On the use of deep learning for computational imaging

Barbastathis, George

George Barbastathis, "On the use of deep learning for computational imaging," Proc. SPIE 11463, Optical Trapping and Optical Micromanipulation XVII, 114631L (2 September 2020); doi: 10.1117/12.2571322

SPIE.

Event: SPIE Nanoscience + Engineering, 2020, Online Only

On the use of Deep Learning for Computational Imaging

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

Abstract

Deep learning[1] has emerged as a class of optimization algorithms proven to be effective for a variety of inference and decision tasks. Similar algorithms, with appropriate modifications, have also been widely adopted for computational imaging.[2] Here, we review the basic tenets of deep learning and computational imaging, and overview recent progress in two applications: super resolution and phase retrieval.

1. COMPUTATIONAL IMAGING

Let f denote an unknown vector, which we seek to establish from optical measurements. We may think of f as the spatial description of a physical object, e.g. the index of refraction $n(\mathbf{r})$ as function of Cartesian coordinates $\mathbf{r}$ sampled appropriately. The optical system forms an intensity pattern, which is also sampled to a vector g and we refer to as the raw image measured on the digital camera. In the absence of noise, the relationship between the object and raw image is denoted as

$$g = Hf, \tag{1}$$

where H is the imaging system's forward operator. H includes the illumination model; the model of light scattering by the object; the propagation to the detector, including the transfer function of the optics in the system; and the detector's photoelectric conversion model. In actuality, (1) is stochastic due to the particle nature of light and the detector's noise statistics; however, to avoid overcomplicating, we will limit the discussion in this paper to the deterministic description.

It is worthwhile to note that the choice of f as being spatially faithful representation of the object isn't the only one. It is possible to structure f as representing higher level questions about the object; for example, we could require each element in f to represent "the 3D coordinates from the scene where a cat is present." This is

Optical Trapping and Optical Micromanipulation XVII, edited by Kishan Dholakia, Gabriel C. Spalding, Proc. of SPIE Vol. 11463, 114631L · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2571322

the broadest definition of Computational Imaging, but it presents the obvious problem that H, in almost all such cases, cannot be expressed explicitly. Recognition of hand-written characters from the output of multi-mode fibers where the spatial information is highly diffuse[3] is a good example of this approach; however, in the interest of confining the content of the present paper somewhat, we will limit discussion to spatially faithful f's.

There are several strategies to invert (1) and obtain an estimate $\hat{f}$ of the object. We discuss them briefly here, together with their properties and caveats.

1.1 The direct inverse

If the matrix H is non-singular, then direct inversion of (1) yields

$$\hat{f} = H^{-1}g. \tag{2}$$

If H represents a spatially invariant optical system, then it is appropriate to refer to the direct inverse as "deconvolution." Unfortunately, this approach seldom yields good results, because the effect of the noise on the inverse estimate is amplified.

To see why, consider a very simple 2×2 linear imaging system with

$$H = \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix}, \quad 0 < \epsilon < 1, \tag{3}$$

and additive noise, *i.e.* $g = Hf + n$, where n is a random process whose statistics are irrelevant to the present discussion. We may see immediately that the direct inverse yields

$$\hat{f} = f + \frac{1}{1-\epsilon^2} \begin{pmatrix} 1 & -\epsilon \\ -\epsilon & 1 \end{pmatrix} n. \tag{4}$$

If the "cross-talk" term ϵ is near 1, e.g. $\epsilon \sim 0.9$, the noise gain is as high as ~ 5. More generally, if μ_n are the eigenvalues of H, then a simple linear algebra argument shows that the noise amplification may be as high as

$$\sim \left(\prod_n \mu_n \right)^{-1}.$$

Therefore, forward operators whose majority of eigenvalues are significantly smaller than the maximum eigenvalue are referred to as "ill-conditioned." The forward operators of diffraction-limited optical systems become invariably ill-conditioned if the attempt is made to sample the object at a spacing finer than the so-called Rayleigh resolution limit $\sim \lambda/(\mathrm{NA})$, where NA is the numerical aperture (more on this in section 2.1.) In the presence of aberrations, ill-conditioning becomes worse.

1.2 Wiener filter and Tikhonov inverse

In the Wiener filter[4–7] approach the estimated inverse is obtained through a linear filter as

$$\hat{f} = \hat{W} g, \quad \text{where} \quad \hat{W} = \underset{W}{\text{argmin}} \; \|Wg - f\|_2^2 . \tag{5}$$

The optimal filter $\hat{W}$ is required to minimize the quadratic (L^2) error between the reconstruction $\hat{W}g$ and the true object f. Additional assumptions are that the noise process n is zero-mean and independent of the object f. Then, solving the resulting quadratic minimization problem yields

$$\hat{W} = C_{\text{ff}} \, H^\dagger \left(H \, C_{\text{ff}} \, H^\dagger + C_{\text{nn}} \right)^{-1}, \tag{6}$$

where C_{ff} and C_{nn} are the correlation operators of the signal and noise, respectively, and $\dagger$ denotes the transpose. In the special case when the correlation operators reduce to $P_0 I$ and $N_0 I$, respectively, where P_0 is the average signal power, N_0 is the average noise power,* and I is the unitary operator; then

$$\hat{W} = H^\dagger \left(H H^\dagger + \frac{I}{(\text{SNR})} \right)^{-1}, \quad \text{where} \quad (\text{SNR}) \equiv \frac{P_0}{N_0}. \tag{7}$$

It can be seen that in the limit $N_0 \ll P_0$ the above expression reduces to $\hat{W} \approx H^{-1}$, *i.e.* the direct inverse. If the noise is non-negligible, then $\hat{W}$ according to (7) is better than H^{-1} because it is designed to minimize the effect of the noise in the reconstruction, at least in the quadratic sense (see eq. 5). However, the linear inverse filter (7) is known to also introduce blur in the reconstructions, as the price to pay for compensating noise amplification.[8]

Interestingly, the same linear inverse filter is obtained from a different principle, originally developed by Tikhonov,[9–11]

$$\hat{f} = \underset{f}{\text{argmin}} \left\{ \; \|Hf - g\|_2^2 + \frac{\kappa}{2} \|f\|_2^2 \; \right\}, \quad \text{with} \quad \kappa = \frac{1}{(\text{SNR})}. \tag{8}$$

The second term in the above functional is called the *regularizer* and κ the regularization parameter, because they are meant to contain (regularize) the energy of the noise that the direct inverse would have otherwise amplified. This is why the regularizer is the L^2 norm, *i.e.* the energy of the object's estimate in the functional. It is also worthwhile to note that the Kalman filter[12] derivation also follows from the Wiener filter principle.

*This implies that both signal and noise are white noise-like. Such an assumption is seldom realistic for typical objects of imaging systems, but it does apply to communication signals that have been pre-compressed; it also leads to a neat result.

1.3 Sparsity-regularized inverse

Tikhonov's functional (8) is generalized as

$$\hat{f} = \underset{f}{\operatorname{argmin}} \left\{ \ \|Hf - g\|_2^2 + \Phi(f) \ \right\}, \tag{9}$$

where $\Phi(f)$ may be designed to still contain the effects of noise in the reconstruction while ameliorating the blur problems of the Wiener-Tikhonov approaches. A particularly effective choice of regularizers results from transforming the object as

$$s = Sf, \tag{10}$$

such that S is invertible and the new representation s is *sparse, i.e.* it contains few non-zero elements. It may then be shown[13] that an equivalent resilient solution is obtained from

$$\hat{s} = \underset{s}{\operatorname{argmin}} \left\{ \ \|As - g\|_2^2 + \alpha \|s\|_1 \ \right\}, \tag{11}$$

where $A = HS^{-1}$, α is another regularization parameter and we can see that the regularizer function is now the L^1 norm. The basic solution to (11) is the iterative shrinkage-thresholding algorithm[14] (ISTA)

$$\hat{s}^{[m+1]} = \mathcal{P}_\varepsilon \left\{ \left(I - \alpha A^\dagger A \right) \hat{s}^{[m]} + \alpha A^\dagger g \right\}, \quad m = 1, 2, \ldots \tag{12}$$

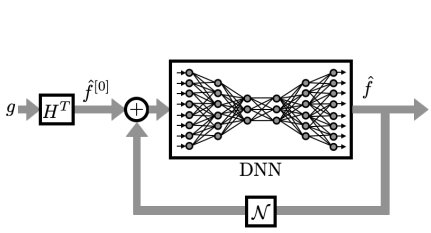
After the iteration has terminated, the object estimate is obtained as $\hat{f} = S^{-1}\hat{s}$. Here, $\mathcal{P}_\varepsilon$ represents the proximal gradient[15,16] (soft thresholding) operator

$$\mathcal{P}_\varepsilon(u) = \begin{cases} u + \varepsilon, & u < -\varepsilon; \\ 0, & -\varepsilon < u < \varepsilon; \\ u - \varepsilon, & u > \varepsilon. \end{cases} \tag{13}$$

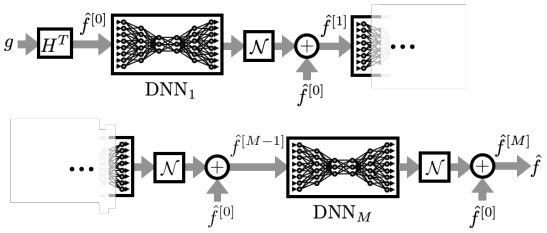
The purpose of soft thresholding is to get around the problem that the L^1 norm regularizer in the functional (11) is non-differentiable. Faster and more numerically stable versions of ISTA have been developed, e.g. TwIST.[17] Alternatively, the problem may instead be formulated as one of *constrained* minimization[18–20]

$$\hat{s} = \underset{s}{\operatorname{argmin}} \ \|s\|_1 \quad \text{subject to} \quad g - As = 0. \tag{14}$$

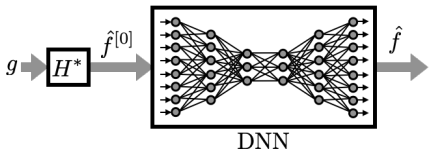
The *sparsifying* transform S is typically chosen as a redundant basis,[13,18–20] e.g. wavelets.[21] Another common regularizer that promotes sharpness in the reconstructions is Total Variation (TV),[19] which requires a separate computation for the proximal gradient (not given here.) Alternatively, S may be *learnt* from data that are statistically similar to typical objects, if available; it is then referred to as a *dictionary.*[22,23]



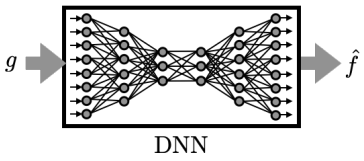
(a) Recurrent architecture.[2,24,25]



(b) Unfolded (cascaded) architecture.[2,24,25]



(c) Pre-processor or "Approximant" architecture.[2,27,28]



(d) End-to-End architecture.[2,26]

Figure 1. Using neural networks to regularize and invert the problem (1). Reprinted from Ref. 2.

1.4 Inversion by neural networks

As an extension of dictionaries, it is natural to suggest replacing the proximal gradient operator itself with a learning algorithm, whose purpose is to learn the correct regularization principle directly from examples on the original object f, without necessarily sparsifying. In that case, (12) becomes

$$\hat{f}^{[m+1]} = \mathrm{DNN}\left\{ \left(I - \alpha H^\dagger H\right) \hat{f}^{[m]} + \alpha H^\dagger g \right\}, \quad m = 1, 2, \ldots, \tag{15}$$

where now $\mathrm{DNN}\{\cdot\}$ represents the input-output relationship of a deep neural network. Among the first uses of this approach were for tomography[24,25,29] but its origins are in an insight by Gregor and LeCun.[30] Equation 15 suggests a recurrent architecture, as shown in Figure 1(a), with m acting as the time step in the dynamical process. In practice, the recurrence is typically unfolded to a cascade of DNNs for better numerical stability, as shown in Figure 1(b). However, the number of parameters to be learnt (the weights of the DNNs) also increases. A generative adversarial approach, where the recurrent scheme acts as generator, has also been implemented.[31] The operator

$$\mathcal{N} = I - H^\dagger H \tag{16}$$

is the *null space projection:* essentially, what the direct inverse of section 1.1 would have missed, if it were naïvely applied.

Figures 1(c-d) are two simpler alternatives. In Figure 1(c) we have a single stage only, with a pre-processor H^*, which we call the Approximant, resulting from an approximate inversion of the original forward operator H. The Approximant's role is to produce an initial estimate $\hat{f}^{[0]}$, which may be crude and prone to noise artifacts yet the subsequent DNN learns how to improve toward a final reconstruction $\hat{f}$ of decent quality. Finally, in the End-to-End architecture of Figure 1(d) we discard the Approximant altogether and train the DNN to receive the raw intensity g as input and produce the estimate $\hat{f}$ as output. Naturally, the learning burden is higher in this last case, because the physical model H and the priors need to both be encoded in the DNN input-output relationship through the training process; because of that, the simple architecture of Figure 1(d) tends to not behave well when the raw images are highly noisy.

2. EXAMPLES

To illustrate the formulation of inverse problems and the use of the machine learning architectures of section 1.4, we develop two examples: spatially incoherent imaging of intensity objects; and spatially coherent imaging of phase objects. We will use the scalar and paraxial approximations; extension to more accurate models is straightforward albeit arduous.

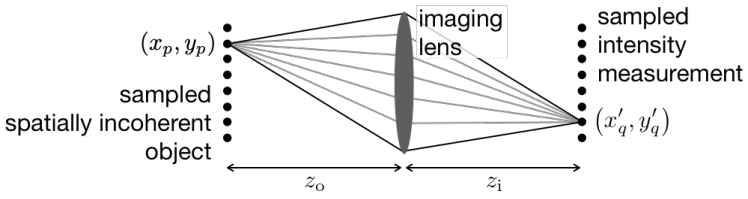


Figure 2. Simple spatially incoherent imaging system with a single imaging lens of focal length f_0.

2.1 Spatially incoherent imaging

The geometry for constructing this problem is in Figure 2. The object-lens distance z_o and lens-image plane distance z_i satisfy $1/z_\mathrm{o} + 1/z_\mathrm{i} = 1/f_0$ and without loss of generality we also assume $z_\mathrm{o} = z_\mathrm{i}$, *i.e.* unit magnification.

The object consists of uncorrelated monochromatic point emitters of wavelength λ, regularly spaced on a $P_x \times P_y$ grid with coordinates (x_p, y_p), $p = 1, \ldots, P = P_x P_y$. The user wishes to retrieve the strengths $I_\mathrm{in}(x_p, y_p)$ of these emitters by sampling the intensity $I_\mathrm{out}(x'_q, y'_q)$ on the camera plane at pixels regularly spaced on a $Q_x \times Q_y$

grid with coordinates (x'_q, y'_q), $q = 1, \ldots, Q = Q_x Q_y$. If $Q > P$ the problem is oversampled, also referred to as overdetermined. If $Q < P$, it is undersampled and the recovery of f from g is sometimes referred to as "super resolution," especially in the community of machine vision.

In the generic notation of (1), the object is represented as $f = \left\{ I_{\text{in}}\left(x_p, y_p\right) \right\}_{p=1,\ldots,P}$, while the measurement is represented as $g = \left\{ I_{\text{out}}\left(x'_q, y'_q\right) \right\}_{q=1,\ldots,Q}$. It is well known[32,33] that the intensity at the output plane of a monochromatic spatially incoherent system is linearly related to the intensity at the input plane. If the optical system is shift invariant the relationship becomes a convolution and if, moreover, it is diffraction-limited with a circular aperture of maximal angular admittance equal to NA (numerical aperture), then is expressed simply as

$$I_{\text{out}}\left(x'_q, y'_q\right) = \sum_p \text{jinc}^2\left(\frac{\rho_{qp}}{b}\right) I_{\text{in}}\left(x_p, y_p\right), \quad \text{where} \tag{17}$$

$$\rho_{qp} \equiv \left\{ \left(x'_q - x_p\right)^2 + \left(y'_q - y_p\right)^2 \right\}^{1/2}, \quad \text{jinc}\left(u\right) \equiv \frac{\text{J}_1(u)}{2u},$$

$\text{J}_1(.)$ is the Bessel function of the first kind and 1^{st} order, and b is the main half-lobe of the Airy diffraction spot $b = \lambda/(\text{NA})$. The simple linear 2×2 operator mentioned in section 1.1 is a special case with $Q = P = 2$ and $\epsilon = \text{jinc}^2\left(\rho_{12}/b\right)$.

If the main lobe width of the Airy function $1.22b \underset{\sim}{>} \min_{p,q} \rho_{qp}$ then H creates in g a blurred version of the object f. This simplified expression for the blur is valid for the diffraction-limited paraxial case. More generally, a linear equation of the form (17) remains valid but the kernel H is influenced by additional factors such as aberrations, motion blur, etc. and it may also become space variant. The situation where the sampling distance $\Delta\rho$ at the object plane is less than b is also referred to, confusingly, as "super resolution," especially in the optical imaging community, even though it is clearly different than the earlier mention of the term.

The undersampling super resolution problem $Q < P$ was one of the earliest instances of using a deep neural network to reconstruct the object at a denser sampling grid,[34,35] using an End-to-End architecture similar to the one in Figure 1(d). The cascaded architecture of Figure 1(b) was later considered, in a generative adversarial scheme, for the same problem.[31,36]

One of the earliest attempts to solve the super resolution problem $\Delta\rho < b$ was by training a neural network to receive images by a low-NA objective as inputs and reproduce images of high-NA quality as outputs.[27,37] In this approach, training of the

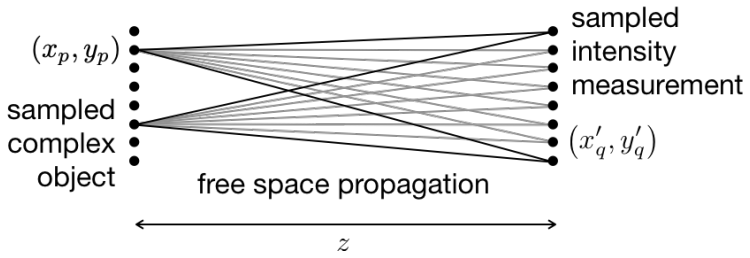


Figure 3. Lensless spatially coherent imaging system with free-space propagation distance z. The presence of a reference beam is optional (see discussion in Example 2.)

DNN takes place experimentally using images obtained by the two objective lenses, high-NA and low-NA. An adversarial training approach has also been developed.[38]

2.2 Lensless quantitative phase imaging

The geometry is shown in Figure 3. Let us suppose that the unknown object is a transparent plate, with uniform index of refraction and variable thickness; and that the user wishes to retrieve the phase delay ϕ at pixels regularly spaced on a $P_x \times P_y$ grid with coordinates (x_p, y_p), $p = 1, \ldots, P = P_x P_y$. Then $f = \left\{\phi\left(x_p, y_p\right)\right\}_{p=1,\ldots,P}$. The plate is illuminated on-axis by a monochromatic plane wave at wavelength λ. Under the additional scalar and paraxial approximation assumptions, the field immediately after the plate is

$$\psi_{\text{in}}(x, y) = \exp\left\{i\phi(x, y)\right\}. \tag{18}$$

After free-space propagation by distance z, the intensity is sampled at pixels regularly spaced on a $Q_x \times Q_y$ grid with coordinates (x'_q, y'_q), $q = 1, \ldots, Q = Q_x Q_y$. In discrete notation, the intensity measurement is

$$g_q = \left|\psi_{\text{out}}\left(x'_q, y'_q\right)\right|^2 = \left|\sum_p \exp\left\{i\phi(x_p, y_p) + i\pi \frac{\left(x'_q - x_p\right)^2 + \left(y'_q - y_p\right)^2}{\lambda z}\right\}\right|^2. \tag{19}$$

Here, we have neglected the effect of spatial integration over the detector pixel area. The measurement vector then is $g = \left\{g_q\right\}_{q=1,\ldots,Q}$. Clearly, the operator H is nonlinear in this example. In the community of x-ray optics, this method is referred to as Coherent Diffraction Imaging (CDI).

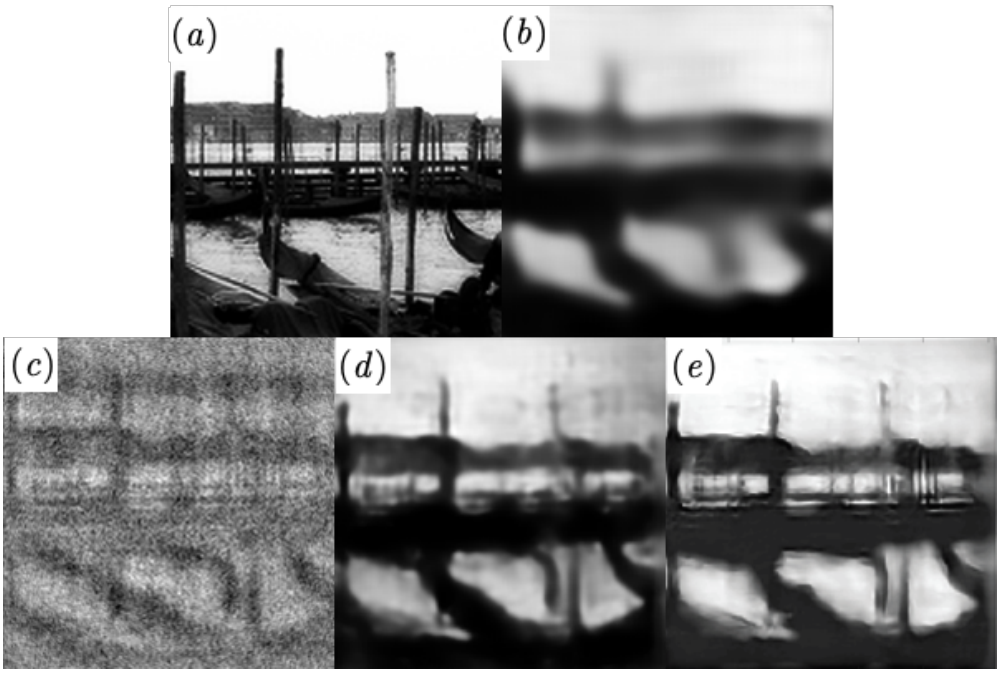


Figure 4. Quantitative phase retrieval in highly noisy conditions, $\sim$ 1 photon/pixel on average. (PhENN[26,28,46]) with limited illumination of *(a)* Ground-truth phase modulation f projected on the Spatial Light Modulator. *(b)* Approximant $\hat{f}^{[0]}$ obtained as single iteration of the Gerchberg-Saxton algorithm. *(c)* Reconstruction by the End-to-End architecture, Figure 1d.[26] *(d)* Reconstruction by the Approximant architecture, Figure 1c with the signal (b) as Approximant.[28] *(e)* Reconstruction by the Learnting-to-Synthesize (LS) architecture which processes low and high spatial frequencies separately.[46] Figures (a-d) are reprinted from Ref. 28 and Figure (e) from Ref. 46.

Classical iterative solutions to this problem are the Gerchberg-Saxton algorithm;[39] the input-output algorithm and its variants proposed by Fienup;[39–43] and the gradient descent[44,45] with its well-known variants such as steepest descent and conjugate gradient algorithms. Prior knowledge about the object such as positivity and support constraints are often imposed at each iteration in the object and measurement domains.

Figure 4 shows CDI reconstruction results obtained with DNNs on the phase retrieval problem under noisy experimental conditions ($\sim$ 1 photon per pixel on the detector.) The End-to-End architecture,[26] Figure 4(b), performs worst, because the highly noisy problem requires strong priors to be learnt and the burden on the DNN to learn the priors as well as the physical model becomes too high. Using the Gerchberg-

Saxton reconstruction of Figure 4(b) as Approximant improves significantly,[28] even though the Approximant itself is quite noisy. Even higher fidelity reconstructions are obtained by splitting the reconstruction job into two separate channels for the high and low spatial frequencies, and training two DNNs to process them respectively plus a third DNN to recombine (synthesize) them into a final reconstruction that has even quality over all spatial frequency bands.[46]

The free-space propagation (CDI) approach is just one of the widely used methods for phase retrieval. DNN reconstructions have been successfully demonstrated in, for example, holographic,[47,48] ptychographic,[49] and coherent modulation imaging (CMI).[50] In these cases, the forward operators H are still nonlinear but they benefit from modulation by the reference beam (in holography) or the wavefront (in ptychography and CMI).

3. CONCLUSIONS AND OUTLOOK

A persistent concern in any engineering enterprise that utilizes machine learning is verifiability. Neural networks typically contain millions of parameters (the interconnection weights) which are trained through an optimization process, typically stochastic gradient descent. Unfortunately, there is no guarantee that this optimization is convex. Worse, there is no good way to ascertain that the examples chosen for training are statistically representative of the intended class and, thus, result in correct learned priors. One way to make progress in this difficult problem is to investigate the relative importance of physical models and priors, such as dataset complexity, as they are "hard-wired" into the DNN during training.[51] Better ways to integrate neural networks into first principle physical models[52] will certainly accelerate the pace of this desirable evolution.

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] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature* **521**, pp. 436–444, 2015.

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

[3] N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, "Learning to see through multimode fibers," *Optica* **5**, pp. 960–966, Aug 2018.

[4] N. Wiener and E. Hopf, "Uber eine Klasse singulaerer Integralgeichungen," *Sitzungsberichte der Preussischen Akademie, Mathematisch-Physikalische Klasse* **31**, pp. 696–706, 1931.

[5] 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.

[6] N. Levinson, "Heuristic exposition of Wiener's mathematical theory of prediction and filtering," *Journal of Mathematics and Physics* **XXVI**(2), pp. 110–119, 1947.

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

[8] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*, Institute of Physics, 1998.

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

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

[11] 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.

[12] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME* **82(D)**, pp. 35–45, 1960.

[13] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory* **51**(12), pp. 4203–4215, 2005.

[14] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.* **57**(11), pp. 1413–1457, 2004.

[15] 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.

[16] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, "A l_1-unified variational framework for image restoration," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, **3024**, pp. 1–13, 2004.

[17] J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image processing* **16**(12), pp. 2992–3004, 2007.

[18] 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.

[19] 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.

[20] 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.

[21] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.* **41**(7), pp. 909–996, 1988.

[22] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing* **54**(11), p. 4311, 2006.

[23] 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.

[24] 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.

[25] 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.

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

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

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

[29] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems* **33**, p. 124007, 2017.

[30] 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.

[31] 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.

[32] G. B. Airy, "On the diffraction of an object-glass with circular aperture," *Transactions of the Cambridge Philosophical Society* **5**(III), pp. 283–291, 1834.

[33] H. H. Hopkins, "The concept of partial coherence in optics," *Proc. Roy. Soc. A* **208**, pp. 263–277, 1951.

[34] Chao Dong, Chen Loy, Kaiming He, and Xiaoou 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.

[35] Chao Dong, Chen Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Machine Intell.* **38**(2), pp. 295–307, 2015.

[36] 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.

[37] Hongda Wang, Y. Rivenson, Zhensong Wei, H. Gunaydin, L. Bentolila, and A. Ozcan, "Deep learning achieves super-resolution in fluorescence microscopy." bioRxiv, https://doi.org/10.1101/309641, 2018.

[38] Tairan Liu, K. de Haan, Y. Rivenson, Zhensong Wei, Xin Zeng, Yibo Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems." arXiv:1810.06611, 2018.

[39] 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.

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

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

[42] 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.

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

[44] 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.

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

[46] Mo Deng, Shuai Li, A. Goy, Iksung Kang, and G. Barbastathis, "Learning to synthesize: robust phase retrieval at low photon counts," *Light: Science and Applications* **9**(36), 2020.

[47] Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light: Sci. Appl.* **7**, p. 17141, 2018.

[48] 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.

[49] Shaowei Jiang, Kaikai Guo, Jun Liao, and Guoan Zheng, "Solving fourier ptychographic imaging problems via neural network modeling and TensorFlow," *Biomed. Opt. Express* **9**(7), pp. 3306–3319, 2018.

[50] Iksung Kang, Fucai Zhang, and G. Barbastathis, "Phase extraction neural network (PhENN) with coherent modulation imaging (CMI) for phase retrieval at low photon counts," *Opt. Express* **28**(15), pp. 21578–21600, 2020.

[51] Mo Deng, Shuai Li, Zhengyun Zhang, Iksung Kang, Nicholas X. Fang, and G. Barbastathis, "On the interplay between physical and content priors in deep learning for computational imaging," *Opt. Express* **28**(16), pp. 24512–24170, 2020.

[52] C. Rackauckas, Yingbo Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, and A. Ramadhan, "Universal differential equations for scientific machine learning." arXiv:2001.04385, January 2020.