

---

# End-to-End Learning for Phase Retrieval

---

Raunak Manekar<sup>1</sup> Kshitij Tayal<sup>1</sup> Vipin Kumar<sup>1</sup> Ju Sun<sup>1</sup>

## Abstract

We consider the end-to-end deep learning approach for phase retrieval, a central problem in scientific imaging. We highlight a fundamental difficulty for learning that previous work has neglected, likely due to the biased datasets they use for training and evaluation. We propose a simple yet different formulation for PR that seems to overcome the difficulty and return consistently better qualitative results.

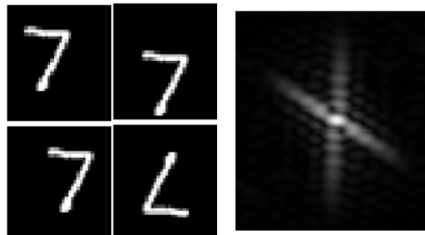


Figure 1. Shift and flipping symmetries in PR. **Left:** shifted and flipped copies of the digit image of 7; **Right:** their common Fourier magnitudes.

## 1. Introduction

Given the 2D Fourier magnitudes  $\mathbf{Y} = |\mathcal{F}(\mathbf{X})|^2 \in \mathbb{C}^{m \times m}$  of a complex-valued matrix  $\mathbf{X} \in \mathbb{C}^{n \times n}$ , is it possible to recover  $\mathbf{X}$ ? This is the (Fourier) phase retrieval (PR) problem that takes a central place in scientific imaging and has fueled numerous revolutions in fields such as X-ray crystallography, biology, optics, astronomy, and signal processing (Bendory et al., 2017; Shechtman et al., 2015). The complex phases of  $\mathcal{F}(\mathbf{X})$  are missing because detectors in practical imaging systems cannot record complex phases.

When the complex phases of  $\mathcal{F}(\mathbf{X})$  are available,  $\mathcal{F}(\mathbf{X})$  is known and recovering  $\mathbf{X}$  is just a matter of inverse 2D Fourier transform. Without the phases, recovery becomes tricky: (1) for any  $\mathbf{X}$ ,  $\mathbf{X}e^{i\theta}$  for all  $\theta \in [0, 2\pi)$  (*global phase symmetry*) and shifted copies of  $\mathbf{X}$  (*shift symmetry*) and the top-down and left-right flipped copy of  $\mathbf{X}$  (*flipping symmetry*) are all mapped to the same  $\mathbf{Y}$  (see Fig. 1), as determined by the properties of the 2D Fourier transform  $\mathcal{F}$ . So the best one can hope for is recovery up to these intrinsic symmetries; and (2) the mapping  $\mathbf{X} \mapsto |\mathcal{F}(\mathbf{X})|^2$  is *generically* injective up to the intrinsic symmetries when  $m \geq 2n - 1$  (Hayes, 1982). So in this paper, we always assume  $m \geq 2n - 1$  and solving PR is up to the intrinsic symmetries.

---

<sup>1</sup>Department of Computer Science and Engineering, University of Minnesota, Twin Cities, USA. Correspondence to: Raunak Manekar <manek009@umn.edu>, Kshitij Tayal <tayal007@umn.edu>.

When  $\mathbf{X}$  is real-valued, and nonnegativity can be enforced (e.g., images), or when support (i.e., locations of nonzero elements) of  $\mathbf{X}$  is provided with reasonable accuracy, PR can often be successfully solved by the classic hybrid input-output (HIO) method (Fienup, 1982), or its recent variants such as RAAR (Bauschke et al., 2002) or difference map (Elser et al., 2007); see a comparison of these methods in Marchesini (2007). When these extra constraints on  $\mathbf{X}$  are not applicable, even the most sophisticated variants can fail to work in practice. The failure is mostly due to stagnation caused by the intrinsic symmetries (Guizar-Sicairos & Fienup, 2012).

Deep learning has brought about new prospects of solving difficult inverse problems, of which PR is an instance. One approach would be phrasing PR as a regularized optimization problem first:  $\min_{\mathbf{X}} \ell(\mathbf{Y}, |\mathcal{F}(\mathbf{X})|^2) + \lambda\Omega(\mathbf{X})$ , and then replacing part or the entirety of  $\ell$ ,  $\Omega$ , and components of specific numerical methods for solving the regularized formulation using data-driven neural network modules. This approach is applied to PR in, e.g., (Metzler et al., 2018; Işıl et al., 2019), where HIO is still needed to produce good initialization and their methods mostly only perform local refinement—for simpler inverse problems, such special initialization is not required (Ongie et al., 2020).

More radical is the end-to-end approach, where a neural network is trained to directly approximate the inverse mapping or its proxies. Goy et al. (2018); Uelwer et al. (2019); Metzler et al. (2020) have taken this approach and shown promising results. Here, we take a critical view of the initial successes.

**Difficulty of learning with symmetries** When solving nonlinear inverse problems with symmetries, the end-to-end approach may face the difficulty of approximating highly oscillatory functions. This issue has been elucidated in our recent work [Tayal et al. \(2020\)](#). We summarize the main argument using the learning square root example: suppose

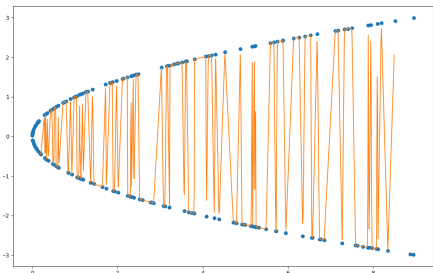


Figure 2. Highly oscillatory function defined by the data set when learning to take square root. Image credit: [Tayal et al. \(2020\)](#).

we randomly sample real values  $x_i$ 's and form a training set  $\{x_i, x_i^2\}$  and try to learn the square-root function using the end-to-end approach, allowing both positive and negative outputs. Now if we think of the function determined by the training set, which the neural network is trying to approximate, it is highly oscillatory (see [Fig. 2](#)): the sign symmetry dictates that in the training set, there are frequent cases where  $x_i^2$  and  $x_j^2$  are close but  $x_i$  and  $x_j$  have different signs and are far apart. Although in theory neural networks with adequate capacity are universal function approximators, in practice they will struggle to learn such irregular functions. For general inverse problems, so long as the forward symmetries can relate remote inputs to the same output, such as all the three symmetries in PR, similar problems can surface.

**Biases in practical image datasets** Strangely, few previous works on PR ([Goy et al., 2018](#); [Uelwer et al., 2019](#)) have discussed this issue, except for [Metzler et al. \(2020\)](#). When we examine the training and test data that previous works use, it becomes clear that the issue is probably covered by intrinsic data biases. Previous experiments typically use images from standard computer vision datasets such as MNIST, ImageNet, CelebA(faces), where the image contents tend to be centralized and naturally oriented (see [Fig. 3](#) (a)–(b)). This helps break the shift and flipping symmetries naturally, as these images are relatively close to each other compared to when mixed up with some of their symmetric copies.<sup>1</sup> Our analysis is confirmed in our later experiment, where we show that augmenting natural datasets to account for symmetries fails a naive end-to-end pipeline, which performs well without the augmentation.

<sup>1</sup>The global phase symmetry is absent as they only deal with real-valued images.

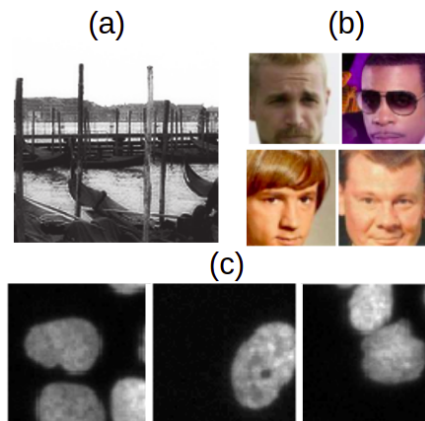


Figure 3. Sample training images used in: (a) [Goy et al. \(2018\)](#) and (b) [Uelwer et al. \(2019\)](#). (c) Sample images of biological cells from [Gustafsdottir et al. \(2015\)](#), which do not have any natural orientation or centering.

In short, the fundamental difficulty has been concealed by biased data which do not reflect the properties of data in PR applications: biological image samples (see [Fig. 3](#) (d)), astronomical objects ([Fienu, 2019](#)), where there is no natural orientation or centering of the image contents. The difficulty of learning with symmetries needs to be addressed for building practical PR systems.

## 2. Passive Symmetry Breaking for PR

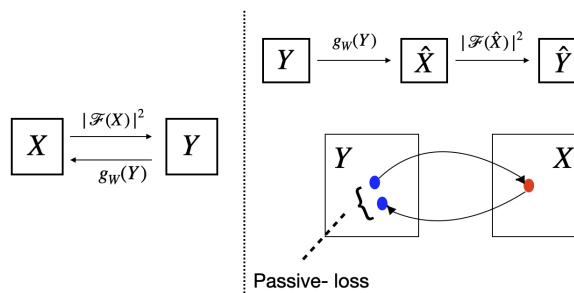


Figure 4. Illustration of PR and the passive symmetry-breaking formulation.

In [Tayal et al. \(2020\)](#), we proposed mathematical ways of preprocessing the training data to break symmetries and hence eliminating the learning difficulty for simplified versions of PR, i.e., Gaussian PR—shift and flipping symmetries are erased. It may be possible to generalize the solution to PR, but here we take a different route that is arguable much simpler but effective.

Let  $\mathcal{T} \doteq \{X_i, Y_i = |\mathcal{F}(X)|^2\}$  be the training dataset we set up for implementing the end-to-end approach, and  $g_w$

be a chosen neural network parametrized by weights  $W$ . Then the naive end-to-end approach will take the form

$$\min_W \sum_i \ell(\mathbf{X}_i, g_W(\mathbf{Y}_i)), \quad (2.1)$$

where  $\ell$  is the loss function. Due to the symmetries,  $\mathbf{Y}_i, \mathbf{Y}_j$  that are close may correspond to  $\mathbf{X}_i$  and  $\mathbf{X}_j$  that are centered at very different locations or flipped and hence far apart. This forces  $g_W$  to approximate a rapidly changing function, i.e., the difficulty that we alluded to above.

The difficulty occurs because we require  $g_W(\mathbf{Y}_i)$  to match  $\mathbf{X}_i$  as possible, where the latter data are problematic and induce mixed symmetries. How about we get rid of  $\mathbf{X}_i$ 's? A natural alternative formulation is

$$\min_W \sum_i \ell(\mathbf{Y}_i, |\mathcal{F} \circ g_W(\mathbf{Y}_i)|^2), \quad (2.2)$$

as whether  $g_W(\mathbf{Y}_i)$  outputs  $\mathbf{X}_i$  or any of its symmetric copies,

$$|\mathcal{F} \circ g_W(\mathbf{Y}_i)|^2 \approx \mathbf{Y}_i \forall i. \quad (2.3)$$

This is illustrated in Fig. 4.

**Why it might work?** We toss away the difficult  $\mathbf{X}_i$ 's, but we also supply less information to the learning model. Now  $g_W$  has much more freedom, and it can still generate distant outputs for nearby inputs. Why is there hope? We draw our inspiration from the growing pile of evidence that neural networks optimized with first-order stochastic methods tend to learn simple functions over complicated ones, known as implicit regularization (Neyshabur et al., 2014). For our problem,  $g_W$  is simple when all the symmetries are broken and complicated when there are symmetries. So if implicit regularization occurs, symmetries are naturally broken without any active intervention. So we call Eq. (2.2) the passive symmetry-breaking formulation for PR, as against the active approach proposed in Tayal et al. (2020).

**Precursors** Formulation (2.3) can be considered as an autoencoder objective with a known decoder. It is also similar to the cycle consistency idea (Zhu et al., 2017; Godard et al., 2017; Zhou et al., 2016) used in many computer vision tasks. A unified theme is to approximate identity maps. But our motivation here is for implicit symmetry breaking taking advantage of implicit regularization, which differ from all other works. The same formulation and its equivalent form in the autocorrelation form has been independently proposed in the recent work Metzler et al. (2020). They have not articulated the learning difficulty caused by symmetries, and they only motivated their formulation using the shift symmetry.

### 3. Experiment

**Goals** In this preliminary experiment, we hope to quickly confirm that (1) the naive formulation (2.1) works well with biased data, and fails when symmetries are accounted for in the dataset; and (2) our passive formulation (2.2) works well with and without the dataset biases.

**Data** We conduct our experiments on the MNIST dataset (LeCun et al., 1998), which is used by several previous works on PR. We take their 60,000 training images and 10,000 test images to construct our training and test sets, respectively. The images are  $28 \times 28$  each and so  $n = 28$ . We take  $m = 64$  here and so the injectivity threshold  $2n - 1 = 55$  is exceeded.

**Model and training** For  $g_W$ , we use the U-Net architecture (Ronneberger et al., 2015), which is a state-of-the-art deep model for image segmentation and other regression-type tasks. For training, we use the Adam optimizer and train all models for a maximum of 300 epochs. The learning rate is set as  $10^{-4}$  by default and training is stopped if the validation loss does not reduce for 20 consecutive epochs.

**Setups** We create 4 variants of the dataset to test the impact of symmetries on learning. We do this by modifying the images as described below, followed by the standard operation of taking Fourier magnitudes.

- **No Symmetry:** i.e., original MNIST dataset; samples shown in Fig. 5 (a)-left;
- **Shift symmetry:** all images placed in a larger dark background and randomly translated; samples shown in Fig. 5 (b)-left;
- **Flipping symmetry:** 50% of randomly selected training and test images are top-down and left-right flipped; samples shown in Fig. 5 (c)-left.
- **Shift and flipping symmetries:** random flipping followed by random translation; samples shown in Fig. 5 (d)-left.

We denote the vanilla formulation Eq. (2.1) and the passive formulation Eq. (2.2) as DNN- $V$  and DNN- $P$ , respectively. They are trained and tested on all the 4 variants.

**Qualitative results** We focus on the qualitative aspects in this preliminary experiment; we hope to include more quantitative comparisons in the future. Results on randomly selected test images are presented in Fig. 5. For results on each variant of the dataset, the left column is the groundtruth image, and the middle and right columns are results produced by DNN- $V$  and DNN- $P$ , respectively.

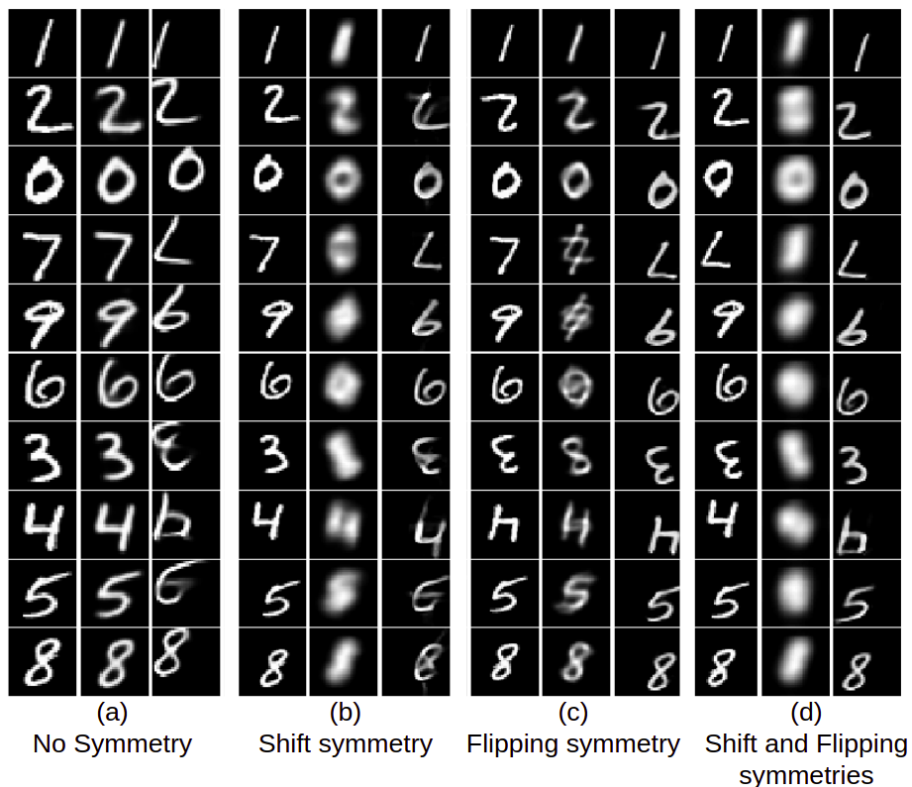


Figure 5. Visualization of recovery results. For each group, the first columns contain the groundtruth images. Second and third columns are reconstructions produced by DNN- $V$  and DNN- $P$ , respectively.

First note that on the original MNIST dataset, DNN- $V$  gives good recovery, but it quickly fails on the variants containing symmetries. This confirms our analysis regarding the dataset bias. The mode of failure is interesting, as the estimated images are almost always the superposition of the symmetric copies of the groundtruth. This is very similar to the failure mode of the classic methods on PR (Guizar-Sicairos & Fienup, 2012). Moreover, for digits that are visually similar between the original and the flipped copy, e.g., “1”, “2”, “0”, the reconstruction results are good with or without the flipping symmetry, consistent with our intuition.

On the other hand, irrespective of the symmetries, DNN- $P$  consistently leads to good recovery. Interestingly, DNN- $P$  can sometimes recover novel symmetric copies of the groundtruth image: e.g., flipped digits “7” and “9” are returned for the no-symmetry and shift-symmetry datasets even when there is no flipping symmetry entering the training data, and shifted digit “0” is returned for the no-symmetry and flipping symmetry dataset even when there is no shift symmetry present. These striking observations are not unexpected, given the fact that the DNN- $P$  loss function works only with  $Y$  and never sees the real images  $X$ . These observations reinforce our intuition that DNN- $P$  can deal with the symmetries automatically and try to learn a low-complexity function.

These preliminary results are encouraging and the next step would be to test this new formulation on other natural image datasets such as faces or objects, and more importantly, datasets that stem from real PR applications, such as the cell dataset (Fig. 3 (c)) or astronomical data. Datasets which are much less dense than MNIST may present a considerable challenge for this formulation. It would also be interesting to see whether this formulation can help break the third type of symmetry, global phase, in complex images such as MRI data.

## 4. Conclusion

Tayal et al. (2020) tackles the symmetry issue in Gaussian PR and explicitly breaks the symmetry in the training set. Generalizing the idea to PR can be mathematically heavy, as three types of symmetries exist for PR whereas only one type exists for Gaussian PR. The passive formulation presented in this work is lightweight and turns out to be effective, making this a first work of this kind on PR. This formulation can be developed into a practical algorithm for end-to-end learning on Fourier-PR.

## Acknowledgments

KT and VK are supported by NSF BIGDATA 1838159. The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper.

## References

- Bauschke, H. H., Combettes, P. L., and Luke, D. R. Phase retrieval, error reduction algorithm, and fienuip variants: a view from convex optimization. *Journal of the Optical Society of America A*, 19(7):1334, jul 2002. doi: 10.1364/josaa.19.001334.
- Bendory, T., Beinert, R., and Eldar, Y. C. Fourier phase retrieval: Uniqueness and algorithms. In *Compressed Sensing and its Applications*, pp. 55–91. Springer International Publishing, 2017. doi: 10.1007/978-3-319-69802-1\_2.
- Elser, V., Rankenburg, I., and Thibault, P. Searching with iterated maps. *Proceedings of the National Academy of Sciences*, 104(2):418–423, jan 2007. doi: 10.1073/pnas.0606359104.
- Fienup, J. R. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758, aug 1982. doi: 10.1364/ao.21.002758.
- Fienup, J. R. Phase retrieval for image reconstruction. In *Imaging and Applied Optics 2019 (COSI, IS, MATH, pCAOP)*. OSA, 2019. doi: 10.1364/cosi.2019.cm1a.1.
- Godard, C., Mac Aodha, O., and Brostow, G. J. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 270–279, 2017.
- Goy, A., Arthur, K., Li, S., and Barbastathis, G. Low photon count phase retrieval using deep learning. *Physical Review Letters*, 121(24), dec 2018. doi: 10.1103/physrevlett.121.243902.
- Guizar-Sicairos, M. and Fienup, J. R. Understanding the twin-image problem in phase retrieval. *Journal of the Optical Society of America A*, 29(11):2367, oct 2012. doi: 10.1364/josaa.29.002367.
- Gustafsdottir, S., Ljosa, V., Sokolnicki, K., et al. Human u2os cells-compound cell-painting experiment. *The Cell Image Library*, 2015.
- Hayes, M. The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(2):140–154, apr 1982. doi: 10.1109/tassp.1982.1163863.
- Işıl, Ç., Oktem, F. S., and Koç, A. Deep iterative reconstruction for phase retrieval. *Applied Optics*, 58(20):5422, jul 2019. doi: 10.1364/ao.58.005422.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- Marchesini, S. Invited article: A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments*, 78(1):011301, jan 2007. doi: 10.1063/1.2403783.
- Metzler, C. A., Schniter, P., Veeraraghavan, A., and Baraniuk, R. G. prdeep: Robust phase retrieval with a flexible deep network. *arXiv preprint arXiv:1803.00212*, 2018.
- Metzler, C. A., Heide, F., Rangarajan, P., Balaji, M. M., Viswanath, A., Veeraraghavan, A., and Baraniuk, R. G. Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging. *Optica*, 7(1):63, jan 2020. doi: 10.1364/optica.374026.
- Neyshabur, B., Tomioka, R., and Srebro, N. In search of the real inductive bias: On the role of implicit regularization in deep learning. *arXiv preprint arXiv:1412.6614*, 2014.
- Ongie, G., Jalal, A., Baraniuk, C. A. M. R. G., Dimakis, A. G., and Willett, R. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 2020.
- Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.
- Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J., and Segev, M. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, may 2015. doi: 10.1109/msp.2014.2352673.
- Tayal, K., Lai, C.-H., Kumar, V., and Sun, J. Inverse problems, deep learning, and symmetry breaking. *arXiv preprint arXiv:2003.09077*, 2020.
- Uelwer, T., Oberstraß, A., and Harmeling, S. Phase retrieval using conditional generative adversarial networks. *arXiv:1912.04981*, 2019.
- Zhou, T., Krahenbuhl, P., Aubry, M., Huang, Q., and Efros, A. A. Learning dense correspondence via 3d-guided cycle consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 117–126, 2016.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.