On the use of Machine Learning for solving Computational Imaging problems

George Barbastathis
Department of Mechanical Engineering and
Singapore-MIT Alliance for Research & Technology (SMART) Centre
Massachusetts Institute of Technology

Abstract

It has recently been recognized\(^1\) that compressed sensing, especially dictionaries\(^2\) and related methods, formally map to machine learning architectures, e.g. recurrent neural networks. This has led to rapid growth in algorithms and methods based on deep neural networks (but not only) for solving a variety of inverse and computational imaging problems.\(^3\) In this paper, we review these developments in the specific context of quantitative phase imaging and emphasizing the impact of object power spectral density and noise properties on the quality of the reconstructions.

1. INVERSION, OPTIMIZATION, AND LEARNING

Even though the terms “Computational Imaging” and “Inverse Problems” are distinct, there is significant overlap in terms of both intellectual content and research communities. In a sense, the narrowest definition is as follows: let \( f \) denote an unknown vector, \( H \) the forward operator of a physical system and \( N \) the noise model transforming \( f \) to a measurement \( g \) according to

\[
g = N \{ Hf \}. \tag{1}\]

This is called the Forward Problem. For example, in the case of pure additive Gaussian noise, the Forward Problem becomes

\[
g = Hf + n, \tag{2}\]

where \( n \) is a Gaussian random variable; whereas in the case of pure Poisson statistics, the Forward Problem becomes

\[
g = P \left( \frac{Hf}{I_0} \right), \tag{3}\]
where $P(.)$ denotes a Poisson random variable with arrival rate equal to the argument, and $I_0$ is a reference mean intensity. The Inverse Problem, naturally, is defined as

$$
\text{given } g, \text{ find } f.
$$

(4)

In this context, the measurement $g$ is sometimes called “raw image,” to differentiate from the final image, which is the solution to the Inverse Problem for $f$.

Eq. 4 can be made more specific, for example, in a maximum likelihood sense as

$$
\text{find the likeliest } f \text{ that produces } g \text{ according to (1)}.
$$

(5)

In the Inverse Problem sense, typically $f(x, y, z, \lambda)$ is an optical attribute of a physical object, for example the index of refraction as function of Cartesian coordinates $(x, y, z)$ and wavelength $\lambda$.

Computational Imaging is more encompassing in the sense that $f$ may represent generalized attributes of an object, e.g. its volume; or even semantic attributes like “is $f$ a cat?” The second type of question is also the domain of Computer Vision, although the latter tends to steer clear of the physical model (1); in other words, Computer Vision tends to assume that $g \approx f$ so that the emphasis is on interpreting the signal measured directly on the camera, rather than on the physics that produced this signal. The relative weight between physics and content in these overlapping communities is depicted in Figure 1.

In addition, Computational Imaging and Inverse Problems approaches have the optimization framework in common. Let us use $\hat{f}$ to denote the estimate of the
solution, thus acknowledging the estimate’s imperfection. One then defines the functional
\[ \Psi(f) = D \left( Hf, g \right) + \Phi(f), \] (6)
where \( D(\cdot, \cdot) \) is a distance metric and \( \Phi(\cdot) \), also known as the “regularizer,” expressing prior knowledge about the class of objects where \( f \) is drawn from. The estimate \( \hat{f} \) is then obtained as
\[ \hat{f} = \arg\min_f \Psi(f). \] (7)

In the case of purely additive Gaussian statistics, the maximum likelihood distance for (6) is \( L^2 \), and (7) becomes
\[ \hat{f} = \arg\min_f \left\{ \|Hf - g\|_2^2 + \Phi(f) \right\}. \] (8)

This formulation, it is generally agreed, was first proposed by Andrey Tikhonov\(^4\)–\(^6\) in the 1906’s. However, the origin of Inverse Problem formulations can be traced even earlier in the work of Norbert Wiener,\(^7\) who proposed the notion that the inverse estimate may be obtained through linear filtering as
\[ \hat{f} = \hat{W}g, \] (9)
where, for additive Gaussian noise, the maximum likelihood filter is obtained as
\[ \hat{W} = \arg\min_W \|Wg - f\|_2^2. \] (10)

The solution is, of course, the celebrated Wiener filter.\(^8\)–\(^10\)

It turns out that for white Gaussian noise and white objects (i.e., uncorrelated across pixels) the solutions to (8) and (9) are identical as long as the regularizer in (8) is chosen as
\[ \Phi(f) = \frac{E \left\{ \|n\|_2^2 \right\}}{E \left\{ \|f\|_2^2 \right\}} \|f\|_2^2 \] (11)
and \( E(.) \) denotes the expectation value. Tikhonov made the more general choice
\[ \Phi(f) = \kappa \|f\|_2^2, \] (12)
with the real-valued \( \kappa \), the regularization parameter, to be determined freely. In the Wiener filter, \( \kappa \) is chosen to equal the NSR, noise-to-signal ratio.
This analogy elucidates the significance of the regularizer as an exclusion principle, to be balanced against fidelity to the forward model. Suppose, as an example, that the regularization term in (8) is entirely omitted; then, the estimate $f$ has to come exclusively from the measurement $g$, which is noisy (see eq. 1). This means that if we minimized the term $\|Hf - g\|^2_2$ alone, i.e. direct deconvolution without any caveats, we would allow the noise to unduly influence $\hat{f}$, our solution estimate—a failure mode that belongs to the class of “overfitting,” as referred to in mainstream optimization theory. We add, therefore, the regularizer term $\Phi(f)$ to (8) to penalize, or exclude, estimates $\hat{f}$ that do not meet certain a priori known criteria. Tikhonov chose energy, i.e. $\|f\|^2_2$, as quantity to penalize, recognizing that deconvolution tends to amplify the noise energy and, hence, by trying to keeping the overall reconstructed energy low, we would be doing more damage to contributions from the noise rather than from the signal. However, it can be shown that the Tikhonov inverse (8) estimates are generally low-pass filtered—the price to pay for avoiding overfitting.

The modern regularizer design approach, resulting from early work by Candès and Tao, and Elad, is to discover an invertible transformation $S$ that sparsifies the signal in the sense that

$$s = Sf$$

has very few non-zero elements. Then, the inverse estimate $\hat{f}$ is obtained from a constrained minimization problem as

$$\hat{f} = S^{-1}\hat{s}, \quad \hat{s} = \arg\min_s \|s\|_1 \quad \text{subject to} \quad g - HS^{-1}s = 0.$$ 

In dictionaries, the transformation $S$ itself results from an optimization process that in effect minimizes $E\{\|Sf\|_0\}$. Thus, it can be said that in dictionary methods the sparsifier itself is learnt.

Either way, Eq. 14 results in an iterative optimization process including proximal gradient steps, since the $L^1$ norm is non-differentiable. Gregor and LeCun’s original insight was that the proximal gradient step in the iteration can be replaced by a (trained) Deep Neural Network (DNN) in the recurrent architecture of Figure 2(a). In a sense, then, the purpose of the training process of the DNN includes, at least partially, learning the priors for regularization. Instead of the recurrent architecture, it is often more practical to unfold the recurrence into a cascade, as shown in Figure 2(b); this stabilizes the training process but has the downside that more DNNs need to be trained. The unfolded architecture has been used in various forms for super-resolution and tomography. It has also been combined with the adversarial model for learning the priors through competition of the generator and discriminator networks.
Figure 2. Ways to obtain the inverse solution \( \hat{f} \) from the raw image \( g \) using Deep Neural Networks to learn the regularizer and/or the forward models. Figures (a–d) reprinted from Ref. 3.

In the notation of Figure 2(b), \( H^T \) is the conjugate of the forward operator, \( \hat{f}^{[m]} \) is the estimate after \( m \) DNN stages, and

\[
\mathcal{N} = 1 - H^T H
\]  

(15)

is referred to as “null space projection.” \( \mathcal{N} \) essentially is the residual what would be missed if one were to simply apply the conjugate operator as inverse, i.e. a naive deconvolution. Thus, in both Figures 2(a & b), the DNNs operate on the sum of the...
reconstruction at the present stage of the iteration plus the residual (except the first stage DNN1 in Fig. 2b).

To avoid the need to train multiple neural networks, it is possible to reduce the cascade to a single stage, as in Figure 2(c). The single pre-processor $H^*$ is the Approximant. For a linear forward operator $H$, one may simply choose $H^* = H^T$. However, this single-stage architecture has been used successfully with nonlinear inverse problems as well; in such situations, $H^*$ may be even a crude approximation to the inverse, as we discuss in Section 2 specifically for phase retrieval. To simplify further, the entire pre-processor may be dropped, leading to the simplest End-to-End architecture of Figure 2(d). In this case, the burden is on the DNN to learn both the forward operator and the priors governing the inverse problem; this overburdening leads to reduced robustness to noise, in particular.

It has also been shown\textsuperscript{20, 21} that improved performance at high spatial frequencies may be obtained by assigning the task of processing low and high spatial frequencies to two separate neural networks DNN-L and DNN-H, respectively, and recombining the outcomes $\hat{f}^{LF}$ and $\hat{f}^{HF}$ after processing the former through a synthesizer network DNN-S. This architecture is shown in Figure 2(e) with an Approximant (though it may also be implemented without one.) The three DNNs L, H, S are trained separately, as follows: the L network is trained with unfiltered examples $f$; the H network is trained with spatially high-pass filtered examples $f^{HF}$; and the S network is trained to match the unfiltered examples to the sum of its own output with the H output. The significance of using the LS architecture against the simpler ones when the raw images are highly noisy will be discussed in the next section in the specific context of phase retrieval; however, the principles are applicable to more general cases of noisy inverse problems.\textsuperscript{20}

2. QUANTITATIVE PHASE RETRIEVAL AND LEARNING

Let $\psi(x, y) = \exp \{ i f(x, y) \}$ denote a pure phase modulation on the electromagnetic field. In this paper, we limit our discussion of phase retrieval to the recovery of $\phi(x, y)$ from the noisy intensity

$$g(x, y) = N \left\{ \left| \int \int \exp \left\{ i f(x', y') + i \pi \frac{(x' - x)^2 + (y' - y)^2}{\lambda z} \right\} dx' dy' \right|^2 \right\}$$

after free-space propagation by distance $z$. Here, we use the paraxial scalar (Fresnel) propagation model, but others are possible as well. This scheme is sometimes referred

Proc. of SPIE Vol. 11249  112490B-6
to as “Coherent Diffraction Imaging” (CDI). In the standard formulation (1) of Section 1, we can see that the forward operator $H$ is the Fresnel propagation integral followed by a magnitude-square operation, i.e. $H$ is nonlinear. Standard approaches to this inverse problem are variants of the basic Gerchberg-Saxton-Fienup (GSF) algorithm$^{29-34}$ and the basic Conjugate Gradient method.$^{35,36}$

This inverse problem was solved for the first time, to our knowledge, by a DNN in the End-to-End architecture that we refer to as “Phase Extraction Neural Network” (PhENN.$^{17}$) A Spatial Light Modulator was used to generate ground truth phase objects $f$, propagated by fixed distance $z$ in free space, and the intensity raw images $g$ were captured on a CMOS camera. During the training phase of the DNN, the SLM was driven by phase images generated using samples from various databases, such as ImageNet;$^{37}$ the inputs to the DNN were the raw images $g$ and the weight optimization procedure minimized the $L^1$ distance (also referred to as Minimum Absolute Error, MAE) between the DNN outputs $\hat{f}$ and their corresponding true objects $f$. It was later found, albeit in the different context of imaging through diffuse media,$^{38}$ that the Negative Pearson Correlation Coefficient,$^{39,40}$ although derided as unsatisfactory metric when it comes to image similarity,$^{41}$ leads to better reconstruction quality when the forward operator $H$ is strongly ill-posed; hence, NPCC became the distance metric of choice in subsequent works.

Testing the trained PhENN consists of projecting on the SLM ground-truth phase images from the database that had never been used in training. PhENN’s output $\hat{f}$ was able to approximate well the true objects $f$, but without presenting any tangible advantage over the GSF algorithm under conditions of relatively ample signal power and numerical aperture. At about the same time, Prof. Ozcan’s group at UCLA showed that a DNN with the holographic back-propagation algorithm as pre-processor in the architecture Figure 2(c) was able to extract the phase from in-line interferograms.$^{42,43}$

PhENN’s performance in the End-to-End architecture significantly degrades with noise. Figure 3(a) is the phase object $f$ (ground-truth) projected onto the SLM as a test pattern, and in the experiment was attenuated down to approximately 1 photon/pixel flux at the camera plane. Figure 3(b) is the reconstruction with the End-to-End architecture, showing severe loss of high spatial frequencies. This can be attributed to the heavy burden on the DNN of learning both the physical model and the database priors. Since in the extremely highly noisy case the priors’ importance for resolving noise-induced ambiguity is accentuated, the learning process cannot cope and lossy reconstructions result.

The burden can be lowered by utilizing the pre-processor architecture, so that the
Figure 3. Quantitative phase retrieval by the Phase Extraction Neural Network (PhENN\cite{17,19,21}) with limited illumination of $\sim 1$ photon/pixel on average. (a) Ground-truth phase modulation $f$ projected on the Spatial Light Modulator. (b) Approximant $\hat{f}^{[0]}$ obtained as single iteration of the Gerchberg-Saxton algorithm. (c) Reconstruction by the End-to-End architecture, Figure 2d.\cite{17} (d) Reconstruction by the pre-processor architecture, Figure 2c with the signal (b) as Approximant.\cite{19} (e) Reconstruction by the Learning-to-Synthesize (LS) architecture of Figure 2e.\cite{21} Figures (a-d) are reprinted from Ref. 19 and Figure (e) from Ref. 21.

The physical model can be at least partially “pre-wired” into the learning architecture. The Approximant operator $H^*$ was a single iteration of the Gerchberg-Saxton algorithm,* in this case, and the resulting Approximant $f^{[0]}$ is shown in Figure 3(c). The reconstruction $\hat{f}$ produced by this improved architecture is shown in Figure 3(d).

One remaining limitation to the reconstruction quality in Figure 3(d) is the spatial frequency distribution of the training examples. For example, if a database of natural

*Running more iterations of GSF in this noisy case does not yield any discernible improvement in the reconstruction quality, neither is it helpful to use these higher iterates as Approximants in the scheme of Figure 2(c). Therefore, we opted to use the first iterate for faster computation.
Figure 4. Power spectral density $S(\nu)$, where $\nu = \sqrt{\nu_x^2 + \nu_y^2}$, for different DNN-based reconstruction approaches.

Images such as ImageNet is used for training, the training examples’ Power Spectral Density (PSD) $S(\nu_x, \nu_y)$ follows the well-known\textsuperscript{44} inverse-quadratic relationship

$$S(\nu_x, \nu_y) = \frac{1}{\nu_x^2 + \nu_y^2},$$

(17)

where $(\nu_x, \nu_y)$ denote the Cartesian spatial frequency coordinates. By the definition of PSD, (17) implies that higher spatial frequencies are less popular among the training examples and, hence, present more difficulty in getting learnt by PhENN. To combat this unfairness, \textit{ad hoc} spatial pre-filtering by the transfer function

$$T(\nu_x, \nu_y) = (\nu_x^2 + \nu_y^2)^{\frac{1}{2}},$$

(18)

flattening the PSD, was proposed.\textsuperscript{45} The LS method of Figure 2(e) is a more systematic approach toward balanced handling of spatial frequencies in the reconstruction. Indeed, as seen in Figure 3(e), the LS with Approximant reconstruction\textsuperscript{21} preserves more fine detail even in the noisy case of $\sim 1$ photon/pixel flux, compared to the simple Approximant, Figure 3(d).\textsuperscript{19} It is also worth noting that, while investigating the LS architecture,\textsuperscript{21} we studied more general transfer functions of the form

$$T(\nu_x, \nu_y) = (\nu_x^2 + \nu_y^2)^q,$$

(19)

but found that, at least for natural images satisfying (17), it is not worthwhile to significantly deviate from the spectrum flattening choice $q \sim 1/2$. 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Power spectral density $S(\nu)$, where $\nu = \sqrt{\nu_x^2 + \nu_y^2}$, for different DNN-based reconstruction approaches.}
\end{figure}
It is instructive to compare the PSDs of the different methods, obtained on an ensemble of $\sim 50$ test examples, as in Figure 4. The ground-truth PSD follows the inverse quadratic relationship (17). Reconstruction by DNN-L, which is essentially identical to the original PhENN, is faithful at low frequencies but attenuates the high frequencies. On the other hand, DNN-H attenuates low frequencies while recovering more high-frequency content—albeit not fully. DNN-S matches the performance of DNN-L at low frequencies and almost matches the performance of DNN-H at high frequencies.

This research was supported by the U. S. Intelligence Advanced Research Projects Activity (IARPA) and Singapore’s National Research Foundation (NRF) through the Campus for Research Excellence and Technological Enterprise (Create) programme.

REFERENCES

[1] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10, pp. 399–406, Omnipress, (USA), 2010.

[2] M. Elad and M. Aharon, “Image denoising via learned dictionaries and sparse representation,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, pp. 895–900, IEEE, 2006.

[3] G. Barbastathis, A. Ozcan, and Guohai Situ, “On the use of deep learning for computational imaging,” Optica 6(8), pp. 921–943, 2019.

[4] A. N. Tikhonov, “On the solution of ill-posed problems and the method of regularization,” Dokl. Acad. Nauk SSSR 151, pp. 501–504, 1963.

[5] A. N. Tikhonov, “On the regularization of ill-posed problems,” Dokl. Acad. Nauk SSSR 153, pp. 49–52, 1963.

[6] A. N. Tikhonov, “On the stability of algorithms for the solution of degenerate systems of linear algebraic equations,” Zh. Vychisl. Mat. i Mat. Fiz. 5, pp. 718–722, 1965.

[7] N. Wiener and E. Hopf, “Uber eine Klasse singulaerer Integralgleichungen,” Sitzungsberichte der Preussischen Akademie, Mathematisch-Physikalische Klasse 31, pp. 696–706, 1931.

[8] N. Levinson, “The Wiener RMS (root mean square) error criterion in filter design and prediction,” Journal of Mathematics and Physics XXV(4), pp. 261–278, 1946.

[9] N. Levinson, “Heuristic exposition of Wiener’s mathematical theory of prediction and filtering,” Journal of Mathematics and Physics XXVI(2), pp. 110–119, 1947.

[10] J. B. Lawrie and L. D. Abrahams, “A brief historical perspective of the Wiener-Hopf technique,” J. Eng. Math. 59, pp. 351–358, 2007.

[11] E. Candès and T. Tao, “Decoding by linear programming,” IEEE Trans. Inform. Theory 51(12), pp. 4203–4215, 2005.
[12] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on information theory 52(2), pp. 489–509, 2006.

[13] E. Candès and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?,” IEEE Trans. Inform. Theory 52(12), pp. 5406–5425, 2006.

[14] E. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” Comm. Pure Appl. Math. 59(8), pp. 1207–1223, 2006.

[15] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing over-complete dictionaries for sparse representation,” IEEE Transactions on signal processing 54(11), p. 4311, 2006.

[16] M. Mardani, Enhao Gong, J. Y. Cheng, S. Vasanawala, G. Zaharchuk, M. Alley, N. Thakur, Song Han, W. Daly, J. M. Pauly, and Lei Xing, “Deep generative adversarial networks for compressed sensing automates MRI.” arXiv:1706.00051, 5 2017.

[17] A. Sinha, Justin Lee, Shuai Li, and G. Barbastathis, “Lensless computational imaging through deep learning,” Optica 4, pp. 1117–1125, 2017.

[18] Y. Rivenson, Z. Gorocs, H. Gunaydin, Yibo Zhang, Hongda Wang, and A. Ozcan, “Deep learning microscopy,” Optica 4, pp. 1437–1443, 2017.

[19] A. Goy, K. Arthur, Shuai Li, and G. Barbastathis, “Low photon count phase retrieval using deep learning,” Phys. Rev. Lett. 121, p. 243902, 2018.

[20] Deng Mo, Li Shuai, and G. Barbastathis, “Learning to synthesize: splitting and recombining low and high spatial frequencies for image recovery.” arXiv:1811.07945, November 2018.

[21] Deng Mo, Li Shuai, A. Goy, Kang Iksung, and G. Barbastathis, “Learning to synthesize: robust phase retrieval at low photon counts.” Light: Science and Applications, accepted for publication.

[22] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, “A l1-unified variational framework for image restoration,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 3024, pp. 1–13, 2004.

[23] J. Eckstein and D. P. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators,” Math. Program. 55, pp. 293–318, 1992.

[24] Chao Dong, Chen Loy, Kaiming He, and Xiaou Tang, “Learning a deep convolutional neural network for image super-resolution,” in European Conference on Computer Vision (ECCV) / Lecture Notes on Computer Science Part IV, 8692, pp. 184–199, 2014.

[25] Chao Dong, Chen Loy, Kaiming He, and Xiaou Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Machine Intell. 38(2), pp. 295–307, 2015.
[26] Kyong Hwan Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” *IEEE Trans. Image Process.* **26**(9), pp. 4509–4522, 2017.

[27] Dongwook Lee, Jaejun Yoo, and Jong Chul Ye, “Deep residual learning for compressed sensing MRI,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 15–18, April 2017.

[28] M. Mardani, H. Monajemi, V. Papyan, S. Vasanawala, D. Donoho, and J. Pauly, “Recurrent generative residual networks for proximal learning and automated compressive image recovery.” arXiv:1711.10046, 11 2017.

[29] R. W. Gerchberg and W. O. Saxton, “Phase determination from image and diffraction plane pictures in electron-microscope,” *Optik* **34**(3), pp. 275–284, 1971.

[30] R. W. Gerchberg and W. O. Saxton, “Practical algorithm for the determination of phase from image and diffraction plane pictures,” *Optik* **35**(2), pp. 237–246, 1972.

[31] J. R. Fienup, “Reconstruction of an object from the modulus of its Fourier transform,” *Optics letters* **3**(1), pp. 27–29, 1978.

[32] J. Fienup and C. Wackerman, “Phase-retrieval stagnation problems and solutions,” *JOSA A* **3**(11), pp. 1897–1907, 1986.

[33] H. H. Bauschke, P. L. Combettes, and D. R. Luke, “Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization,” *JOSA A* **19**(7), pp. 1334–1345, 2002.

[34] V. Elser, “Phase retrieval by iterated projections,” *J. Opt. Soc. Am. A* **20**(1), pp. 40–55, 2003.

[35] M. R. Hestenes and E. Stiefel, “Method of conjugate gradients for solving linear systems,” *Journal of Research of the National Bureau of Standard* **49**(6), pp. 409–436, 1952.

[36] J. R. Fienup, “Phase retrieval algorithms: a comparison,” *Appl. Opt.* **21**(15), pp. 2758–2769, 1982.

[37] O. Russakovsky, Jia Deng, Hao Su, J. Krause, S. Satheesh, Sean Ma, Zhiheng Huang, A. Karpathy, A. Khosla, and M. Bernstein, “ImageNet large scale visual recognition challenge,” *Int. J. Comput. Vis.* **115**(3), pp. 211–252, 2015.

[38] Shuai Li, Mo Deng, Justin Lee, A. Sinha, and G. Barbastathis, “Imaging through glass diffusers using densely connected convolutional networks,” *Optica* **5**(7), pp. 803–813, 2018.

[39] K. Pearson, “Contributions to the mathematical theory of evolution. note on reproductive selection,” *Proc. Royal Society* **59**, pp. 300–305, 1896.

[40] J. L. Rodgers and W. A. Nicewander, “Thirteen ways to look at the correlation coefficient,” *The American Statistician* **42**(1), pp. 59–66, 1988.

[41] Eugene K. Yen and R. G. Johnston, “The ineffectiveness of the correlation coefficient for image comparisons,” Tech. Rep. LA-UR-96-2474, Los Alamos National Laboratory, 1996.
[42] Y. Rivenson, Yibo Zhang, H. Gunaydin, Da Teng, and A. Ozcan, “Phase recovery and holographic image reconstruction using deep learning in neural networks,” *Light: Science and Applications* 7, p. 17141, 2018.

[43] Yichen Wu, Y. Rivenson, Yibo Zhang, Zhensong Wei, H. Gunaydin, Xing Lin, and A. Ozcan, “Extended depth-of-field in holographic image reconstruction using deep learning based auto-focusing and phase-recovery,” *Optica* 5(6), pp. 704–710, 2018.

[44] A. Van der Schaaf and J. H. van Hateren, “Modelling the power spectra of natural images: statistics and information,” *Vision Research* 36(17), pp. 2759–2770, 1996.

[45] Shuai Li and G. Barbastathis, “Spectral pre-modulation of training examples enhances the spatial resolution of the phase extraction neural network (PhENN),” *Opt. Express* 26(22), pp. 29340–29352, 2018.