MITLibraries | DSpace@MIT

MIT Open Access Articles

Spectral pre-modulation of training examples enhances the spatial resolution of the phase extraction neural network (PhENN)

Massachusetts Institute of Technology

Spectral pre-modulation of training examples enhances the spatial resolution of the phase extraction neural network (PhENN)

SHUAI LI[1,*] **AND GEORGE BARBASTATHIS**[1,2]

[1]*Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
[2]*Singapore-MIT Alliance for Research and Technology (SMART) Centre, One Create Way, Singapore 117543, Singapore*
[*]shuaili@mit.edu

Abstract: The phase extraction neural network (PhENN) [Optica **4**, 1117 (2017)] is a computational architecture, based on deep machine learning, for lens-less quantitative phase retrieval from raw intensity data. PhENN is a deep convolutional neural network trained through examples consisting of pairs of true phase objects and their corresponding intensity diffraction patterns; thereafter, given a test raw intensity pattern, PhENN is capable of reconstructing the original phase object robustly, in many cases even for objects outside the database where the training examples were drawn from. Here, we show that the spatial frequency content of the training examples is an important factor limiting PhENN's spatial frequency response. For example, if the training database is relatively sparse in high spatial frequencies, as most natural scenes are, PhENN's ability to resolve fine spatial features in test patterns will be correspondingly limited. To combat this issue, we propose "flattening" the power spectral density of the training examples before presenting them to PhENN. For phase objects following the statistics of natural scenes, we demonstrate experimentally that the spectral pre-modulation method enhances the spatial resolution of PhENN by a factor of 2.

1. Introduction

The use of machine learning architectures is a relatively new trend in computational imaging and rapidly gaining popularity. Originally it was proposed for imaging through scatter using support vector regression (SVR) [1], a version of a two-layer regression network. Subsequently, contemporary Deep Neural Networks (DNNs) have been applied successfully to the same problem of imaging through scatter [2, 3], as well as tomography [4], lensless quantitative phase retrieval [5, 6], microscopy [7], GHOST imaging [8], imaging through fiber bundles [9], holographic image processing [10–12], and imaging at extremely low light levels [13].

The main motivation for the use of machine learning is to overcome certain deficits of traditional computational imaging approaches. The latter are based on convex optimization, structured so that the optimal solution is as close as possible to the true object. The functional to be minimized is specified by the physical model H of the imaging process, also referred to as forward operator; and by prior knowledge $\Phi(f)$ about the class of objects being imaged, also known as regularizer. The inverse (estimate) $\hat{f}$ of an object f is obtained from a measurement g as

$$\hat{f} = \text{argmin}_f \left\{ \|Hf - g\|^2 + \alpha\Phi(f) \right\} \tag{1}$$

The regularization parameter α expresses the imaging system designer's relative belief in the measurement *vs.* belief in the available prior knowledge about the object class.

Clearly, the performance of Eq. (1) in terms of producing acceptable inverses is crucially dependent upon correct and explicit knowledge of both H and Φ, and judicious selection of the

parameter α [14]. In situations where this knowledge is questionable or not explicitly available, deep machine learning approaches become appealing as an effort to *learn* the missing knowledge implicitly through examples. Instead of Eq. (1), the object estimate is then obtained as

$$\hat{f} = \mathrm{DNN}(g), \tag{2}$$

where DNN(.) denotes the output of the trained deep neural network.

Eq. (2) may be used for other, non-deep machine learning structures even though they are generally less effective. However, strictly applied, Eq. (2) is limited to the special "end-to-end" design where the measurement g from the camera is fed directly to the DNN. In some cases g first goes through a physical pre-processor, and the pre-processor's output is fed into the DNN [6, 13]; whereas in other cases g is fed multiple times into a cascade of generator DNNs [15] to assess the outputs at each step. Developing notation for and carrying out a full debate on the relative merits of these different approaches is beyond the scope of the present paper, where, in any case, we used the end-to-end method Eq. (2) only.

Just as the performance of minimization principle Eq. (1) depends upon knowledge of the operators H and Φ, performance of the DNN principle Eq. (2) depends on the specific DNN architecture chosen (number of layers, connectivity, etc.) and the quality of the training examples. It is the latter aspect of DNN design that we focus on in the present paper. More specifically, we are concerned with the spatial resolution that the DNN can achieve, depending on the spatial frequency content of the examples the DNN is trained with.

We chose to study this question in the specific context of quantitative phase retrieval. This is a classical problem in optical imaging, because by virtue of its challenge it evokes elegant solutions and also because it has important applications in biology, medicine, and inspection for manufacturing and security. Traditional approaches include digital holography (DH) [16–20] and the related phase shifting interferometry method [21], propagation based methods such as the Transport of Intensity Equation (TIE) [22–30] and iterative methods such as the Gerchberg-Saxton-Fienup algorithm [31–35].

The end-to-end residual convolutional DNN solution to lens-less quantitative phase retrieval is PhENN [5], shown to be robust to errors in propagation distance and fairly well able to generalize to test objects from outside the databases used for training. In the present paper, we implemented PhENN in a slightly different optical hardware configuration, described in Section 2.1. The computational architecture, described in Section 2.2, was similar to the original PhENN except here we used the Negative Pearson Correlation Coefficient (NPCC) as training loss function. This has a small beneficial effect in the reconstructions, but necessitates a histogram calibration procedure, described in Section 2.3, to remove linear amplification and bias in the reconstructed phase images.

From the point of view of the original inverse problem formulation Eq. (1), PhENN in effect has to learn both the forward operator H and the prior Φ at the entire range of spatial frequencies of interest. The examples presented to PhENN during training establish the spatial frequency content that is stored in the network weights contributing to the retrieval operation Eq. (2). In principle, this should be sufficient because, if the training examples are representative enough of the object class, then retrieval of each spatial frequency should be learnt proportionally to that spatial frequency's presence in the database. In practice, however, we found that spatial frequencies with relatively low representation in the database tend to be overshadowed by the more popular spatial frequencies, perhaps due to the nonlinearities in the network training process and operation.

Invariably, high spatial frequencies tend to be less popular in most available databases. ImageNet, in particular, exhibits the well-known inverse-square power spectral density of natural images [36], as we verify in Fig. 6. This means that high spatial frequencies are inherently under-represented in PhENN training. Compounded by the nonlinear suppression of the less

popular spatial frequencies due to PhENN nonlinearities, as mentioned above, this results in low-pass filtering of the estimates and loss of fine detail. Detailed analysis of this effect is presented in Section 3.

To better recover high spatial frequencies in natural objects then, one should emphasize high spatial frequencies more during training; this may be achieved, for example, by flattening the power spectral density of the training examples *before* they are presented to the neural network. It would appear that this spectral intervention violates the object class priors: PhENN does not learn the priors of ImageNet itself, it rather learns an edge-enhanced version of the priors. Yet, in practice, again probably because of nonlinear PhENN behavior, we found this spectral pre-modulation strategy to work quite well. The detailed approach and results are found in Section 4.

It is worth mentioning here that the first, to our knowledge, explicit experimental analysis of a DNN's spatial resolution was conducted on IDiffNet in the context of imaging through diffuse media [3]. We chose to pursue the issue further in the present paper but on a different optical problem because spatial resolution in quantitative phase retrieval, in addition to also being worthwhile, is not impacted by the extreme ill-posedness of diffuse media. Even though we have not tried extensively beyond phase retrieval, pre-processing of training examples by spectral manipulation might have merit for several other challenging imaging problems.

2. Imaging system architecture

2.1. Optical configuration

Our optical configuration is shown in Fig. 1. Unlike [5], a transmissive spatial light modulator (SLM) (Holoeye, LC2012, pixel size 36μm) is used in this system as a programmable phase object f representing the ground truth. The transmissive SLM is coherently illuminated by a He-Ne laser light source (Research Electro-Optics, Model 30995, 633nm). The light is transmitted through a spatial filter consisting of a microscope objective (Newport, M-60X, 0.85NA) and a pinhole aperture ($D = 5\mu$m) and then collimated by a lens (focal length 200mm) before illuminating the SLM. A telescope consisting of two plano-convex lenses L_1 and L_2 is placed between the SLM and a CMOS camera (Basler, A504k, pixel size 12μm). The CMOS camera captures the intensity g of the diffraction pattern produced by the SLM at a defocus distance $\Delta z = 50$mm. The focal lengths of L_1 and L_2 are set to $f_1 = 150$mm and $f_2 = 50$mm, respectively. As a result, this telescope demagnifies the object by a factor of 3, consistent with the ratio between SLM and CMOS camera pixel sizes. An iris with diameter 5mm is placed at the pupil plane of the telescope to keep the 0th diffracted order of the SLM and filter out all the other orders.

The modulation performance of the SLM depends on the input and output polarizations, which are controlled by the polarizer P and the analyzer A, respectively. In order to realize phase-mostly modulation, we set the incident beam to be linearly polarized at 310° with respect to the vertical direction and also set the analyzer to be oriented at 5° with respect to the vertical direction. The specific calibration curves for the SLM's modulation performance can be found in [37]. In the present paper, all the training and testing objects are of size 256×256. They are zero-padded to the size 1024×768, before being uploaded to the SLM. For the diffraction patterns captured by the CMOS camera, we crop the central 256×256 region for processing.

2.2. Neural network architecture and training

Similar to [5], the phase extraction neural network (PhENN) that we implement in this paper follows the U-net architecture [38] and utilizes residuals to facilitate learning (ResNet [39].) The detailed architecture is shown in Fig. 2. PhENN input is the intensity g, and successively passes through 4 down-residual blocks (DRBs) for feature extraction. The extracted feature map then successively passes through 4 up-residual blocks (URBs) and 2 residual blocks (RBs)
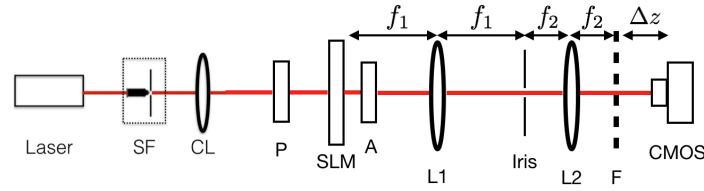
Fig. 1. Optical configuration. SF: spatial filter; CL: collimating lens; P: linear polarizer; A: analyzer; SLM: spatial light modulator; L1 and L2: plano-convex lenses; F: focal plane of L2.

for pixel-wise regression and at the last layer outputs the estimate $\hat{f}$ of the object phase. Skip connections are used in the architecture to pass downstream local spatial information learnt in the initial layers. More details about the structures of the DRBs, URBs and RBs can be found in [5].
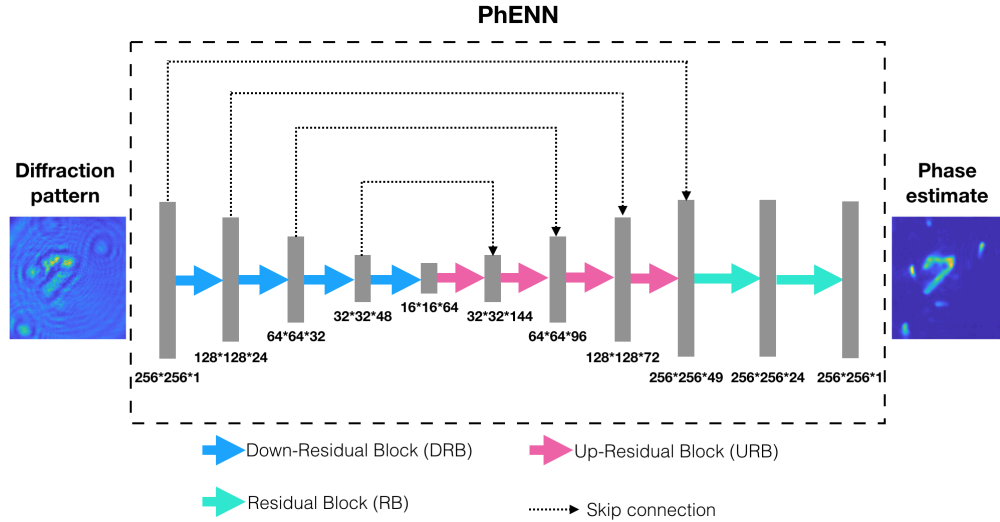

Fig. 2. Phase extraction neural network (PhENN) architecture.

Unlike [5], here we use the Negative Pearson Correlation Coefficient (NPCC) as loss function [3] to train PhENN. The NPCC loss function is defined as

$$\mathcal{L} = \sum_k \mathcal{E}_{\text{NPCC}}\left(f_k, \hat{f}_k\right), \qquad \text{where} \tag{3}$$

$$\mathcal{E}_{\text{NPCC}}\left(f_k, \hat{f}_k\right) \equiv (-1) \times \frac{\sum_{i,j}\left(f_k(i,j) - \langle f_k \rangle\right)\left(\hat{f}_k(i,j) - \langle \hat{f}_k \rangle\right)}{\sqrt{\sum_{i,j}\left(f(i,j) - \langle f_k \rangle\right)^2}\sqrt{\sum_{i,j}\left(\hat{f}(i,j) - \langle \hat{f}_k \rangle\right)^2}}; \tag{4}$$

f and $\hat{f}$ are the true object and the object estimate according to Eq. (2), respectively; the summations take place over all pixels (i,j) and training example labels k; and $\langle . \rangle$ denotes spatial averaging. We have found the NPCC to generally result in better DNN training in the problems that we examined, especially for objects that are spatially sparse [3]. However, some care needs to be taken when the estimate $\hat{f}$ is not affine-invariant; we discuss this immediately below.

2.3. Calibration of PhENN output trained with NPCC

From the definition Eq. (4) it follows that for any function ψ and arbitrary real constants a and b representing linear amplification and bias, respectively,

$$\mathcal{E}_{\text{NPCC}}(\psi, a\psi + b) = -1. \tag{5}$$

In other words, a DNN trained with NPCC as loss function can only produce affine transformed estimates; there is no way to enforce the requirement $a = 1$, $b = 0$ which would guarantee linear amplification- and bias-free reconstruction and is especially important for quantitative phase imaging. Neither does there exist a way that we know of to predetermine the values of a and b through specific choices in DNN training.

Therefore, after DNN training a calibration step is required to determine the values of a and b that have resulted so that they can be compensated. This is realized by histogram matching according to the process shown in Fig. 3. Given a set of calibration data, we compute the cumulative distribution functions (CDFs) for the ground truth values as well as the PhENN output values, as shown in Figs. 3(a) and 3(b). For an arbitrary value f in the ground truth, we find its corresponding PhENN output value $\hat{f}$ that is at the same CDF level; and repeat the process for several $(f, \hat{f})$ samples. Subsequently, the values of a and b are determined by linear fitting of the form $\hat{f} = af + b$, as shown in Fig. 3(c).
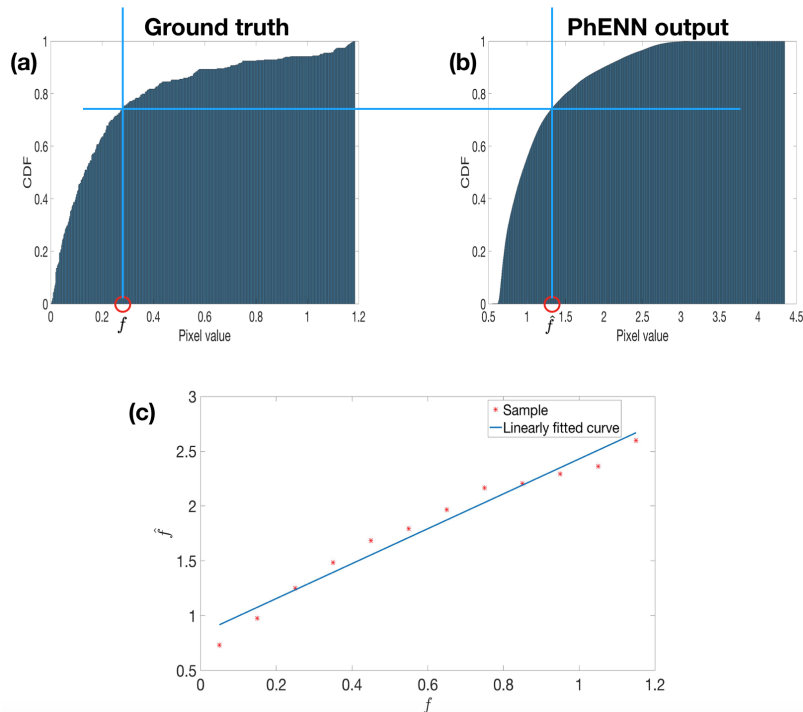

Fig. 3. Calibration process. (a) Cumulative distribution function (CDF) of the ground truth. (b) Cumulative distribution function (CDF) of the PhENN output. (c) Linear curve fitting.

3. Resolution analysis of ImageNet-trained PhENN

In [5], we trained separate PhENNs using the databases Faces-LFW [40] and ImageNet [41] and found that both PhENNs generalize to test objects within and outside these two databases. In the present paper, we restrict our analysis to the ImageNet database only.

In the PhENN training phase, a total of 10, 000 images selected from the ImageNet database are uploaded to the SLM and the respective diffraction patterns are captured by the CMOS. For testing, we use a total of 471 images selected from several different databases: 50 Characters, 40 Faces-ATT [42, 43], 60 CIFAR [44], 100 MNIST [45], 100 Faces-LFW, 100 ImageNet, 20 resolution test patterns [3], and 1 all-zero (dark) image. The diffraction pattern corresponding to the all-zero image is used as the background. For every test diffraction pattern that we capture, we first subtract the background and then normalize, before feeding into the neural network.

3.1. Reconstruction results

The phase reconstruction results are shown in Fig. 4. Here, we use 100 ImageNet test images as calibration data to compensate for the unknown affine transform effected by the NPCC-trained PhENN (Section 2.3). As expected, PhENN is not only able to quantitatively reconstruct the phase objects within the same category as its training database (ImageNet), but also to retrieve the phase for those test objects from other databases. This indicates that PhENN has indeed learned a model of the underlying physics of the imaging system or at the very least a generalizable mapping of low-level textures between the phase objects and their respective diffraction patterns.
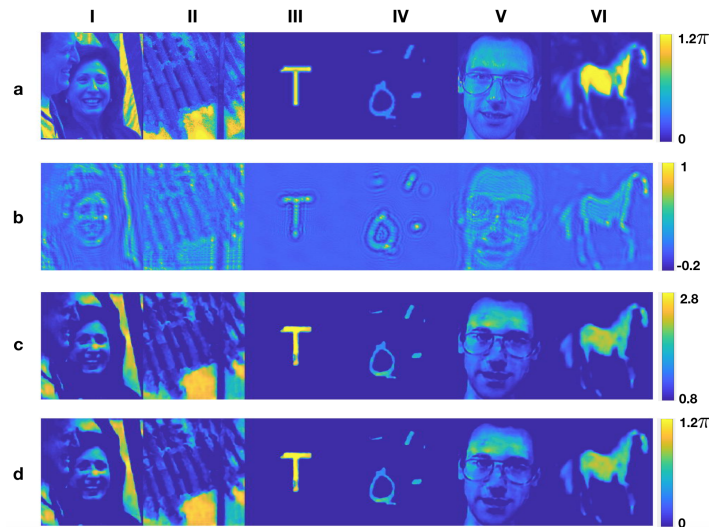

Fig. 4. Reconstruction results of PhENN trained with ImageNet. (a) Ground truth for the phase objects. (b) Diffraction patterns captured by the CMOS (after background subtraction and normalization). (c) PhENN output. (d) PhENN reconstruction after the calibration shown in Section 2.3. Columns (i-vi) correspond to the dataset from which the test image is drawn: (i) Faces-LFW [40], (ii) ImageNet [41], (iii) Characters, (iv) MNIST Digits [45], (v) Faces-ATT [42, 43], or (vi) CIFAR [44], respectively.

3.2. Resolution test

In order to test the spatial resolution of our trained PhENN, we use dot patterns as test objects [3], shown in Fig. 5(a). Altogether 20 dot patterns are tested, with spacing D between dots gradually increasing from 2 pixels to 21 pixels. From the resolution test results shown in Fig. 5 it can be observed that the PhENN trained with ImageNet is able to resolve two dots down to $D = 6$ pixels but fails to distinguish two dots with spacing $D \leq 5$ pixels. Thus, $D \approx 6$ pixels can be considered as the Rayleigh resolution limit of this PhENN for point-like phase objects.
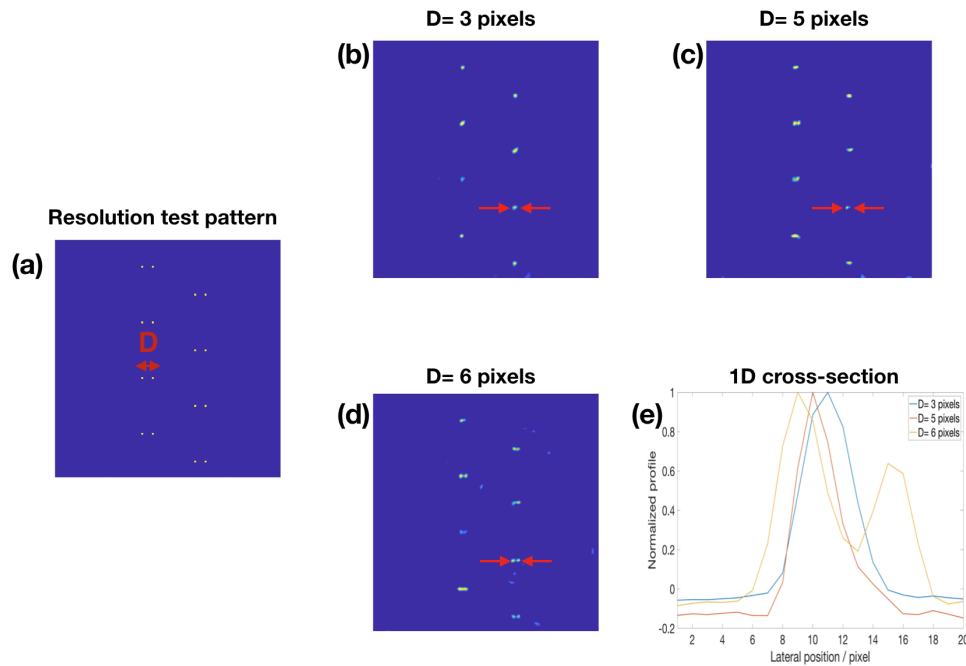
Fig. 5. Resolution test for PhENN trained with ImageNet. (a) Dot pattern for resolution test. (b) PhENN reconstructions for dot pattern with $D = 3$ pixels. (c) PhENN reconstructions for dot pattern with $D = 5$ pixels. (d) PhENN reconstructions for dot pattern with $D = 6$ pixels. (e) 1D cross-sections along the lines indicated by red arrows in (b)-(d).

4. Resolution enhancement by spectral pre-modulation

In our imaging system, the SLM pixel size limits the spatial resolution of the trained PhENN since the minimum sampling distance in all the training and testing objects displayed on the SLM equals one pixel $d_p = 36\mu m$, or maximum spatial frequency [46] 13.9mm^{-1}. However, as we saw in Section 3.2, the resolution achieved by our PhENN trained with ImageNet database is merely 6 pixels (216μm), much worse than the theoretical value.

The additional factor limiting the spatial resolution of the trained PhENN is the spatial frequency content of the training database. Generally, databases of natural objects, such as natural images, faces, hand-written characters, etc. do not cover the entire spectrum up to $1/(2d_p)$. For example, below we analyze the ImageNet database and show that it is dominated by low spatial frequency components, with the prevalence of higher spatial frequencies decreasing quadratically.

During training, the neural network learns the particular prevalence of spatial frequencies in the training examples as prior Φ, in addition to learning the physical forward operator H. What this implies is that the less prevalent spatial frequencies are actually learnt *against*, meaning that by presenting them less frequently we may be teaching PhENN to suppress or ignore them. In the rest of this section, we present evidence to corroborate this fact, and suggest as solution a pre-processing step that edge enhances the training examples as a way to impress their importance better upon PhENN.

4.1. Spectral pre-modulation

The 2D power spectral density (PSD) $S(u, v)$ as function of spatial frequencies u and v for the $10,000$ images in the ImageNet is shown in Figs. 6(a) and 6(b) in linear and logarithmic scales, respectively; and in cross-section along the spatial frequency u in Figs. 6(c) and 6(d). Not surprisingly [36], the cross-sectional power spectral density follows a power law of the form $|u|^p$ with $p \approx -2$.
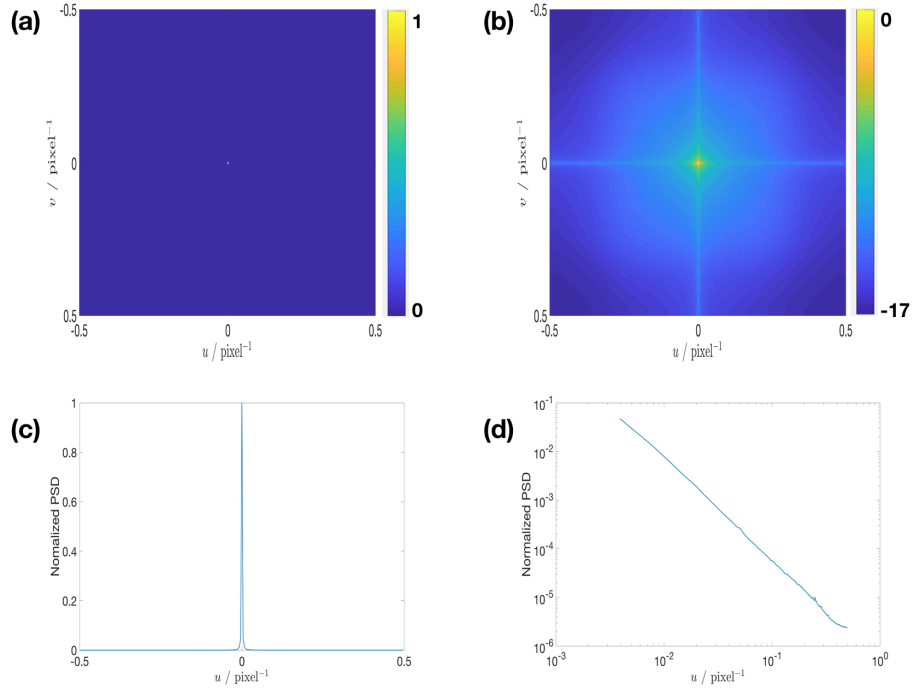

Fig. 6. Spectral analysis of the ImageNet database. (a& b) 2D normalized power spectral density (PSD) of the ImageNet database in linear and logarithmic scale. (c& d) 1D cross-sections along the spatial frequency u of (a& b), respectively.

Therefore, we may approximately represent the 2D PSD of ImageNet database as

$$S(u, v) \propto \left(\sqrt{u^2 + v^2} \right)^{-2} = \frac{1}{u^2 + v^2}. \tag{6}$$

This is flattened by the inverse filter

$$G(u, v) = \sqrt{u^2 + v^2}. \tag{7}$$

As expected, the high spatial frequency components in the image are amplified after the modulation, as can be seen, for example, in Fig. 7.

4.2. Resolution enhancement

We trained a new PhENN using training examples that were spectrally pre-modulated according to Eq. (7). That is, we replaced every training example $f(i, j)$ with $f_e(i, j)$, where

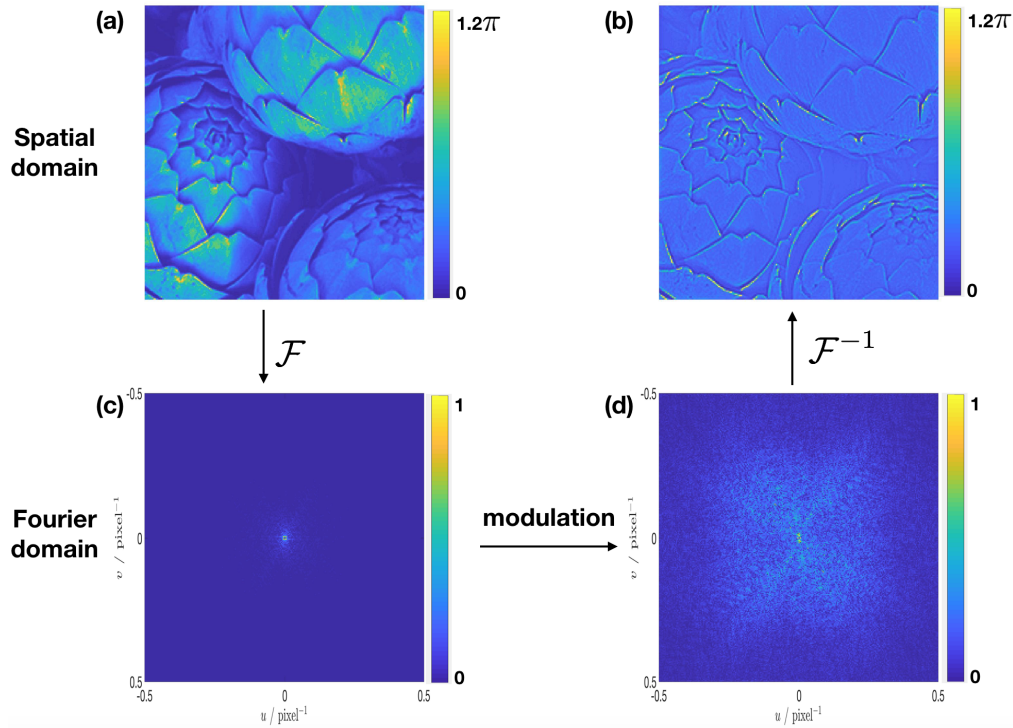$$F_e(u, v) = G(u, v)F(u, v) \tag{8}$$

Fig. 7. Spectral pre-modulation. (a) Original image [41]. (b) Modulated image. (c) Fourier spectrum of the original image. (d) Fourier spectrum of the modulated image.

and F, F_e are the Fourier transforms of f, f_e, respectively. We also collected the corresponding diffraction patterns $g_e(i, j)$. The test examples were left without modulation, *i.e.* the same as in the original use of PhENN described in Section 3. All the training parameters were also kept the same. Both dot pattern and ImageNet test images were used to demonstrate the resolution enhancement, shown in Figs. 8 and 9, respectively.

From Fig. 8, we find that with spectral pre-modulation of the training examples according to Eq. (8), PhENN is able to resolve two dots with spacing $D = 3$ pixels. Compared with the resolution test results shown in Fig. 5, it can be said that the spatial resolution of PhENN has been enhanced by a factor of 2 with the spectral pre-modulation technique. In Fig. 9, for the same test image selected from ImageNet database, more details are recovered by the PhENN that was trained with spectrally pre-modulated ImageNet, albeit at the cost of amplifying some noisy features of the object, near edges most notably.
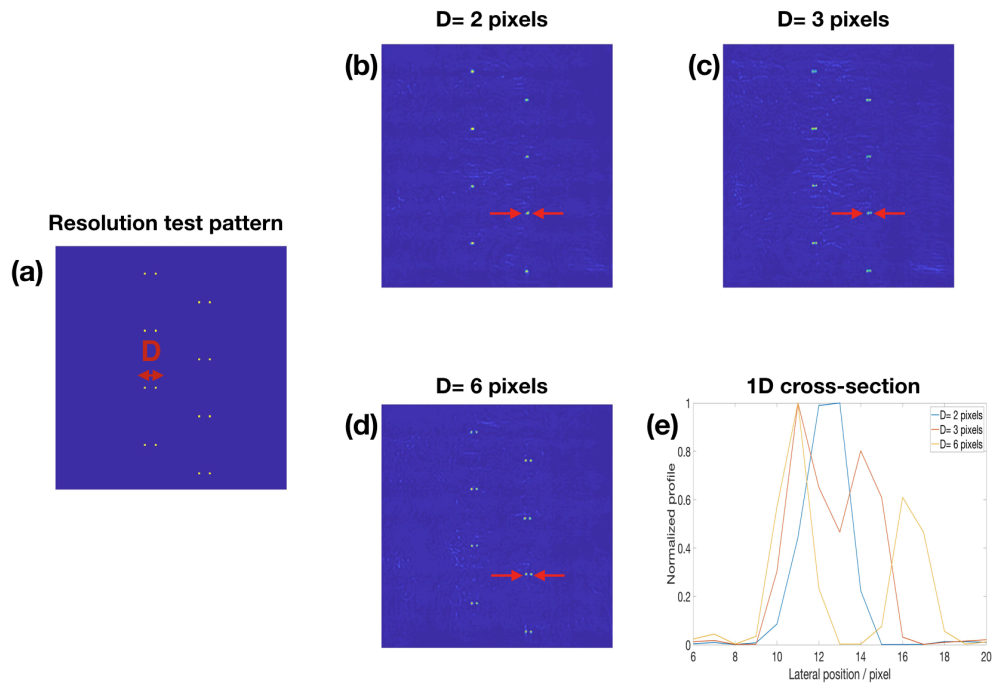
Fig. 8. Resolution test for PhENN trained with examples from the ImageNet database with spectral pre-modulation according to Eq. (8). (a) Dot pattern for resolution test. (b) PhENN reconstructions for dot pattern with $D = 2$ pixels. (c) PhENN reconstructions for dot pattern with $D = 3$ pixels. (d) PhENN reconstructions for dot pattern with $D = 6$ pixels. (e) 1D cross-sections along the lines indicated by red arrows in (b)-(d).
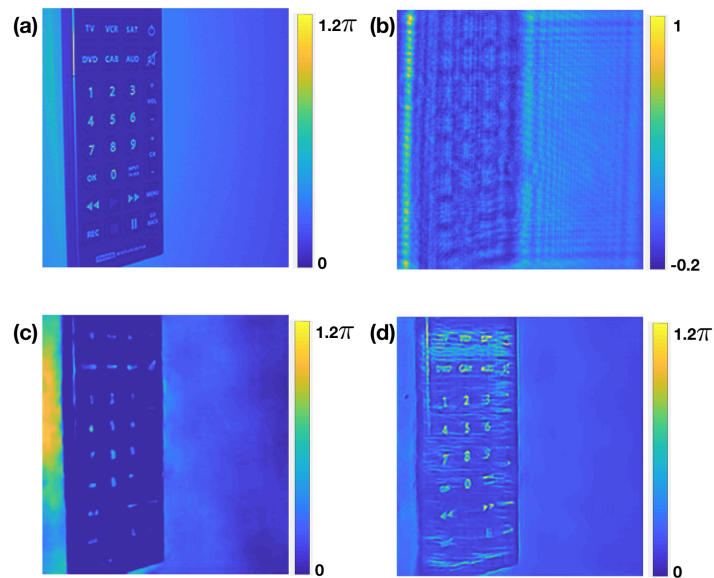

Fig. 9. Resolution enhancement demonstration. (a) Ground truth for a phase object [41]. (b) Diffraction pattern captured by the CMOS (after background subtraction and normalization). (c) Phase reconstruction by PhENN trained with ImageNet examples. (d) Phase reconstruction by PhENN trained with ImageNet examples that were spectrally pre-modulated according to Eq. (8).

We also investigated the effect of spectral post-modulation in the original PhENN; that is, if we use a PhENN trained without spectral pre-modulation, and modulate the PhENN output $\hat{f}(i, j)$ according to

$$\hat{F}_e(u, v) = G(u, v)\hat{F}(u, v) \tag{9}$$

and $\hat{F}$, $\hat{F}_e$ are the Fourier transforms of $\hat{f}$, $\hat{f}_e$, respectively, do we obtain a similar resolution enhancement? The answer is no, as can be clearly verified from the results of Fig. 10.

This negative result illustrates that in the original training scheme (without spectral pre-modulation) the fine details are indeed lost and not recoverable by simple means, e.g. linear post-processing. It also highlights the effect of the nonlinearity in PhENN's operation and bolsters our claim that spectral pre-modulation does something non-trivial: it teaches PhENN a prior, namely how to recover high spatial frequency content.
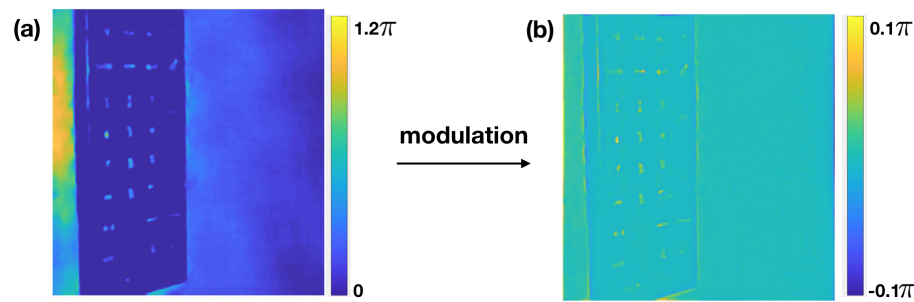

Fig. 10. Spectral post-modulation. (a) Output of PhENN trained with ImageNet. The same as Fig. 9 (c). (b) Modulated output.

5. Conclusions

The spectral flattening approach Eq. (8) as pre-modulation is a simple approach that we found to be effective in enhancing PhENN's resolution by a factor of 2 when trained and tested on ImageNet examples. We have not investigated the performance of other (non-flattening) filters; indeed, it would be an interesting theoretical question to ask: given a particular form of the PSD in the training examples, what is the optimal spectral pre-modulation for improving spatial resolution?

It is also worth repeating the concern about the priors that PhENN is learning from the spatially pre-modulated examples that we pointed out in Section 1. The amplification of certain noise artifacts, clearly seen in the result of Fig. 9(d), shows that, in addition to learning how to resolve fine details in the artifact, PhENN has learnt, somewhat undesirably, to edge enhance (since all the examples it was trained with were also edge enhanced.) These observations should present fertile ground for further improvements upon the work presented here.

References

1. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," Opt. Express **24**, 13738–13743 (2016).
2. M. Lyu, H. Wang, G. Li, and G. Situ, "Exploit imaging through opaque wall via deep learning," arXiv preprint arXiv:1708.07881 (2017).

3. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," Optica **5**, 803–813 (2018).

4. K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," IEEE Transactions on Image Process. **26**, 4509–4522 (2017).

5. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**, 1117–1125 (2017).

6. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light. Sci. & Appl. **7**, 17141 (2018).

7. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**, 1437–1443 (2017).

8. M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging," Sci. Reports **7**, 17865 (2017).

9. N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, "Learning to see through multimode fibers," Optica **5**, 960–966 (2018).

10. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," Optica **5**, 337–344 (2018).

11. T. Nguyen, V. Bui, V. Lam, C. B. Raub, L.-C. Chang, and G. Nehmetallah, "Automatic phase aberration compensation for digital holographic microscopy based on deep learning background detection," Opt. Express **25**, 15043–15057 (2017).

12. S. Jiao, Z. Jin, C. Chang, C. Zhou, W. Zou, and X. Li, "Compression of phase-only holograms with jpeg standard and deep learning," arXiv preprint arXiv:1806.03811 (2018).

13. A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," arXiv preprint arXiv:1806.10029 (2018).

14. H. Liao, F. Li, and M. K. Ng, "Selection of regularization parameter in total variation image restoration," J. Opt. Soc. Am. A **26**, 2311–2320 (2009).

15. M. Mardani, H. Monajemi, V. Papyan, S. Vasanawala, D. Donoho, and J. Pauly, "Recurrent generative adversarial networks for proximal learning and automated compressive image recovery," arXiv preprint arXiv:1711.10046 (2017).

16. J. W. Goodman and R. Lawrence, "Digital image formation from electronically detected holograms," Appl. Phys. Lett. **11**, 77–79 (1967).

17. Y. Rivenson, A. Stern, and B. Javidi, "Compressive fresnel holography," J. Disp. Technol. **6**, 506–509 (2010).

18. J. H. Milgram and W. Li, "Computational reconstruction of images from holograms," Appl. Opt. **41**, 853–864 (2002).

19. D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive holography," Opt. Express **17**, 13040–13049 (2009).

20. L. Williams, G. Nehmetallah, and P. P. Banerjee, "Digital tomographic compressive holographic reconstruction of three-dimensional objects in transmissive and reflective geometries," Appl. Opt. **52**, 1702–1710 (2013).

21. K. Creath, "Phase-shifting speckle interferometry," Appl. Opt. **24**, 3053–3058 (1985).

22. M. R. Teague, "Deterministic phase retrieval: a green's function solution," J. Opt. Soc. Am. **73**, 1434–1441 (1983).

23. S. S. Kou, L. Waller, G. Barbastathis, and C. J. Sheppard, "Transport-of-intensity approach to differential interference contrast (ti-dic) microscopy for quantitative phase imaging," Opt. Lett. **35**, 447–449 (2010).

24. D. Paganin and K. A. Nugent, "Noninterferometric phase imaging with partially coherent light," Phys. Rev. Lett. **80**, 2586 (1998).

25. J. A. Schmalz, T. E. Gureyev, D. M. Paganin, and K. M. Pavlov, "Phase retrieval using radiation and matter-wave fields: Validity of teague's method for solution of the transport-of-intensity equation," Phys. Rev. A **84**, 023808 (2011).

26. L. Waller, S. S. Kou, C. J. Sheppard, and G. Barbastathis, "Phase from chromatic aberrations," Opt. Express **18**, 22817–22825 (2010).

27. L. Waller, M. Tsang, S. Ponda, S. Y. Yang, and G. Barbastathis, "Phase and amplitude imaging from noisy images by kalman filtering," Opt. Express **19**, 2805–2815 (2011).

28. L. Tian, J. C. Petruccelli, Q. Miao, H. Kudrolli, V. Nagarkar, and G. Barbastathis, "Compressive x-ray phase tomography based on the transport of intensity equation," Opt. Lett. **38**, 3418–3421 (2013).

29. A. Pan, L. Xu, J. C. Petruccelli, R. Gupta, B. Singh, and G. Barbastathis, "Contrast enhancement in x-ray phase contrast tomography," Opt. Express **22**, 18020–18026 (2014).

30. Y. Zhu, A. Shanker, L. Tian, L. Waller, and G. Barbastathis, "Low-noise phase imaging by hybrid uniform and structured illumination transport of intensity equation," Opt. Express **22**, 26696–26711 (2014).

31. R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35**, 237–246 (1972).

32. J. R. Fienup, "Reconstruction of an object from the modulus of its fourier transform," Opt. Lett. **3**, 27–29 (1978).

33. R. Gonsalves, "Phase retrieval from modulus data," J. Opt. Soc. Am. **66**, 961–964 (1976).

34. J. Fienup and C. Wackerman, "Phase-retrieval stagnation problems and solutions," J. Opt. Soc. Am. A **3**, 1897–1907 (1986).

35. H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization," J. Opt. Soc. Am. A **19**, 1334–1345 (2002).

36. v. A. Van der Schaaf and J. v. van Hateren, "Modelling the power spectra of natural images: statistics and information," Vis. Res. **36**, 2759–2770 (1996).

37. S. Li, A. Sinha, J. Lee, and G. Barbastathis, "Quantitative phase microscopy using deep neural networks," in *Quantitative Phase Imaging IV,* vol. 10503 (International Society for Optics and Photonics, 2018), p. 105032D.

38. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention,* (Springer, 2015), pp. 234–241.

39. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition,* (IEEE, 2016), pp. 770–778.

40. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007).

41. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," Int. J. Comput. Vis. **115**, 211–252 (2015).

42. F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the Second IEEE Workshop on Applications of Computer Vision,* (IEEE, 1994), pp. 138–142.

43. AT&T Laboratories Cambridge, "AT&T Database of Faces," (1994). Data retrieved from https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

44. A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Tech. rep., University of Toronto (2009).

45. Y. LeCun, C. Cortes, and C. J. Burges, "Mnist handwritten digit database," AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist **2** (2010).

46. L1 determines the aperture stop with diameter 25.4mm, *i.e.* a numerical aperture NA= 12.7/150 = 0.0847. The nominal diffraction-limited resolution should be $d_0 = \lambda/(2\text{NA}) = 3.74\mu$m. That calculation is irrelevant to PhENN, since objets of that spatial frequency are never presented to it during training.