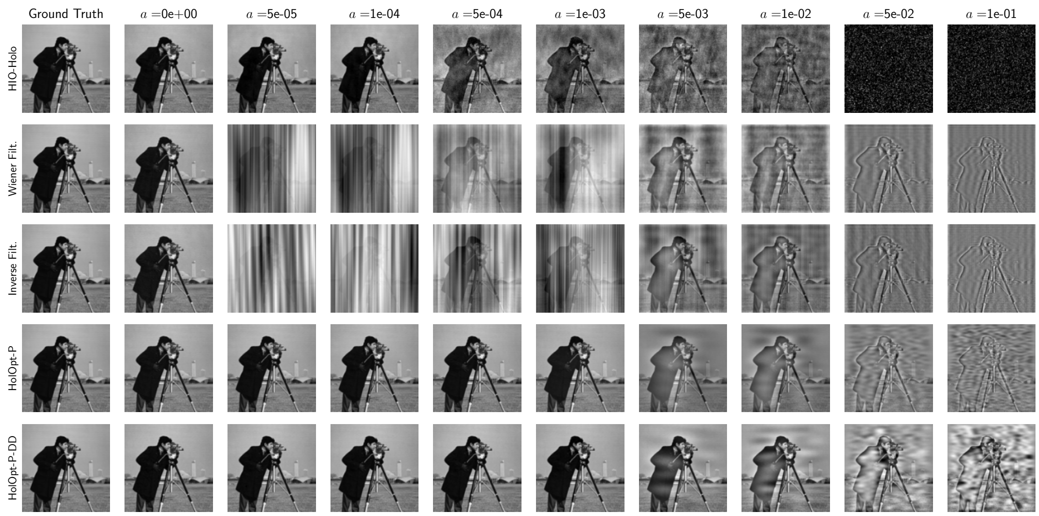


Figure 27: Reconstructed CAMERA image with varying beamstop at $N_p = 1000$ photons.

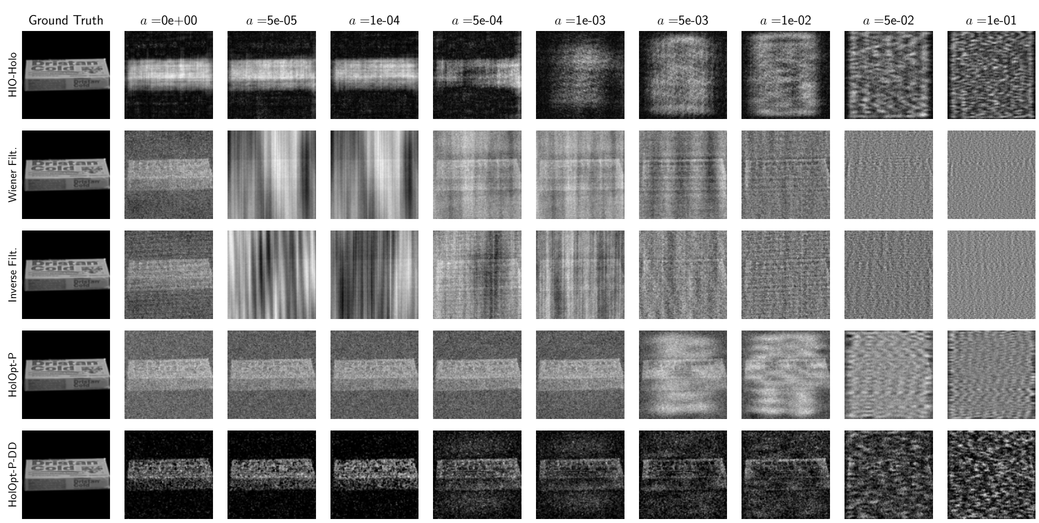


Figure 28: Reconstructed COIL image with varying beamstop at $N_p = 1$ photons.

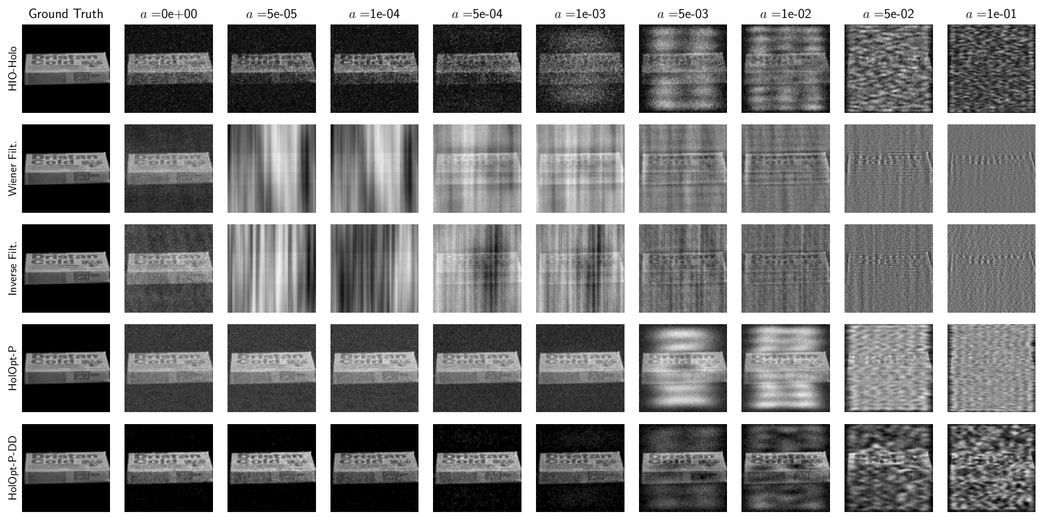


Figure 29: Reconstructed COIL image with varying beamstop at $N_p = 10$ photons.

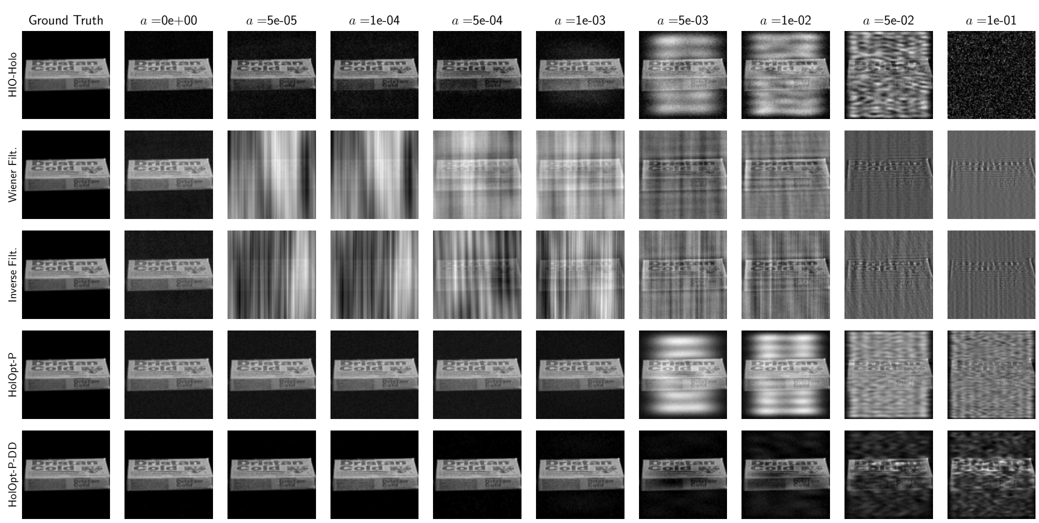


Figure 30: Reconstructed COIL image with varying beamstop at $N_p = 100$ photons.

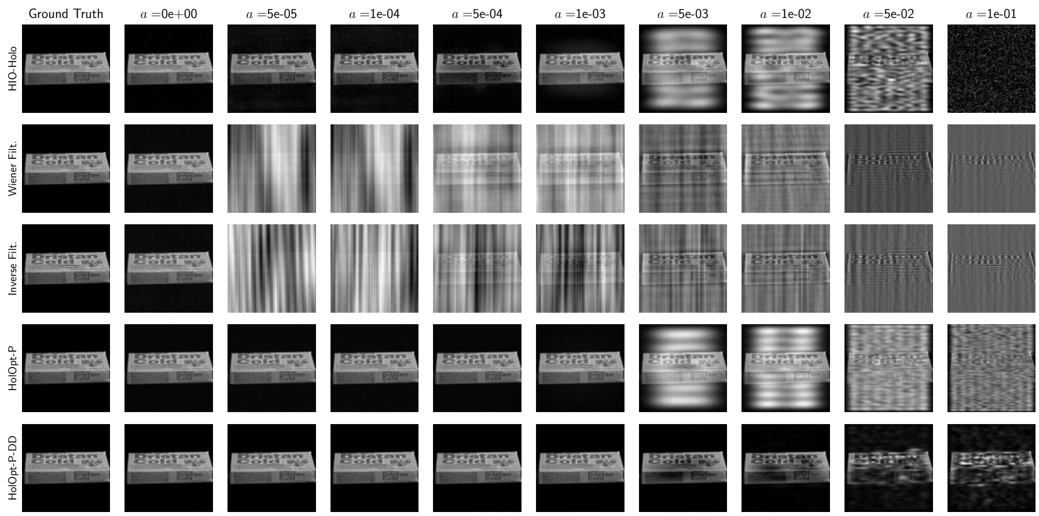


Figure 31: Reconstructed COIL image with varying beamstop at $N_p = 1000$ photons.

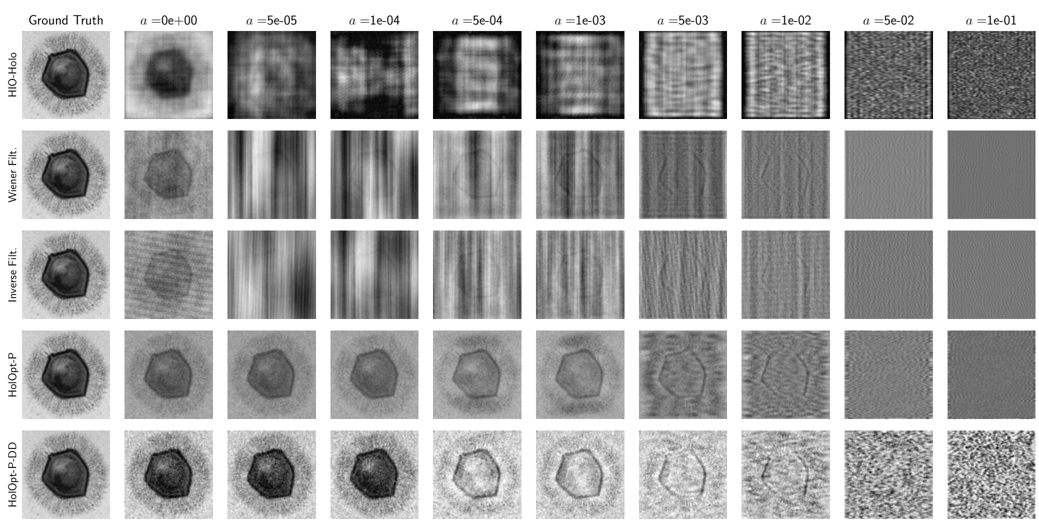


Figure 32: Reconstructed VIRUS image with varying beamstop at $N_p = 1$ photons.

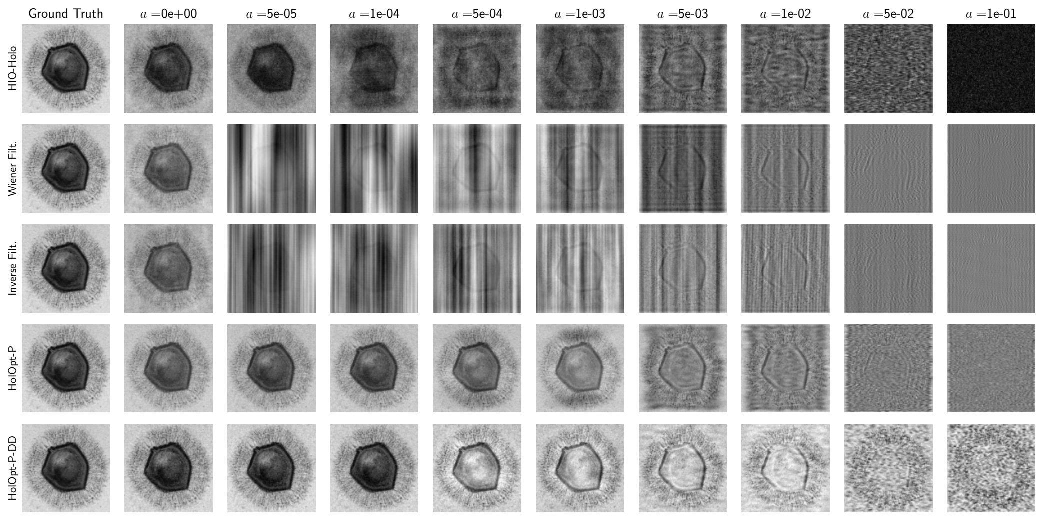


Figure 33: Reconstructed VIRUS image with varying beamstop at $N_p = 10$ photons.

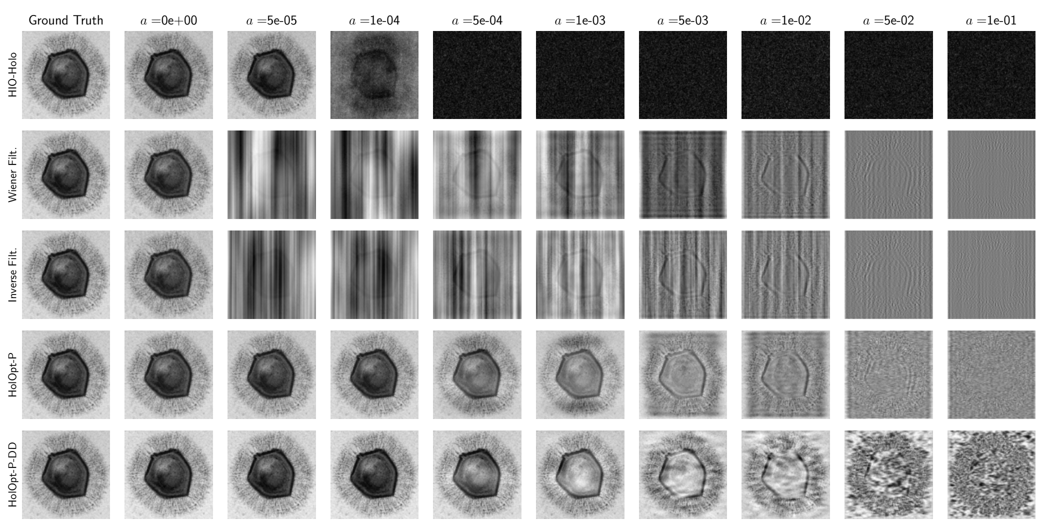


Figure 34: Reconstructed VIRUS image with varying beamstop at $N_p = 100$ photons.

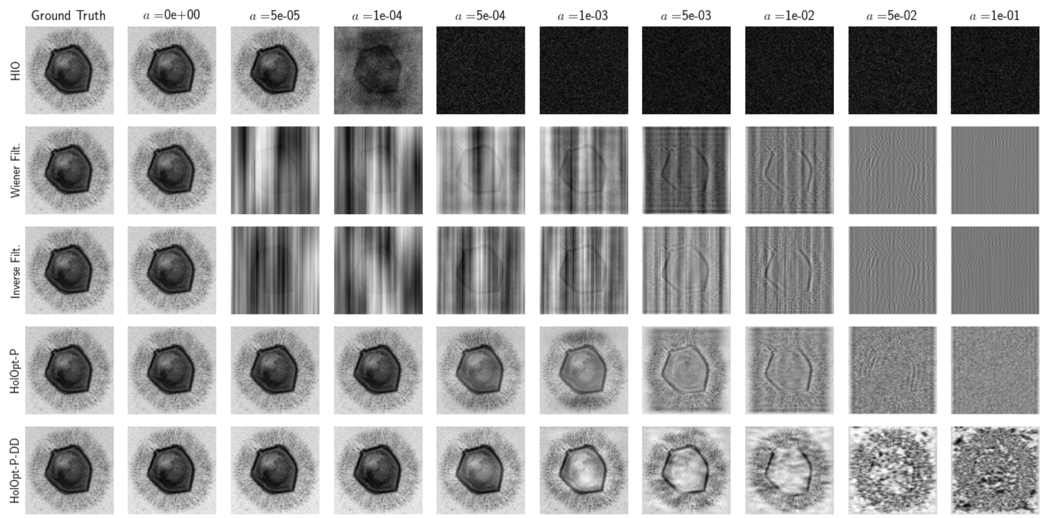


Figure 35: Reconstructed VIRUS image with varying beamstop at $N_p = 1000$ photons.

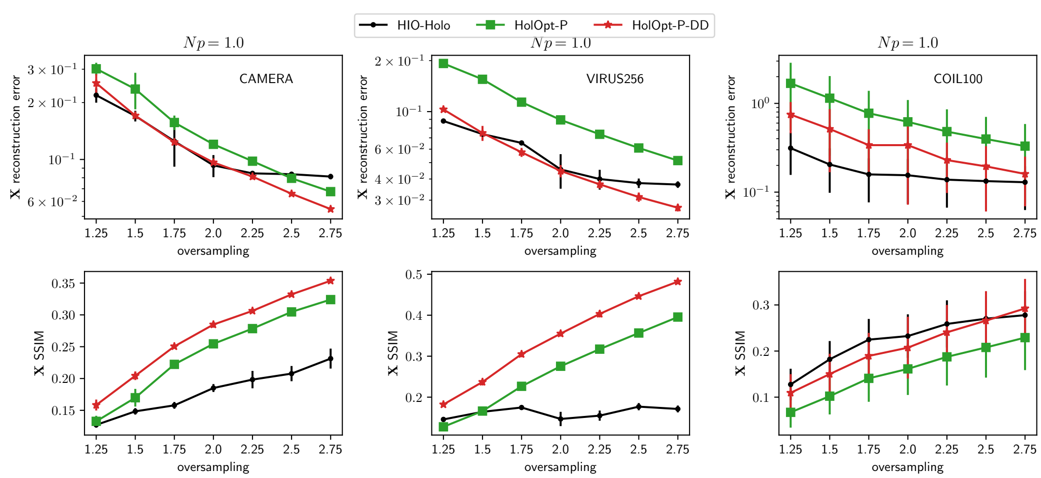


Figure 36: Reconstruction errors and SSIM for CAMERA, VIRUS and their averages over the COIL100 dataset with varying oversampling factor at $N_p = 1$ photon/pixel.

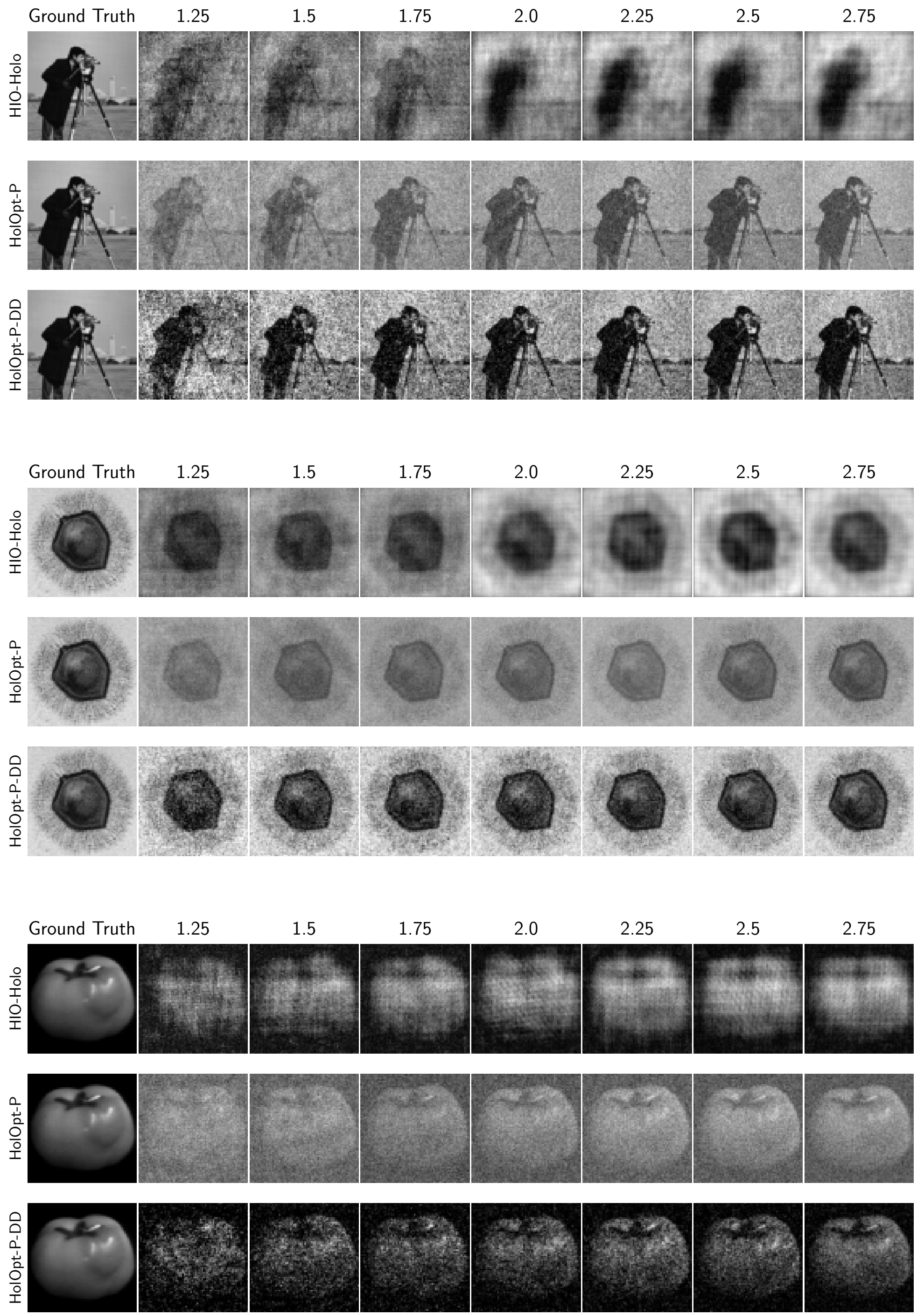


Figure 37: Reconstructed images for CAMERA, VIRUS and their averages over the COIL100 dataset with varying oversampling factors (numbers above each column) at $N_p = 1$ photon/pixel.