Learning Conditional Generative Models for Phase Retrieval

Tobias Uelwer   TOBIAS.UELWER@TU-DORTMUND.DE
Department of Computer Science
Technical University of Dortmund
Otto-Hahn-Straße 12, 44227 Dortmund, Germany

Sebastian Konietzny   SEBASTIAN.KONIETZNY@TU-DORTMUND.DE
Department of Computer Science
Technical University of Dortmund
Otto-Hahn-Straße 12, 44227 Dortmund, Germany

Alexander Oberstrass   ALEXANDER.OBERSTRASS@HHU.DE
Department of Computer Science
Heinrich Heine University Düsseldorf
Universitätsstraße 1, 40225 Düsseldorf, Germany

Stefan Harmeling   STEFAN.HARMELING@TU-DORTMUND.DE
Department of Computer Science
Technical University of Dortmund
Otto-Hahn-Straße 12, 44227 Dortmund, Germany

Abstract

Reconstructing images from magnitude measurements is an important and difficult problem arising in many research areas, such as X-ray crystallography, astronomical imaging and more. While optimization-based approaches often struggle with the non-convexity and non-linearity of the problem, learning-based approaches are able to produce reconstructions of high quality for data similar to a given training dataset. In this work, we analyze a class of methods based on conditional generative adversarial networks (CGAN). We show how the benefits of optimization-based and learning-based methods can be combined to improve reconstruction quality. Furthermore, we show that these combined methods are able to generalize to out-of-distribution data and analyze their robustness to measurement noise. In addition to that, we compare how the methods are impacted by missing measurements. Extensive ablation studies demonstrate that all components of our approach are essential and justify the choice of network architecture.

Keywords: Phase retrieval, conditional generative adversarial networks, image reconstruction.

1. Introduction

Phase retrieval is the process of reconstructing images from magnitude-only measurements. This is a relevant problem in different research areas, e.g., in X-ray crystallography (Millane, 1990), astronomical imaging (Fienup and Dainty, 1987), optics (Walther, 1963), array imaging (Bunk et al., 2007), or microscopy (Zheng et al., 2013). For simplicity, we define the problems for the one-dimensional case but all methods discussed in this work are also

applicable to images with more dimensions. In this paper, we consider reconstructing an image x from measurements

$$y = |Ax|, \tag{1}$$

where A is a known linear operator that depends on the specific application.

A relevant instance of this problem is Fourier phase retrieval, where the operator is the discrete Fourier transform $\mathcal{F}$. In that case, the measurements are given as

$$y = |\mathcal{F}x|. \tag{2}$$

For color images the Fourier transform is applied to each color-channel separately. In typical real-world settings, such as X-ray crystallography, the Fourier measurements are not oversampled, i.e., the image x is not zero-padded. Thus, the measurements y have the same size as the unknown image x. This setup makes the Fourier phase retrieval problem particularly difficult to solve, which is why we consider this setup in our work.

Another instance of the problem is compressive Gaussian phase retrieval, where the measurements are obtained by multiplying the image with a random rectangular matrix M that has entries sampled from a Gaussian distribution, i.e.,

$$y = |Mx|. \tag{3}$$

Fourier measurements will be discussed in Section 7.1 and Gaussian measurements in Section 7.7.

2. Our Contributions

Note that this paper extends our previous conference paper (Uelwer et al., 2021b). Our contributions can be summarized as follows:

1. We describe the PRCGAN, which combines a conditional generative adversarial network and a subsequent optimization to solve different phase retrieval problems.

2. We study three variants of the PRCGAN: (i) in an end-to-end mode (PRCGAN-D), (ii) in combination with latent optimization (PRCGAN-L), and (iii) in combination with weight optimization of the network (PRCGAN-W).

3. We extensively evaluate all variants of the PRCGAN on the Fourier phase retrieval problem using openly available benchmark datasets.

4. We perform detailed ablation studies to examine the impact of each of the components of our method. We also experiment with different loss functions and model architectures.

5. We analyze how our trained models can generalize to out-of-distribution data. In order to do so, we evaluate the performance of the models on datasets that differ from the training set. Furthermore, we create a novel dataset that contains MNIST-like symbols. We also use this dataset to evaluate our models.

6. We investigate the robustness of the different models: we study the impact of Poisson measurement noise and additive Gaussian measurement noise and analyze the impact of missing measurement entries on the reconstruction process.

7. While this work focuses on Fourier phase retrieval, we show in additional experiments that the proposed methods are also applicable to the Gaussian phase retrieval problem.

The remainder of this work is structured as follows: Section 3 gives an overview of existing optimization-based and learning-based phase retrieval methods. Section 4 discusses baseline learning-based approaches that form the basis of our proposed method. In Section 5 we propose our hybrid method that combines end-to-end learning with a subsequent optimization. Furthermore, we discuss how to classify the learning-based approaches. In Section 6 we discuss our experimental setup. Section 7 presents the results of our experiments for the Fourier and Gaussian phase retrieval problem. Furthermore, we analyze the out-of-distribution generalization of our method and perform an extensive ablation study. Section 8 summarizes our method and results.

3. Related Work

Existing methods to solve the phase retrieval problem can be classified into two categories: optimization-based and learning-based approaches. While optimization-based approaches are especially appealing when the number of measurements is larger than the number of pixels (i.e., the measurements are oversampled), these methods usually fail in the non-oversampled regimes. Learning-based methods use additional information about the distribution of the target images to solve the phase retrieval problem (sometimes also in the non-oversampled case).

There are several other setups that are related to phase retrieval, e.g., ptychography, where we reconstruct an image given a sequence of magnitude measurements of many partially overlapping frames (Xu et al., 2018). Solving ptychography with conditional generative adversarial networks has been done by Boominathan et al. (2018) and Kumar et al. (2019). Another related problem is compressed sensing which asks to reconstruct images from linear measurements. Kim et al. (2021) recently also proposed approaching the compressed sensing problem with conditional GANs. Generative priors with sparsity constraints for compressed sensing have been discussed by Killedar et al. (2021).

However, these are different from the problems discussed in this paper. To the best of our knowledge, we are the first to apply the conditional GAN framework to solve the Fourier and Gaussian phase retrieval problem.

3.1 Optimization-based Approaches

One of the first phase retrieval algorithms is the Gerchberg-Saxton (GS) algorithm (Gerchberg, 1972), which starts with a random image and iteratively enforces a magnitude constraint in the Fourier domain and a positivity constraint of pixel intensities in the object domain. Based on the GS algorithm Fienup (1982) proposed three extensions: the input-output, the output-output and the hybrid input-output (HIO) algorithm, where the last one is the most popular since it usually produces the best results among the three.

Luke (2004) proposed another iterative phase retrieval algorithm which is based on relaxed averaged alternating reflections (RAAR).

3.2 Learning-based Approaches

Learning-based methods are used for the oversampled Fourier phase retrieval problem and for the non-oversampled Fourier phase retrieval problem.

3.2.1 OVERSAMPLED PROBLEM

Deep neural network approaches for the simpler oversampled Fourier phase retrieval problem have, e.g., been discussed by Manekar et al. (2020), who propose a passive loss for end-to-end learning which is invariant to symmetries, and by Cha et al. (2020), who formulate a novel loss function based on the PhaseCut algorithm (Waldspurger et al., 2015). The regularization-by-denoising framework for oversampled Fourier phase retrieval is discussed by Metzler et al. (2018), Wu et al. (2019) and Wang et al. (2020).

3.2.2 NON-OVERSAMPLED PROBLEM

Deep neural networks for the non-oversampled Fourier phases retrieval problem have first been studied by Nishizaki et al. (2020). Their end-to-end learning approach has later been extended to a deep neural network cascade by Uelwer et al. (2021a).

Learning-based methods for the Gaussian phase retrieval problem include deep generative priors (Hand et al., 2018), deep generative priors with sparsity constraints (Killedar and Seelamantula, 2022) and untrained neural network priors (Jagatap and Hegde, 2019).

4. Baseline Methods

In our experiments, we will compare against several baseline methods, including the well-known HIO and the RAAR. In addition to these, there exist learning-based approaches, which are the basis for our method PRCGAN. The idea of learning-based methods is that ill-posed problems are easier to solve if properties of the particular application domain can be exploited. Concretely, we assume to have access to example images $x_1, \ldots, x_N$ which allow us to learn a specialized reconstruction procedure.

First, we formulate in Section 4.1 an end-to-end approach, that directly learns the mapping H_ϕ from magnitudes to the images. Next we discuss deep generative priors (DPR) from Hand et al. (2018) who use a generative model G that was already trained on the images $x_1, \ldots, x_N$ without using magnitude measurements. The image is reconstructed by searching in the latent space.

For our method PRCGAN we combine both ideas and learn a conditional generative model which has access to the magnitude information during training and is subsequently used in an optimization procedure to recover the unknown image (similar to DPR).

4.1 End-to-End (E2E)

An end-to-end approach (E2E) to reconstruct images from their magnitude measurements is to train a deep neural network H_ϕ with learnable parameters ϕ that maps from the space

of the magnitudes directly to the image space. This is achieved by minimizing the error between the true image (from the given dataset) and its reconstruction (from the neural network)

$$\mathcal{L}_{\text{rec}}(\phi) = \frac{1}{N} \sum_{i=1}^{N} \|x_i - H_\phi(|Ax_i|)\|_p^p, \tag{4}$$

where for $p = 1$ this is the mean absolute error (MAE) and for $p = 2$ it is the mean squared error (MSE). At test time the reconstruction is performed using a single forward pass through the network H_ϕ. This approach has been studied by Nishizaki et al. (2020) using a ResNet architecture and by Uelwer et al. (2021a) using a fully connected network.

4.2 Deep Generative Priors (DPR)

Hand et al. (2018) use a generative model G that was already trained on the example images, and then use G to solve the phase retrieval problem by minimizing the error between the measured magnitudes and the magnitudes of the reconstruction

$$z^* = \arg\min_z \|y - |AG(z)|\|_2^2. \tag{5}$$

The generative model G restricts the search space to images that come from the distribution that G was trained on. Hand et al. (2018) solve the optimization problem stated in Equation (5) by gradient descent and then obtain the reconstruction as $\hat{x} = G(z^*)$. They call their method deep phase retrieval (DPR).

5. PRCGAN: Combining End-to-End Learning and DPR

To increase the quality of the reconstructions we replace the generative model G of DPR with a conditional GAN (CGAN) that is conditioned on the magnitude measurements. In that way, our approach, called PRCGAN, is a hybrid between the E2E and the DPR approach. By doing so, we learn a CGAN that is tailored to the phase retrieval reconstruction process. The PRCGAN consists of a discriminator network D_θ with parameters θ and a generator network G_ϕ with parameters ϕ. Both networks are conditioned on the given magnitude measurement. For the latent variable z we choose a multivariate Gaussian with zero mean and unit covariance matrix and denote the latent distribution by q. After training, the generator network G_ϕ takes the role of the reconstruction network.

5.1 PRCGAN: Training

The PRCGAN is trained by optimizing a combination of two losses: an adversarial loss

$$\mathcal{L}_{\text{adv}}(\theta, \phi) = \frac{1}{N} \sum_{i=1}^{N} \log D_\theta(x_i, y_i) + \frac{1}{N} \sum_{i=1}^{N} \log\left(1 - D_\theta(G_\phi(z_i, y_i), y_i)\right), \tag{6}$$

where $y_i = |Ax_i|$ and $z_i \sim q$ for $i = 1, \ldots, N$ and a reconstruction loss

$$\mathcal{L}_{\text{rec}}(\phi) = \frac{1}{N} \sum_{i=1}^{N} \|x_i - G_\phi(z_i, y_i)\|_p^p, \tag{7}$$

5

where p is either 1 or 2. During training the parameters θ and ϕ of D and G are optimized to solve the min-max-problem

$$\min_{\phi} \max_{\theta} \mathcal{L}_{\text{adv}}(\theta, \phi) + \lambda \mathcal{L}_{\text{rec}}(\phi), \tag{8}$$

where $\lambda > 0$ is a hyperparameter that balances the two losses.

5.2 PRCGAN: Reconstructions

A trained PRCGAN gives us several options for the reconstruction which span the spectrum between end-to-end learning-based and optimization-based image reconstruction. Besides directly using the output of the PRCGAN, we can further enhance the results by optimizing the latent variable or even the weights of the generator itself.

5.2.1 PRCGAN-D: DIRECT RECONSTRUCTION

The PRCGAN can be directly used to reconstruct images with a single forward pass. To do so, the given magnitude measurement y and a randomly sampled value z from the latent distribution are fed into the generator G_ϕ. This gives us the reconstruction

$$\hat{x} = G_\phi(z, y). \tag{9}$$

While $\hat{x}$ is often a reasonable solution, we can greatly improve the reconstruction quality by employing additional optimization procedures, as we will describe next.

5.2.2 PRCGAN-L: LATENT OPTIMIZATION

In addition to feeding the magnitude y into the generator network, we can tune the latent variable z to ensure that the generated image has the correct magnitude y, or expressed as an optimization problem, we solve

$$z^* = \arg\min_{z} \|y - |AG_\phi(z, y)|\|_2^2. \tag{10}$$

Given an optimal z^*, the reconstruction is the output of the generator $\hat{x} = G_\phi(z^*, y)$. Note that since the parameters of G_ϕ are fixed, we can solve this optimization in parallel for several magnitudes by passing batches through the generator. We denote this approach as PRCGAN-L.

5.2.3 PRCGAN-W: WEIGHT OPTIMIZATION

Instead of optimizing the latent variable z, we can also search for optimal weights ϕ^* for the generator network G_ϕ to match the given magnitude y for a randomly sampled and fixed z, i.e.,

$$\phi^* = \arg\min_{\phi} \|y - |AG_\phi(z, y)|\|_2^2. \tag{11}$$

A similar idea was used by Hussein et al. (2020) and Ulyanov et al. (2018) for linear inverse problems. Plugging the fine-tuned weights ϕ^*, the fixed latent variable z and the given magnitudes y into the generator G_ϕ yields the reconstructed image $\hat{x} = G_{\phi^*}(z, y)$. Different

(a) E2E

(b) DPR

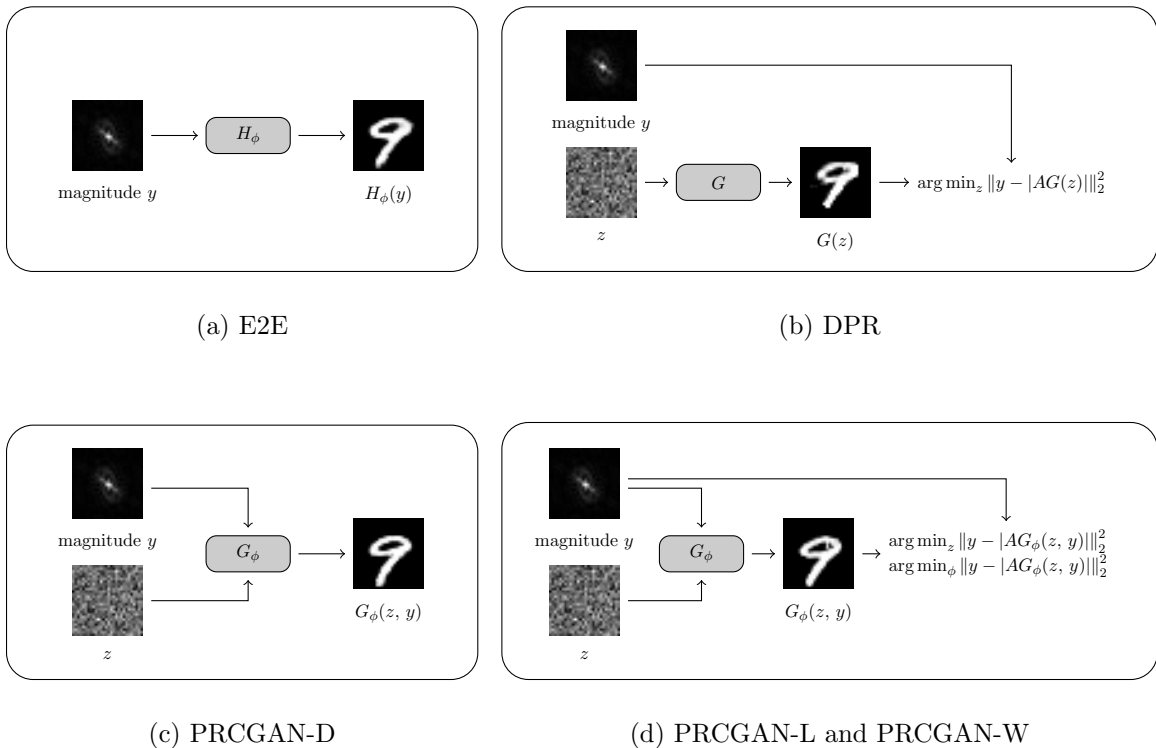(c) PRCGAN-D

(d) PRCGAN-L and PRCGAN-W

Figure 1: Learning-based approaches for phase retrieval: after training the learned-models are used in different ways to reconstruct the image. We propose three variants of our method which differ in the reconstruction process: PRCGAN-D, PRCGAN-L, and PRCGAN-W.

from the latent optimization the weight optimization can only be performed for a single magnitude at a time which makes this approach computationally more expensive. However, we hypothesize that the space of the network weights offers more flexibility to find a good reconstruction. We denote this approach as PRCGAN-W.

5.3 Taxonomy of Learning-based Approaches for Phase Retrieval

The learning-based methods discussed in the previous sections can be divided into two categories:

Unsupervised Learning: Some methods like DPR are only trained on images $x_1, \ldots, x_N$. Thus no retraining is necessary when the measurement process is changed. We call these methods unsupervised.

Supervised Learning: In contrast to unsupervised methods, other methods are not only trained on the images $x_1, \ldots, x_N$ but also have access to the corresponding magnitude measurements $y_1, \ldots, y_N$ during training. This allows the model to specialize in that measurement process and usually results in better image reconstruction quality. We call these methods supervised. In this paper, E2E and all variants of the PRCGAN are examples thereof.
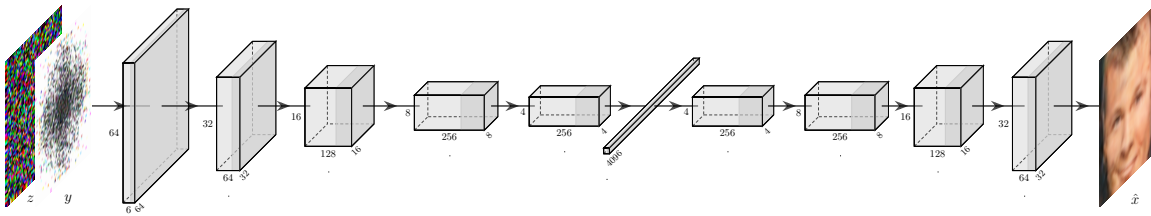
Figure 2: Architecture of the generator used to reconstruct the CelebA images with a resolution of 64×64 from their magnitude.

Note that the term end-to-end learning is in this work exclusively used for methods that reconstruct images in a single forward-pass through the network, having no additional optimization step during reconstruction. Therefore, E2E and PRCGAN-D are considered as end-to-end learning approaches. Figure 1 gives an overview of the different learning-based methods discussed in this work.

6. Experimental Setup

6.1 Datasets

For our experiments, we consider six different datasets. Four of these datasets consist of 28×28 grayscale images, namely, MNIST (LeCun et al., 1998), FMNIST (Xiao et al., 2017), EMNIST (Cohen et al., 2017) and KMNIST (Clanuwat et al., 2018). Although these datasets are considered toy-datasets for classification tasks, solving phase retrieval on these datasets is a non-trivial problem. The other two datasets consist of color-images: the CelebA dataset (Liu et al., 2015) and the well-known CIFAR-10 dataset (Krizhevsky et al., 2009). We rescaled both datasets to 64×64 resolution. Fourier magnitudes were calculated without oversampling, i.e., the magnitudes had the same dimensionality as the images from the dataset.

6.2 Architecture

For the grayscale image datasets (MNIST, FMNIST, KMNIST, and EMNIST) we use a multilayer perceptron (MLP) consisting of five layers each having 2048 hidden units. Empirically, we found out that the fully connected layers are better suited for the global structure of the Fourier phase retrieval problem than convolutional layers, if the image size allow the necessary computation. We use batch normalization (Ioffe and Szegedy, 2015) and ReLU nonlinearities for the intermediate layers and a sigmoid function for the final layer as the pixel intensities of the images are assumed to be normalized between 0 and 1.

Since the MLP architecture is no longer feasible for the increased pixel count of the color-image datasets CIFAR-10 and CelebA, we use a convolutional neural network (CNN) with two fully connected intermediate layers for the generator network. Figure 2 gives an overview of the used architecture.

For the DPR implementation, we follow Hand et al. (2018). That means a variational autoencoder (VAE; Kingma and Welling, 2013) is used for the MNIST-like datasets, where

the dimension of the latent variable z is chosen to be 128 and two ReLU activated fully connected layers with 500 hidden units each are used. For the CelebA and CIFAR-10 datasets a DCGAN architecture (Radford et al., 2015) is used with batch normalization (Ioffe and Szegedy, 2015) and ReLU activation functions.

6.3 Baselines

We also compare our PRCGAN approach with classical methods that do not have a learning component. The most commonly used phase retrieval algorithm is the hybrid-input-output algorithm (HIO) which was proposed by Fienup (1982). The optimal hyperparameters of HIO (1000 iterations and step size $\beta = 0.8$) were determined using a validation dataset, and two random restarts (out of three reconstructions we used the one with the lowest magnitude error). Another optimization-based method that we evaluate is the relaxed-averaged-alternating-reflections algorithm (RAAR) proposed by Luke (2004). Here, we also ran 1000 iterations with two random restarts. We set the step size to $\beta = 0.87$ which was reported to perform best in the original work.

For the learning-based E2E approach, we considered different architectures for the generator and different loss functions. The best performance was obtained with the MSE for the MNIST-like datasets, and with the MAE for the color-image datasets.

For the DPR approach, we tried a non-conditional GAN and a VAE for the generator. For the MNIST-like datasets the VAE performed better, so we only report results for the VAE-based DPR method. For the other datasets, we use a DCGAN (Radford et al., 2015) as the underlying generative model which produced better results. After training each model, we initialized the latent variable with samples from a standard Gaussian distribution and performed $10,000$ iterations of Adam with a step size of 0.1, which we found to perform best. In order to get useful results, we allowed multiple restarts and used the reconstruction with the lower magnitude error.

6.4 Training and Optimization

We determined all hyperparameters by using a separate validation dataset. We trained all previously mentioned learning-based models with a batch size of 32 for the MNIST-like datasets and 64 for the color-images, using the Adam optimizer (Kingma and Ba, 2014). We trained the PRCGAN for 100 epochs for all datasets except for CIFAR-10, where we increased the number of epochs to 250. We set $\lambda = 100$ for MNIST, EMNIST, and KMNIST and used $\lambda = 1000$ for FMNIST, CelebA, and CIFAR-10.

Analogous to the DPR approach, we optimized the latent variable z using 10,000 steps with a learning rate of 0.1. We observed that even without restarting, our approach outperformed DPR, so we decided to eliminate the random restarts to keep the computational effort limited. In the weight optimization of PRCGAN-W, we also used 10,000 steps but with a decreased learning rate of 10^{-6}.

6.5 Evaluation

For grayscale datasets MNIST, FMNIST, EMNIST and KMNIST we compare the mean squared error (MSE), the mean absolute error (MAE) and the structural similarity index

measure (SSIM; Wang et al., 2004). For datasets consisting of color-images (CelebA and CIFAR-10) we report the MSE, the LPIPS (Zhang et al., 2018) and the SSIM. For the grayscale images with black background, we observe that the reconstructions are sometimes flipped and translated after performing the latent optimization. This is to be expected as flipping and translating the image does not change the magnitudes in the case of Fourier phase retrieval. Therefore, we tread these reconstructions to be equally correct and perform image registration with the target image before calculating the errors. We use cross-correlation (Guizar-Sicairos et al., 2008) to align the reconstruction and the flipped reconstruction with the target image and report the better metric. For the color images from CelebA and CIFAR-10 datasets, we did not observe any flips or shifts and therefore omitted the registration step. We use 1024 images in each test set to limit the computational time.

7. Experiments

7.1 Results for Fourier Phase Retrieval

In the following, we discuss the results of the optimization-based and learning-based methods for Fourier phase retrieval.

HIO. On the MNIST-like datasets, the HIO algorithm reconstructs images but creates many artifacts. For the other datasets, HIO is not successful at all, resulting in fragmented, blurry reconstructions. Furthermore, for the color-images of the CIFAR-10 and the CelebA datasets, HIO does not produce any useful results.

RAAR. Overall, we observe similar results as when using HIO, sometimes even slightly worse. For the color-image datasets RAAR also does not succeed at reconstructing the images.

End-to-End. In comparison to the optimization-based methods, the E2E approach does not produce fragmented parts. Although some reconstructions are still blurry, E2E performs better in all six datasets than HIO and RAAR.

DPR. On MNIST, the latent optimization approach produces better visual appearance of the digits. To avoid local optima, we did multiple random restarts as detailed in Section 6.3. However, sometimes DPR still does not work as well as E2E, as can be seen in the first and second image shown in Figure 4. While DPR got slightly better MNIST reconstructions than the E2E approach, it is having difficulties with the other datasets.

PRCGAN. The PRCGAN combines end-to-end learning with a subsequent optimization, so we are expecting better results than the E2E and the DPR approach. The most basic PRCGAN-D achieves similar performance to the E2E method, as one can see in Table 1 and 2. However, due to the adversarial loss, the PRCGAN alleviates the problem of blurriness and gets much more realistic reconstructions. As one can see in the second and eighth image in Figure 4 (columns 2 and 8) the reconstructions show finer texture components like the text and the checkered pattern on the shirts, respectively. Quantitatively, the blurry E2E reconstructions are better than the reconstructions of PRCGAN-D, since the MSE, MAE, and SSIM punish misplaced sharp edges more than blurriness.
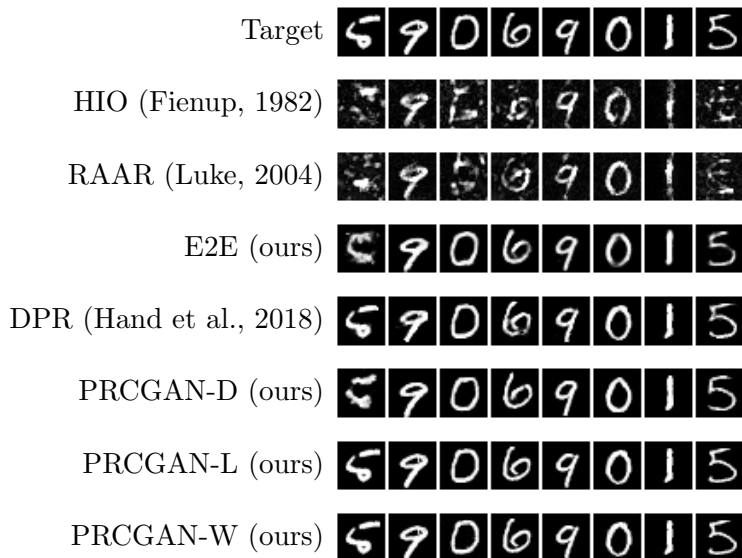
10

Figure 3: Registered reconstructions from the Fourier magnitudes of samples from the MNIST test dataset for each model.
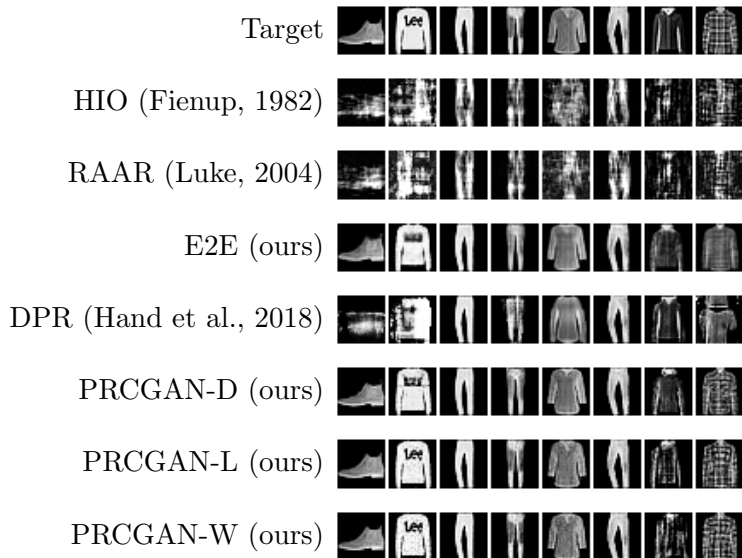

Figure 4: Registered reconstructions from the Fourier magnitudes of samples from the FMNIST test dataset for each model.
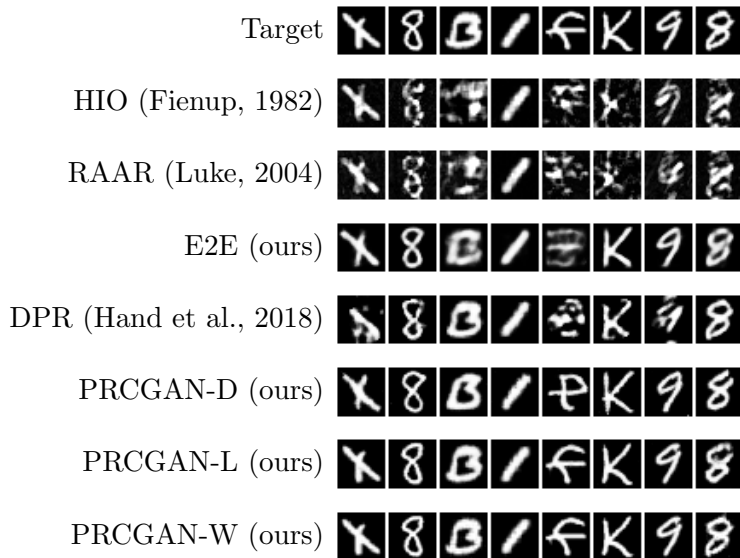
Figure 5: Registered reconstructions from the Fourier magnitudes of samples from the EMNIST test dataset for each model.

The variations PRCGAN-L and PRCGAN-W, which optimize latent variables or weights with respect to the magnitude, produce the best reconstructions regarding both qualitative and quantitative performance on all datasets except for CIFAR-10. It is remarkable that on FMNIST we were even able to reconstruct the text shown on the second image in Figure 4, where all other baseline methods failed. While our presented optimization approaches do not significantly differ in quality, we observe that PRCGAN-L with a few minor exceptions achieves the best reconstructions. One exception are the results of PRCGAN-W, which achieves a lower MAE for MNIST and EMNIST. CIFAR-10 shows the limits of learning-based methods for phase-retrieval: due to the high variation of the dataset, none of the methods recovers good images. Quantitatively, E2E is slightly better than the other approaches. E2E produced blurry images, while our PRCGAN-D is trying to create images with edges.

7.2 Computational Runtime

Reconstructing a single image of shape $3\times64\times64$ using RAAR or HIO takes approximately 3.70 seconds on an AMD EPYC 7742 CPU (including two restarts). The end-to-end approaches E2E and PRCGAN-D, which reconstruct the images in a single forward pass, are the fastest methods discussed in this work. They take 0.16 and 0.02 seconds, respectively. The latent optimization of the DPR approach takes 137.24 seconds, as it used two random restarts. In contrast to that PRCGAN-L takes 85.76 seconds but does not use any restarts. The weight optimization approach (PRCGAN-W) runs for 138.89 seconds to reconstruct a single image. In contrast to DPR and PRCGAN-L, which can process a dataset of images batchwise, PRCGAN-W needs to process each image separately. Thus it scales linearly with the number of images and not with the number of batches as is the case for DPR
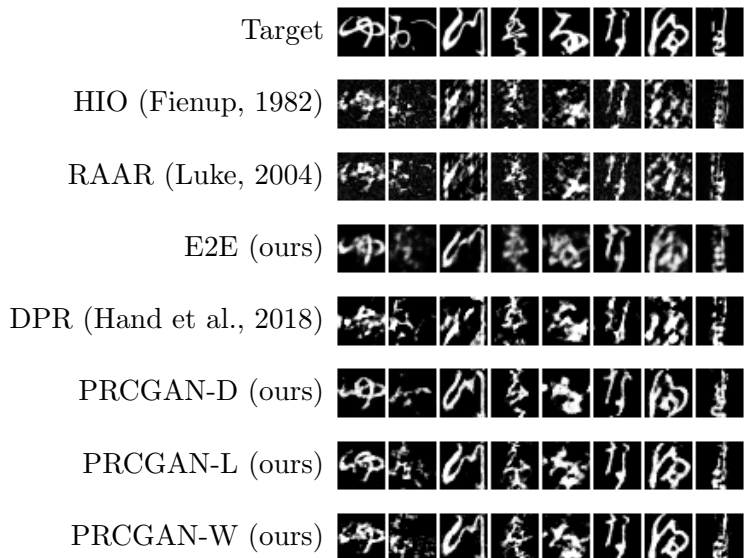
Target

HIO (Fienup, 1982)

RAAR (Luke, 2004)

E2E (ours)

DPR (Hand et al., 2018)

PRCGAN-D (ours)

PRCGAN-L (ours)

PRCGAN-W (ours)

Figure 6: Registered reconstructions from the Fourier magnitudes of samples from the KMNIST test dataset for each model.

Target

HIO (Fienup, 1982)

RAAR (Luke, 2004)

E2E (ours)

DPR (Hand et al., 2018)

PRCGAN-D (ours)
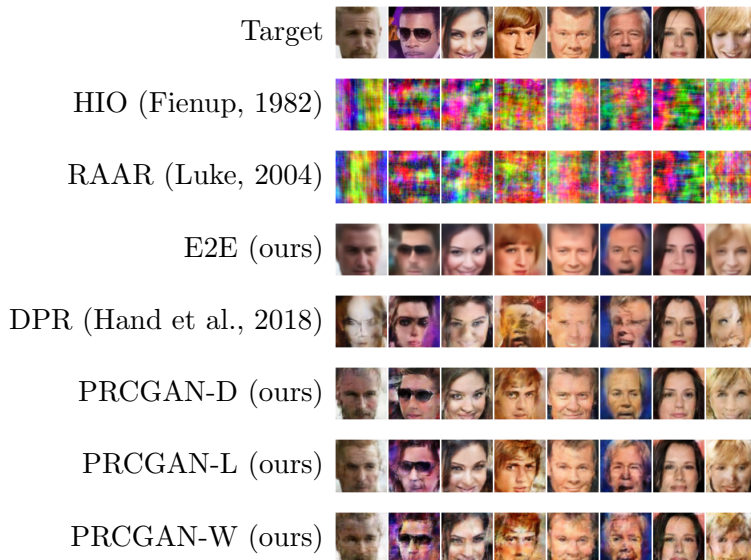
PRCGAN-L (ours)

PRCGAN-W (ours)

Figure 7: Registered reconstructions from the Fourier magnitudes of samples from the CelebA test dataset for each model.

13

Table 1: Quantitative evaluation for MNIST, FMNIST, EMNIST and KMNIST for the registered reconstructions from the Fourier magnitudes. MSE, MAE: lower is better. SSIM: higher is better.

	MNIST			FMNIST		
Method	MSE	MAE	SSIM	MSE	MAE	SSIM
HIO (Fienup, 1982)	0.0440	0.1016	0.5274	0.0649	0.1608	0.4019
RAAR (Luke, 2004)	0.0489	0.1150	0.4879	0.0668	0.1673	0.3491
E2E (ours)	0.0164	0.0429	0.8191	0.0129	0.0564	0.7400
DPR (Hand et al., 2018)	0.0139	0.0302	0.8685	0.0288	0.0855	0.6093
PRCGAN-D (ours)	0.0185	0.0415	0.8196	0.0149	0.0569	0.7414
PRCGAN-L (ours)	**0.0009**	0.0045	**0.9890**	**0.0084**	**0.0403**	**0.8376**
PRCGAN-W (ours)	0.0010	**0.0034**	0.9867	0.0090	0.0429	0.8266
	EMNIST			KMNIST		
Method	MSE	MAE	SSIM	MSE	MAE	SSIM
HIO (Fienup, 1982)	0.0645	0.1364	0.4942	0.0920	0.1721	0.3592
RAAR (Luke, 2004)	0.0668	0.1425	0.4767	0.0931	0.1774	0.3773
E2E (ours)	0.0223	0.0656	0.7598	0.0538	0.1193	0.5510
DPR (Hand et al., 2018)	0.0326	0.0685	0.7480	0.0884	0.1446	0.4504
PRCGAN-D (ours)	0.0318	0.0692	0.7546	0.0694	0.1192	0.5535
PRCGAN-L (ours)	0.0066	0.0255	0.9416	**0.0393**	**0.0775**	**0.7384**
PRCGAN-W (ours)	**0.0063**	**0.0218**	**0.9436**	0.0402	0.0787	0.7259

Table 2: Quantitative evaluation for CelebA and CIFAR-10 for the reconstructions from the Fourier magnitudes. Note that we do not register the reconstructions for these datasets. MSE, LPIPS: lower is better. SSIM: higher is better.

|	CelebA			CIFAR-10		
Method	MSE	SSIM	LPIPS	MSE	SSIM	LPIPS
HIO (Fienup, 1982)	0.1005	0.0510	0.8228	0.0814	0.0881	0.7827
RAAR (Luke, 2004)	0.1011	0.0537	0.8183	0.0808	0.0927	0.7779
E2E (ours)	0.0123	0.6367	0.2683	**0.0390**	**0.2735**	0.5852
DPR (Hand et al., 2018)	0.0388	0.4185	0.3529	0.0707	0.1713	0.5819
PRCGAN-D (ours)	0.0155	0.5653	0.2655	0.0402	0.2297	0.5403
PRCGAN-L (ours)	**0.0093**	**0.6846**	**0.2182**	0.0489	0.2219	**0.5401**
PRCGAN-W (ours)	0.0115	0.6405	0.2835	0.0492	0.2173	0.5487
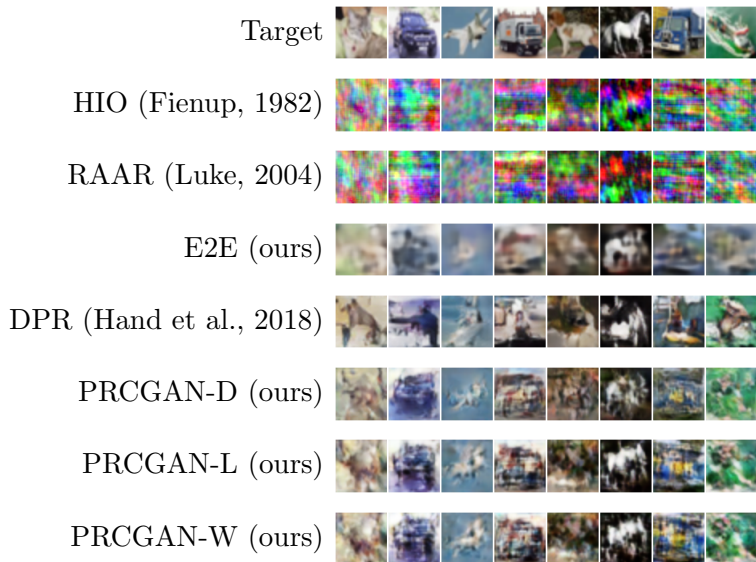
Figure 8: Registered reconstructions from the Fourier magnitudes of samples from the CIFAR-10 test dataset for each model.

and PRCGAN-L. The runtimes of the learning-based methods are measured on an NVIDIA A100 GPU. In conclusion, PRCGAN-L gives excellent reconstructions in most cases and has reasonable runtime.

7.3 Dissecting the PRCGAN: Ablation Experiments

In this section, we explain what the PRCGAN has learned by presenting the results of different ablation experiments.

7.3.1 Is the magnitude passed to the CGAN being used for the reconstruction?

In the basic PRCGAN-D approach, the magnitudes are processed by the CGAN to recover the image. However, for the other PRCGAN variants we are also employing the magnitudes in the subsequent optimization. To evaluate the influence of these two roles of the magnitude, we run experiments where we used wrong magnitudes as inputs to the model on purpose while afterwards solving the optimization problem stated in Equation 10 with the correct magnitudes. Table 3 shows that the performance completely drops, which shows that for PRCGAN-L, the magnitude input to the generator is essential.

7.3.2 Is it sufficient to condition on the label?

Instead of feeding magnitudes to the CGAN, one could argue that just the labels should be sufficient for successful reconstruction. To answer this question we train a CGAN conditioned on the labels and attempt to reconstruct the image given the correct label information. While the label might not be available in practice, this experiment helps us to

Table 3: Quantitative evaluation of the different ablation experiments. MSE, MAE: lower is better. SSIM: higher is better.

	MNIST		
Method	MSE	MAE	SSIM
PRCGAN-L	**0.0009**	**0.0045**	**0.9890**
PRCGAN-L (conditioned on labels)	0.0160	0.0361	0.8442
PRCGAN-L (trained on wrong magnitudes)	0.0244	0.0455	0.7801
PRCGAN-L (trained w/o adversarial loss)	0.0289	0.0523	0.7363

Table 4: Comparison of reconstruction performance for different choices of reconstruction loss functions $\mathcal{L}_{\mathrm{rec}}$. MSE, LPIPS: lower is better. SSIM: higher is better.

		CelebA			CIFAR-10		
Method	$\mathcal{L}_{\mathrm{rec}}$	MSE	SSIM	LPIPS	MSE	SSIM	LPIPS
PRCGAN-D	MAE	0.0155	0.5653	0.2655	0.0402	0.2297	**0.5403**
PRCGAN-D	MSE	0.0149	0.5535	0.3028	0.0422	**0.2432**	0.5878
PRCGAN-D	LPIPS	**0.0143**	**0.5856**	**0.2418**	**0.0385**	0.2390	0.5409
PRCGAN-L	MAE	**0.0093**	**0.6846**	0.2182	0.0489	0.2219	0.5401
PRCGAN-L	MSE	0.0125	0.6110	0.2811	0.0534	0.2133	0.5692
PRCGAN-L	LPIPS	0.0096	0.6804	**0.2126**	**0.0479**	**0.2279**	**0.5359**

understand what information is relevant. As in the previous study, the results of only using label information are inferior to using the magnitude (see Table 3).

7.3.3 CAN WE DROP THE ADVERSARIAL LOSS?

Next, we train a PRCGAN by only minimizing the reconstruction loss $\mathcal{L}_{\mathrm{rec}}$. Note that this approach is not identical to E2E since it has a latent noise variable and a subsequent optimization of it. Again, the performance worsens (see Table 3).

7.3.4 WHICH IMPACT DOES THE CHOICE OF THE RECONSTRUCTION LOSS HAVE?

We are interested whether the choice of the reconstruction loss function $\mathcal{L}_{\mathrm{rec}}$ impacts the performance of the PRCGAN-D and PRCGAN-L model. In addition to the MAE and the MSE we also consider LPIPS, which is a perceptual loss function. Results are shown in Table 4. Overall, the performance of the PRCGAN-D is improved by using LPIPS as reconstruction loss, however when considering PRCGAN-L the effect decreases. This justifies our initial choice of the reconstruction loss.

Table 5: Ablation study for the intermediate fully connected layer used in the neural network architecture for the CelebA and CIFAR-10 dataset. MSE, LPIPS: lower is better. SSIM: higher is better.

| | CelebA | | | | | |
| | with fully connected layer | | | without fully connected layer | | |
Method	MSE	SSIM	LPIPS	MSE	SSIM	LPIPS
PRCGAN-D	0.0155	0.2655	0.5653	0.0260	0.4902	0.2903
PRCGAN-L	0.0093	0.2182	0.6846	0.0153	0.5899	0.2697
PRCGAN-W	0.0115	0.2835	0.6405	0.0156	0.5834	0.3021

7.3.5 CAN WE DROP THE FULLY CONNECTED INTERMEDIATE LAYER?

For the color images, the PRCGAN consists of several convolutional layers, two intermediate fully connected layers and several transposed convolutional layers (see Figure 2). The motivation for the fully connected layers is that it helps to model the global structure of the phase retrieval problem. To measure its influence, we train the same model without the intermediate fully connected layers. The results are shown in Table 5 and confirm our design choice to include a fully connected intermediate layer. Also, the model without it was more difficult to train due to numerical instabilities.

7.4 Robustness Against Distributional Shifts

In real-world applications, we often do not have examples from the true image distribution. Instead, we might be able to train a model on a dataset that is only similar to some degree. To simulate this situation, we use models trained on the MNIST digits and compare to what extent we can reconstruct letters from the EMNIST dataset. Since the MNIST dataset is a subset of the EMNIST dataset, we ensure that only images showing letters are used for evaluation. Table 6 shows that all three variants of the PRCGAN are the best approaches for this setup. For several exemplary reconstructions refer to Figure 10. To further confirm these generalization abilities, we expanded our experiments on the four grayscale datasets. We evaluate E2E, DPR, PRCGAN-D, PRCGAN-L, and PRCGAN-W trained on MNIST, FMNIST, EMNIST and KMNIST images, respectively, with each of the other datasets. Figure 9 summarizes the reconstruction performance of the learned methods on out-of-distribution data. Note that PRCGAN-L and PRCGAN-W perform best in eight out of twelve cases (only considering those where training and testing domains differ).

Carrying on the idea of reconstructing arbitrary shapes we have created a small dataset consisting of 32 MNIST-like symbols and benchmark the different methods on this dataset. Here, we consider models trained on MNIST and EMNIST, and evaluate both using Fourier measurements. Refer to Table 7 and Figure 11. Note that although the results are generally better for the models trained on EMNIST, the PRCGAN variants always perform better than E2E and DPR.

Table 6: Training on MNIST digits leads to reasonable results on letters of EMNIST. Quantitative evaluation for the registered reconstructions from the Fourier magnitudes. MSE, MAE: lower is better. SSIM: higher is better.

	EMNIST (only letters)		
Method (trained on MNIST)	MSE	MAE	SSIM
E2E (ours)	0.0535	0.1026	0.5486
DPR (Hand et al., 2018)	0.0427	0.0814	0.6670
PRCGAN-D (ours)	0.0567	0.1025	0.5710
PRCGAN-L (ours)	**0.0269**	**0.0604**	**0.7751**
PRCGAN-W (ours)	**0.0269**	0.0636	0.7492

Table 7: Evaluation of 32 MNIST-like symbols for the registered reconstructions from the Fourier magnitudes, with the specified methods trained on MNIST and EMNIST, respectively. MSE, MAE: lower is better. SSIM: higher is better.

	trained on MNIST			trained on EMNIST		
Method	MSE	MAE	SSIM	MSE	MAE	SSIM
E2E (ours)	0.0916	0.1583	0.3974	0.0653	0.1413	0.4812
DPR (Hand et al., 2018)	0.0712	0.1192	0.6221	0.0499	0.0940	0.7280
PRCGAN-D (ours)	0.1095	0.1610	0.3770	0.0967	0.1522	0.4422
PRCGAN-L (ours)	0.0498	0.0942	0.7337	0.0452	0.0902	0.7282
PRCGAN-W (ours)	**0.0484**	**0.0930**	**0.7356**	**0.0347**	**0.0751**	**0.7910**

(a) Trained on MNIST

(b) Trained on FMNIST
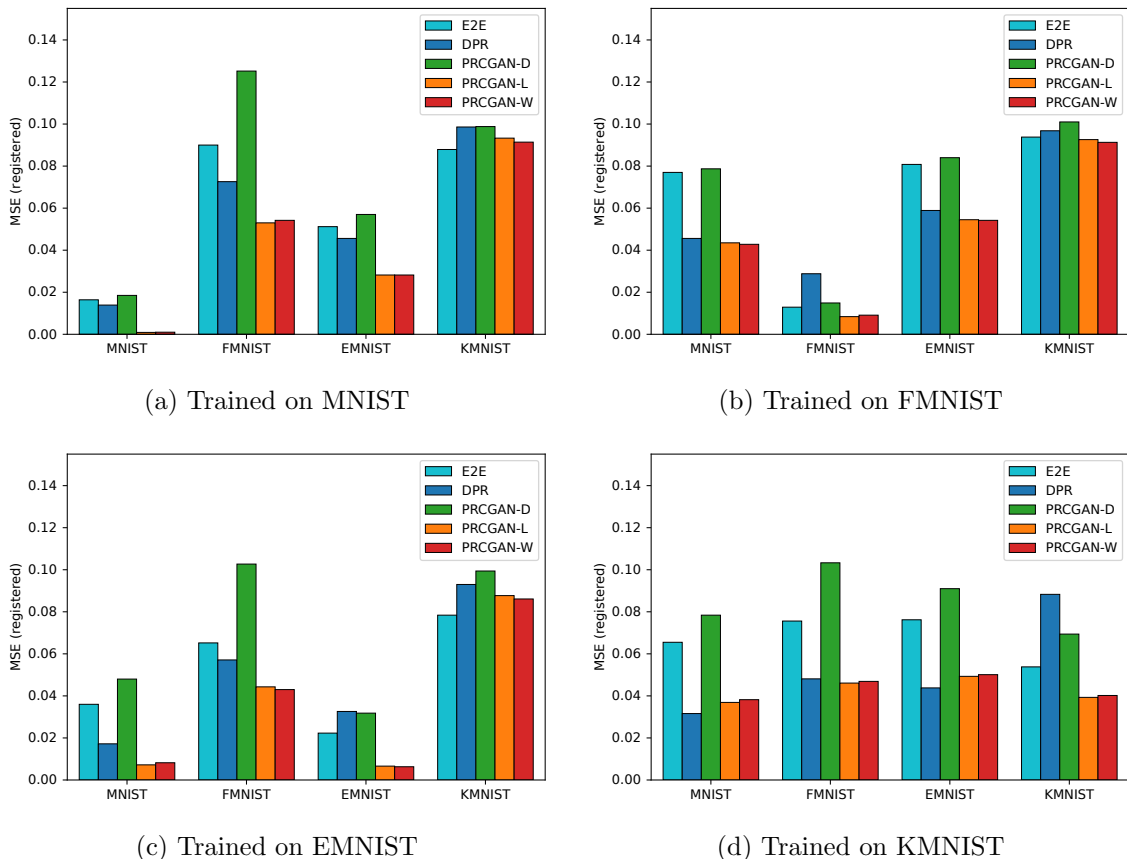
(c) Trained on EMNIST

(d) Trained on KMNIST

Figure 9: Generalization to out-of-distribution data: MSE of registered reconstructions from noisy Fourier magnitudes from the grayscale datasets.

7.5 Robustness Against Noise

All learning-based phase retrieval approaches discussed in this work were trained on synthetic noiseless measurements. However, in practice Fourier magnitude measurements often exhibit different kinds of noise which may influence the reconstruction process. In this section, we study the robustness of these models to different types of measurement noise. First, we take a look at Poisson noise, which is the type of noise usually present in many phase retrieval applications (Yeh et al., 2015). The noisy measurements are then given as

$$\tilde{y} = \alpha\sqrt{s}, \text{ with } s \sim \text{Poisson}\left(y^2/\alpha^2\right), \tag{12}$$

where the parameter α controls the amount of noise (larger values for α correspond to stronger noise). A similar measurement process has been used by Metzler et al. (2018). Second, we consider additive white Gaussian noise, i.e., the measurements are given as

$$\tilde{y} = y + \alpha s, \text{ with } s \sim \mathcal{N}\left(0, 1\right), \tag{13}$$

where, again, α controls the amount of noise.

Target

E2E (ours)

DPR (Hand et al., 2018)

PRCGAN-D (ours)

PRCGAN-L (ours)

PRCGAN-W (ours)

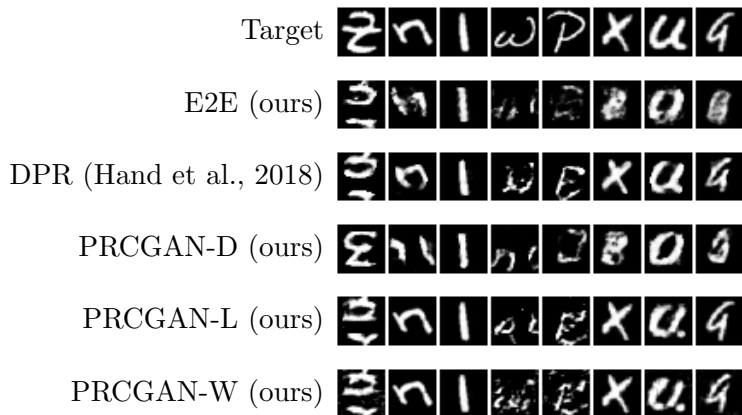Figure 10: Generalization to out-of-distribution data: registered reconstructions of the letters from EMINST. All models were trained on MNIST.

Target

E2E (ours)

DPR (Hand et al., 2018)
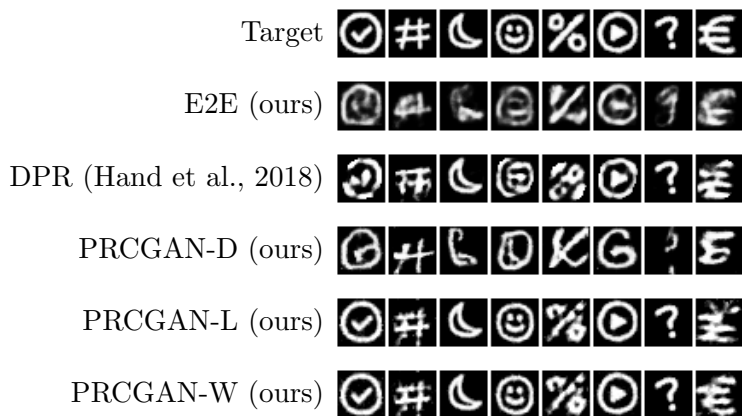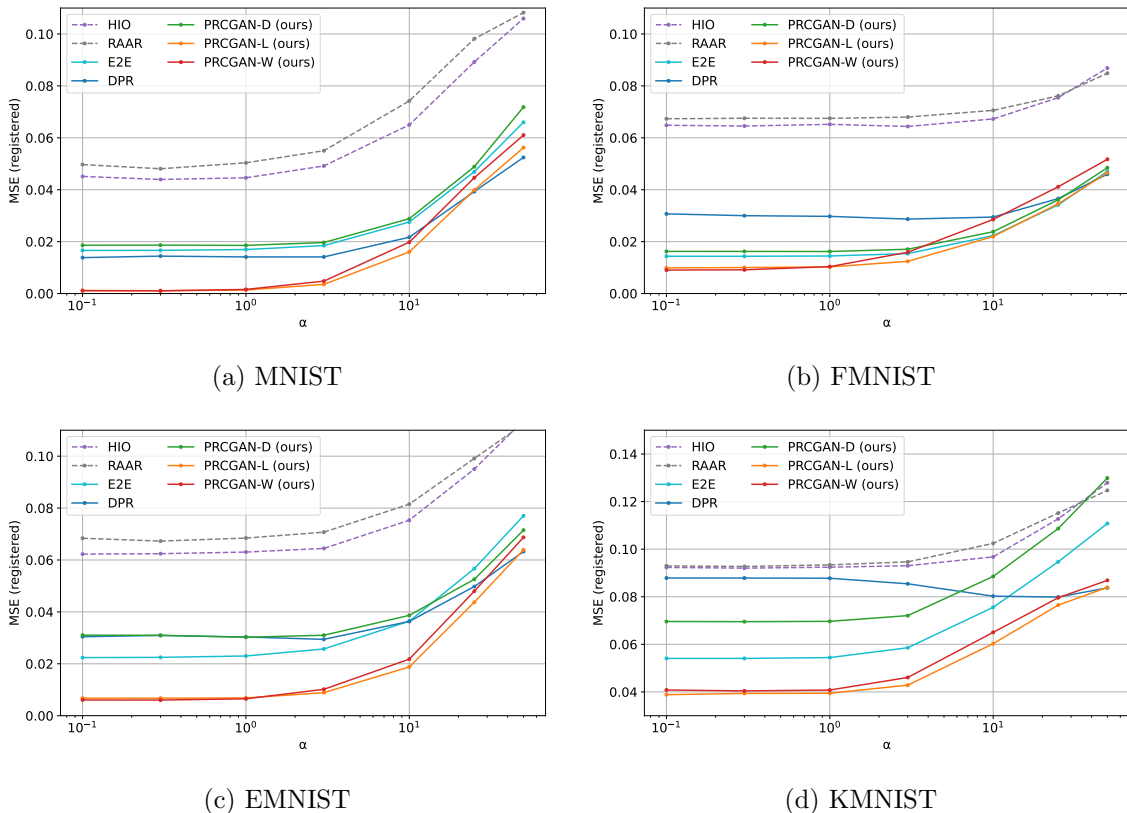
PRCGAN-D (ours)

PRCGAN-L (ours)

PRCGAN-W (ours)

Figure 11: Generalization to out-of-distribution data: registered reconstructions of images from our MNIST-like symbols dataset. All models were trained on EMNIST.

Figure 12: Noise robustness: MSE of registered reconstructions from Fourier magnitudes perturbated with Poisson noise. Larger α implies stronger noise.

Figure 12 compares the performance of all methods regarding the robustness against noise in the measurements used for the reconstruction. More precisely, we plot the MSE of 1024 reconstructions against different noise levels, i.e., values of α on MNIST, FMNIST, EMNIST and KMNIST. On MNIST, the proposed methods are robust to noise up to $\alpha = 3$. On FMNIST, however, PRCGAN-W is less robust compared to PRCGAN-L. While for small amounts of noise our approaches give the best results, for larger amounts of noise, DPR is equally good or even slightly better. This might be due to the fact, that DPR uses random restarts, which might have helped coping with local optima in the large noise regime. Also, PRCGAN-L and PRCGAN-W perform best on EMNIST and KMNIST.

7.6 Robustness Against Randomly Missing Measurements

Furthermore, we analyze how the different methods are impacted by missing measurements. In this experiment, we consider measurements that are given as

$$y = b \odot |\mathcal{F}x|, \tag{14}$$

where b is a binary mask which exhibits the same symmetries as the magnitude measurement $|\mathcal{F}x|$ and $\odot$ denotes elementwise multiplication. By multiplying the measurements
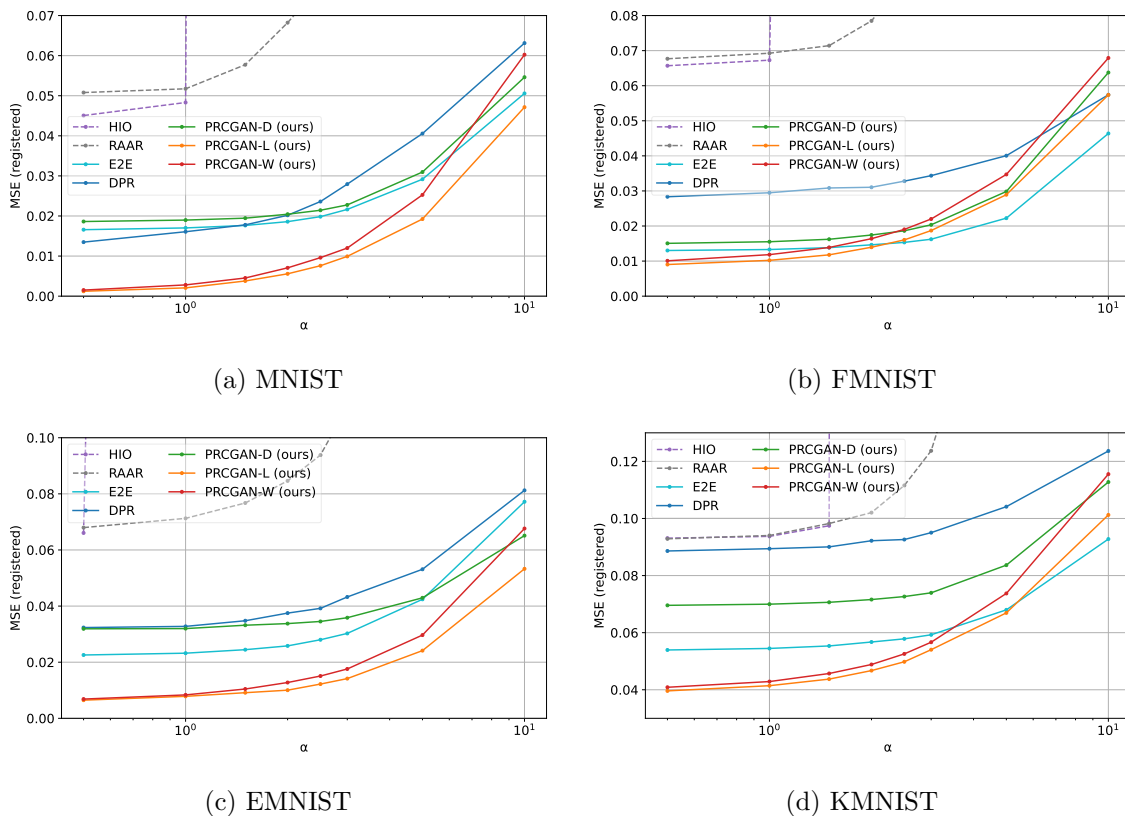
(a) MNIST

(b) FMNIST

(c) EMNIST

(d) KMNIST

Figure 13: Noise robustness: MSE of registered reconstructions from Fourier magnitudes perturbed with additive white Gaussian noise. Larger α implies stronger noise.
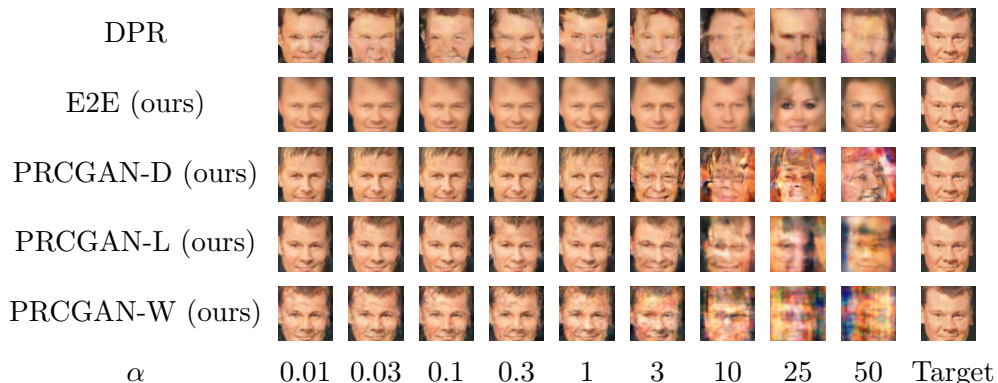

Figure 14: Noise robustness: reconstruction of an image from the CelebA dataset for different levels of Poisson noise.
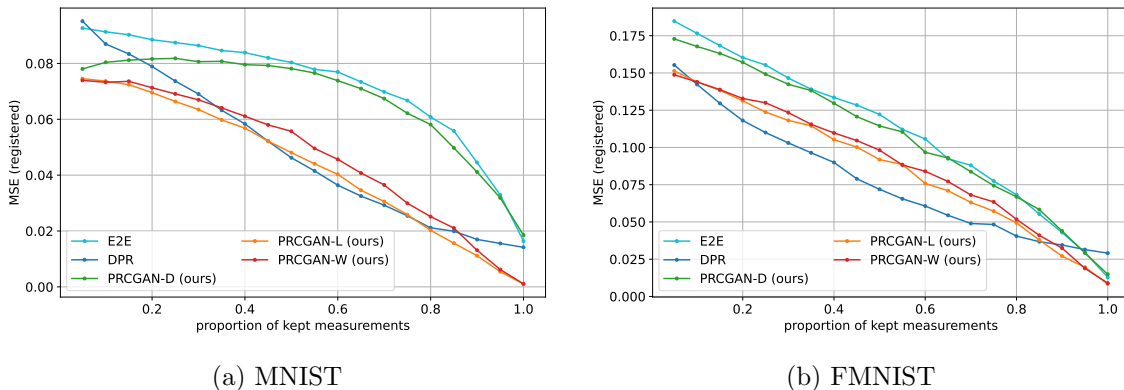
(a) MNIST

(b) FMNIST

Figure 15: Reconstructions from partial Fourier magnitudes: MSE of registered reconstructions from partial Fourier magnitudes.
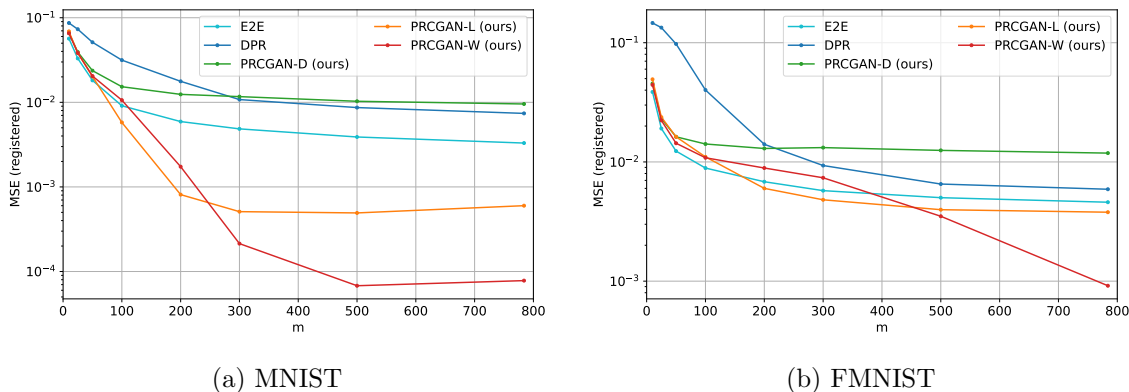

(a) MNIST

(b) FMNIST

Figure 16: Reconstruction from Gaussian measurements: comparison of the PRCGAN, the E2E and generative prior approaches for different numbers of measurements m.

with the binary mask b, we randomly set some of the entries to zero to simulate missing measurements.

Both unsupervised and supervised methods use the magnitudes for the reconstruction error, where the terms of the missing entries are omitted, i.e., images were reconstructed by minimizing the loss $\|y - b \odot |\mathcal{F}G(z)|\|_2^2$ for DPR and $\|y - b \odot |\mathcal{F}G_\phi(z, y)|\|_2^2$ for our method PRCGAN-L. We used the models that were trained on complete magnitude measurements, i.e., we did not retrain the models.

Figure 15 shows the MSE on the MNIST and the FMNIST dataset. Interestingly, methods that get the measurements as input perform worse than the DPR approach. The supervised methods are further affected by missing magnitudes since they also use them as input for the network. Further training on masked Fourier magnitudes could help mitigating this effect.
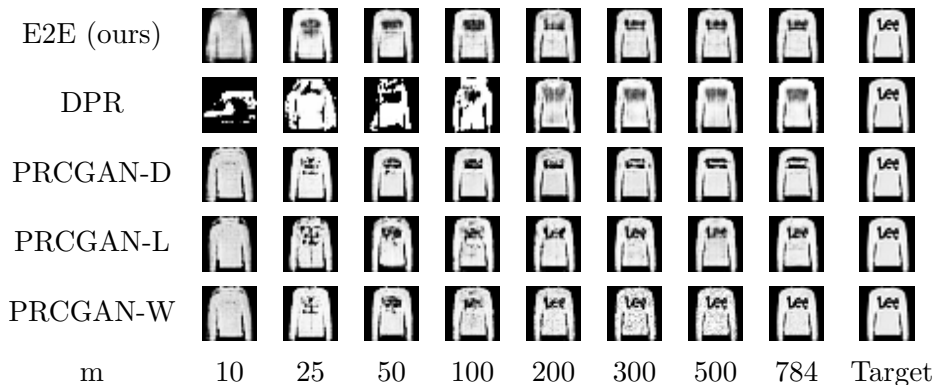
Figure 17: Reconstructions from Gaussian measurements of images from FMNIST for a varying number of measurements m.

7.7 Results for Compressive Phase Retrieval

Besides the Fourier phase retrieval problem, a commonly studied problem is compressive Gaussian phase retrieval. For a vectorized image $x \in \mathbb{R}^n$ the measurement matrix $M \in \mathbb{R}^{m \times n}$ is elementwise sampled from a Gaussian distribution $\mathcal{N}(0, 1/m)$ for different numbers of measurements m. By increasing the number of measurements m we can smoothly adjust the difficulty of the problem. For a comprehensive evaluation, we consider eight different values for m. We start with a small number of 10 measurements and increase it up to the dimension of the flattened image ($m = 784$, number of pixels in MNIST). We sample a measurement matrix M for each m and we keep it fixed to train the different approaches. Analogous to the Fourier phase retrieval, we take the same model architectures as described in Section 6.2. In Figure 16 we plot the MSE of 1024 reconstructions against different values of m. For the PRCGAN variants and the E2E approach, we retrain the model for each value of m. Note that the training of the underlying VAE for DPR is independent of the chosen measurement matrix M, so we can optimize the latent space of the same VAE that we also use in the Fourier phase retrieval experiments. However, for MNIST we observe worse results using the latent dimension of 128, so in this case we keep the latent dimension of 20 from the original work (Hand et al., 2018). As expected, reducing the number of measurements results in higher errors for all methods as one can see from the plots in Figure 16. Note that DPR has the worst performance and is strongly influenced by the number of measurements on both datasets. Lastly, we show an example for the reconstruction performance of the proposed methods on FMNIST in Figure 17. DPR completely fails to reconstruct the letters and E2E only achieves useful results for the maximum number of measurements. In contrast to that, PRCGAN-L and PRCGAN-W successfully reconstruct the images even for a small number of measurements.

8. Discussion and Conclusion

In this work, we propose the PRCGAN for solving Fourier phase retrieval and compressive Gaussian phase retrieval problem. Our method can be seen as an end-to-end learning approach augmented with an additional optimization procedure combining the best of both

worlds. On the one hand, the learning component allows our method to reconstruct images for particularly difficult instances of the phase retrieval problem, and on the other hand, the subsequent latent or weight optimization produces high quality reconstructions. Our ablation study shows that each of the components of our model is necessary to achieve this. Furthermore, we show that our method is robust to Poisson noise and additive Gaussian noise and (at least to some extent) generalizes better to out-of-distribution data than other methods.

References

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

Oliver Bunk, Ana Diaz, Franz Pfeiffer, Christian David, Bernd Schmitt, Dillip K Satapathy, and J Friso Van Der Veen. Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels. *Acta Crystallographica Section A: Foundations of Crystallography*, 63(4):306–314, 2007.

Eunju Cha, Chanseok Lee, Mooseok Jang, and Jong Chul Ye. Deepphasecut: Deep relaxation in phase for unsupervised fourier phase retrieval. *arXiv preprint arXiv:2011.10475*, 2020.

Tarin Clanuwat, Mikel Bober-Irizar, Asanobu Kitamoto, Alex Lamb, Kazuaki Yamamoto, and David Ha. Deep learning for classical japanese literature. *arXiv preprint arXiv:1812.01718*, 2018.

Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. EMNIST: Extending MNIST to handwritten letters. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2921–2926. IEEE, 2017.

James R Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, 1982.

James R Fienup and Chris Dainty. Phase retrieval and image reconstruction for astronomy. *Image Recovery: Theory and Application*, 231:275, 1987.

Ralph W Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

Manuel Guizar-Sicairos, Samuel T Thurman, and James R Fienup. Efficient subpixel image registration algorithms. *Optics Letters*, 33(2):156–158, 2008.

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

Shady Abu Hussein, Tom Tirer, and Raja Giryes. Image-adaptive gan based reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3121–3129, 2020.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456. PMLR, 2015.

Gauri Jagatap and Chinmay Hegde. Phase retrieval using untrained neural network priors. In *NeurIPS 2019 Workshop on Deep Learning and Inverse Problems*, 2019.

Vinayak Killedar and Chandra Sekhar Seelamantula. Compressive phase retrieval based on sparse latent generative priors. In *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1596–1600. IEEE, 2022.

Vinayak Killedar, Praveen Kumar Pokala, and Chandra Sekhar Seelamantula. Learning generative prior with latent space sparsity constraints. *arXiv preprint arXiv:2105.11956*, 2021.

Kyung-Su Kim, Jung Hyun Lee, and Eunho Yang. Compressed sensing via measurement-conditional generative models. *IEEE Access*, 2021.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Alex Krizhevsky et al. Learning multiple layers of features from tiny images. Citeseer, 2009.

Abhinau Kumar, Shashank Gupta, and Sumohana S. Channappayya. Perceptually driven conditional GAN for fourier ptychography. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 1267–1271. IEEE, 2019.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

D Russell Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse problems*, 21(1):37, 2004.

Raunak Manekar, Zhong Zhuang, Kshitij Tayal, Vipin Kumar, and Ju Sun. Deep learning initialized phase retrieval. In *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*, 2020.

Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, and Richard Baraniuk. prdeep: Robust phase retrieval with a flexible deep network. In *International Conference on Machine Learning*, pages 3501–3510. PMLR, 2018.

Rick P Millane. Phase retrieval in crystallography and optics. *JOSA A*, 7(3):394–411, 1990.

Yohei Nishizaki, Ryoichi Horisaki, Katsuhisa Kitaguchi, Mamoru Saito, and Jun Tanida. Analysis of non-iterative phase retrieval based on machine learning. *Optical Review*, 27 (1):136–141, 2020.

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

Tobias Uelwer, Tobias Hoffmann, and Stefan Harmeling. Non-iterative phase retrieval with cascaded neural networks. In Igor Farkaš, Paolo Masulli, Sebastian Otte, and Stefan Wermter, editors, *Artificial Neural Networks and Machine Learning – ICANN 2021*, pages 295–306, Cham, 2021a. Springer International Publishing.

Tobias Uelwer, Alexander Oberstraß, and Stefan Harmeling. Phase retrieval using conditional generative adversarial networks. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 731–738. IEEE, 2021b.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

Irene Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1):47–81, 2015.

Adriaan Walther. The question of phase retrieval in optics. *Optica Acta: International Journal of Optics*, 10(1):41–49, 1963.

Yaotian Wang, Xiaohang Sun, and Jason Fleischer. When deep denoising meets iterative phase retrieval. In *International Conference on Machine Learning*, pages 10007–10017. PMLR, 2020.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

Zihui Wu, Yu Sun, Jiaming Liu, and Ulugbek Kamilov. Online regularization by denoising with applications to phase retrieval. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Rui Xu, Mahdi Soltanolkotabi, Justin P Haldar, Walter Unglaub, Joshua Zusman, Anthony FJ Levi, and Richard M Leahy. Accelerated wirtinger flow: A fast algorithm for ptychography. *arXiv preprint arXiv:1806.05546*, 2018.

Li-Hao Yeh, Jonathan Dong, Jingshan Zhong, Lei Tian, Michael Chen, Gongguo Tang, Mahdi Soltanolkotabi, and Laura Waller. Experimental robustness of fourier ptychography phase retrieval algorithms. *Optics Express*, 23(26):33214–33240, 2015.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

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