Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

SHIRAZ S. KADERUPPAN[1,2], **EUGENE WAI LEONG WONG**[1,2,3], **(Member, IEEE),**
ANURAG SHARMA[1,2,3], **(Member, IEEE), AND WAI LOK WOO**[2,4], **(Senior Member, IEEE)**

[1]Newcastle Research and Innovation Institute Pte Ltd, (NewRIIS), Devan Nair Institute for Employment and Employability, Singapore 609607
[2]Faculty of Science, Agriculture and Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.
[3]Newcastle University in Singapore, SIT Building @ Nanyang Polytechnic, Singapore 567739
[4]Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K.

Corresponding author: Anurag Sharma (anurag.sharma@newcastle.ac.uk)

ABSTRACT The field of optical *nanoscopy*, a paradigm referring to the recent cutting-edge developments aimed at surpassing the widely acknowledged 200nm-diffraction limit in traditional optical microscopy, has gained recent prominence & traction in the 21st century. Numerous optical implementations allowing for a new frontier in traditional confocal laser scanning fluorescence microscopy to be explored (termed *super-resolution fluorescence microscopy*) have been realized through the development of techniques such as stimulated emission and depletion (STED) microscopy, photoactivated localization microscopy (PALM) and stochastic optical reconstruction microscopy (STORM), amongst others. Nonetheless, it would be apt to mention at this juncture that optical nanoscopy has been explored since the mid-late 20th century, through several computational techniques such as deblurring and deconvolution algorithms. In this review, we take a step back in the field, evaluating the various *in silico* methods used to achieve optical nanoscopy today, ranging from traditional deconvolution algorithms (such as the Nearest Neighbors algorithm) to the latest developments in the field of computational nanoscopy, founded on artificial intelligence (AI). An insight is provided into some of the commercial applications of AI-based super-resolution imaging, prior to delving into the potentially promising future implications of computational nanoscopy. This is facilitated by recent advancements in the field of AI, deep learning (DL) and convolutional neural network (CNN) architectures, coupled with the growing size of data sources and rapid improvements in computing hardware, such as multi-core CPUs & GPUs, low-latency RAM and hard-drive capacities.

INDEX TERMS Super-resolution microscopy, computational nanoscopy, high-resolution microscopical imaging, optical microscopy, deep learning.

I. INTRODUCTION

Optical microscopy has proven to be a ubiquitous tool and a gold standard for biological, geological and materials science research, as well as industrial quality control processes. Nonetheless, traditional optical microscopy suffers from numerous limitations, including (but not being restricted to) blurring/haze, lateral and axial resolution limitations [1], poor signal-noise ratio (SNR) and poor contrast at higher magnifications. In particular, an oft-cited equation describing the inherent resolution limitation faced by the compound optical microscope is the Abbe equation (proposed by Ernst

Abbe in 1873 [1]), which may be defined as follows:

$$\text{Abbe Lateral Resolution}, \quad d_{x,y} = \frac{\lambda}{2n\sin\theta} \quad (1)$$

$$\text{Abbe Axial Resolution}, \quad d_z = \frac{2\lambda}{(n\sin\theta)^2} \quad (2)$$

where λ is the wavelength of the irradiating light, n is the refractive index of the imaging medium and θ is the aperture angle of the light cone [1]. The product $n\sin\theta$ is also sometimes defined as the numerical aperture (NA) of the lens.

In 1896, Rayleigh extrapolated (1) to include an additional contribution by the condenser optics, thereby contrasting episcopic (traditionally epifluorescence) microscopy with
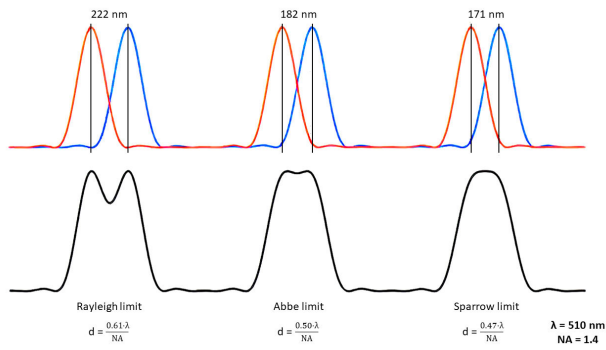
The associate editor coordinating the review of this manuscript and approving it for publication was Chao Zuo.

IEEE Access·

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

FIGURE 1. The 3 widely-utilized formulae (i.e. Rayleigh, Sparrow and Abbe) for resolution computation. The red and blue curves represent the individual intensity variations at different points in a specimen where the vertical (y-) axis is the intensity and the horizontal (x-) axis is the lateral separation between the points. The top plots describe the said individual contributions to the intensity distribution while the bottom plots illustrate a super-imposed intensity profile formed by each of the individual components in the respective top plots. The Abbe limit is based on the full width at half maximum (FWHM) of the 2 overlapping Airy disks, the Rayleigh limit is used when the central maximum of one Airy disk overlaps with the first minimum of another (producing a superimposed trough of 20-30% of the peak intensity) and the Sparrow limit when the 2 Airy disks overlap so that there is no visible difference in their superimposed intensities across the entire resolution distance. Airy disk patterns are further described in Fig. 3. Figure reprinted with permission from [5] and adapted.

conventional diascopic brightfield illumination [2]:

$$\text{Rayleigh Criterion, } d_{x,y} = \frac{0.61\lambda}{\text{NA}_{obj}} \text{ (for fluorescence) } \textbf{\textit{or}}$$

$$d_{x,y} = \frac{1.22\lambda}{\text{NA}_{cond} + \text{NA}_{obj}} \text{ (for brightfield)}$$

$$(3)$$

where NA_{cond} and NA_{obj} refer to the numerical apertures of the condenser and objective optics respectively [2].

Evaluation of the Abbe and Rayleigh equations above yield a minimum lateral resolution of $\sim$174nm (for Rayleigh) and 143nm (for the Abbe equation), when considering the shortest possible wavelength of visible light (400nm) and the highest possible NA attainable by most microscope objectives and condenser lenses today ($\sim$1.4) [3] (lenses having NAs beyond this value are known to be manufactured but are excluded from the current context, as their applications are rather constrained and specialized in nature, e.g. TIRF [4]). A third method for lateral resolution computation (the Sparrow criterion) may be defined as $d_{x,y} = \frac{0.47\lambda}{n\sin\theta}$, which evaluates to a value much closer to that obtained from the Abbe lateral resolution formula [1]. A diagram illustrating each of these 3 methods of resolution computation (and their differences) is depicted in Fig. 1:

As such, numerous researchers globally have sought to circumvent these limitations through the development and exposition of both optical and computational approaches, prominently exemplified through the emergence of *super-resolution fluorescence microscopy* (as an optical enhancement of existing fluorescence microscopy methods) which
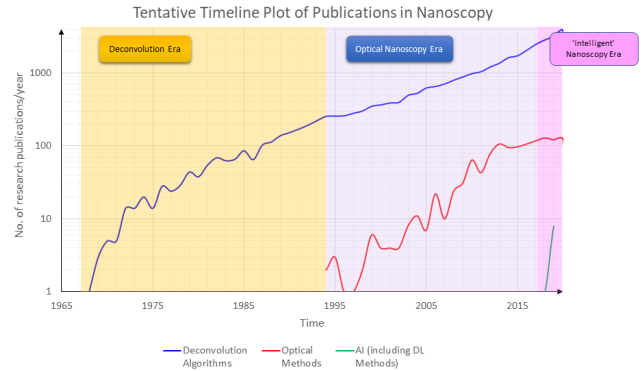

FIGURE 2. The 3 eras of nanoscopy (inferred from publication count as obtained from ScienceDirect). In this aspect, the publication count may be employed as a tentative measure to indicate the research interest in the said area.

culminated in the Nobel Prize in Chemistry being awarded to its developers (Moerner, Betzig and Hell) in 2014 [6]. Nonetheless, in this succinct (yet desirably impactful) review, we seek to evaluate some of the recently-employed *computational* advancements in the field of optical microscopy, with the intent that researchers worldwide would be inspired to address some of the existing limitations through further advancements in these *in silico* methodologies. This would inadvertently aid in potentially pushing the envelope of optical microscopy into the nanoscopy domain. In doing so, it would also be imperative to highlight the need for exploring the principles of ***image deconvolution***, which is exemplified within the present review as well. In this light, Fig. 2 depicts a general timeline plot of publications which illustrates the advancements made in the field of nanoscopy over the decades:

At this juncture, it would be noteworthy to mention that optical nanoscopical procedures (such as STED [7], GSDIM [8], dSTORM [9], Lattice SIM [10], etc) are **not** discussed in the present review, as the focus of this study is to assess the *computational* aspects of nanoscopy. Nonetheless, as optical approaches are complementary to the computational aspects in most nanoscopy applications, the interested reader is encouraged to explore the afore-mentioned references (or [11]) for a detailed discussion on each of these techniques. The review is thus structured in the following manner: Section II presents an overview of some popular deconvolution algorithms utilized in microscopical imaging today, Section III discusses some commonly-used noise removal methods, while Section IV delves into AI and its current role in computational optical nanoscopy. The use of deep learning for image denoising is further presented in Section V and this is coupled with an exploration of the commercialized applications of current AI-based image enhancement approaches in Section VI. Section VII details the present limitations and potential future advancements in the field of computational nanoscopy, with the study being concluded in Section VIII.

II. DECONVOLUTION IN OPTICAL MICROSCOPY
Deconvolution methods have long been a source of image refinement and sharpening, although it would be prudent to
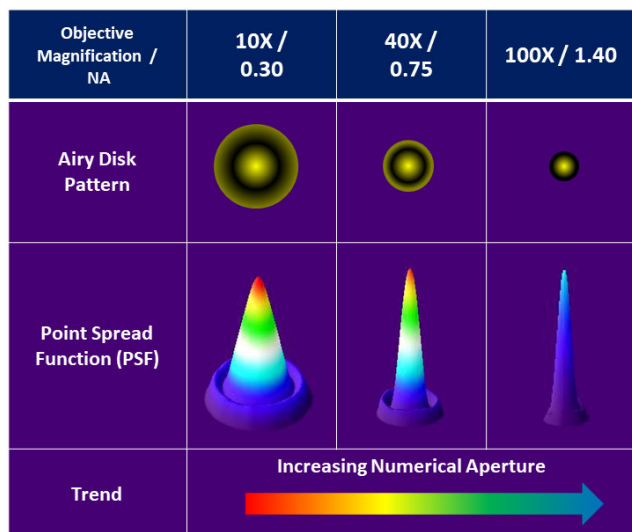
S. S. Kaderuppan et al.: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE Access

FIGURE 3. Airy disk patterns and their corresponding PSFs (which represent a Gaussian distribution). PSFs are widely used in the optics domain to refer to the blurring kernel (imposed by the limitation of the imaging optics, mismatch between the exact specifications of the objective and the mountant, deviations from ideal imaging conditions, etc). A higher resolution is attained by using a lens having a higher numerical aperture [11], as the FWHM of the PSF (and consequently the diameter of the centroid maximum of the Airy disk) is reduced.
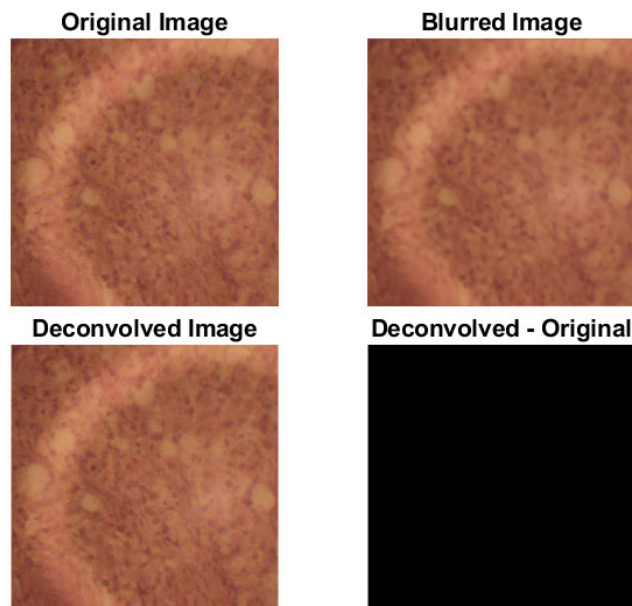

FIGURE 4. A micrograph of an ovary section blurred with a Gaussian kernel and subsequently deconvolved (using Wiener filtering) in MATLAB (The MathWorks, Inc.) Notice the computed difference between the deconvolved and raw image (the image on the bottom right) which is almost negligible, implying the high restorative efficiency of Wiener filtering in this context.

mention that deconvolution differs from sharpening, in that the former is intended to increase the *resolution* of an image, while the latter enhances *edges* in an image. Nonetheless (in some instances), the consequence of a resolution enhancement may very well be a sharpened image. Initially developed for temporal analysis by Wiener [12], deconvolution has gained prominence in the field of optical microscopy due to its ability to reassign out-of-focus light spots to their centroid maxima – made plausible due to the symmetric nature of the point spread function (PSF) of the light rays, as shown in Fig. 3:

Generally, the emergence of light rays through the optical train may be expressed by the mathematical convolution operation ($\circledast$) as described in (4):

$$(h \circledast x) + \varepsilon = y \qquad (4)$$

where h refers to the PSF (convolved with x as a consequence of the optical limitations of the system in question), x is the actual signal to be recovered, y is the detected signal/impulse response of the system and ε is the noise which has interacted with the convolved signal ($h \circledast x$) (adapted from [13]).

Notably, the convolution operator would be translated to multiplication in Fourier space, hence deconvolution may be perceived as division following a Fourier transform (FT) of the acquired impulse response y, where the FT of the PSF is known as the Optical Transfer Function (OTF) [1]. Nonetheless (and as expected in most instances), the introduction of noise ε makes it exceedingly difficult to accurately obtain the unaltered signal x, although numerous techniques (both optical and computational) have been proposed to counter and mitigate the effects of ε on the detected impulse response y.

In this regard, optimal performance in deconvolution has been reported for thin ($<50 \mu$m) sections, or optically-transparent material with little fluorescence, posing a challenge for live cell imaging applications due to motion blurring and enhanced spherical aberration effects [14]. Fig. 4 illustrates an image subjected to a Gaussian blur kernel/PSF and its corresponding deconvolved image, while Fig. 5 shows the diffraction limitations imposed by the OTF (referred to as the Abbe limiting frequency) [1]:

As such, it would be noteworthy to explore the various current computational approaches to resolving this dilemma, although one should be aware that none of these methods provide a perfect solution in reality. Moreover, the problem is further exacerbated by spatiotemporally-variant PSFs (partially elucidated in [15] which describes a non-linear variation in the PSF across different axial planes), coupled with complications such as the variable optical density of the specimen and mountant at different locations. In this regard, it would be essential to holistically evaluate the various deconvolution and denoising algorithms being utilized today, emphasizing on their underlying principles, advantages and shortcomings (where appropriate). Past studies (such as [16]) have attempted to assay and categorize the numerous deconvolution algorithms presently available into 3 main classes, namely (i) deblurring, (ii) inverse filtering and (iii) constrained iterative protocols, with both (ii) and (iii) being regarded as image restoration algorithms [14]. Further details on each of these algorithms and their respective classes are as depicted in Table 1 below (information adapted from [14] and [17]):
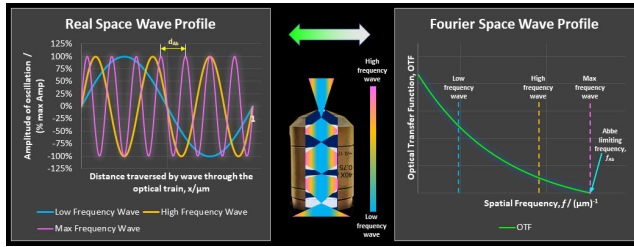
FIGURE 5. A simplified diagram illustrating the conversion of spatial displacements in real space into spatial frequency variations in Fourier space (the PSF in real space is translated into the OTF in Fourier space). The low frequency (blue) wave in real space corresponds to large features, while the high frequency (yellow) and max frequency (magenta) waves correspond to much smaller features (interference caused by super-positioning of these waves result in the final image as observed through the microscope eyepiece). In the Fourier space, waves having frequencies beyond the Abbe limiting frequency (f_{Ab}) are not captured by the lens, implying that features smaller than the Abbe diffraction limit (d_{Ab}) cannot be effectively resolved. The diagram in the middle depicts how the light rays enter the objective lens – as the frequency of the incident light ray increases, its angle of deviation from the optical axis of the objective lens increases as well, since higher frequency waves are refracted more than their lower frequency counterparts when traversing the boundary between 2 optically different media. Figure adapted from [1].

TABLE 1. Categorization of current image deconvolution procedures (Adapted from [14] and [17]).

Main Class	Sub Class	Filters / Algorithms
Deblurring	Subtractive	• Nearest Neighbors • No neighbors • Unsharp mask
Inverse Filtering	Linear Inverse	• Tikhonov-Regularized Inverse Filter (TRIF) • Naïve Inverse Filter (NIF)
	Regularized Inverse	• (Laplacian) Regularized Inverse Filter [(L)RIF] (includes Wiener filter)
Iterative	Poisson Noise	• Maximum Likelihood Estimation [e.g. Richardson-Lucy (RL) and RL with Total Variation] • Blind / Adaptive blind / Non-blind deconvolution
	Gaussian Noise	• Jansson van Cittert • Tikhonov-Miller / Iterative constraint Tikhonov-Miller • Linear Least Squares (Landweber) deconvolution • Wavelet deconvolution (including ISTA, FISTA, SURE-LET, PURE-LET)
	Constrained (Non-negative)	• Blind / Adaptive blind / Non-blind deconvolution • Least squares deconvolution (including NNLS, BVLS)

A deeper insight into the various deconvolution methodologies highlighted above is provided in the subsequent sections (to cater to the reader's interest), although greater emphasis is placed on the constrained iterative methods in light of their current popularity and enhanced capabilities (as compared to the deblurring or inverse filtering protocols).

A. DEBLURRING

Deblurring algorithms (as the precursor to deconvolution) seek to remove haze (or blurring) in the image, although this might **not** result in a significant improvement in image resolution. Hence, such algorithms are often utilized for quick image inspections (e.g. defect identification), rather than in-depth analysis. Common deblurring approaches include the **No Neighbors** or **Nearest Neighbors** algorithms [18] – the former referring to the projection of the blur kernel from the image plane itself, while the latter utilizes the impulse response from a point source above and below the image plane as a reference to predict the extent of blurring introduced in the image. In addition, the **Unsharp mask** [19] (an image sharpening tool) also represents a popular dehazing tool employed in professional photography and image-editing applications, such as Adobe Photoshop. Further details on the mathematical principles underpinning each of these methods are discussed in the paragraphs which follow.

The No Neighbors deconvolution algorithm assumes that the blurring is caused by out-of-focus light originating from the *same* image plane as the object being imaged and is described by lower spatial frequencies [18]. Hence, the No Neighbors deconvolution algorithm seeks to eliminate this out-of-focus light by implementing *high-pass* filters, thereby amplifying the proportion of high spatial frequencies in the image. Mathematically, this may be expressed as follows (adapted from [18]):

$$\widbreve{I^m} = I^m \left(I^m \circledast H_g \right) \tag{5}$$

where $\widbreve{I^m}$ is the sharpened image, I^m is the original blurred image, H_g is the Gaussian blur kernel (derived from the PSF) and $\circledast$ is the 2D convolution operator.

In addition, [18] also states how the afore-mentioned equation may be improved through the incorporation of a weighted factor (herein represented by β) so that the above equation becomes:

$$\widbreve{I^m} = \beta \cdot I^m \left[(1 - \beta) \cdot \left(I^m \circledast H_g \right) \right] \tag{6}$$

where the range of $\beta = [0.6, 0.85]$ to yield optimal results for $\widbreve{I^m}$ in most instances [18].

In contrast, the Nearest Neighbors algorithm utilizes the impulse response from the optical planes both immediately before and after the image plane in an acquired Z-stack to be convolved with a suitable blur kernel and the result subtracted from the blurred image plane to obtain the sharpened image [18]. In this respect, the Nearest Neighbors algorithm is based on the premise that the out-of-focus blur in the image plane is primarily generated from the optical planes both immediately above and below the image plane, so that elimination of the defocused optical planes from the image plane would yield the sharpened image. As previously, this may be expressed mathematically by the following equation (adapted from [18]):

$$\widbreve{I^m} = I^m \gamma \cdot [(I^m_{-1} \circledast H_g) + (I^m_{+1} \circledast H_g)] \tag{7}$$

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy
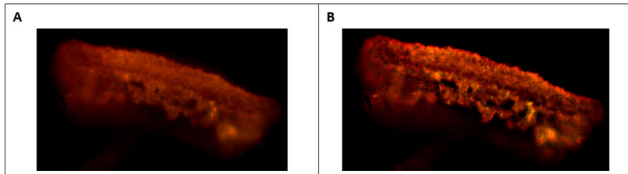
IEEE *Access*

FIGURE 6. A Raw image of an anther. **B** Unsharp mask applied to the raw image in **A** using Adobe PhotoShop CS5 (© 1990-2010 Adobe Systems Inc.). Notice the highlighted structural details evident upon application of the Unsharp mask algorithm. Filter parameters used are 200% (Amount), 33.6 pixels (Radius), 2 levels (Threshold).

where I^m_{-1} and I^m_{+1} are the images acquired in the optical planes immediately before and after the evaluated image plane within a Z-stack and γ is the additional weighting factor. In this respect, an apparent drawback of the Nearest Neighbors algorithm refers to the step-size in the acquired Z-stack, which may be perceived as a modulated function of the depth-of-field (DoF) of the microscope objective lens being used. An axial step-size greater than the DoF would result in an inaccurate approximation of the blur introduced by the image matrices I^m_{-1} and I^m_{-1}, which would consequentially result in the computation of a non-optimal raw image I^m.

A third deblurring algorithm often utilized in image processing is the Unsharp mask (which is akin to the No Neighbors deconvolution algorithm). However, in Unsharp masking, the edge-enhancing kernel is obtained by subtracting a smoothed/blurred version of the original image from the original image [19]. This kernel is then added to the original image to obtain a sharpened image [19]. Mathematically, this may be illustrated by the following matrix computation (adapted from [19]):

$$M = I^m - (I^m \circledast H) = I^m - \widetilde{I^m} \qquad (8)$$

where H is the smoothing kernel, $\widetilde{I^m}$ is the smoothed (blurred) version of I^m, M is the mask image and

$$\widecheck{I^m} = \min[(I^m + \alpha \cdot M), 255] \qquad (9)$$

where $\widecheck{I^m}$ is the sharpened image of I^m and α is the weight controlling the extent of sharpening by the mask M.

Visually, an image deblurred through implementation of the Unsharp mask algorithm is presented in Fig. 6 below:

B. INVERSE FILTERING
Inverse filtering represents another set of frequently employed image deconvolution algorithms. However, as the principle of inverse filtering is simple (to introduce a deconvolution kernel h^{-1} which restores the blurred impulse response y to its original state x, where $h \circledast x = y$, $\circledast$ representing the convolution operator), the effective use of inverse filtering is determined by the level of noise (both photon and detector noise) in the image – the higher the noise, the less effective the deconvolution. Commonly used inverse filters include Naïve inverse filtering (NIF), the Tikhonov-Miller filter (an example of a linear inverse filter) and the Wiener

filter (a regularized inverse filter). An additional approach couples Tikhonov regularization with NIF, minimizing the latter's contribution of noise [20]. The specific details on each of these inverse filtering algorithms are outlines in the following sub-sections for the interested reader.

1) NAÏVE INVERSE FILTERING (NIF)
NIF represents one of the most basic deconvolution protocols, which seeks to minimize a least-squares cost function [20]. However, in so doing, NIF also accentuates measurement noise, resulting in an increased number of high-frequency waveforms detected in the impulse response of the system. Mathematically, NIF may be represented by the following equation:

$$\widetilde{x} = \arg\min_x \xi(x) \qquad (10)$$

where $\xi(x) = \|y\text{-}Hx\|^2$, y is the observed data, H is the PSF matrix and x is the underlying fluorescence signal [20].

When Gaussian noise is present, NIF reduces to the concept of maximum-likelihood estimation, which may be determined via a quotient of coefficients in Fourier space as follows:

$$\widehat{\widetilde{x}} = \frac{\widehat{y}}{\max(\widehat{h}, \epsilon)} \qquad (11)$$

where $\widehat{\widetilde{x}}, \widehat{y}$ and $\widehat{h}$ represent the discrete FT coefficients of $\widetilde{x}$, y and the PSF, max represents the element-wise maximum and ϵ is a constant term, included to circumvent divisions by zero (adapted from [20]).

Subsequently, the inverse FT of $\widehat{\widetilde{x}}$ is used to obtain the final solution, where a regularization parameter defined as the squared Euclidean norm of x ($\|x\|^2_2$) may be added to $\xi(x)$ to accord large values with a penalty (a procedure known as Tikhonov-regularized NIF/TRNIF) as follows [20]:

$$\xi(x) = \|y\text{-}Hx\|^2 + \lambda \|x\|^2_2 \qquad (12)$$

where λ is a weighting factor determining the contribution of the 2 terms in the equation [20].

The above equation may then be minimized by applying the following relation, which may be considered as a *maximum a posteriori* (MAP) model, since λ is used to introduce prior information about x to facilitate its estimation:

$$x = \left(H^{\mathrm{T}}H + \lambda I\right)^{-1} H^{\mathrm{T}}y \qquad (13)$$

where I is the identity matrix [20].

2) TIKHONOV-MILLER (TM) AND ITERATIVELY CONSTRAINED TIKHONOV-MILLER (ICTM) FILTER
Tikhonov-Miller (TM) filtering (as an example of linear inverse filtering) is often utilized as a preliminary strategy for image deconvolution prior to iterative deconvolution. This may be attributed to it being computationally economical and rapid, although TM filtering is simultaneously susceptible to artifact generation, allowing noisy effects to be effectively

transferred and thus unsuited for enhancing an image's resolution [14]. The TM algorithm uses an iterative gradient descent approach to minimize the regularized inverse filter cost, allowing the implementation of a positivity constraint at each iteration [17]. An alternative form of the TM filter refers to the *iteratively constrained* Tikhonov–Miller (ICTM) filter, which is formed with repeated incremental projections of TM onto the set $\left(\mathbb{R}^+\right)^K$, where K represents the dimensionality of the fluorescence signal x (i.e. $x \in \mathbb{R}^K$) [20]. Mathematically, the iterative construct of ICTM may be expressed by the following equation from [20]:

$$x^{(k+1)} = P_{(\mathbb{R}^+)^K}\left\{x^{(k)} + \gamma\left(H^T y - \left(H^T H + \lambda L^T L\right)x^{(k)}\right)\right\}$$

(14)

where $P_{(\mathbb{R}^+)^K}\{a\} = \max(a,\, 0)$ and indicates the component-wise projection of a onto the set $\left(\mathbb{R}^+\right)^K$, γ is a weighting factor and L represents the matrix corresponding to the discretization of a differential operator [20].

3) REGULARIZED INVERSE FILTERING (RIF) AND WIENER FILTERS

Yet another regularized approach to inverse filtering (other than TRNIF) involves subjecting x to a smoothness constraint through reducing the impact of its derivative – a procedure known as *regularized inverse filtering* (RIF). The cost function for RIF may be described by the following expression [20]:

$$\xi(x) = \left\| y - H\left(H^T H + \lambda L^T L\right)^{-1} H^T y \right\|^2$$
$$+ \lambda \left\| \left(H^T H + \lambda L^T L\right)^{-1} H^T y \right\|_2^2 \quad (15)$$

The above expression may also be effectively reduced to (12) when the explicit minimizer $x = (H^T H + \lambda L^T L)^{-1} H^T y$ [20], thereby proving (15) to be a reformulation of TRNIF.

According to [20], defining λ as the inverse of the noise variance and coupling it with $L^T L$ filtering poses a whitening effect on x, thereby effectively converting RIF to Wiener filtering. Wiener filtering shall be discussed in greater detail in a subsequent section in this review (**Noise removal**), where it is being used to both deconvolve and restore a blurred, noisy image. Nonetheless, [21] has also represented the Wiener filter via the mathematical relationship below and regarded it as the "golden linear deconvolution trade-off":

$$y = \frac{H^* |x|}{|H|^2 \cdot |x| + |\varepsilon|}$$

(16)

where H^* is the complex conjugate of H and $|H|$ refers to the magnitude of H (adapted from [21]).

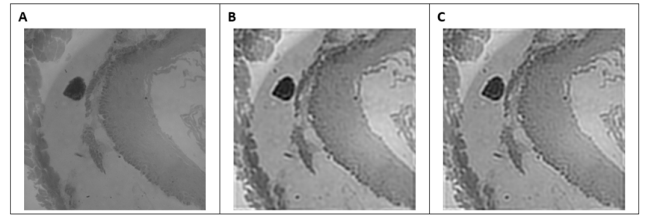For illustration purposes, a single image deconvolved using ICTM and RIF is depicted in Fig. 7 below:

FIGURE 7. A Raw image of an oesophagus section; B The raw image in Adeconvolved using ICTM ($N = 20$, y-step = 0.7, $\lambda = 1.000E-05$); C The same image in A processed using RIF ($\lambda = 5$). All deconvolution processes were performed using DeconvolutionLab2 (© 2018 EPFL) [17].
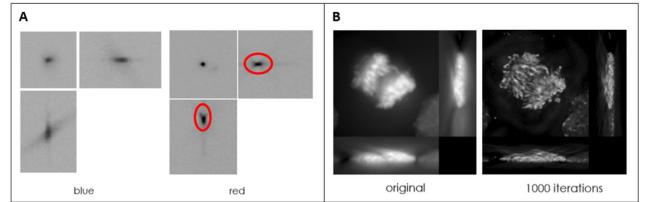

FIGURE 8. A The image of a sub-diffraction-sized bead (corresponding to the PSF) determined for both blue and red channels. Notice the PSF shift in the red channel circled above, potentially due to chromatic aberration. B Raw image of a dividing cell's chromosomes (left) and following application of a NBD algorithm over 1000 iterations (right). Figures adapted with permission from [22].

C. NON-BLIND DECONVOLUTION

Non-blind deconvolution (NBD) algorithms utilize an empirically determined PSF to deconvolve an image [22]. Most suited for constant blurring across a specified region of interest, NBD algorithms require the user to image a sample containing sub-diffraction-sized beads [22] (often ≈80-150nm in diameter, or $0.61\lambda/3$ NA, where $\lambda = 550$nm) to compute the PSF of the optical train and subsequently factor the FT of this value (as a divisor) into the FT of the impulse response to obtain the FT of the deconvolved image. Ideally, the sample containing the beads should be the same sample imaged under identical conditions to minimize deviations in the acquired PSF caused by differences in refractive index, optical density and spherical aberration [22]. A sample image of the PSF measured using this mode of deconvolution (as well as the associated images generated via application of the NBD algorithm) is shown in Fig. 8 as follows:

NBD algorithms (generally) have a number of constraints, including the need for sample clarity and absence of dirt, the axial focal range used for acquiring the Z-stack, the objective type and matching mountant being used, the temperature of the sample being imaged, etc [22]. Despite satisfying these conditions, one may still encounter setbacks through the employment of a NBD algorithm, due to the varying PSF caused by sample-coverslip distance deviations, noise or inherent differences between the imaging conditions of the beads and the sample [22]. However, NBD algorithms are still widely utilized today in popular scientific imaging application suites such as Huygens Professional (© Scientific Volume Imaging B.V.) [23] and AutoQuant X3 (© Media Cybernetics, Inc.) [24], as they minimize the num-

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

ber of iterations required to attain an optimally-deconvolved image while accounting for individual setup-specific deviations/aberrations [22]. In this regard, NBD algorithms are particularly suited for imaging fixed samples, where most of the confounders, such as varying sample-coverslip distances due to the movement of the sample (for live cell imaging) are generally absent.

D. BLIND DECONVOLUTION

Blind deconvolution (as its name implies) refers to the deconvolution of an image *without* an exact understanding of the blur kernel/PSF employed. Sporadically, this problem may be further amplified with the introduction of noise in the blurred image. Blind deconvolution (BD) may be performed in either an iterative or a non-iterative manner (as expounded in [25]), with non-iterative methods such as SeDDaRA [26] estimating the PSF through a comparison of the spatial frequency of the blurred *vs* the target image. In contrast, a common iterative approach adopted by numerous researchers ([3], [4], [5]) refers to the *maximum a posteriori* (MAP) distribution as described mathematically in [27] below (expectation-maximization algorithms are also exploited as iterative BD methods):

$$\arg \max_{u,k} p\left(u, k \mid f\right) = \arg \max_{u,k} p\left(f \mid u, k\right) p\left(u\right) p\left(k\right) \quad (17)$$

where $p(f|u, k)$ refers to the noise distribution in the blurred image, while $p(u)$ and $p(k)$ refer to the original (unblurred) image and the blur kernel respectively (Equation adapted from [27]).

In addition, according to [27], although BD has been conceived several decades ago, it has garnered recent mounting interest particularly in resolving the issue of *motion blurring* (caused by mobile image acquisition devices [28]) with significant progress being illustrated in numerous studies conducted in this domain (see [29]–[31] and [32] for details). BD has also been described in [27] to often utilize *priors*, the most common of which refer to total variation (TV) ([33], [34]) due to its ability to resolve sparse signals [35], although a recent study ([36]) indicated the preferential implementation of an alternative approach – the joint optimization of the image and the blur kernel which depicts convergence when initialized with the typical (no-blur) solution (i.e. the blurry image and its Dirac delta pair) [27]. [27] then proceeds to investigate this anomaly, leading them to discover that total variation should *not* be utilized as a prior for deriving a sharp image since it never actually converges to the global minimum. The authors of [27] further propose an alternative methodology to circumvent this issue – the Alternating Minimization (AM) and the Projected AM (PAM) Algorithms.

AM seeks to resolve both (18) and (19), which involve an unconstrained convex problem in u and a constrained convex problem in k (adapted from [27]) as follows:

$$u^{t+1} \leftarrow \arg \min_{u} \sum_{x \in F} \left(\left(k^t \circ u\right) [x] - f[x]\right)^2$$
$$+ \sum_{x \in U} \|\nabla u[x]\|_2 \quad (18)$$

$$k^{t+1} \leftarrow \arg \min_{k} \sum_{x \in F} \left(\left(k \circ u^{t+1}\right) [x] - f[x]\right)^2 \quad (19)$$

where $k \succeq 0$, $\|k\|_1 = 1$

Here, the authors of [27] have defined $\circ$ to be the valid convolution (which is also employed in MATLAB), with $k \circ u \neq u \circ k$ ($u \circ k$ being undefined if the support of k is large) [27].

By doing so and utilizing the findings of [37], [38] and [39] through implementation of the AM algorithm for a 1D signal domain, [27] has derived an expression for λ based on a clear (but linearly-transformed) u^0 as the solution of the following denoising problem expressed mathematically:

$$\widehat{u}[x] = \arg \min_{u} \frac{1}{2} \sum_{x=-L_1+1}^{L_2-1} (u[x] - f[x])^2$$
$$+ \lambda \sum_{x=-L_1}^{L_2-1} |u[x+1] - u[x]| \quad (20)$$

Nonetheless, the scaling of u^0 (induced by the use of the total variation regularizer) was claimed by [27] to be the primary factor accounting for the non-convergence of the AM algorithm towards a global minimum, resulting in the postulation of PAM (an iterative construct similar to AM in u, but alternating with an unconstrained convex problem in k) being described by the following equations:

$$k^{t+\frac{1}{3}} \leftarrow \arg \min_{k} \sum_{x \in F} ((k \circ u)[x] - f[x])^2 \quad (21)$$

$$k^{t+\frac{2}{3}} \leftarrow \max\left\{k^{t+\frac{1}{3}}, 0\right\}, k^{t+1} \leftarrow \frac{k^{t+\frac{2}{3}}}{\left\|k^{t+\frac{2}{3}}\right\|_1} \quad (22)$$

PAM implementation (based on the following protocol from [27]) and its results are depicted in Fig. 9:

The significant improvement of images deconvolved using PAM (over traditional TV-based regularizers in blind deconvolution) may thus hold promise in the utilization (and potential further development) of PAM to achieve nanoscopic super-resolution in the field of optical microscopy.

E. RICHARDSON-LUCY (RL) DECONVOLUTION

In 1972 and 1974, 2 researchers ([40], [41]) independently proposed an approach to address the issue of deblurring images captured off the sensing plane with the addition of noise. This approach came to be known as the Richardson-Lucy (RL) model for image deconvolution, which is still being widely used today.

Fundamentally, the RL deconvolution algorithm seeks to solve for x given the values of h and y in (4) previously. Elucidation of x is made possible through an understanding of the Poissonian nature of photoelectron distribution and image formation [42] exemplified by the Poisson probability mass function $P(n)$ as follows (adapted from [43] for non-mathematically-inclined readers):

$$P(n) = \frac{v^n e^{-v}}{n!} \quad (23)$$

where $P(n)$ (in the current context) refers to the probability of a *single* pixel receiving n photoelectrons and v is the expected (*average*) number of photoelectrons collected per pixel.
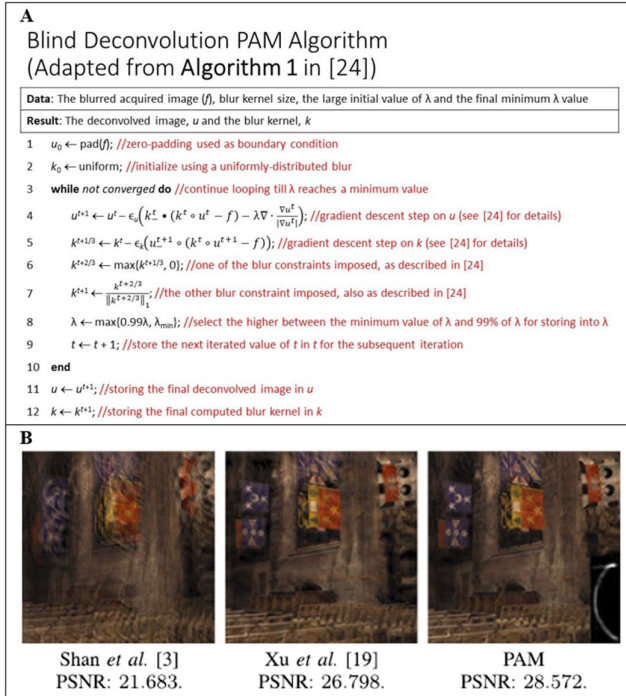
A

Blind Deconvolution PAM Algorithm
(Adapted from Algorithm 1 in [24])

Data: The blurred acquired image (*f*), blur kernel size, the large initial value of λ and the final minimum λ value

Result: The deconvolved image, *u* and the blur kernel, *k*

1   $u_0 \leftarrow \text{pad}(f);$ //zero-padding used as boundary condition
2   $k_0 \leftarrow \text{uniform};$ //initialize using a uniformly-distributed blur
3   **while** *not converged* **do** //continue looping till λ reaches a minimum value
4   $u^{t+1} \leftarrow u^t - \epsilon_o\left(k_-^t \bullet (k^t \circ u^t - f) - \lambda\nabla\cdot\frac{\nabla u^t}{|\nabla u^t|}\right);$ //gradient descent step on *u* (see [24] for details)
5   $k^{t+1/3} \leftarrow k^t - \epsilon_k\left(u_-^{t+1} \circ (k^t \circ u^{t+1} - f)\right);$ //gradient descent step on *k* (see [24] for details)
6   $k^{t+2/3} \leftarrow \max[k^{t+1/3}, 0];$ //one of the blur constraints imposed, as described in [24]
7   $k^{t+1} \leftarrow \frac{k^{t+2/3}}{\|k^{t+2/3}\|_1};$ //the other blur constraint imposed, also as described in [24]
8   $\lambda \leftarrow \max\{0.99\lambda, \lambda_{min}\};$ //select the higher between the minimum value of λ and 99% of λ for storing into λ
9   $t \leftarrow t + 1;$ //store the next iterated value of *t* in *t* for the subsequent iteration
10 **end**
11 $u \leftarrow u^{t+1};$ //storing the final deconvolved image in *u*
12 $k \leftarrow k^{t+1};$ //storing the final computed blur kernel in *k*

B

| Shan *et al.* [3] | Xu *et al.* [19] | PAM |
| PSNR: 21.683. | PSNR: 26.798. | PSNR: 28.572. |

FIGURE 9. **A** The algorithm pseudocode describing the iterative procedure of PAM. **B** A comparison of PAM deblurring with 2 other methods (highlighted in separate studies) referenced in [27]. Notice the significantly higher PSNR of PAM (28.572) as compared to the other 2 approaches, with more examples being provided in [27] (including motion de-blurring) but which are not included here. Figure A is based on the algorithm described in [27] (further details are expounded therein), while Figure B is reprinted from [27], © 2015 IEEE.

From [42], one can thus deduce the probability of the collected photoelectrons contributing to the formation of the entire image I^m as follows (adapted from [42]):

$$p\left(I^m|o\right) = \prod_a \left(\frac{[(h \circledast x)]^y \, e^{-(h\circledast x)}}{y!}\right) \quad (24)$$

where an object *o* is projected as an image I^m through an optical system having a PSF *h* and subject to Poisson noise. Notice that the Poisson parameter *v* in (23) has been replaced by the convolution (⊛) of *x* with *h*, so that $p(I^m|o) = \prod_a [P_a(n)]$, where $P_a(n)$ refers to $P(n)$ at pixel *a* {*a* being a 2-D or a 3-D vector, represented as (X,Y) or (X,Y,Z) respectively [42]}.

In order to maximize the value of $p(I^m|o)$ [a process known as *maximum likelihood estimation* (MLE) which represents a common paradigm employed in machine learning] so that the image I^m closely resembles *o*, we may reduce (24) to the following (also from [42]) which we seek to *minimize*:

$$
\begin{aligned}
J_1(o) &= -\ln\left[p\left(I^m \mid o\right)\right] \\
&= \int_x \left[(h \circledast x) - y \cdot \ln(h \circledast x) + \ln(y!)\right] dx \\
&\Leftrightarrow \int_x \left[(h \circledast x) - y \cdot \ln(h \circledast x)\right] dx \quad (25)
\end{aligned}
$$

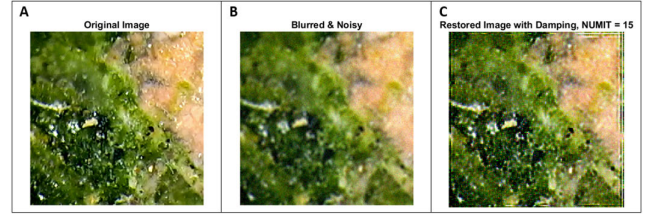**FIGURE 10.** **A** The original image; **B** The image blurred with Gaussian PSF (hsize = 12, sigma = 6) and subjected to Gaussian noise (mean = 0; variance = 0.0021); **C** The post-processed image following RL deconvolution (with damping) in MATLAB (NUMIT = number of iterations of the RL algorithm) (Script source: [44]).

where $J_1(o)$ is the first-order Bessel function while $\ln(y!)$ is a constant (with respect to *o*).

[43] then further proceeds to elaborate on how (25) may be converted (through a series of steps not described here) to the following final equation, which defines the RL deconvolution algorithm:

$$x_{k+1} = x_k\left[\dot{h} \circledast \frac{y}{(h \circledast x_k)}\right] \quad (26)$$

where *k* is the iteration count, $\frac{x_{k+1}}{x_k} = 1$ at convergence and $\dot{h}$ represents the PSF *h* at a complementary point about the origin.

Notably, (26) above represents the *multiplicative* form of the RL deconvolution algorithm. [42] also provides an *additive* gradient-descent variant of the RL deconvolution algorithm as follows:

$$o_{k+1} = o_k + \delta t\left(1 - \dot{h} \circledast \frac{y}{(h \circledast x_k)}\right) \quad (27)$$

where δt represents the gradient descent step size.

Nonetheless, [42] has observed that the vanilla RL algorithm may not always converge to a particular solution, as no prior information on *o* has been supplied to it, highlighting the need for a prior model to be constructed on *o* (a process termed *regularization*) [42]. [42] further described 2 common regularizers commonly-utilized in conjunction with RL deconvolution – (i) TM regularization [indicated in (28) below] and (ii) Total Variation (TV) regularization [indicated in (29) below], both of which are sourced from [42]:

$$x_{k+1} = \left\{\left[\frac{y}{(x_k \circledast h)}\right] \circledast \dot{h}\right\} \cdot \frac{x_k}{1 + 2\lambda_{TM}\Delta x_k} \quad (28)$$

where λ_{TM} is the TM regularization parameter.

$$x_{k+1} = \frac{x_k}{1 - \lambda \cdot \text{div}\left(\frac{\nabla o_k}{|\nabla o_k|}\right)}\left[\dot{h} \circledast \frac{y}{(h \circledast x_k)}\right] \quad (29)$$

where λ is the TV regularization parameter and div(**F**) is the divergence of **F**.

A sample image implementing the RL deconvolution algorithm is depicted in Fig. 10 below:

As with PAM [27] discussed in the previous section, further improvements made in the RL deconvolution algorithm seem promising for achieving nanoscopic resolution in optical microscopy in the near future.
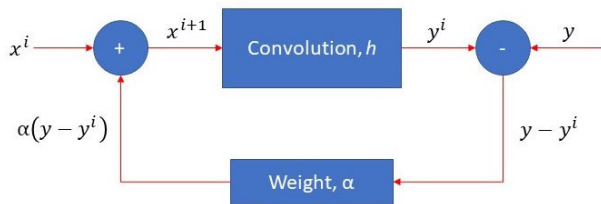
S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

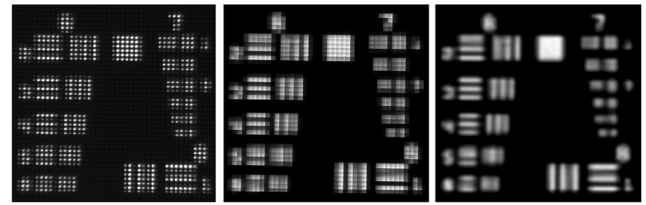**FIGURE 11.** The JVD algorithm schema. Adapted from [45].

FIGURE 13. The USAF 1951 target imaged using a light field microscope (left), a comparative deconvolution method discussed in [58] and used for image reconstruction (center), the anti-aliased image following application of the presently reviewed approach (right). Notice the image produced in [48] depicting reduced aliasing (*increased smoothing*) between neighboring pixels. Figures adapted with permission from [48], © The Optical Society.

FIGURE 12. Lunar neutron maps implementing the JVCD algorithm. The maps A1 and B1 depict the implementation of the JVCD algorithm with a small $p_{17}psf$ on O_A and O_B respectively, while A2 and B2 depict the implementation of the JVCD algorithm with a large $p_{31}psf$ on O_A and O_B respectively. p_{17} = [17° × 17°, N(8,2.8), fwhm = 6.6°]; p_{31} = [31° × 31°, N(15, 4), fwhm = 9.4°]. Figures reprinted with permission (and adapted) from [46], copyright Lunar and Planetary Institute.

F. JANSSON-VAN CITTERT DECONVOLUTION

The Jansson-Van Cittert deconvolution (JVCD) algorithm was well-described by Sastry [45] and illustrated in the flow diagram (adapted from [45]) as follows:

Generally, an initial guess x^0 is first provided of the actual signal x and subjected to the convolution kernel h, yielding a convolved signal y^0. This is then compared with the actual acquired signal y and the difference between these 2 impulse responses computed [45]. If this difference evaluates to 0, the prior used (x^0) is the desired 'deconvolved' signal, else the difference ($y - y^0$) is evaluated and a weight α applied to this difference to be added to the input signal x^0 to produce x^1 (the first iteration) [45]. This process repeats itself, until ($y - y^i$) approaches 0 for the ith iteration. A sample of an image (though not a photomicrograph) implementing the JVCD algorithm [46] is shown in the following figure (Fig. 12):

Despite the relative popularity of the JVCD algorithm, numerous studies (including [45] and [47]) have sought to improve it, through increasing its resilience to noise and infusing other constraints in the algorithm, while minimizing potential errors in deriving the convolution (blur) kernel (as compared to other non-iterative deconvolution protocols).

G. ARTIFACT-FREE DECONVOLUTION

Artifact-free deconvolution was proposed by [48] for use with the light field microscope (LFM) – an imaging modality developed by [49] which has been utilized in numerous biomedical and cellular imaging applications (see [50], [51], [52], [53] for details). Here, the authors of [48] claim that the LFM has a depth-dependent sampling pattern, resulting in variable lateral resolution at different optical planes across the sample. According to [48], the LFM utilizes micro-lenses for 3D fluorescent microscopical imaging, thereby eliminating the inherent scanning present in traditional confocal laser-scanning fluorescence microscopy, while allowing the vectorial light field to be acquired in a single exposure for topographical reconstruction of the imaged sample [48]. The LFM was thus conceived by [48] to belong to a family of *plenoptic* devices [54] which allow 3D imaging [49] and refocusing of the acquired images ([55], [56], [57]). In this respect, [48] highlights the numerous past studies seeking to improve the LFM resolution (namely via multi-view reconstruction, ray-based and wave-based methods), with a particular study [58] depicting laterally well-resolved images. Nonetheless, [58] also reported that the improvement rate was inconsistent at different imaged depths – a phenomenon attributed to the variable sampling patterns at different axial planes coupled with aliasing effects. As such, [48] sought to resolve these artifacts by proposing "depth-dependent anti-aliasing filters" [48, p. 31645] applied to an iterative "aliasing-aware deconvolution method for artifact-free 3D reconstruction" [48, p. 31645] via an expectation maximization protocol. A sample image of the USAF 1951 resolution target imaged via raw LFM, the method described by [48] and a comparative approach are depicted in Fig. 13:

In deriving the afore-mentioned algorithms, the authors of [48] first investigated the relationship between the image size *at* the microlens array with respect to that *under* a microlens, before considering the depth-dependent variation in the blur radius, leading them to derive the following equations:

$$\gamma_z = \frac{d_{mla}^{sens}}{d_{tl}^{mla}} \left| \frac{z''}{d_{tl}^{mla} - z''} \right| \tag{30}$$

$$b_z = r_{ml} \left| \frac{1}{z'''} - \frac{1}{d_{mla}^{sens}} \right| \tag{31}$$

where γ_z is the scaling factor expressing the relationship between the image size at the microlens array (MLA) with

IEEE Access

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy
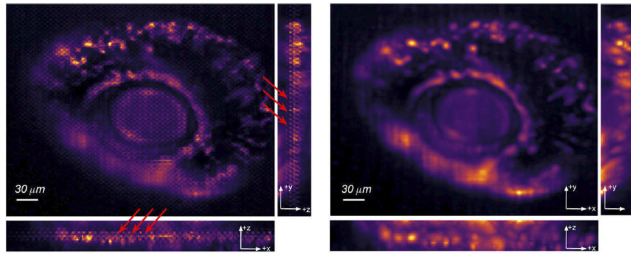
FIGURE 14. A comparison between the 3D reconstruction of a cardiomyocyte organoid via the comparative approach (as proposed in [58]) (left) and that of the presently-reviewed study (right). Figures adapted with permission from [48], © The Optical Society.

that under a microlens, d_{mla}^{sens} is the distance between the sensor and the MLA planes, d_{tl}^{mla} is the distance between the tube lens and the MLA planes, z'' is the conjugate image plane formed by the tube lens, z''' is the conjugate image plane formed by the MLA, r_{ml} is the micro-lens radius and b_z is the blur radius (Adapted from [48]).

Subsequently, the authors of [48] proceeded to evaluate a probable radius of the ideal filter kernel, $w_{obj_z} = \frac{w_{sens_z}s}{p_{ml}}$, prior to subsequently proposing the depth-dependent ideal filter $h_{f_{w_z}}$, with individual kernels at different depths (defined by w_{obj_z} as previously) and the use of a Lanczos2-constrained sinc kernel as a non-ideal anti-aliasing filter (further details of these are expounded in [48] for the interested reader). Following this, the authors of [48] then derived an equation describing a 3D aliasing deconvolution schema (based on an expectation-maximization-smoothing approach) which depicted significant improvement in artifact removal when compared with a different study [58] as shown in Fig. 14:

3D Aliasing Schema (From [48]):

$$v^{q+1} = h_{f_{w,z}} \circledast \frac{v^q}{\mathrm{A^T 1}} \left[\mathrm{A}^T \frac{\boldsymbol{m}}{Av^q} \right] \qquad (32)$$

where v refers to the imaging volume to be constructed, q is the iteration count, $h_{f_{w,z}}$ is the axially-based anti-aliasing filter, A defines the forward operation of the LFM and **m** refers to the noisy measurements acquired through the LFM [48].

H. LEAST SQUARES DECONVOLUTION
The Least squares deconvolution algorithms category comprises of 3 main classes, namely (i) Linear Least Squares (Landweber), (ii) Non-Negative Least Squares and (iii) Bounded-Variable Least Squares deconvolution. In this regard, we seek to evaluate each of these 3 algorithms in greater depth as follows:

The Landweber deconvolution (LD) approach utilizes similar principles as NIF, i.e. it seeks to minimize the same least squares cost function $C(x) = \|y - Hx\|^2$ (adapted from [20]), although unlike the latter, it does this in an iterative fashion using gradient descent ([59] from [20]). Specifically, [20] utilizes the iterative feature of LD as a means of ensuring each iteration yields positive results, which may be expressed mathematically as follows (sourced from [20]):

$$\boldsymbol{x}^{(k+1)} = \mathcal{P}_{(\mathbb{R}^+)^K} \left\{ \boldsymbol{x}^{(k)} + \gamma H^T \left(\boldsymbol{y} - H\boldsymbol{x}^{(k)} \right) \right\} \qquad (33)$$

where γ is the step size, k is the iteration count, $\boldsymbol{y}$ is the observed volume, H is the PSF matrix and $\mathcal{P}_{(\mathbb{R}^+)^K} \{x\} = \max(\boldsymbol{x}, 0)$, which represents the projection of $\boldsymbol{x}$ onto the domain $(\mathbb{R}^+)^K$ in a component-wise fashion.

Notably, [20] mentions that the number of iterations is only suggestive – iterative convergence of the computed solution results in over-fitting of the noise in the input data, while an optimal SNR may be achieved through premature truncation of the algorithm. Additionally, [20] highlights that the number of iterations is a *pseudo regularizer* common to all MLE algorithms, of which the LD is an example. Other studies ([60] and [61]) have also sought to encapsulate the Landweber algorithm with Shannon wavelets [60] or Shannon shearlets [61], the former demonstrating an accelerated algorithmic convergence (by a single order of magnitude) while the latter exhibited improvements for motion deblurring.

The Non-Negative Least Squares (NNLS) algorithm extends the LD algorithm to include a constraint of non-negativity, while seeking to minimize C(x) [17]. Similarly, the Bounded Variable Least Squares (BVLS), also known as the Stark-Parker (SP) [17], algorithm represents an alternative approach to NNLS, by imposing limits (both an upper- and a lower-bound active set) as an additional constraint on C(x) [62]. Mathematically, the problem which the BVLS algorithm seeks to solve may be expressed as follows:

$$\min_{l \leq x \leq u} \|Ax\text{-}b\|_2 \qquad (34)$$

where A is a 2-dimensional ($m \times n$) matrix, $l, x, u \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $\|x\|_2$ refers to the Euclidean (ℓ_2) norm of x.

It would be prudent to highlight at this juncture that BVLS incorporates some differences from NNLS, including (but unrestricted to) a re-computation of the QR decomposition each time the following calculation is performed (in order to improve numerical stability) [62]:

$$z = \arg \min \left\| A'z - b' \right\|_2^2 \qquad (35)$$

for some variable z, where A' is a matrix of the columns in A whose indices are 'free' components of x between l and u and b' is a data vector excluding the bound variable predictions [62]. By doing so, the BVLS algorithm allows a return of the last free and bound sets in a sequence of relations (where the upper and lower limits are well-defined) in a computationally-efficient manner.

Visually, image deconvolution performed using LD, NNLS and BVLS in DeconvolutionLab2 may be described by Fig. 15:

I. WAVELET DECONVOLUTION
Wavelet deconvolution refers to a field of image processing which primarily involves the use of wave kernels (often a single oscillation) to convolve with the detected impulse response of a system so as to identify the presence of a frequency corresponding to the mother wavelet in the said impulse response. The popularity of wavelets has resulted in the development of a large family of wavelets, which may
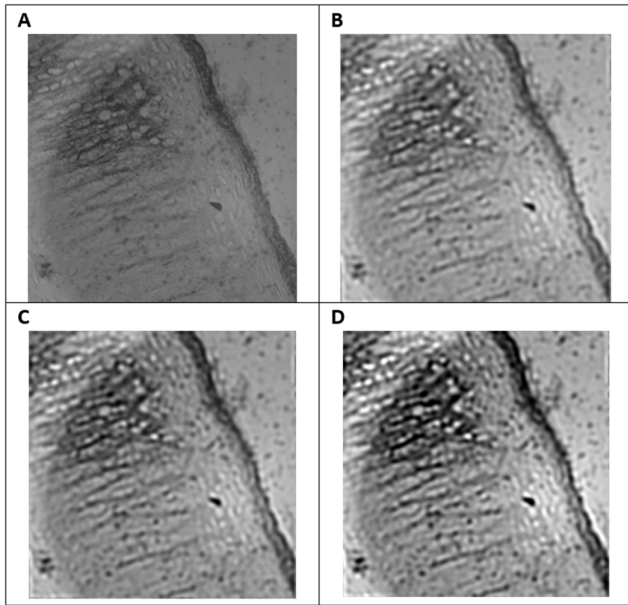
S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

FIGURE 15. A The original image (T.S. of cartilaginous tissue); B The image in A deconvolved using the LD algorithm; C The image in A processed using the NNLS algorithm; D The image in A deblurred using the BVLS algorithm. In B, C and D, a total of 9 iterations with a y-step of 1 was used.
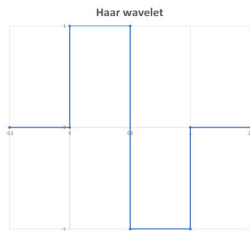

FIGURE 16. The Haar wavelet as a basic waveform kernel belonging to the family of discrete wavelets.

be classified as either *discrete wavelets* (a popular member of which refers to the Haar wavelet) or *continuous wavelets* (a popular example of which is the Spline wavelet). Haar wavelets [63] refer to a square wave kernel, having a mother wavelet $\psi(t)$ defined mathematically in (36) and graphically (in Fig. 16):

$$\psi(t) = \begin{cases} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \\ 0 & otherwise \end{cases} \tag{36}$$

In contrast, the spline wavelet is based on a spline function [64], with some spline wavelets (as discussed in [65]) being orthogonal and having an unbound (non-compact) support. Nonetheless, the term 'spline wavelets' has been popularly characterized (in most instances) to refer to a specific class of B-spline wavelets which have a compact support and are not orthogonal, termed B-spline wavelets or cardinal B-spline wavelets ([66], [67]). Cardinal B-spline wavelets have numerous properties, although they all seek to solve the following mathematical equation (where N_m refers to
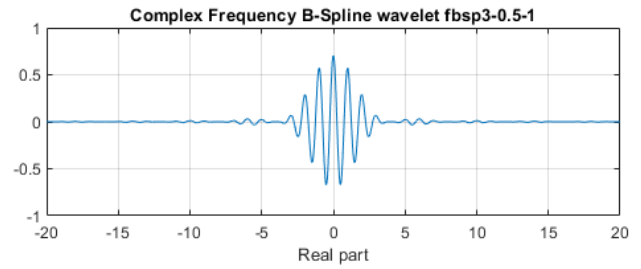

FIGURE 17. A plot of the real component of a complex frequency B-spline wavelet ($m = 3$, $FB = 0.5$, $FC = 1$). The figure was plotted in MATLAB (© 1994-2020 The MathWorks, Inc.)

the B-spline wavelets of order m having knots in the set $\mathbb{Z}$) from [66]:

$$N_1(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & otherwise \end{cases} \tag{37}$$

$$N_m(x) = \int_0^1 N_{m-1}(x\text{-}t)\,dt, \quad \text{for } m > 1, m \in \mathbb{Z}^+ \tag{38}$$

Essentially, such cardinal B-splines may span multiple orders, from $m = 1$ (for constant B-splines) to $m = 5$ (for quintic B-splines), as an example.

Conversely, compactly supported B-spline wavelets having a support of $[0, m]$ are based on the following generalized scaling and B-spline wavelet functions [represented as $\varphi_m(x)$ and $\psi_m(x)$ in (39) – (40) respectively], exhibiting up (but not being restricted) to the sextic order [68]:

$$\varphi_m(x) = \sum_{k=0}^{m} p_k \varphi_m(2x - k) \tag{39}$$

where the two-scale sequence $p_k = 2^{1-m} \binom{m}{k}$ (for $0 \le k < m$), and

$$\psi_m(x) = \sum_{k=0}^{3m-2} q_k \varphi_m(2x - k) \tag{40}$$

where $q_k = (-1)^k 2^{1-m} \sum_{l=0}^{m} \binom{m}{l} \varphi_{2m}(k - l + 1)$.

Here, [68] defines the order m of a B-spline wavelet as being 1 more than the highest exponent variable in its scaling function $\varphi_m(x)$. The interested reader would henceforth be encouraged to refer to [68] for a further detailed mathematical treatment of the decomposition function $\varphi_m(2x - k)$.

On a separate note, a visual representation of the real part of a complex frequency B-Spline wavelet having order $m = 3$, a bandwidth FB of 0.5 and center frequency FC of 1 is depicted in the following diagram (Fig. 17):

A third class of B-spline wavelets is the Battle-Lemarie (BL) wavelets [69] which is defined from the cardinal B-spline wavelets, but have their expressions derived in the Fourier (i.e. frequency) domain of the mother wavelet. Here, the BL wavelet of mth order $\{\psi_{BL,m}(t)\}$ has the following FT (denoted as $\widehat{\psi}_{BL,m}(\omega)$ and adapted from [69]):

$$\widehat{\psi}_{BL,m}(\omega) = -\frac{e^{-i\omega/2}\overline{\widehat{\varphi_m}(\omega + 2\pi)}\,\widehat{\varphi_m}\left(\frac{\omega}{2}\right)}{\widehat{\varphi_m}\left(\frac{\omega}{2} + \pi\right)} \tag{41}$$

IEEE Access

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

where $\widehat{\varphi_m}(\omega)$ is the FT of the scaling function of $\psi_{BL,m}(t)$ (i.e. $\varphi_m(t)$) and $\bar{x}$ represents the complex conjugate of x [69]. Here too, the interested reader is encouraged to refer to [69] for a complete treatment and proof of the BL wavelets through a series of fast FT (FFT) computations.

1) ITERATIVE SHRINKAGE THRESHOLDING ALGORITHM (ISTA) AND FAST ISTA (FISTA)

The ISTA and FISTA algorithms represent 2 other methods employing wavelets which are coupled with an expectation-maximization (EM) algorithm that is founded on a likelihood penalization approach in the wavelet domain for the derivation of a maximum penalized likelihood estimator (MPLE) [70]. Here, both ISTA and FISTA utilize the property of *sparsity* for wavelets to preserve image details and discontinuities [20], being particularly adapted to evaluating linear inverse problems [71]. Nonetheless, the use of a non-smooth Manhattan (l_1) norm for FISTA allows efficient solving of the cost function up to several orders of magnitude faster than ISTA [71]. Mathematically, the equations defining ISTA are indicated as follows (adapted from [70]):

Expectation (E)-Step:

$$Q\left(\theta, \widehat{\theta}^{(t)}\right) = E\left[\log p\left(y, z \mid \theta\right) \mid y, \widehat{\theta}^{(t)}\right] \quad (42)$$

Maximization (M)-Step:

$$\widehat{\theta}^{(t+1)} = \arg\min_{\theta}\left\{Q\left(\theta, \widehat{\theta}^{(t)}\right) - pen\left(\theta\right)\right\} \quad (43)$$

where $pen(\theta)$ refers to the penalized value of θ.

Essentially, the EM algorithm in ISTA alternates between an FFT-derived E-step and a M-step centered on the Discrete Wavelet Transform (DWT), thereby increasing the operational efficiency of the iteration to achieve a linear logarithmic complexity, i.e. O(Nlog$N$) [70]. In addition, [70] has also shown ISTA to be convergent towards a global maximum under suitable conditions, while potentially out-performing other leading deconvolution methods in benchmark tests. Conversely, FISTA may be employed with a (i) constant or (ii) backtracking step-size, the former of which utilizes an iterative shrinkage operator $p_L(\cdot)$ on y_k (a linear combination of the previous 2 points x_{k-1} and x_{k-2}) [71]. In this regard, FISTA is able to achieve an improved computational complexity of O($1/k^2$) as opposed to O($1/k$) attainable by ISTA [71]. Further details underlying the FISTA algorithm may be accessed from [71], for the interested reader.

A photomicrograph of a copepod, individually subjected to each of these wavelet deconvolution algorithms (ISTA and FISTA) as implemented within DeconvolutionLab2 (© EPFL) [17], is shown in Fig. 18:

2) MULTI-WIENER SURE-LET AND PURE-LET DECONVOLUTION

In 2013, Multi-Wiener SURE-LET Deconvolution was proposed by [72] as a new approach to image deconvolution integrating Wiener filters with undecimated Haar-wavelet
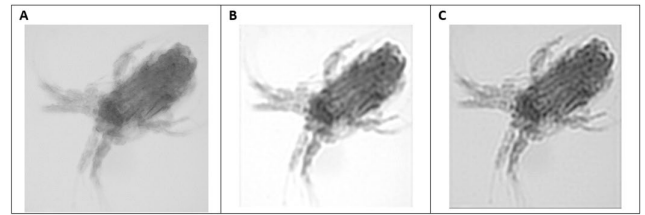

FIGURE 18. **A** The original raw acquired image (a copepod); **B** The image from A deblurred using ISTA (λ: 0.1) in DeconvolutionLab2 (© EPFL) [17]; **C** The same image from A deblurred using FISTA (λ: 1.00E-06) in DeconvolutionLab2 (© EPFL) [17]. Images B and C were artificially brightened in Microsoft Word (© Microsoft Corporation), with the input parameters used in DeconvolutionLab2 (© EPFL) [17] being 12 iterations, y-step = 1, Haar wavelets with Scale = 3, and convolved with the default simulated Airy PSF.

thresholding so as to minimize the regularized Stein's unbiased risk estimate (SURE) as an unbiased mean squared error (MSE) estimate (assuming Gaussian noise), was proposed by [72] and named [72]. Here, SURE (being solely dependent on empirical data) has been noted by [72] as a feasible parameter for solving linear problems, with potential use-cases including (i) parameter optimization (as in Tikhonov regularization [73], (ii) non-local means (NLM) denoising [74], (iii) monitoring PSNR increments during IST iterations [72], or (iv) being utilized as a minimization measure for denoising algorithms implementing the linear expansion of thresholds (LET) approach, leading to the integrated procedure named SURE-LET ([75], [76]).

Mathematically, SURE may be expressed by ϵ in (44) as follows:

$$\epsilon = \frac{1}{N}\left(\|f(y)\|^2 - 2y^T H^{-T}f(y) + 2div_y\left\{CH^{-T}f(y)\right\} + \|x\|^2\right) \quad (44)$$

where N is the number of pixels in the image, y is the convolved impulse response of the system, H is a square matrix representing a linear distortion, x is the uncorrupted and unknown signal, C is the covariance matrix and div_y is the divergence of y [72]. Here, ϵ represents the unbiased estimate of the MSE $= \frac{1}{N}E\left\{\|f(y) - x\|^2\right\}$ where $E(x)$ refers to the expectation of x, which translates into the following expression:

$$\epsilon = \frac{1}{N}\left(\|f(y)\|^2 - 2y^T H^{-T}f(y) + 2\sigma^2 div_y\left\{H^{-T}f(y)\right\} + \|x\|^2\right) \quad (45)$$

with variance σ^2 and $C = \sigma^2 I$ (where I is the identity matrix).

Further details underlying the mathematical proof for SURE, as well as visual representations of images deconvolved using SURE-LET are discussed in [72], although generally, a significant improvement was noted in the results obtained from SURE (as expressed in [72]). Notably, the average PSNR (over 10 cycles) of the 256×256 House image deconvolved via SUER-LET was 25.20, as compared to other leading deconvolution methods such as ForWaRD,

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

which exhibited an average PSNR of 24.27 (Gaussian blurred image with variance, $\sigma = 30$) [72].

An alternative approach (named PURE-LET) was proposed by [77] in 2017 to resolve 3D epifluorescence microscopy images, which sought to extend SURE-LET to Poisson and mixed Poisson-Gaussian noise cases. Here, PURE-LET utilizes a Poisson unbiased risk estimate (hence its name, PURE) to solve a set of linear equations both rapidly and accurately, with promising results reported for different convolution kernels and noise levels [78]. Mathematically, the PURE-LET model is based on the following equation [78]:

$$y = \alpha P \left(\frac{\mathbf{H}x}{\alpha} \right) + N \left(\mathbf{0}, \sigma^2 \text{Id} \right) \qquad (46)$$

where y represents the distorted representation of the true d-dimensional image x, $N = \prod_{i=1}^{d} N_i$, $\mathbf{H} \in \mathbb{R}^{N^2}$ represents a convolution matrix of the PSF h (where $\mathbf{H}x$ is used to impose a positivity constraint on the Poisson noise intensities), $P(\cdot)$ represents the effect of Poisson noise, α is the scaling factor controlling the strength of the noise and σ^2 is the variance of the additive-white-Gaussian-noise (AWGN) [78].

PURE-LET seeks to resolve this problem by considering 2 cases – (i) the Poisson Noise case and (ii) the mixed Poisson-Gaussian Noise case (full mathematical details are discussed in [78]). In so doing, [78] proposed defining the fundamental deconvolution functions using Wiener filtering followed by transform-domain denoising. The Haar wavelet transform was also employed by [78] following prior studies confirming its efficacy in minimizing numerous types of noise degradations [78]. Visual realization of the improvements garnered through the utilization of PURE-LET for image deconvolution (as compared to other techniques, e.g. PIDAL or GILAM) are depicted in [78], where (for instance) a 256 × 256 Galaxy image deconvolved with PURE-LET had a PSNR of 27.74 (for $\alpha = 4$) while PIDAL and GILAM facilitated deconvolutions reported a PSNR of 27.38 and 26.97 respectively. Separately, [77] has demonstrated a further use case of PURE-LET for deconvolving autofluorescence microscopical images of mature pollen grains, with a similar stellar performance reported for PURE-LET as compared to other techniques (such as ParalterDecon and MitivDecon) [PSNR$_{\text{PURE-LET}}$ = 28.77, PSNR$_{\text{ParalIterDecon}}$ = 26.11, PSNR$_{\text{MitivDecon}}$ = 27.56 for $\alpha = 0.5$], while being significantly faster than these methods as well (t$_{\text{PURE-LET}}$ = 11.01s, t$_{\text{ParalIterDecon}}$ = 67.39s, t$_{\text{MitivDecon}}$ = 27.29s) [77]. An image deconvolved using PURE-LET (from [77]) is indicated in Fig. 19 below:

III. NOISE REMOVAL
Noise in microscopical imaging represents an inherent problem faced by most algorithm developers today in the never-ending quest to push the envelope of optical resolution to greater heights. Noise may arise from numerous optical components (such as the field lens, condenser and objective optics, the tube lens, the slide on which the specimen is
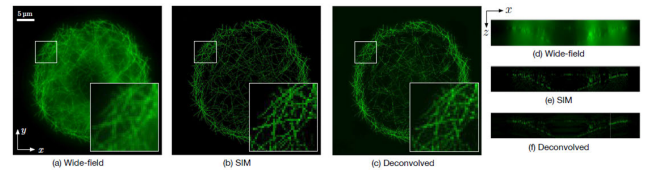

FIGURE 19. A comparison of epifluorescence micrographs of microtubules in a *Drosophila* S2 cell imaged under (a) widefield, (b) SIM and (c) deconvolved using PURE-LET, as according to [76]. (d) – (f): the *x-z* sections for (a) – (c) respectively. Figures adapted from [77], © 2017 IEEE.

mounted, etc), although primary sources of noise are stray background light, delaminated microscope optics (for older scopes), as well as dark current of the sensor or camera. Hence, it would be prudent to mention that out-of-focus blur (generated by optical planes other than image plane) is *not* considered as noise in this context, although the blurring contributes to the background signal to be removed in the deconvolution of an acquired image. A prior in-depth analysis on numerous popular cutting-edge denoising algorithms was conducted by [79], and is as highlighted in Table 2 below:

Despite the wide array of denoising algorithms holistically described in Table 2, a key emphasis in the current context is placed on *other* image denoising algorithms which are not surfaced in Table 2 (but commonly utilized in optical microscopy), as follows:

A. WIENER FILTERING
The Wiener filter remains one of the most popular denoising algorithms employed in the image processing workspace today, having been developed and proposed over 7 decades ago by [12]. Mathematically, the Wiener filter may be represented (in the frequency domain and assuming standard white Gaussian noise having variance σ^2) by the following equation (from [92]):

$$W = \frac{R(\omega) \cdot S_{pp}(\omega)}{|R(\omega)|^2 S_{pp}(\omega) + \sigma_n^2} \qquad (47)$$

where $R(\omega)$ is the FT of the convolved input signal and $S_{pp}(\omega)$ is the power spectrum of the input projection [92].

Nonetheless, this indicates that the input projection S_{pp} is unknown, necessitating a viable estimate of S_{pp} to be used instead. Here, [92] proposes the use of the relation $S_{ss}(\omega) = S_{pp}(\omega) \cdot |R(\omega)|^2$, allowing (47) to be simplified into the following:

$$W = \frac{S_{ss}(\omega)}{S_{ss}(\omega) + \sigma_n^2} \cdot \frac{1}{R} \qquad (48)$$

which represents 2 filters (denoising and inverse filtering) operating in tandem in the Fourier domain [92].

Wiener filtering is often employed when noise (high frequency signals) is present in the image and seeks to silence/attenuate these frequencies based on their signal-noise ratio (SNR), while simultaneously implementing image deconvolution [92]. A single image subjected to motion-blurring (linear motion: 12, angle of camera

IEEEAccess·

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

TABLE 2. Categorization of Current Image Denoising Procedures (Adapted from [79]).

Signal Model	Noise Model	Acronym	Underlying Algorithm	Advantages	Setbacks
Deterministic	Additive	i) TV [80] ii) TV-ICE [81] iii) NLM [82] iv) Analysis K-SVD [83] v) Wavelet Thresholding [84] vi) BM3D [85]	i) Convex optimization ii) Iterative local update of pixel values iii) Non-local filtering iv) Dictionary learning v) Wavelet domain filtering vi) Non-local wavelet-based filtering	i) Highly efficient on piecewise constant images ii) Works well for small areas iii) Good for textured surfaces iv) Same as NLM v) Moderately good performance for *HeLa* cells vi) Optimum when tested with *HeLa* cells	i) Staircase effect observed with real images ii) Unable to discriminate between irregularities for similar ROIs iii) Contrast affected, textural patterns depicted for synthetic image iv) Slow and unable to resolve high noise images v) Edge artifacts identified vi) Visual artifacts found for synthetic image
	Multiplicative	vii) MIDAL [86]	vii) Variable splitting convex optimization	vii) Universally good	vii) Contrast affected, unable to resolve texture
	High Frequency	viii) FSR [79]	viii) Combination of various sparsity-based reconstructions	viii) Depicts very good results over all tested image types in [79]	viii) Identifiable staircase effects and hazing of edges for synthetic image
Probabilistic	Poisson	ix) TV-MAP Poisson [87] x) Poisson EM-TV [88] xi) NLM-Poisson [89] xii) NLPCA [90] xiii) PNLW [91]	ix) Bayesian x) Convex Optimization xi) Non-local filtering xii) Principal component analysis xiii) Non-local filtering	ix) Overshadows TV-ICE on large ROIs x) Surpasses TV for Poisson noise xi) Efficient for synthetic image in the presence of Poisson noise xii) Good for *HeLa* cells xiii) Good for Poisson noise	ix) Contrast affected for synthetic cell image [79] x) Observed staircase effect xi) Quality affected when Gaussian noise is present xii) Textures are visible, but structures blurred for synthetic image xiii) Contrast loss when noise models are mixed

shake: 11°) and Gaussian noise (filter size: 4, $\sigma = 0.75$) in MATLAB (© 1984-2019 The MathWorks, Inc.), followed by Wiener deconvolution is portrayed in Fig. 20 below:

B. MEDIAN FILTERING

The median filter is well-adapted to removing salt-and-pepper/granular noise in images, which often results from the amplification of dark current/background caused by using high gain settings to compensate for poor S/N ratios. Median filtering involves replacing the RGB values of a pixel with the median RGB values of its neighboring pixels, following the ranking of their intensities [93]. Fundamentally based
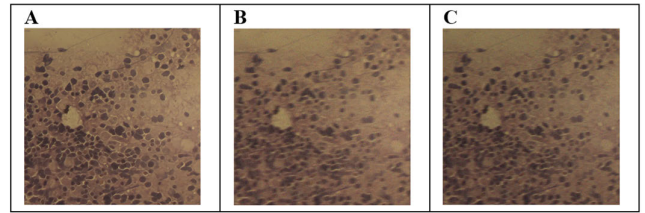

FIGURE 20. **A** The original image (bone marrow smear); **B** the motion-blurred and noisy image; **C** the Wiener-deconvolved image. Image processing was done in MATLAB (© 1984-2019 The MathWorks, Inc.).
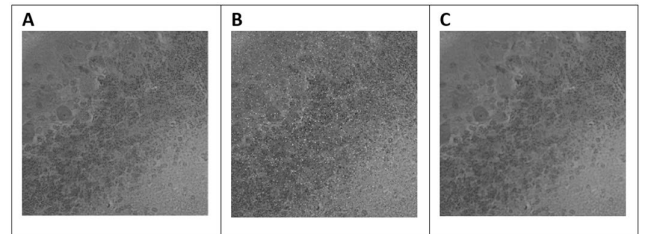

FIGURE 21. **A** The original image (neuronal tissue); **B** The noisy image, ingrained with 'salt-and-pepper' noise; **C** The median-filtered image. Image processing was done in MATLAB (© 1984-2019 The MathWorks, Inc.).
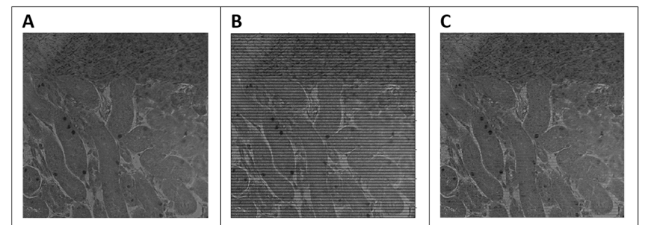

FIGURE 22. **A** The original image (mammalian adrenal gland section); **B** The noisy image (exhibiting a periodic signal artifact); **C** the FFT-denoised image. Image processing was done in MATLAB (© 1984-2019 The MathWorks, Inc.).

on the intensity distribution of the neighboring pixels (rather than a single pixel), median filtering is resilient to statistical outliers, while also being relatively easy to implement with a reduced probability of blurring (which is dependent on the size of the filter kernel used) [93]. An image denoised using median filtering is shown in Fig. 21:

C. FFT FILTERING

FFT filtering for denoising is another approach often utilized when the noise present in the image is relatively well-distributed across the image and can be reduced through the use of a low pass filter to exclude the high frequency components corresponding to the noise [93]. In addition, FFT is also commonly employed for removing a periodic signal (contributing to an *artifact*) in the image [94]. The latter implementation of the FFT denoising filter MATLAB code (as discussed in [94]) on a single noisy image and its output is described by Fig. 22:

IV. ARTIFICIAL INTELLIGENCE (AI) AND ITS ROLE IN OPTICAL NANOSCOPY

Artificial Intelligence (AI) represents a relatively new domain to have its concepts infused into the realm of

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

optical microscopy. Although AI is an encompassing concept which incorporates numerous machine learning (ML) and deep learning (DL) algorithms which may be further sub-classed into supervised, semi-supervised, unsupervised and reinforcement learning algorithms [95] and [96], a particular class of DL algorithms known as *convolutional neural networks* (CNNs) has been successfully deployed in computational imaging and object detection applications. Some examples of research studies and applications utilizing CNNs in this respect include [97] and [98].

In the employment of deep neural networks (DNNs) for computer vision and image processing, several factors play an essential role in the configuration and fine-tuning of such a network. These include the loss function [99], the number of epochs [100], the convolution kernel used ([101] and [102]), backpropagation, etc. Further details pertaining to each of these aspects and their roles in DNN-training for image analysis are discussed in [103], with some additional contributors (as a complement to [103]) being described in Table 3:

As DNNs represent an emerging field with an ever-increasing utility in image analytics, numerous current studies are focused on developing DNN models which seek to extend the current functionality of deconvolution algorithms for image processing. Notably, a similar trend has been observed in optical microscopical imaging, with some prominent research efforts in this respect being exemplified in [117], [118] and [119]. This is in addition to other recent *in silico* (albeit non-DNN) approaches (such as [120] and [121]) to achieve super-resolution nanoscopy. It would be prudent at this juncture to emphasize the potential relevance of AI-based approaches in breaching the Abbe diffraction limit (as opposed to the traditional deconvolution algorithms as discussed in Section 1 previously) since the PSF is often a variable function across the image volume, hence DNNs trained to recognize these variations would be better adapted in resolving an image. In this context, we evaluate 3 recent research efforts in the present review, with particular emphasis on their employed DNN architectures, as well as the efficacy of their proposed models in extrapolating the bounds of optical microscopy into the nanoscopy domain. Further details on each of these approaches are provided in the following sub-sections.

A. ANNA-PALM [122]

An eminent study in *smart nanoscopy* refers to ANNA-PALM [122] which (as its name implies) refers to the utilization of AI-based approaches to achieve super-resolution (SR) fluorescence microscopical imaging akin to that of the empirical technique known as PALM [123]. According to [122], the developers of ANNA-PALM utilized a special conditional GAN (cGAN) named A-net which was based on the pix2pix architecture and comprised of 3 ANNs as follows (from [122]):
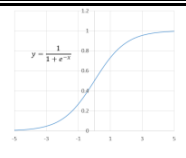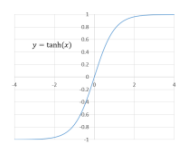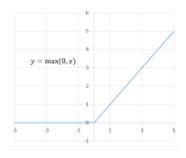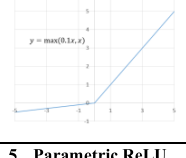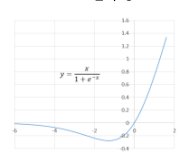
TABLE 3. Hyperparameters often associated with DNN training (adapted from [112]).

DNN Aspect	Description
Optimization functions (including Gradient Descent, SGD, Mini Batch Descent, NAG and Adam optimizer)	Optimization functions are commonly employed during the training of an ANN to minimize the computed error between the predicted values and the target values. There exist numerous examples of popular optimization functions, which may be classed into *first-order*, *high-order* and *heuristic derivative-free* optimizers [104]. First-order optimizers utilize the gradient of the loss function, a popular example of which is Stochastic Gradient Descent (SGD) [104], while high-order optimizers involve the Hessian (second order derivative) matrix (H) which, though being more challenging caused by estimating its inverse (H^{-1}), provides quicker convergence through analysis of the surface curvature [104]. Heuristic derivative-free optimizers (such as coordinate descent [104]) are employed when the derivative of the objective function f (and its limits) are inaccessible due to f being difficult to compute or subject to noise [105].
	Gradient descent (GD) is a mathematical process used to discover the minimum turning point in a plot of a parabolic function. GD operates by iteratively updating the input variables in the opposite direction to f during backpropagation, thereby minimizing the cost function. The step size in each iteration is determined by the learning rate η, which (in turn) influences the number of iterations required to reach the desired global optimum for f [104]. Mathematically, this algorithm may be represented as follows (from [104]):

$$\frac{\partial L(\theta)}{\partial \theta_j} = -\frac{1}{N}\sum_{i=1}^{N}\left(y^i - f_\theta(x^i)\right)x_j^i \qquad (49)$$

$$\theta_j' = \theta_j + \eta \cdot \frac{1}{N}\sum_{i=1}^{N}\left(y^i - f_\theta(x^i)\right)x_j^i \qquad (50)$$

given $L(\theta) = \frac{1}{2N}\sum_{i=1}^{N}\left(y^i - f_\theta(x^i)\right)^2$ and $f_\theta(x) = \sum_{j=1}^{D}\theta_j x_j$ and where $L(\theta)$ is the loss function, $f_\theta(x)$ is the original function, θ is the optimization parameter, N and D are the number of training samples and input features respectively, x^i is the independent variable and y^i is the target output [104].

Nonetheless, up-scaling of the GD algorithm is computationally expensive and time-consuming, popularizing the use of a variant known as *stochastic* GD (SGD) [104]. Here, a *single* data point is randomly-sampled from the set at each iteration, with an alternative approach (termed *mini-batch* GD) sampling a *subset* of data points (instead of individual points) for gradient computation and update [104]. Another popular GD algorithm, called the *Nesterov Accelerated Gradient* (NAG), incorporates the use of momentum as a means to accelerate the process of gradient descent initially, but slow it down as the error function approaches its minimum [104].

Other optimization algorithms (such as AdaGrad and AdaDelta) are also sometimes used – these are able to *adapt* (hence the prefix 'Ada') the learning rate η based on the previous gradients of the input parameter [104]. AdaGrad however has a couple of limitations – (i) the initial value of η has to be manually determined and (ii) the learning rate diminishes with each iteration of the algorithm due to an increasing gradient of descent, the latter being effectively resolved through imposing a constraint on the *number* of past learning rates to be considered in AdaDelta [104]. Another popular alternative optimizer, termed Adaptive Moment Estimation (Adam) computes and stores the adaptive learning rates for each parameter (including averages of the past squared gradients V_t and past gradients m_t) [104]. Through this approach, Adam is able to outperform other optimizers in its convergence rate (being comparable to SGDNesterov in terms of speed) while being able to adapt the learning rate range for *individual layers*, rather than manual selection (as in SGD) [106]. A mathematical representation of the Adam optimizer (based on the input parameter θ, $\alpha = 0.001$, exponential decay rates $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$) is provided as follows (adapted from [106]):

IEEE Access

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{51}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \tag{52}$$

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{m_t/(1-\beta_1^t)}{\sqrt{v_t/(1-\beta_2^t)}+\epsilon} \tag{53}$$

where m_t and v_t are estimates of the moments of the mean and uncentered variance of the gradients respectively [106], and (55) is the final equation for updating the input parameter θ [106].

Nonetheless, a recent study [107] has noted current adaptive optimizers (including Adam) to exhibit poor generalization as compared to SGD and are even unable to converge when η is extreme and unstable. In this regard, [107] has proposed 2 other alternative adaptive optimizers (AdaBound and AMSBound) which the authors claim are immune to extreme learning rates, while being able to smoothly converge to SGD (or with momentum) when the time step is increased. The interested reader is thus encouraged to refer to [107] for further details.

Normalization	Normalization procedures are often employed in training an ANN where the input values may have different ranges. Through normalization, the output of a neuron has a distribution with zero mean and unit variance [108]. Normalization may be carried out via either (i) rescaling each value to a specific range (e.g. from 0 to 1) or (ii) by an approach known as *batch normalization* (BN), which increases training efficacy through smoothing the network optimization landscape [109]. BN has also been characterized to alleviate the vanishing/exploding gradient issue, while posing as a regularizer [110]. In a single *perceptron* (the building block of a NN), the BN algorithm may be expressed mathematically as follows (from [110]): $$y = g(\hat{h}) \tag{54}$$ $$\hat{h} = \gamma \frac{h-\mu_\beta}{\sigma_\beta} + \beta \tag{55}$$ $$h = \mathbf{w}^\mathrm{T}\mathbf{x} \tag{56}$$ where y is the neuron output, $g(\cdot)$ the activation function, h and $\hat{h}$ are hidden values before and after BN, $\mathbf{w}$ and $\mathbf{x}$ are the weight vector and input to the NN respectively, μ_β and σ_β are the estimated mean and standard deviation of h respectively within an independent batch of samples, γ and β are the scale and shift parameters respectively [110]. (58) is the final normalization equation [110]. Nonetheless, it would be essential to note that BN may also be prone to issues such as intra-batch dependence of sample loss and a discrepancy between the batch and network normalization parameters when using the batch for computing the mean and variance of the dataset [111]. In this regard, other approaches (such as weight normalization or layer normalization) as discussed in [111] may be utilized. In addition, the interested reader is also encouraged to explore [112] and [113] for several other normalization algorithms (such as spectral normalization, switchable normalization, etc).
Activation Function	Activation functions are often employed as 'mathematical gates' regulating the outflow from an input node, either by (i) switching the neuron output on or off (depending on a defined rule/threshold being met) or (ii) mapping the input signals (via a suitable transformation function) into output values required for proper functioning of the ANN [114]. Activation functions may be grouped into 3 main classes – (i) binary step (which switches on/off the output based on the defined threshold being met), (ii) linear and (iii) non-linear (the only type allowing backpropagation and multiple layers to be stacked to form a DNN) [114]. Currently, there exist numerous activation functions, of which 7 commonly utilized non-linear variants (as detailed in [114]) have been consolidated and outlined as follows (all graphs plotted using Desmos [115]):

Activation Function Types	Description
1. **Sigmoid / Logistic** $$\sigma(x) = \frac{1}{1+e^{-x}}$$	Allows a smooth transition of output values in the range [0, 1] but is prone to vanishing gradients (when the input value is very high or low).

2. **TanH / Hyperbolic Tangent** $$f(x) = \tanh(x)$$		Is zero-centered and similar to the sigmoid function but is also prone to the same setbacks faced by the sigmoid activation function.
3. **Rectified Linear Unit (ReLU)** $$f(x) = \max(0, x)$$		Is computationally efficient and non-linear but is unable to perform backpropagation when inputs are negative or close to zero (the dying ReLU problem).
4. **Leaky ReLU** $$f(x) = \max(0.1x, x)$$		Has the advantages of ReLU (without the dying ReLU problem), but predictions are inconsistent when the input is negative.
5. **Parametric ReLU** $$f(x) = \max(a\delta, \delta)$$		Allows the negative input values to be handled effectively while permitting backpropagation with all the benefits of ReLU. However, it is inconsistent for variable inputs.
6. **Softmax** $$\sigma(z)j = \frac{ezj}{\sum_{k=1}^k ezk}$$ for $j = 1, \ldots, k$		Is able to handle multi-classed variables (having several probable outputs), thereby being particularly useful for the final output layer of an ANN/DNN.
7. **Swish** $$f(x) = \frac{x}{1+e^{-x}}$$		Developed by researchers at Google, reportedly outperforming ReLU with similar computational efficiency. Further details of Swish are discussed in [114]. (***Authors' Note***: We propose that this function may be further generalized as follows: $$f(x) = \frac{x}{1+e^{-\beta x}}$$ for some trainable parameter β).

i) A generator network G (principled on U-net with skip connections [124] and 16 convolutional layers) responsible for creating the SR image,

ii) A "low-resolution estimator" network Q (having 4 convolutional layers) responsible for generating the error map for low-resolution images (reducing the input image resolution by a factor of 4 in each dimension), and
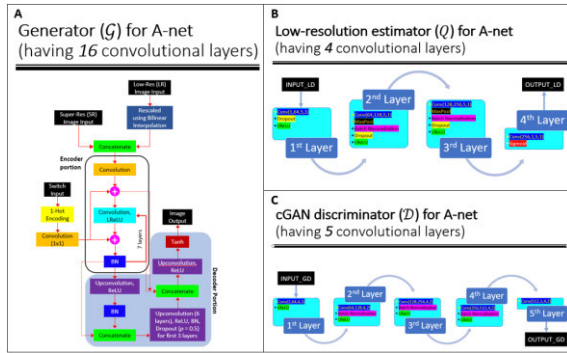
S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

FIGURE 23. The A-net architecture developed by [122] and employed in ANNA-PALM [122]. A The generator network *G*; B The "low-resolution estimator" network *Q*; C The cGAN network *D* [122]. All figures are drawn by the authors of the current review and are based on the original A-net architecture as described in [122].

FIGURE 24. A comparison of microtubule images obtained using Widefield, PALM and ANNA-PALM. Notice the agreement between images using ANNA-PALM with *both* widefield (a) and sparse PALM (b) inputs, as well as with the ground truth/dense PALM output (c). Adapted from [122].

iii) A cGAN adversarial network *D* (with 5 convolutional layers) which produces the loss and decides if the input image is obtained empirically or artificially-generated by *G*

Training of the A-net was conducted using randomly undersampled (*sparse*) PALM images (having resolutions of $256m \times 256m$, with $m \in \mathbb{Z}$) (as an input), and their associated dense PALM variants (for the desired output) [122]. The *G* network also contains a computational selector (switch) allowing for different structures (e.g. nuclear pores or microtubules) to be resolved [122], while the error map may be defined by the following equation:

$$E_Q(A, W) = (1MS\_SSIM(Q(A), W)) \times (Q(A) + W) \quad (57)$$

where E_Q refers to the error generated by the network *Q*, *A* is the reconstructed SR image and *W* is the widefield input image [122].

The DNN architecture of the A-net (further details of which are elaborated in [122]) is illustrated in Fig. 23:

According to the authors of [122], the convolutional layers in the ANN are coupled with batch normalization procedures, while the dropout layers (dropout rate = 50%) are only utilized during training but disabled during image inference [122]. The activation functions used are ReLUs [$f(x) \rightarrow \sup(x,0)$] or 'leaky' ReLUs [$f(x) \rightarrow \sup(x, 0) + \inf(\varepsilon x, 0)$, where $\varepsilon = 0.2$], except the final layer of *G* (which uses a tanh function) and the last layer of *Q* (which uses a sigmoid function) [122]. The cGAN network in (iii) utilized least squares loss functions (being empirically superior to the log loss functions often employed in GANs [122]), which are defined mathematically as follows:

$$L_D(G, D) = E_{(c,x) \sim p_{data}(c,x)}[D(c,x) - 1]^2$$
$$+ E_{c \sim p_{data}(c), z \sim p_z(z)}[D(c, G(c,z))]^2 \quad (58)$$
$$L_{cGAN}(G, D) = E_{c \sim p_{data}(c), z \sim p_z(z)}[1 - D(c, G(c,z))]^2 \quad (59)$$

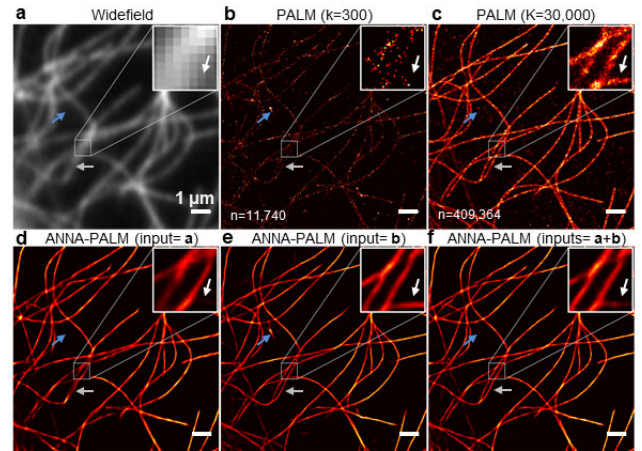Visually, a comparison of the results derived from ANNA-PALM with empirical PALM suggests a high degree of correlation between these 2 image sets (as indicated in Fig. 24), implicating the evident capability of the A-net in performing *in silico* nanoscopy. Nonetheless, a potential limitation of this approach refers to its mapping of learnt features to unknown images (as surfaced by [125]), thereby obscuring the detection of otherwise present anomalies, which might occur at a relatively low percentage in endogenous tissue.

B. DEEP-STORM [126]
Deep-STORM represents a then-novel approach in providing a rapid, precise way to obtain SR images from randomly-emitting fluorophores without the need to acquire any additional data inputs (e.g. the convolution PSF of an optical system, etc) [126]. As its name implies, Deep-STORM utilizes a DNN coupled with widefield epifluorescent microscopy to *directly* generate SR images without localizing individual emitters, unlike traditional empirical approaches (such as STORM or PALM) [126]. The CNN-based architecture of Deep-STORM comprises of an encoder (*convolution*) and a decoder (*deconvolution*) segment, the former composed of three 3×3 convolutional layers (with *increasing* depth) alternated with 2×2 max pooling layers, while the latter consists of 2×2 upsampling layers alternated with 3×3 convolutional layers (with *decreasing* depth) [126]. Batch normalization and ReLU were also incorporated within the CNN architecture (with 1.3 million nodes for training), appended with a 1×1 depth-reducing convolutional filter (having a linear activation function) for the final SR image generation [126]. Visually, the architecture of Deep-STORM may be described by Fig. 25 (from [126]):

Deep-STORM was trained using 10k pairs of simulated data points, first generated as twenty 64×64 pixel frames using the ThunderSTORM plugin [127] within ImageJ ([128] and [129]). This was subsequently processed by extracting 500 random 26×26 pixel ROIs from each image, which are then up-sampled by 8-fold, forming 208×208 pixel regions [126]. The training was done over 100 epochs (having

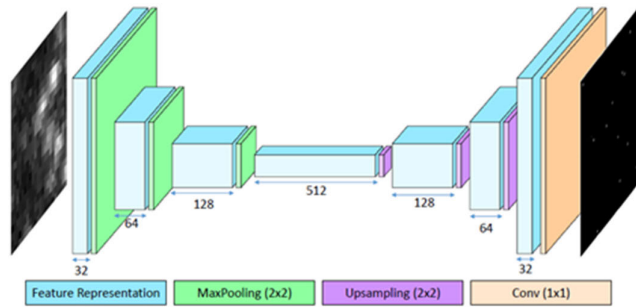**FIGURE 25.** The CNN architecture as utilized in Deep-STORM [126]. Figures reprinted with permission from [126], © The Optical Society.

FIGURE 26. A visual comparison on the performance of Deep-STORM (the last image) against the raw input image (the first image) and CEL0 (the middle image) algorithms, using A a simulated image set, and empirical images of B quantum dots and C fluorescently-labelled microtubules. In all 3 comparisons, Deep-STORM surpasses that of CEL0, being closer to the ground truth. D The runtimes for each method. Here (too) Deep-STORM clearly surpasses the other methods. Figures adapted with permission from [126], © The Optical Society.

a batch size of 16 frames) with an Adam optimizer (Gaussian kernel, $\sigma = 1$, learning rate $\varepsilon = 0.001$) [126]. The loss function used for training was based on regression analysis, utilizing the squared Euclidean distance between the 2D Gaussian-convolved network's predicted points and its corresponding ground truth, while having an l_1 penalizer [126]. A mathematical representation of the loss function is indicated as follows (adapted from [126]):

$$\ell(x, \widehat{x}) = \frac{1}{N} \sum_{i=1}^{N} \left\| \widehat{x_i} \circledast h_g - x_i \circledast h_g \right\|_2^2 + \left\| \widehat{x_i} \right\|_1 \quad (60)$$

where $\widehat{x_i}$ refers to the predicted values, x_i refers to the ground truth, h_g is the Gaussian kernel, N is the size of the training dataset, and $\circledast$ refers to the convolution operator (adapted from [126]).

Verification of the network accuracy was performed using both empirically-obtained datasets and simulated data points {as (i) samples of quantum dots imaged under different laser powers and composited and (ii) obtained from [130] respectively}. In all instances, Deep-STORM exhibited significantly-reduced run-times while maintaining the overall integrity of the SR image. A visual representation comparing the output from (and runtimes of) Deep-STORM with CEL0 [131] is shown in Fig. 26:

From the results gleaned from [126], the authors have emphasized the key strengths of Deep-STORM as being able to generate SR *videos* of fluorescently-labelled structures (without much intervention from the user), although they have also acknowledged that it did not provide localization-specific information of the resolved molecules, despite using this information for creating the SR image [126].

C. ISONET-1 AND ISONET-2 [132]

In 2017, a couple of DNN architectures (named IsoNet-1 and IsoNet-2) were proposed by [132] for restoring isotropic resolution in the axial plane (induced primarily by the severe anisotropy exhibited by the oblique PSF in the z-axis). Both IsoNet-1 and IsoNet-2 are based on a CNN architecture, with the main difference between the two being that IsoNet-1 utilizes a traditional CNN kernel-downsampling approach (with a ReLU activation function) while IsoNet-2 is based off a U-Net architecture with skip connections [124]. In the
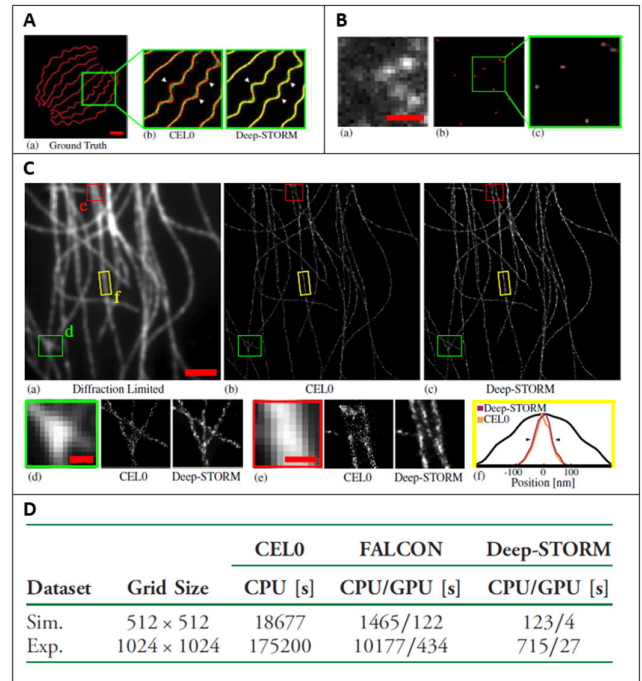
development of IsoNet, the authors of [132] considered 2 primary factors – (i) the process of formation of a stereographic volume in fluorescence microscopy and (ii) the anisotropic PSF volume, leading them to propose a DNN model based on sparsity-induced super-resolution coupled with image deconvolution. Nonetheless, [132] has also highlighted the absence of ground truth data for training their DNN, leading them to utilize the same dataset for training (a concept described as *self super-resolution*). In this regard, [132] has emphasized a key difference between conventional methods of image restoration and their proposed approach – the former utilizes generalized iterative signal-decoding procedures (without any information on the sample) while the latter seeks to understand the formation of the observed blurred image from its actual uncorrupted impulse response and potentially achieve the resolution attainable in the *xy* image plane for the *z*-axis, from the equation $y = S_\sigma\left(\widetilde{h}*x\right)$ (where $\widetilde{h}$ is the 3d rotated PSF and S_σ is the downsampling function in the *z*-axis for a defined factor σ). This, according to [132], was done through inverse mapping from assembled lateral image ROIs, with the authors of [132] choosing to deconvolve the orthogonally-rotated PSF (in the *xy*-plane) by the average PSF volume (assuming a spherically-bounded PSF), which they referred to as h_{iso}. This method has a clear advantage of minimizing the PSNR loss as opposed to directly deconvolving the observed impulse response with a standardized

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

function, although one may surmise that it is founded on the assumption of a uniform PSF being distributed across the observed volume, which (in most instances) would be untrue. The results obtained by [132] seem promising, with relatively high PSNR values reported for both IsoNet-1 and IsoNet-2 – $\text{PSNR}_{\text{IsoNet}-1}$: 32.47, $\text{PSNR}_{\text{IsoNet}-2}$: 35.61 for simulated nuclei samples with a gaussian split PSF, as compared to traditional RL deconvolution (PSNR: 27.48), SRCNN (PSNR: 25.89) or the original blurred input image (PSNR: 25.84). To further verify the efficacy of their approach, [132] applied a 3D watershed algorithm on cellular boundaries, with a reported SEG of 0.913 (for IsoNet-2) as opposed to 0.742 (for the blurred input image) and 0.923 (assuming an isotropic PSF). With real data, equally promising results are depicted in [132], leading the developers of IsoNet to propose how utilization of their approach would aid in eliminating phototoxic cell damage. However, the authors of [132] acknowledged (as well) a key limitation of their approach – that a sampling rate greater than the Shannon limit is required for effective axial super-resolution.

D. DEEP-Z [133]

As a putative successor to IsoNet-1 and IsoNet-2 [132], Deep-Z was conceived by [133] as a way to generate 3D optical microscopical stacks from a 2D optical image plane. Deep-Z uses a conditional GAN (cGAN) trained with a Z-stack of 2D fluoromicrographs [each coupled with a digital propagation matrix (DPM) specifying the Z-distance between the target surface and the input image plane] and their corresponding ground-truth fluoromicrographs acquired at the target plane as specified by the DPM [133]. Architecturally, the least-square GAN employed in Deep-Z is composed of the following parts:

i) A U-Net-based *generator G* consisting of 5 downsampling blocks in its descending arm and 4 upsampling blocks in its ascending arm, with each block containing 2 convolutional layers. The mathematical equations governing the blocks in each of these arms are described as follows:

$$x_{k+1} = x_k + \text{ReLU}\left[\text{CONV}_{k_2}\left\{\text{ReLU}\left[\text{CONV}_{k_1}\left\{x_k\right\}\right]\right\}\right] \quad (61)$$

where CONV refers to the convolution operator (with bias), '+' indicates a residual connection, k_1 and k_2 (as subscripts of the CONV operator) refer to the number of channels {where

$$k_m = \begin{cases} 25, & \text{when } m, n = 1 \\ 6\,(2+m)\,(2^n), & \text{where m} = 1\text{or}2, \text{n} \in \mathbb{Z}, \\ & 1 \leq n \leq 5 \end{cases}$$

[133]}. According to [133], mismatches between channel numbers of input and output tensors were rectified through zero padding, with a 2×2 max pooling layer having a stride of 2×2 (for a $2\times$ downsampling) between 2 adjacent downsampling blocks [133]. The

5th downsampling block is connected to the upsampling path [133].

$$y_k = \text{ReLU}\left[\text{CONV}_{k_4}\left\{\text{ReLU}\left[\text{CONV}_{k_3}\right.\right.\right.$$
$$\left.\left.\left.\{\text{CAT}\,(x_{k+1}, y_{k+1})\}\right]\right\}\right] \quad (62)$$

where CAT is used to join the tensors along the channel direction, i.e. $\text{CAT}(x_{k+1}, y_{k+1})$ joins tensor x_{k+1} to tensor y_{k+1}. Here, $k_m = 12(6 - \text{m})(2^\text{n})$ for $m = 3$ or 4, $n \in \mathbb{Z}$, $1 \leq n \leq 4$ [133]. An up-convolution (convolution transpose) block connects adjacent upsampling blocks, upsampling the image by $2\times$, with the last block (a convolutional layer) combining the 48 channels into a single output channel [133].

ii) A 6-block CNN for the *discriminator D*, with each block expressing the following mathematical transformation:

$$z_{i+1} = \text{LReLU}\left[\text{CONV}_{i_2}\left\{\text{LReLU}\left[\text{CONV}_{i_1}\left\{z_i\right\}\right]\right\}\right] \quad (63)$$

where z_i refers to the input value, z_{i+1} refers to the output value (for some level i), LReLU (leaky ReLU) has a slope of 0.01, i_1 and i_2 (as subscripts of the CONV operator) refer to the number of channels [where $i_m = 3(2^{m+n+2})$ for $m = 1$ or 2, $n \in \mathbb{Z}$, $1 \leq n \leq 6$] [133].

Following the discriminator, mean pooling is used to downscale the parameter set size to 3072, appended with fully connected (FC) layers (size: 3072×3072) employing leaky ReLU functions, and a single FC layer (size: 3072×1) using a sigmoid activation function [133]. The convolutional kernels used are 3×3 matrices with a 1-pixel stride in each dimension, except for the second CONV in (63) which has a 2-pixel stride in each dimension (for a resolution reduction factor of 2). Xavier initialization was used for the weights, with bias set to 0.1. The output (a discriminator score) lies in the range [0, 1] with '0' being false and '1' being true. A detailed visual representation of the Deep-Z architecture is provided in [133] and demonstrated in Fig. 27 below:

Deep-Z was trained in 2 phases – input images to G were $256 \times 256 \times 2$ (the second channel corresponding to the DPM) while those to D were either (i) outputs from the generator (256×256) or (ii) the target $z^{(i)}$. The loss functions employed for the generator L_G and the discriminator L_D are as follows (from [133]):

$$L_G = \frac{1}{2N}\sum_{i=1}^{N}\left[D\left(G\left(x^{(i)}\right)\right) - 1\right]^2$$
$$+ \alpha\frac{1}{2N}\sum_{i=1}^{N}\text{MAE}\left(x^{(i)}, z^{(i)}\right) \quad (64)$$

$$L_D = \frac{1}{2N}\sum_{i=1}^{N}\left[D\left(G\left(x^{(i)}\right)\right)\right]^2$$
$$+ \frac{1}{2N}\sum_{i=1}^{N}\left[D\left(z^{(i)}\right) - 1\right]^2 \quad (65)$$

where N is the batch size, $G(x^{(i)})$ refers to the generator output for $x^{(i)}$, $z^{(i)}$ is the target label, MAE refers to mean absolute error, α (the regularization parameter for the GAN and MAE loss in L_G) = 0.02 [133].

IEEE *Access*

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy
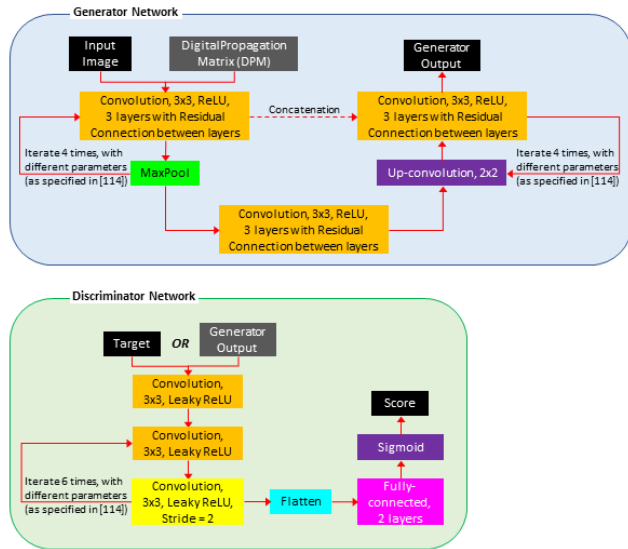

FIGURE 27. The CNN architecture as utilized in Deep-Z [133]. All figures are drawn by the authors of the current review and are based on the original Deep-Z architecture as described in [133].
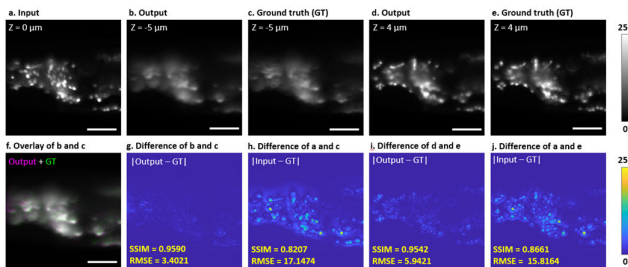

FIGURE 28. Digitally refocused *C. elegans* fluoromicrographs. Notice the improved SSIM and RMSE of the Z = −5µm and Z = 4µm layers elucidated through Deep-Z (when compared with the GT images at these positions), as opposed to the original input image (panel a). Figure adapted from [133].

An Adam optimizer was utilized during the training (learning rate, $\varepsilon_{L_D} = 3 \times 10^{-5}$, $\varepsilon_{L_G} = 10^{-4}$), with validation performed every 50 iterations and the optimal network chosen based on minimizing MAE loss. The DPM values were rescaled by a factor of 0.1, due to the refocusing distance spanning from −10µm to 10µm. On testing, the results obtained from Deep-Z bear a significantly close resemblance to the ground truth (GT) image with the *difference* between these 2 sets of images exhibiting a much higher structural similarity index (SSIM) and lower root mean square error (RMSE) as compared to the input image, implicating the high predictive accuracy of Deep-Z as an *in silico* refocusing method. Sample images indicating this correlation (obtained from [133]) are described in the following figure (Fig. 28):

According to [133], Deep-Z may be putatively extended in future applications as an initial module for neuronal imaging (reducing the need to acquire multiple image stacks), while also potentially incorporating additional acquired information at different optical sections. The advantages of Deep-Z (from our perspective) are clear and represent a well-planned step in achieving computational nanoscopy – axial resolution

limitations (which are more severe than lateral resolution limits for widefield microscopy) coupled with haze (generated from out-of-focus image planes) place a severe limitation and obstacle in breaching the Abbe diffraction limit (as defined in (2) previously). Nonetheless, we also believe that Deep-Z exhibits further growth potential, especially where image corruption by noise is evident (particularly in the presence of low fluorescence signals).

V. DEEP LEARNING FOR IMAGE DENOISING

The use of DL algorithms has long found a niche for solving an ill-posed problem in image denoising, yielding comparatively better results than traditional denoising algorithms (such as Wiener or median filtering as discussed previously). Through developments in the use of ANNs for denoising (as expounded in [134] and [135]), other studies (such as [136]) have proceeded to explore using DNNs for image denoising with promising results – [136] illustrated the use of a 4 hidden-layer CNN (each layer comprising 24 nodes) coupled with a 5×5 convolution kernel to denoise images subjected to Gaussian noise with an unknown variance, producing results which surpassed traditional non-blind denoising algorithms then (CN1/CN2 PSNR = 24.12/24.25 *vs* BLS-GSM PSNR = 23.78, FoE PSNR = 23.02) and at much higher rates ($\sim$42 times quicker than FoE and between 68% – 133% of BLS-GSM). Elemental in this context is that [136] highlighted the need for (i) an *extremely small learning rate for the final layer* of the CNN (0.001) as opposed to a larger learning rate for all other layers (r = 0.1) and (ii) the utilization of a *gradient learning algorithm* to tune multiple parameters based on the input images. Similarly, [137] proposed a smart denoising algorithm termed *content-aware image restoration* (CARE), which seeks to denoise fluorescence microscopical images while simultaneously achieving axial super-resolution with 20-fold faster image acquisition and at 60-fold lower light intensities, although [137] also indicated CARE as being constrained by its image-specific training dataset. In this regard, it would be prudent to further evaluate the utility of other DNN models for image denoising, of which 3 such architectures are described in greater detail in the following sections. For a more elaborate discussion on the various denoising approaches, the interested reader is also encouraged to refer to [138], which seeks to provide a holistic overview of the image denoising algorithms currently available (the following 3 algorithms not being discussed in this survey).

A. MULTI-LEVEL WAVELET-CNN (MWCNN) [139]

Multi-level Wavelet-CNN (MWCNN) is based on an integration of 2D Haar wavelets with a traditional CNN architecture, as a means of optimizing the size of the receptive field while potentially reducing computational complexity [139]. Here, the MW packet transform (as the skeletal framework of MWCNN) is alternated with a convolutional block, the former involving the use of 2D DWT for splitting of the sub-band images $\mathbf{x}_i$ into i^n images for a n-level MW

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy
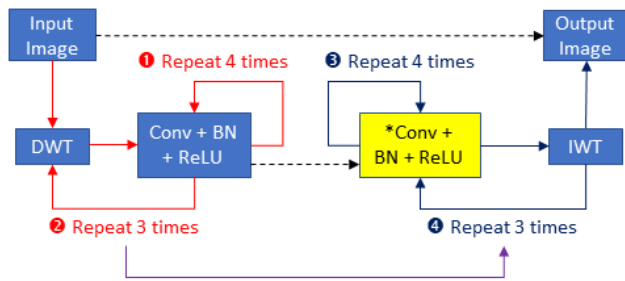
IEEE *Access*

FIGURE 29. The U-Net architecture used as a skeletal framework for development of the MWCNN algorithm (as discussed in [139]). The figure shown here describes a 3-level wavelet-CNN architecture. The yellow-filled box implements the convolution kernel followed by a batch normalization (BN) and ReLU activation for 4 times in all layers, except for the penultimate layer, where only convolution is applied. Figure adapted from [139].

transform. Downsampling (as a component of the DWT) is then used to pool the responses/outputs from each layer (prior to its upsampling and recombination) followed by an inverse wavelet transform (IWT) for image reconstruction [139]. A generalized visual map of the MWCNN architecture (based on the U-Net structure [124] and as deployed in [139]) is shown in Fig. 29:

Each of the 4 CNN layers in a single block is comprised of a 3×3 convolution kernel, batch normalization and ReLU activation, except for the penultimate CNN layer (where only a convolution operation is employed) [139]. Variations made in the U-Net framework (for this MWCNN model) are as follows: (i) using a DWT and IWT instead of max-pooling and up-convolution respectively, (ii) an increase in the number of feature map channels caused by downsampling and (iii) combining feature maps through element-wise addition in MWCNN, as opposed to concatenation for U-Net [139]. The authors of [139] utilized the ADAM optimizer for training the MWCNN with the following loss function:

$$L(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \| F(\mathbf{y}_i, \Theta) - \mathbf{x}_i \|_F^2 \qquad (66)$$

where $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^{N}$ is the training set having $\mathbf{x}_i$ as the ith ground truth image and $\mathbf{y}_i$ as the corresponding input image. Θ and $F(\mathbf{y}, \Theta)$ refer to the network parameters and the output of the MWCNN model respectively [139].

According to [139], application of the MWCNN algorithm reported significantly high PSNR (dB) / SSIM results for almost all assayed datasets when compared with other algorithms (e.g. 33.17/0.9357 for MWCNN *vs* 32.49/0.9244 for IRCNN on the Urban100 image having noise level $\sigma = 15$, or 32.23/0.8999 for MWCNN *vs* 31.85/0.8942 for DnCNN on the BSD100 image with a scale factor, S of 2) [139]. This clearly demonstrated the efficacy of MWCNN in image denoising, single image SR (SISR) and removal of JPEG image artifacts. Nonetheless, [139] also noted that the image denoising was only performed on grayscale (and not color) images, while only the luminance (Y) channel (in a YCbCr model) was processed with SISR. The measured computational efficiency of MWCNN however was comparable to

other in-class algorithms, exhibiting a run-time of 0.3575s for denoising a 1024×1024-pixel image, as compared to 15.77s for a similar image denoised by RED30, although DnCNN and TNRD depicted run-times of 0.1688s and 0.116s respectively [139]. A similar trend was observed for SISR (0.3167s for MWCNN *vs* 25.23s for DRRN or 0.1411s for LapSRN) and JPG artifacts removal (0.2931s for MWCNN *vs* 14.69s for MemNet or 0.095s for TNRD), all of which were performed on 1024×1024-pixel images [139]. In all these instances, non-bulk image processing pointed to the potential usage of MWCNN for optimal post-processed image quality, although batch processing may favor other comparable alternatives.

In addition to the above, [139] had also compared the use of different wavelets [traditional Haar, *Daubechies-2* (DB2) as well as an amalgamation of Haar in the contracting arm and DB2 in the expanding arm] embedded within their proposed MWCNN framework, coupled with additional frameworks such as WaveResNet and deep convolutional framelets (DCF). The findings from these experiments indicate (i) the significant impact of mismatched pixel information caused by binning of incongruent pixels (termed '*the gridding effect*' [139]) on image restoration, (ii) that a sum (rather than a concatenation) operator should be deployed in the U-Net architecture of the MWCNN and (iii) the preferable use of Haar wavelets throughout the U-Net architecture (as compared to DB2 or Haar/DB2) for optimal image quality (denoted by the PSNR) and efficiency (run-time) of the MWCNN model [139].

B. EDGE-PRESERVING DNN FOR IMAGE DENOISING [140]
In 2018, [140] proposed a DNN/CNN incorporating a Canny filter for edge-preservation as a means of achieving image denoising. The rationale of doing this was (according to [140]) (i) ensuring that detailed morphological variations detected in the specimen are maintained, (ii) discriminating between features and noise in areas where there is a low signal/noise (S/N) ratio, (iii) flexibility afforded through the use of CNNs for region-specific denoising (as opposed to traditional algorithms) and (iv) automated identification and characterization of features by the convolution kernel employed in the DNN. In their DNN, [140] utilized a non-subsampled shearlet transform (NSST) as it is able to effectively map out variations in structural features while incurring less computational cost than the non-subsampled contourlet transform (NSCT), and (ii) resolve sparse signals without requiring additional filters for determining the direction to be resolved (unlike NSCT). Training of the CNN was performed using Canny-determined noiseless edge maps, with noise subsequently introduced and the maps stacked to form a noisy 3D image volume. The resolved edges were used to discriminate between signal and noise, with pixels contributing to edge formation considered as the impulse response signal (the converse being regarded as noise). The CNN architecture of the proposed method comprised a total of 6 layers (2 FC layers, 1 subsampling layer and 3 convolutional layers), with
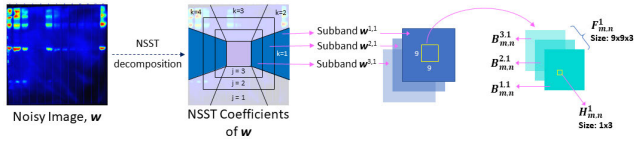
FIGURE 30. 3D block construction for training the DNN ($J = 3$, $\Theta = 4$). NSST decomposition is applied to the noisy image, followed by stacking of 2D blocks ($B_{m,n}^a$) into a 3D block ($F_{m,n}^a$) having $H_{m,n}^a$ as its central vector ($a = 1$ in this figure). Figure adapted from [140].

FIGURE 31. The generalized architecture of MP-DCNN. The figure incorporates CNN layers from the MP-DCNN model, as well as the loss functions used to tune the network [141].

the convolutional layer and the sub-sampling layer equations (the former incorporating a 3D convolution function nested within a traditional ReLU activation) denoted as follows:

$$q_j = \max\left(0, b_j + \sum_i k_{ij} \circledast p_i\right) \qquad (67)$$

$$q_i^{r,c} = \max_{0 \leq m,n < s}\left\{p_i^{r.s+m, c.s+n}\right\} \qquad (68)$$

where b_j is the bias term, k_{ij} is the convolution kernel ($\circledast$ represents the 2D convolution operator), $q_i^{r,c}$ is the (r, c) pixel in the ith output subsampled feature map, p_i is the ith input feature map (having a size of $s \times s$) (adapted from [140]).

A visual description of the DNN architecture is portrayed in Fig. 30 below:

Subsequently, a softmax activation function coupled with SGD having an update rule $z_{i+1} = 0.9z_i - 0.0005bw_i - \beta \cdot \left(\frac{\delta L}{\delta w_i}\right)$, where $w_{i+1} = w_i + z_{i+1}$ was used to refine the weights at each node, allowing the DNN to distinguish noise from signal impulses with the proposed method demonstrating significant improvement in image denoising – for e.g., in the House (256×256) image, the proposed method exhibited a $\text{PSNR}_{\sigma_n=10} = 37.53$ and a $\text{PSNR}_{\sigma_n=70} = 28.96$, as compared to NLFMT ($\text{PSNR}_{\sigma_n=10} = 34.87$ and $\text{PSNR}_{\sigma_n=70} = 25.62$), although [140] acknowledged that NLFMT did not generate artifacts. Another assayed method (BM3D) was also unable to surpass the performance achieved by the proposed method (BM3D: $\text{PSNR}_{\sigma_n=10} = 36.71$ and $\text{PSNR}_{\sigma_n=70} = 27.91$), although BM3D reportedly executed ~4 times more rapidly than the proposed method for the Lena and Barbara images ($t_{\text{Lena}} = 6.54s$ and $t_{\text{Barbara}} = 6.12s$ for BM3D, as compared to $t_{\text{Lena}} = 28.79s$ and $t_{\text{Barbara}} = 26.49s$ for the proposed algorithm) respectively. Nonetheless, the authors of [140] highlighted the inability of the presently-proposed algorithm for denoising ultrasound and color images, which they intend to incorporate in future versions of their method.

C. MP-DCNN [141]
Developed by the authors of [141], MP-DCNN represents a means of achieving image denoising without edge-induced artifacts in a substantially noisy environment through the employment of a joint loss function in a CNN [141]. Here, MP-DCNN utilizes a 3×3 convolution kernel in order to maximize the dimensions of the receptive field to $(2f+1)^2$, where f refers to the CNN depth [141]. This serves to optimize the chances of recovering pixels which have a significantly low S/N ratio. Visually, the architecture of the MP-DCNN network may be described by Fig. 31 below:
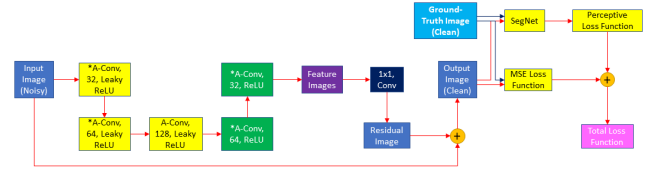
The MP-DCNN architecture consists of 10 convolution layers (and an output layer), the first 5 of which utilize the leaky ReLU activation function and have feature map outputs numbering (32, 32, 64, 64, 128) respectively, while the next 4 convolutional layers use a ReLU activation function, with feature maps numbering (64, 64, 32, 32) respectively [141]. The final convolution layer uses a 1×1 convolution kernel to generate c feature maps, with an added residual unit being employed to speed up the process while improving the overall network performance [141].

The authors of [141] have utilized a couple of loss functions within their developed MP-DCNN network architecture (namely MSE and perceptual loss) combined into a *joint loss function* (i.e. $L_{joint} = L_{MSE} + \lambda L_{SegNet}$), with each of these component losses (L_{MSE} and L_{SegNet}) being defined in (69) and (70) as follows (from [141]):

$$\text{L}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{w \times h}\sum_{j=1}^{w}\sum_{k=1}^{h}\|f_i(j,k) - X_i(j,k)\|^2\right) \qquad (69)$$

$$L_{SegNet} = \frac{1}{w_i h_i}\|C_i(Y - R(Y) - C_i(X))\|^2 \qquad (70)$$

where f is the denoised output image, X is the input image, n is the sample size, w and h are the dimensions (width and height) of the sample image respectively [141].

The results of implementation of MP-DCNN proved promising, with the former depicting relatively superior results when compared to other image denoising algorithms such as WNNM, TNRD and BM3D, amongst others (PSNR values for the BSD68 dataset when utilizing $\text{MP-DCNN}_{\sigma=15} = 32.05$ *vs* $\text{BM3D}_{\sigma=15} = 31.07$ or $\text{TNRD}_{\sigma=15} = 31.42$, $\text{MP-DCNN}_{\sigma=50} = 26.68$ *vs* $\text{WNNM}_{\sigma=50} = 25.87$ or $\text{DnCNN}_{\sigma=50} = 26.23$, etc) [141]. In addition, structural similarity (SSIM) comparisons between MP-DCNN with other image denoising algorithms reveal a similar trend – MP-DCNN clearly outperforms even its closest assayed predecessor (DnCNN) with more significant improvement margins for noisier datasets – SSIM values for the BSD68 dataset when utilizing $\text{MP-DCNN}_{\sigma=15} = 0.8829$ *vs* $\text{DnCNN}_{\sigma=15} = 0.8826$, while $\text{MP-DCNN}_{\sigma=50} = 0.7104$ *vs* $\text{DnCNN}_{\sigma=50} = 0.7076$, with a similar stellar performance reported for MP-DCNN when using other test images [141]. In addition, the developers of MP-DCNN have also found their proposed algorithm to be much more efficient than most other tested algorithms – the run-time of MP-DCNN for a 1024×1024 image was 10.9 ± 0.18 s, as compared to 12.1 ± 0.2s for DnCNN [141]. All these

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE Access

results reported by [141] seemingly suggest the superiority of MP-DCNN over other tested algorithms (such as DnCNN, etc), especially in denoising highly noisy images, although the authors of [141] also acknowledged potential areas for development of their proposed framework (i.e. MP-DCNN) by addressing non-Gaussian noise models, denoising large image datasets, as well as color images and videos [141]. In our perspective, one may also consider evaluating how different activation functions (such as leaky ReLU or sigmoid) may also be utilized in the MP-DCNN framework (particularly in its 6th–9th convolutional layers) rather than the traditional ReLU presently expounded in [141].

VI. COMMERCIAL APPLICATIONS OF AI-BASED DEHAZING / DEBLURRING ALGORITHMS

Although traditional deconvolution algorithms (e.g. Nearest Neighbors, Richardson-Lucy deconvolution, etc) have long been harnessed in the commercial sector within applications such as Huygens Professional (Scientific Volume Imaging B.V.) [23], Imaris 9.3 (© Oxford Instruments 2019) [142] and AutoQuant X3 (© Media Cybernetics, Inc.) [24] amongst others, the recent emphasis on the infusion of AI in optical nanoscopy has spurred some organizations to develop imaging applications utilizing such intelligent nanoscopy techniques, which are the subject of this review and discussed in the subsequent sub-sections. Hence, although the commercial implementation of conventional deconvolution algorithms is not being discussed here, the interested reader is encouraged to explore the widespread uses of these algorithms by visiting these manufacturer websites directly.

A. ADAPTIVE DECONVOLUTION AND LIGHTNING

Adaptive deconvolution (AD) represents a novel approach in computational super-resolution and image deconvolution, through utilizing a decision mask to automatically extract deconvolution parameters for individual voxels within the imaging volume [143]. This eliminates the need for the user to provide a PSF/blurring kernel manually, either by computation or determined empirically through supplying images of sub-diffraction-sized beads (as is utilized in NBD algorithms). AD is integrated as an optional deconvolution method within LIGHTNING (© Leica Microsystems), combining the speed and efficiency derived from parallel GPU processing with the decision mask (used in AD) to provide a fully automated, voxel-specific super-resolved image [143] with near-real-time efficiency, thereby seeking to extend 4 of the 6 vertices of the imaging octahedron exemplified in Fig. 32:

A comparison between AD and traditional/classical deconvolution is depicted in Fig. 33 (adapted from [143]):

The initial pre-processing step of AD involves determining the background (Bg) where the signal-noise ratio (SNR) is computed by approximating the grayscale values $g(x, y)$ for each pixel based on its neighborhood and a smoothing kernel f_b as follows:

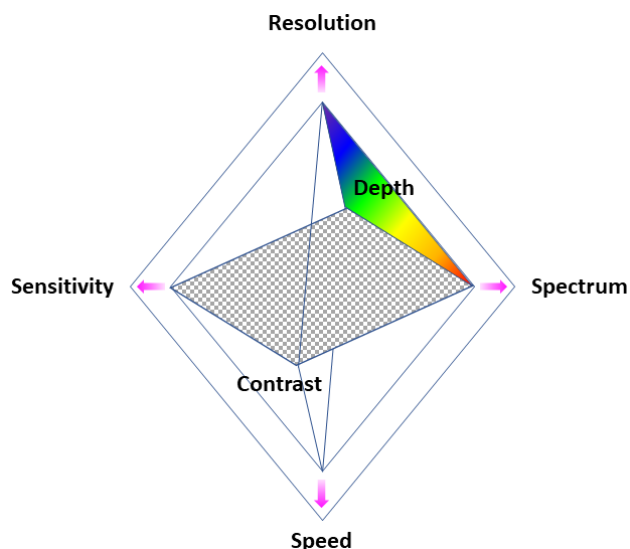$$SNR = S[g(x, y), f_b(x, y)]/N \tag{71}$$

FIGURE 32. The imaging octahedron of optical microscopy. LIGHTNING (© Leica Microsystems) seeks to extend 4 of the 6 vertices (namely Resolution, Sensitivity, Speed and Spectrum).
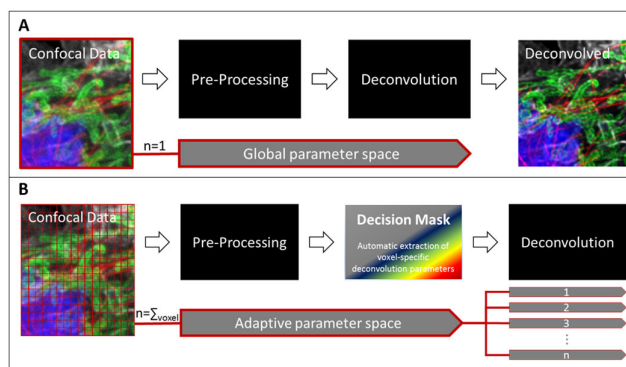

FIGURE 33. A Traditional deconvolution; B LIGHTNING workflows. Notice the additional decision mask present in LIGHTNING, which is absent in classical deconvolution processes. Figures adapted with permission from [143], © Leica Microsystems, 2018.

where N $=$ N[$g(x, y)$] and Bg $=$ $b(x, y)$ $\propto$ max[b_{global}(SNR)] [143].

A general assumption here is that the signal is minimally impacted upon by noise, agreeing with well-defined image processing algorithms [143]. In the next stage a decision mask is created from individual voxels using the SNR and Bg values obtained for each voxel and an image quality map generated for the entire image as shown in Fig. 34 (from [143]):

The image quality is then used to compute the adaptation coefficient (regularization parameter) in an inverse fashion for use in deconvolution in the following step [143].

The third (and final) stage in AD involves actual deconvolution of the image using the regularization parameters obtained from the decision mask previously and the MAP (maximum a posteriori)-based deconvolution algorithm modified according to the imaging method used (e.g. STED, multiphoton, etc) and fine-tuned to the optics used by Leica Microsystems [143]. The iteration count for the MAP-based
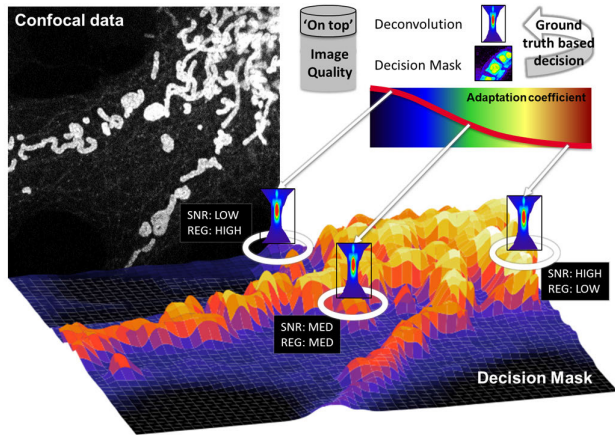
IEEE *Access*

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

FIGURE 34. An image quality map depicting the SNR for different voxels in the decision mask. Figure reprinted with permission from [143], © Leica Microsystems, 2018.
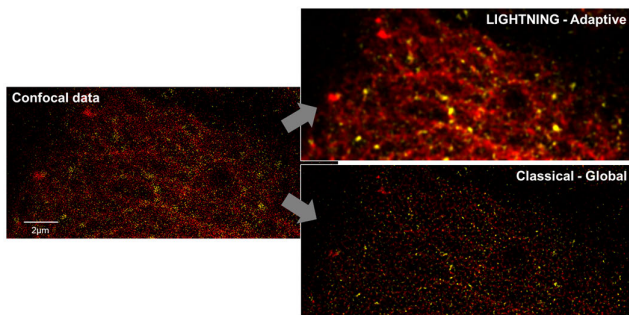

FIGURE 35. Differences exhibited between LIGHTNING & traditional deconvolution algorithms. LIGHTNING utilizes an adaptive mask to locally adapt the deconvolution parameters, while classical deconvolution uses a globally-defined set of parameters for the entire image. The image shown here is that of p24-YFP (a protein that cycles between the Endoplasmic Reticulum and the Golgi). Sample courtesy of Dr. Juan Jung, Pepperkok group, European Molecular Biology Laboratory, Heidelberg, Germany. Reproduced with permission from Leica Microsystems and the Pepperkok Team, © LIGHTNING White paper, 2018 [143].

FIGURE 36. Molecular nanorulers (separated by a distance of 120nm) are resolved via LIGHTNING (but not confocal) demonstrating the superior resolution of LIGHTNING over traditional confocal microscopy. According to the author of [143], the results obtained here are accurate as of Sep 2018, while recent improvements in the algorithm are capable of increasing the effective resolution of the image. Figure reprinted with permission from [143], © Leica Microsystems, 2018.
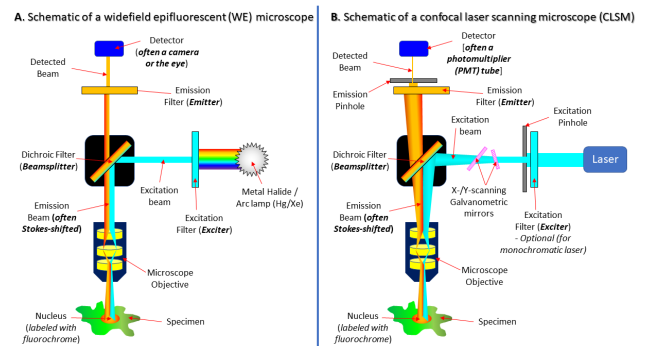
FIGURE 37. Differences in the optical train of A a widefield epifluorescent (WE) and B a confocal laser scanning microscope (CLSM). Notice the presence of an excitation and an emission pinhole in the CLSM, which is absent in the WE microscope. The detected beam in the WE microscope thus incorporates *both* the in-focus & out-of-focus impulse responses (from multiple axial/Z-sections), while in the CLSM, only the in-focus signals are present in the detected beam.

deconvolution algorithm is also determined automatically, with the iterations being ceased when there is no significant differences between the key features of the images at the $(n + 1)^{th}$ iteration as compared to the nth iteration. As such, this automation uniquely characterizes LIGHTNING from traditional deconvolution, with the differences between the outputs of these procedures distinctly visualized in Fig. 35:

For the accustomed user, LIGHTNING also incorporates non-AD (i.e. traditional) deconvolution algorithms, which utilize a globally determined set of parameters for image deconvolution as compared to AD (which computes local voxel-based parameters) [143]. Nonetheless, [143] has demonstrated LIGHTNING to be able to achieve a resolution of up to 120nm (accurate as of Sep 2018, with improved algorithms being able to further enhance the effective resolution of the image) as shown in Fig. 36 below:

B. COMPUTATIONAL CLEARING AND THE THUNDER IMAGER BY LEICA MICROSYSTEMS GMBH

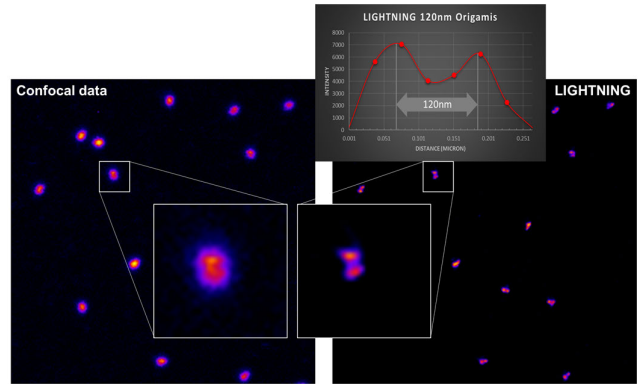An alternative contrast-enhancement methodology also developed by Leica Microsystems, computational

clearing (CC) is utilized in Leica's THUNDER Imagers to remove haze (originating from out-of-focus optical planes) while preserving focused features of interest [144] in the image plane. In this respect, CC is particularly adapted to dehazing widefield microscopical images (particularly widefield epifluorescence (WE) microscopy), where the absence of a pinhole allows light rays emerging from different optical sections to interfere with the image plane as described in Fig. 37:

Here, [144] describes a widefield epifluorescent image $I(r)$ to be estimated mathematically as follows:

$$I(r) \approx psf_{of}(r) \circledast f(r) + psf_{if}(r) \circledast f(r) \qquad (72)$$

where $f(r)$ is the fluorophore intensity distribution and $psf_{of/if}(r)$ refers to the PSF of the in-focus (*if*) and out-of-focus (*of*) light rays [144].

However (and as specifically mentioned in [144]), the PSF of the out-of-focus blur has a wider lateral spread and

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy
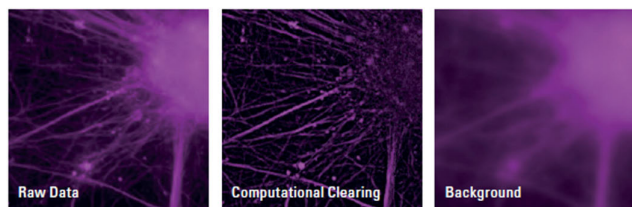
IEEE *Access*

FIGURE 38. An image of Beta III Tubulin Rat Neurons labelled with Cy5. Notice the removal of haze in the CC image. Figure sourced from [144], © Leica Microsystems, 2015.

FIGURE 39. Imaging at depths of 140 to 150 μm. Figure sourced from [144], © Leica Microsystems, 2015.

FIGURE 40. A Original image. Images processed using B Adjust AI (Structured setting); C Sharpen AI (Mode: Focus, Remove Blur: 0.53, Suppress Noise: 0.59); D Gigapixel AI (Scale: 4x).

Full Width at Half Maximum (FWHM) than the PSF of the in-focus signal, allowing scale discriminating operations (such as wavelet transforms) to be effectively utilized in this context. This formed the basis for an iterative algorithm to be constructed, which may be expressed mathematically as follows (from [144]):

$$I_{out} = \text{argmin}_{\widehat{I}_{out}} \left[\left\| I - \widehat{I}_{out} \right\|^2 \right] \tag{73}$$

where the structural length scale (i.e. the lateral (full-width) distance at $\sim$13.5% of the maximum (peak) intensity) of the predicted out-of-focus blur I_{out}) (referred to as $S\left[\widehat{I}_{out}\right]$) > r_0 (the LAS X-defined "feature" scale) [144].

In this respect, CC is used to *reveal* (rather than generate) a deblurred image, allowing the original relative intensities of the key features in the image to be preserved [144]. A CC-processed image is thus depicted in Fig. 38 as follows:

Due to their intensity conservation of the individual features, CC-processed images may be further utilized for quantitative analysis (e.g. in colocalization microscopy) although proper calibration protocols and care has to be exercised when attempting such procedures. Further details on quantifying CC-images are discussed in [144], which cites relatively close correlations between the signal intensity distributions of the fluorescent sample and its background both before and after application of CC (also termed Instant CC/ICC), with a Kolgomorov-Smirnov distance of 0.05 ± 0.02 [144]. This clearly depicts the ability of CC to allow direct fluorescent quantitation without the need for implementation of additional background removal algorithms [144].

An alternative mode of CC in THUNDER, termed Large Volume CC (LVCC) which combines Adaptive Deconvolution (AD) with CC, has also been noted in [144] to allow imaging up to a depth of 140 to 150 μm, as compared to traditional widefield imaging experiments (which can only resolve up to a depth of 50μm). Such deep field imaging allows LVCC to be used with thicker samples (negating the need for microtomy in some instances) although it should be emphasized that the maximum imaging depth depends very much on the refractive index of the sample and its light-scattering ability [144]. Further details on LVCC imaging are discussed in [144], with a figure extracted from [144] (indicating the utility of LVCC in neuronal imaging) being depicted in Fig. 39:
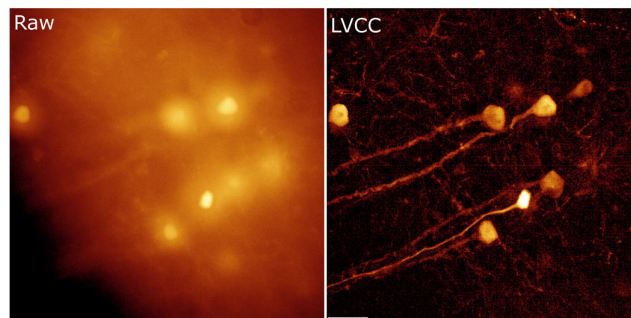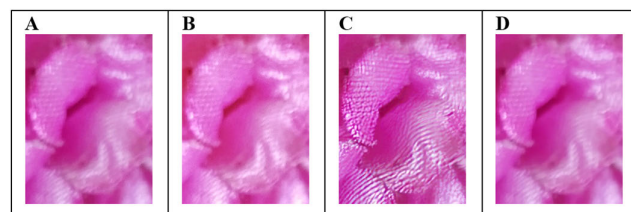
Finally, a third implementation of CC in THUNDER named Small Volume CC (or SVCC) is similar to LVCC, but which is intended for imaging thin ($\sim$80μm) samples. Here, [144] has indicated how SVCC may be used to improve the resolution of an acquired image, by utilizing a single 40nm-diameter bead and achieving resolution improvements of up to $\sim$2 times laterally (FWHM$_{\text{Raw}}$/SVCC = 1.961) and >2.5 times axially (FWHM$_{\text{Raw}}$/SVCC = 2.5641) [144]. Nonetheless, [144] highlights that these improvements in resolution were measured using the size of the spot corresponding to the image bead, instead of the actual distance between 2 structures separated by these distances (which was deemed by [144] to be not empirically possible).

C. TOPAZ LABS AI-BASED IMAGE ENHANCEMENT

The Topaz Labs AI Bundle (comprising of Topaz Sharpen AI, Topaz DeNoise AI, Topaz JPEG to Raw AI, Topaz Gigapixel AI, Topaz Mask AI, Topaz Studio 2 and Topaz Adjust AI) (© Topaz Labs) [145] represents a recent commercial software innovation in the macrophotography domain which has garnered significant interest due to its ability to produce eye-catching and vibrant images. For comparison purposes, a single image processed using Adjust AI, Sharpen AI and Gigapixel AI is being shown in the following diagram (Fig. 40):

From the above, it can be observed that Adjust AI (© Topaz Labs) is particularly adept at dehazing an acquired image, coupled with hue and contrast enhancement of the deblurred image. This may be presumed to be akin to traditional deblurring and deconvolution algorithms, such as Wiener filtering. Similarly, Gigapixel AI (© Topaz Labs) seeks to recover lost signal information caused by zooming into an image

IEEE Access

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

(an aspect of digital magnification) allowing a more highly resolved image to be generated. On a different note, Sharpen AI (© Topaz Labs) [146] utilizes an ANN trained with millions of images to recall and generate a sharpened image from an out-of-focus image [147]. Each of these approaches represent novel conceptions in the field of AI-mediated image processing, which may (upon further improvements) find some relevance in biomedical and geological research in the near future. In particular, the DNN within Sharpen AI could be subjected to transfer learning for the infusion of new learnt features (potentially from inherent optical aberrations in the Fourier domain) to enhance its capability in facilitating biomedical research.

VII. CURRENT LIMITATIONS AND POTENTIAL FUTURE ADVANCEMENTS IN COMPUTATIONAL NANOSCOPY

Despite there having been numerous recent developments in the field of optical nanoscopy, extensive unexplored terrains in this respect exist, being mediated through the utilization of *in silico*-based approaches combined with existing optical microscopical techniques and enhancements. In particular, one such dilemma refers to noise coupled with haze, and while significant progress has been made with the introduction of AI-based concepts and DNN architectures in microscopical imaging, the road towards a total recovery of the spatial information contained in the image remains relatively unexplored. Imperatively, it would be crucial to understand the basis of signal impulses constituting noise – that of ultra-high frequencies unresolvable by the microscope objective lens amalgamating to form a blurred signal. Resolution of the individual components of noise (and thus their discernment) for elimination of 'true noise' (arising from aberrations in the optical train) as opposed to *pseudo*-noise (due to sub-resolved structures in the image) would thus be crucial in fueling the drive towards super-resolved diascopic nanoscopy. In this respect, the reader would be encouraged to pursue a true definition of image noise and its causative factors, developing potential filters to discern and selectively eliminate 'true noise', while attempting to resolve *pseudo*-noise.

Another aspect worthy of further investigation refers to the resolution of Fourier transform waveforms (when coupled with AI) as a means of minimizing the mapping of learnt features from previous super-resolved images (during the training of the DNN) to novel images during deployment. This would be particularly favorable in light of the fact that DNNs have been found to be computationally superior to ANNs [148] due to the multiple paths to the output layer (mathematically represented by the exponential growth factor $\prod n_i$, where n refers to the number of nodes in the layer i) which consequentially permit higher resolutions of input feature deviations to be realized in output actualization while also accounting for spatiotemporal variations in the PSF during live-cell imaging. An exemplification of this aspect is currently being investigated by our research team as a potential approach to achieving *true* super-resolution microscopy via

in silico methods. It would be prudent to mention that this may also be supplemented through existing studies conducted in digital holographic microscopy (DHM) [149] and tomography [150], which are seeking to explore the potential of phase-shifted tomograms to reveal previously-unidentifiable information.

A third possibility for future research in this domain refers to how super-resolution nanoscopic approaches may be integrated into coherent Raman spectroscopy (CRS), as highlighted by [151]. Here, [151] states that current Raman scattering detection capabilities are less sensitive than those of fluorescence, even when utilizing coherent laser sources, posing a major hurdle to be overcome in this respect. Other avenues currently being explored in this domain (with promising future advancements) include super-resolved magnetic resonance imaging (MRI) microscopy [152], nanoscale image resolution via defocused Z-stack acquisition [153], *in-situ* PSF retrieval for 3D single-molecule localization microscopy (SMLM) imaging [154] and 3D super-resolution time-lapse microscopy [155], all of which represent a push towards computational super-resolution in advanced medical imaging processes. Complementarily (& although not being the subject of this review), optical (non-computational) super-resolution microscopy has experienced a leap ahead as well, with the introduction of microspheres for non-fluorescence nanoscopy applications (see [156] for details).

On the andragogical front, a merger of augmented reality (AR) or virtual reality (VR) with nanoscopy would bridge the nanoscopy-immunology gap while being particularly relevant in scientific education (as discussed in [157] and [158]). Here, researchers (both new and experienced) would find such immersive experiences particularly helpful in understanding key processes with the potential to even uncover insights (such as biomolecular interactions) that were previously unfathomed.

Computational nanoscopy would also find a niche (which is in fact being currently explored) in pathological examinations and diagnostics, especially in the detection of cancerous tissue or diseased cells (which would often necessitate human expertise). Some studies expounding research conducted in this field include [159], [160] and [161]. This would potentially also lead to the future (presently unexplored) terrain of *non-fluorescent in vivo* optical **nanoscopy** and **molecular imaging**, for the visualization (and potentially nanomanipulations) of molecular motions, metabolic pathways and dynamics of biochemical reactions in real-time and *in situ*.

In the manufacturing sector too, computational nanoscopy also holds promise in nanomanufacturing processes and their corresponding quality analysis procedures (especially for silicon wafers and chipsets utilized in the semiconductor industry). With rapid advancements made in computing hardware and future CPU fabrication processes reaching single-digit nm dimensions (such as Intel's Meteor Lake architecture using 7nm chips [162]), it would be sensible to explore the utilization of computational nanoscopy in
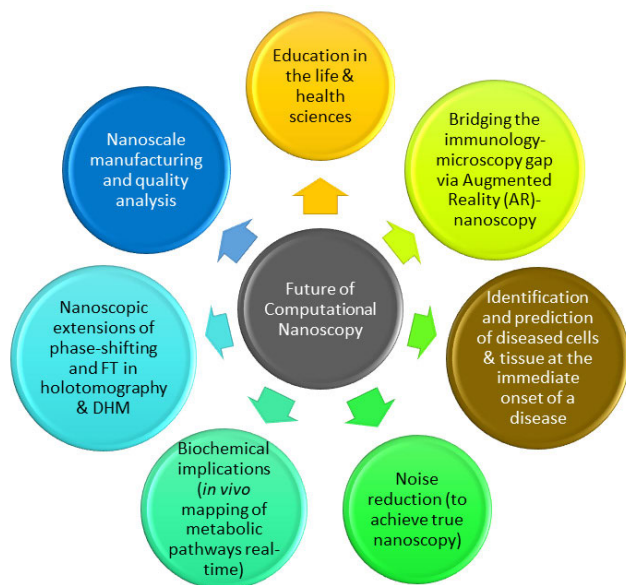
S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

FIGURE 41. Potential future advancements which can be realized in the field of computational nanoscopy within the coming decades. Some of these are currently being explored, while others remain relatively out-of-reach with present day technology. It would be prudent to note that these are only *some* of the potential areas for exploration of the reach of computational nanoscopy within the next few decades, while the authors of the present review surmise numerous other unexplored areas in this respect.

this sector and how it may facilitate the execution of such processes.

The following infographic (Fig. 41) summarizes some of these aspects for easier identification:

Nonetheless, most of these algorithms are plagued with numerous limitations arising as a consequence of imperfections in the optical train. A prominent problem surfaced by [163] refers to what [163] deemed as the "hallucination problem", although we presume that this may not be a significant concern in computational nanoscopy, especially where an extensive dataset (comprising varied samples which addresses the issue being investigated) is utilized, except where highly specific post-acquisition image processing and analysis protocols (such as feature detection and object tracking) are desired, in which instance, an over-fitted DNN (prone to such hallucination issues) may be utilized. In addition, fuzzy logic approaches (such as that employed in [164], [165] & [166]) may also contribute to network hallucination (despite outperforming even the state-of-the-art super-resolution methods as mentioned in [166]), since fuzzy logic *approximates* the mapping of image patches to highly-resolved images, while this approximation may (at times) result in incorrect image mappings. In this respect, it would also be noteworthy to mention that future approaches in computational super-resolution microscopy should seek to integrate ***both*** deconvolution and deep learning principles, by using deep learning to accurately discern the optical PSF and noise of an optical system, which may vary spatiotemporally across the sample, especially in living cells.

The elucidated PSF may then be coupled to a deconvolution method for obtaining the impulse response at individual locations in the sample, with pixel-wise accuracy. Only by doing so, can one achieve *true* super-resolution computational nanoscopy, which would be immune to network hallucinations and optical aberrations, amidst a variable PSF. This may also be fueled through the advent and recent on-going developments in the field of *quantum computing* (QC) [167], which utilizes advanced future-cutting edge hardware to drive computational nanoscopy to unprecedented heights.

VIII. CONCLUSION

Although we have sought to holistically evaluate the recent developments in the field of computational nanoscopy in the present review, recent developments spurred by ongoing research (particularly in artificial intelligence and deep neural networks) are resulting in new findings in the field of computational nanoscopy even at the time of authoring this review. Hence, the field of intelligent computational nanoscopy represents an exciting new phase in the future of optical microscopy, as we enter an era merging traditional microscopical imaging modalities with recent advancements in machine learning and computing, such as QC. It is desired (through the readership of this review) that one would be inspired to further explore and potentially contribute to future advancements in this sector, spurring the advancement of optical nanoscopy to greater (and previously unfathomed) heights in scientific research and visualization.

REFERENCES

[1] J. S. Silfies, S. A. Schwartz, and M. W. Davidson. *The Diffraction Barrier in Optical Microscopy*. Nikon Instruments Inc. Accessed: Nov. 11, 2019. [Online]. Available: https://www.microscopyu.com/techniques/super-resolution/the-diffraction-barrier-in-optical-microscopy

[2] K. D. Hauch and B. D. Ratner, "Microscopy for biomaterials science," in *Biomaterials Science* (An Introduction to Materials in Medicine), 3rd ed. B. D. Ratner, A. S. Hoffman, F. J. Schoen and J. E. Lemons, Eds. New York, NY, USA: Academic, 2013, pp. 677–692.

[3] M. W. Davidson. *Numerical Aperture*. Nikon Instruments Inc—MicroscopyU. Accessed: Nov. 17, 2019. [Online]. Available: https://www.microscopyu.com/microscopy-basics/numerical-aperture

[4] K. N. Fish, "Total internal reflection fluorescence (TIRF) microscopy," *Current Protocols Cytometry*, vol. 50, no. 1, p. 12, Oct. 2009.

[5] (2011). *Super-Resolution Microscopy Tutorial*. The University of Utah—Advanced Microscopy. Accessed: Nov. 17, 2019. [Online]. Available: https://advanced-microscopy.utah.edu/education/super-res/

[6] *The Nobel Prize in Chemistry 2014*. NobelPrize.org. Accessed: Apr. 24, 2019. [Online]. Available: https://www.nobelprize.org/prizes/chemistry/2014/press-release/

IEEE *Access*

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

[7] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.*, vol. 19, no. 11, pp. 780–782, 1994.

[8] F. C. Zanacchi and A. Diaspro, "Ground state depletion imaging," in *Encyclopedia of Biophysics*, G. C. K. Roberts, Ed. Berlin, Germany: Springer, 2013.

[9] U. Endesfelder and M. Heilemann, "Direct stochastic optical reconstruction microscopy (dSTORM)," in *Advanced Fluorescence Microscopy. Methods in Molecular Biology (Methods and Protocols)*, vol. 1251. New York, NY, USA: Humana Press, 2014, pp. 263–276.

[10] J. Siebenmorgen, Y. Novikau, R. Wolleschensky, K. Weisshart, and I. Kleppe, *Introducing Lattice SIM for ZEISS Elyra 7 - Structured Illumination Microscopy with a 3D Lattice for Live Cell Imaging*. Oberkochen, Germany: Carl Zeiss AG, 2018.

[11] L. Schermelleh, A. Ferrand, T. Huser, C. Eggeling, M. Sauer, O. Biehlmaier, and G. P. C. Drummen, "Super-resolution microscopy demystified," *Nature Cell Biol.*, vol. 21, no. 1, pp. 72–84, Jan. 2019.

[12] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA, USA: MIT Press, 1942.

[13] E. W. Weisstein. *Deconvolution*. MathWorld–A Wolfram Web Resource. Accessed: Nov. 17, 2019. [Online]. Available: http://mathworld.wolfram.com/Deconvolution.html

[14] R. M. Parton and I. Davis, "Lifting the fog: Image restoration by deconvolution," in *Cell Biology—A Laboratory Handbook*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2006, pp. 187–200.

[15] M. J. Nasse and J. C. Woehl, "Realistic modeling of the illumination point spread function in confocal scanning optical microscopy," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 27, no. 2, pp. 295–302, 15 11 2010.

[16] P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 32–45, May 2006.

[17] (Oct. 26, 2018). *DeconvolutionLab2*. Biomedical Imaging Group (EPFL). Accessed: Dec. 2, 2019. [Online]. Available: http://bigwww.epfl.ch/deconvolution/deconvolutionlab2/

[18] *Chapter 3.1: Image Deconvolution*. Development of Bioinformatics Tools to Track Cancer Cell Invasion Using 3D in vitro Invasion Assays. Accessed: Dec. 2, 2019. [Online]. Available: https://repositorio-aberto.up.pt/bitstream/10216/60494/1/000142383.pdf

[19] W. Burger and M. J. Burge, "6.6.2 unsharp masking," in *Digital Image Processing—An Algorithmic Introduction Using Java*, 2nd ed. London, U.K.: Springer-Verlag, 2016, pp. 142–143.

[20] D. Sage, L. Donati, F. Soulez, D. Fortun, G. Schmit, A. Seitz, R. Guiet, C. Vonesch, and M. Unser, "DeconvolutionLab2: An open-source software for deconvolution microscopy," *Methods*, vol. 115, pp. 28–41, Feb. 2017.

[21] C. Zhang. (Mar. 15, 2016). *Image Restoration by Deconvolution: Concepts and Applications*. SIMBioSys, Depertment of Information and Communication Technologies Universitat Pompeu Fabra. Accessed: Nov. 20, 2019. [Online]. Available: https://www.upf.edu/documents/4181972/7971893/%5Bslides%5D_Image_Restoration_by_Deconvolution/6b1aa1ac-70d7-2494-6192-2af0f95f17cb

[22] M. Spitaler. *Microscopic Image Restoration by Deconvolution*. Imperial College London—Facility for Imaging by Light Microscopy. Accessed: Nov. 28, 2019. [Online]. Available: https://www.imperial.ac.uk/media/imperial-college/medicine/facilities/film/Deconvolution-training-140219.pdf

[23] *Huygens Professional: Discover the Truth Behind Your Image*. Scientific Volume Imaging. Accessed: Nov. 1, 2019. [Online]. Available: https://svi.nl/Huygens-Professional

[24] (2019). *AutoQuant X3: Image Deconvolution and Restoration Software*. Media Cybernetics, Inc. Accessed: Dec. 1, 2019. [Online]. Available: https://www.mediacy.com/autoquantx3#features

[25] L. Justen and R. Ramlau, "A non-iterative regularization approach to blind deconvolution," *Inverse Problems*, vol. 22, no. 3, pp. 771–800, Jun. 2006.

[26] J. N. Caron, N. M. Namazi, and C. J. Rollins, "Noniterative blind data restoration by use of an extracted filter function," *Appl. Opt.*, vol. 41, no. 32, pp. 6884–6889, 2002.

[27] D. Perrone and P. Favaro, "A clearer picture of total variation blind deconvolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1041–1055, Jun. 2016.

[28] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, Jul. 2006.

[29] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, no. 3, p. 73, 2008.

[30] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–8, Dec. 2009.

[31] L. Xu and J. Jiaya, "Two-phase kernel estimation for robust motion deblurring," in *Proc. 11th Eur. Conf. Comput. Vis.*, vol. 1, 2010, pp. 157–170.

[32] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. CVPR*, Jun. 2011, pp. 2657–2664.

[33] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.

[34] Y.-L. You and M. Kaveh, "Anisotropic blind image restoration," in *Proc. 3rd IEEE Int. Conf. Image Process.*, Sep. 1996, pp. 461–464.

[35] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1999, pp. 541–547.

[36] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, Dec. 2011.

[37] L. Condat, "A direct algorithm for 1-D total variation denoising," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1054–1057, Nov. 2013.

[38] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Problems*, vol. 19, no. 6, pp. S165–S187, Dec. 2003.

[39] P. L. Davies and A. Kovac, "Local extremes, runs, strings and multiresolution," *Ann. Statist.*, vol. 29, no. 1, pp. 1–48, 2001.

[40] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, pp. 55–59, Jan. 1972.

[41] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical J.*, vol. 79, p. 745, Jun. 1974.

[42] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, *3D Microscopy Deconvolution using Richardson-Lucy Algorithm with Total Variation Regularization*, document [Research Report] RR-5272, INRIA, 2004, p. 71.

[43] E. W. Weisstein. *Poisson Distribution*. MathWorld—A Wolfram Web Resource. Accessed: Feb. 20, 2020. [Online]. Available: http://mathworld.wolfram.com/PoissonDistribution.html

[44] *Deblurring Images Using the Lucy-Richardson Algorithm*. The MathWorks, Inc. Accessed: Nov. 27, 2019. [Online]. Available: https://www.mathworks.com/help/images/deblurring-images-using-the-lucy-richardson-algorithm.html

[45] K. Sastry. (Apr. 2002). *Iterative Constrained Deconvolution Methods*. Plexar Associates, Inc. Accessed: Dec. 1, 2019. [Online]. Available: http://atom.lylver.org/AstroSurf/PDF/interactive_cons_decon_rev102.pdf

[46] T. P. McClanahan, J. I. Trombka, I. G. Mitrofanov, R. Z. Sagdeev, and M. H. Loew, "Application of image restoration (Jansson Van-Cittert) to planetary remote sensing neutron count rate maps," in *Proc. 38th Lunar Planetary Sci.*, 2007.

[47] G. Thomas, "An improvement of the van-Cittert's method," in *Proc. ICASSP IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6, Apr. 1981, pp. 47–49.

[48] A. Stefanoiu, J. Page, P. Symvoulidis, G. G. Westmeyer, and T. Lasser, "Artifact-free deconvolution in light field microscopy," *Opt. Exp.*, vol. 27, no. 22, pp. 31644–31666, 2019.

[49] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 924–934, 2006.

[50] H. Li, C. Guo, D. Kim-Holzapfel, W. Li, Y. Altshuller, B. Schroeder, W. Liu, Y. Meng, J. B. French, K.-I. Takamaru, M. A. Frohman, and S. Jia, "Fast, volumetric live-cell imaging using high-resolution light-field microscopy," *Biomed. Optics Express*, vol. 10, no. 1, pp. 29–49, 2019.

[51] R. Prevedel, Y.-G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden, and A. Vaziri, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nature Methods*, vol. 11, no. 7, pp. 727–730, Jul. 2014.

[52] L. Cong, Z. Wang, Y. Chai, W. Hang, C. Shang, W. Yang, L. Bai, J. Du, K. Wang, and Q. Wen, "Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio)," *eLife*, vol. 6, Sep. 2017, Art. no. e28158.

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE*Access*

[53] N. Wagner, N. Norlin, J. Gierten, G. de Medeiros, B. Balázs, J. Wittbrodt, L. Hufnagel, and R. Prevedel, "Instantaneous isotropic, volumetric imaging of fast biological processes," *Nature Methods*, vol. 16, no. 6, pp. 497–500, 2019.

[54] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[55] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford Univ. Comput. Sci., Stanford, CA, USA, Tech. Rep. CSTR 2005-02, 2005.

[56] R. Ng, "Fourier slice photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744, Jul. 2005.

[57] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1027–1034.

[58] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-D deconvolution for the light field microscope," *Opt. Exp.*, vol. 21, no. 21, pp. 25418–25439, 2013.

[59] L. Landweber, "An iteration formula for fredholm integral equations of the first kind," *Amer. J. Math.*, vol. 73, no. 3, pp. 615–624, Jul. 1951.

[60] C. Vonesch and M. Unser, "A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 539–549, Mar. 2008.

[61] P. Grohs, Ž. Kereta and U. Wiesmann, "A shearlet-based fast thresholded Landweber algorithm for deconvolution," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 14, no. 5, pp. 1–19, 2016.

[62] P. B. Stark and R. L. Parker, "Bounded-variable least-squares: An algorithm and applications," *Comput. Statist.*, vol. 10, no. 2, pp. 1–13, 1995.

[63] A. Haar, "Zur theorie der orthogonalen funktionensysteme," *Mathematische Annalen*, vol. 69, no. 3, pp. 331–371, Sep. 1910.

[64] M. Unser, "Ten good reasons for using spline wavelets," *Proc. SPIE*, vol. 3169, pp. 422–431, Oct. 1997.

[65] C. K. Chui and J.-Z. Wang, "A cardinal spline approach to wavelets," *Proc. Amer. Math. Soc.*, vol. 113, no. 3, pp. 785–793, Nov. 1991.

[66] C. K. Chui and J.-Z. Wang, "On compactly supported spline wavelets and a duality principle," *Trans. Amer. Math. Soc.*, vol. 330, no. 2, pp. 903–915, Feb. 1992.

[67] C. K. Chui, *An Introduction to Wavelets*. New York, NY, USA: Academic, 1992.

[68] A. Katunin, "The construction of high-order B-spline wavelets and their decomposition relations for fault detection and localisation in composite beams," *Sci. Problems Mach. Operation Maintenance*, vol. 3, no. 167, pp. 43–59, 2011.

[69] J.-F. Leu, J.-C. Jang, and C. Hwang, "Computation of battle-lemarie wavelets using an FFT-based algorithm," *J. Science Comput.*, vol. 13, pp. 485–504, Dec. 1998.

[70] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[71] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[72] F. Xue, F. Luisier, and T. Blu, "Multi-Wiener SURE-LET deconvolution," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1954–1968, May 2013.

[73] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.

[74] D. Van De Ville and M. Kocher, "SURE-based non-local means," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 973–976, Nov. 2009.

[75] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778–2786, Nov. 2007.

[76] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant representations," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1342–1352, Aug. 2008.

[77] J. Li, F. Luisier, and T. Blu, "Pure-let deconvolution of 3D fluorescence microscopy images," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 723–727.

[78] J. Li, F. Luisier, and T. Blu, "PURE-LET image deconvolution," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 92–105, Jan. 2018.

[79] W. Meiniel, J.-C. Olivo-Marin, and E. D. Angelini, "Denoising of microscopy images: A review of the state-of-the-art, and a new sparsity-based method," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3842–3856, Aug. 2018.

[80] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[81] C. Louchet and L. Moisan, "Total variation denoising using iterated conditional expectation," in *Proc. 22nd IEEE Eur. Signal Process. (EUSIPCO)*, Sep. 2014, pp. 1592–1596.

[82] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, Jan. 2005.

[83] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[84] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[85] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[86] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1720–1730, Jul. 2010.

[87] R. Abergel, C. Louchet, L. Moisan, and T. Zeng, "Total variation restoration of images corrupted by Poisson noise with iterated conditional expectations," in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis.*, May 2015, pp. 178–190.

[88] A. Sawatzky, C. Brune, J. Müller, and M. Burger, "Total variation processing of images with Poisson statistics," in *Comput. Anal. Images Patterns (CAIP), Berlin, Germany, Springer*, vol. 2009, pp. 533–540.

[89] C.-A. Deledalle, F. Tupin, and L. Denis, "Poisson NL means: Unsupervised non local means for Poisson noise," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 801–804.

[90] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *J. Math. Imag. Vis.*, vol. 48, no. 2, pp. 279–294, Feb. 2014.

[91] A. A. Bindilatti, M. A. C. Vieira, and N. D. A. Mascarenhas, "Poisson Wiener filtering with non-local weighted parameter estimation using stochastic distances," *Signal Process.*, vol. 144, pp. 68–76, Mar. 2018.

[92] D. Miller and W. Scott. (Jun. 7, 2006). *Deconvolution with Inverse and Weiner Filters*. OpenStax CNX. Accessed: Dec. 2, 2019. [Online]. Available: https://cnx.org/contents/BCzesWfo@2/Deconvolution-with-Inverse-and-Weiner-Filters

[93] S. Dey, *Hands-on Image Processing with Python*. Birmingham, U.K.: Packt Publishing Limited, 2018, pp. 127–439.

[94] I. Analyst. (Oct. 26, 2015). *Filtering Using FFT in Images*. MATLAB Answers. MathWorks. Accessed: Dec. 29, 2019. [Online]. Available: https://www.mathworks.com/matlabcentral/answers/251071-filtering-using-fft-in-images

[95] A. Dey, "Machine learning algorithms: A review," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016.

[96] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[97] (2019). *Affectiva*. Affectiva. Accessed: Dec. 27, 2019. [Online]. Available: https://www.affectiva.com/

[98] (Sep. 18, 2019). *Overview: Visual Recognition in Watson Studio*. IBM. Accessed: Dec. 12, 2019. [Online]. Available: https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/visual-recognition-overview.html

[99] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," 2017, *arXiv:1702.05659*. [Online]. Available: http://arxiv.org/abs/1702.05659

[100] S. Do, K. D. Song, and J. W. Chung, "Basics of deep learning: A radiologist's guide to understanding published radiology articles on deep learning," *Korean J Radiol.*, vol. 21, no. 1, pp. 33–41, 2020.

[101] J. Ludwig. *Image Convolution*. Portland State University. Accessed: Dec. 12, 2019. [Online]. Available: http://web.pdx.edu/~jduh/courses/Archive/geog481w07/Students/Ludwig_ImageConvolution.pdf

[102] S. H. Ahn. *Convolution*. Accessed: Dec. 27, 2019. [Online]. Available: http://www.songho.ca/dsp/convolution/convolution.html

[103] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, 2019.

IEEE Access

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

[104] S. Sun, Z. Cao, H. Zhu, and J. Zhao, "A survey of optimization methods from a machine learning perspective," 2019, *arXiv:1906.06821*. [Online]. Available: http://arxiv.org/abs/1906.06821

[105] L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: A review of algorithms and comparison of software implementations," *J. Global Optim.*, vol. 56, no. 3, pp. 1247–1293, Jul. 2013.

[106] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[107] L. Luo, Y. Xiong, Y. Liu, and X. Sun, "Adaptive gradient methods with dynamic bound of learning rate," 2019, *arXiv:1902.09843*. [Online]. Available: http://arxiv.org/abs/1902.09843

[108] V. Chiley, I. Sharapov, A. Kosson, U. Koster, R. Reece, S. Samaniego de la Fuente, V. Subbiah, and M. James, "Online normalization for training neural networks," 2019, *arXiv:1905.05894*. [Online]. Available: http://arxiv.org/abs/1905.05894

[109] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" 2018, *arXiv:1805.11604*. [Online]. Available: http://arxiv.org/abs/1805.11604

[110] P. Luo, X. Wang, W. Shao, and Z. Peng, "Towards understanding regularization in batch normalization," 2018, *arXiv:1809.00846*. [Online]. Available: http://arxiv.org/abs/1809.00846

[111] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*. [Online]. Available: http://arxiv.org/abs/1607.06450

[112] P. Luo, Z. Peng, J. Ren, and R. Zhang, "Do normalization layers in a deep ConvNet really need to be distinct?" 2018, *arXiv:1811.07727*. [Online]. Available: http://arxiv.org/abs/1811.07727

[113] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*. [Online]. Available: http://arxiv.org/abs/1802.05957

[114] *7 Types of Neural Network Activation Functions: How to Choose.* Missinglink. Accessed: Dec. 29, 2019. [Online]. Available: https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/

[115] *Desmos | Graphing Calculator.* Desmos. Accessed: Feb. 7, 2020. [Online]. Available: https://www.desmos.com/calculator/zdv2kras4n

[116] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*. [Online]. Available: http://arxiv.org/abs/1710.05941

[117] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica*, vol. 4, no. 11, pp. 1437–1443, 2017.

[118] J. A. Grant-Jacob, B. S. Mackay, J. A. G. Baker, Y. Xie, D. J. Heath, M. Loxham, R. W. Eason, and B. Mills, "A neural lens for super-resolution biological imaging," *J. Phys. Commun.*, vol. 3, no. 6, 2019, Art. no. 065004.

[119] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, Jan. 2019.

[120] A. Descloux, K. S. Grußmayer, and A. Radenovic, "Parameter-free image resolution estimation based on decorrelation analysis," *Nature Methods*, vol. 16, no. 9, pp. 918–924, Sep. 2019.

[121] Y. Zhang, E. L. Nichols, A. M. Zellmer, I. H. Guldner, C. Kankel, S. Zhang, S. S. Howard, and C. J. Smith, "Generating intravital super-resolution movies with conventional microscopy reveals actin dynamics that construct pioneer axons," *Development*, vol. 146, no. 5, Mar. 2019, Art. no. dev171512.

[122] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature Biotechnol.*, vol. 36, no. 5, pp. 460–468, May 2018.

[123] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution.," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[124] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2015.

[125] J. AKST. (Dec. 17, 2018.)*AI Networks Generate Super-Resolution from Basic Microscopy*. The Scientist Magazine. Accessed: Apr. 24, 2019. [Online]. Available: https://www.the-scientist.com/news-opinion/ai-networks-generate-super-resolution-from-basic-microscopy-65219

[126] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, 2018.

[127] M. Ovesný, P. Křížek, J. Borkovec, Z. Švindrych, and G. M. Hagen, "ThunderSTORM: A comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging," *Bioinformatics*, vol. 30, no. 16, pp. 2389–2390, Aug. 2014.

[128] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri, "ImageJ2: ImageJ for the next generation of scientific image data," *BMC Bioinf.*, vol. 18, no. 1, p. 529, Dec. 2017.

[129] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: An open-source platform for biological-image analysis," *Nature Methods*, vol. 9, no. 7, pp. 676–682, Jul. 2012.

[130] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser, "Quantitative evaluation of software packages for single-molecule localization microscopy," *Nature Methods*, vol. 12, no. 8, pp. 717–724, Aug. 2015.

[131] S. Gazagnes, E. Soubies, and L. Blanc-Feraud, "High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, p. 4.

[132] M. Weigert, L. Royer, F. Jug, and G. Myers, "Isotropic reconstruction of 3D fluorescence microscopy images using convolutional neural networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, M. E. A. Descoteaux, Ed. Springer, 2017, pp. 126–134.

[133] Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, L. A. Bentolila, C. Pritz, and A. Ozcan, "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning," *Nature Methods*, vol. 16, no. 12, pp. 1323–1331, Dec. 2019.

[134] Y. Zhou, R. Chellappa, and B. Jenkins, "A novel approach to image restoration based on a neural network," in *Proc. Int. Conf. Neural Netw.*, San Diego, CA, USA, Jun. 1987, pp. 269–276.

[135] Ş. Sağiroğlu and E. Beşdok, "A novel approach for image denoising based on artificial neural networks," *J. Polytech.*, vol. 15, no. 2, pp. 71–86, 2012.

[136] V. Jain and H. S. Seung, "Natural image denoising with convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2008, pp. 769–776.

[137] M. Weigert *et al.*, "Content-aware image restoration: Pushing the limits of fluorescence microscopy," *Nature Methods*, vol. 15, no. 12, pp. 1090–1097, Dec. 2018.

[138] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, "Image denoising review: From classical to state-of-the-art approaches," *Inf. Fusion*, vol. 55, pp. 220–244, Mar. 2020.

[139] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-CNN for image restoration," 2018, *arXiv:1805.07071*. [Online]. Available: http://arxiv.org/abs/1805.07071

[140] H. R. Shahdoosti and Z. Rahemi, "Edge-preserving image denoising using a deep convolutional neural network," *Signal Process.*, vol. 159, pp. 20–32, Jun. 2019.

[141] S. Gai and Z. Bao, "New image denoising algorithm via improved deep convolutional neural network with perceptive loss," *Expert Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112815.

[142] (2020). *Imaris 9.3—3D/4D Image Visualization and Analysis with Deconvolution | Microscopy Software - Imaris - Oxford Instruments*. Oxford Instruments. Accessed: Mar. 4, 2020. [Online]. Available: https://imaris.oxinst.com/versions/9-3

[143] J. Reymann, *Lightning White Paper—Image Information Extraction by Adaptive Deconvolution*. Wetzlar, Germany: Leica Microsystems, 2018.

[144] J. Schumacher and L. Bertrand, *Thunder Imagers: How Do They Really Work*. Wetzlar, Germany: Leica Microsystems, 2015.

[145] (2019). *AI Bundle—Topaz Labs*. Topaz Labs. Accessed: Dec. 15, 2019. [Online]. Available: https://topazlabs.com/ai-bundle/

[146] (2019). *Sharpen AI—Topaz Labs*. Topaz Labs. Accessed: Dec. 15, 2019. [Online]. Available: https://topazlabs.com/sharpen-ai/

[147] H. Salyer. (May 14, 2019). *Getting Started with Topaz Sharpen AI*. Accessed: Dec. 15, 2019. [Online]. Available: https://topazlabs.com/getting-started-with-topaz-sharpen-ai/

[148] I. Goodfellow, Y. Bengio, and A. Courville, "Chapter 6 deep feedforward networks," in *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016, pp. 164–223.

S. S. Kaderuppan *et al.*: Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy

IEEE *Access*

[149] Y. Wu, Y. Luo, G. Chaudhari, Y. Rivenson, A. Calis, K. de Haan, and A. Ozcan, ''Bright-field holography: Cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram,'' *Light, Sci. Appl.*, vol. 8, no. 1, pp. 1–7, Dec. 2019.

[150] Y. Jo, H. Cho, S. Yun Lee, G. Choi, G. Kim, H.-s. Min, and Y. Park, ''Quantitative phase imaging and artificial intelligence: A review,'' *IEEE J. Sel. Topics Quantum Electron.*, vol. 25, no. 1, pp. 1–14, Feb. 2019.

[151] M. May (Dec. 24, 2019). *Analyzing the Real World with Label-Free Technology*. Biocompare. Accessed: Dec. 27, 2019. [Online]. Available: https://www.biocompare.com/Editorial-Articles/558656-Analyzing-the-Real-World-with-Label-Free-Technol/

[152] C. L. Degen, M. Poggio, H. J. Mamin, C. T. Rettner, and D. Rugar, ''Nanoscale magnetic resonance imaging,'' *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 5, pp. 1313–1317, 2009.

[153] M. V. Ryabko, S. N. Koptyaev, A. V. Shcherbakov, A. D. Lantsov, and S. Y. Oh, ''Method for optical inspection of nanoscale objects based upon analysis of their defocused images and features of its practical implementation,'' *Opt. Exp.*, vol. 21, pp. 24483–24489, Oct. 2013.

[154] F. Xu, D. Ma, K. P. MacPherson, S. Liu, Y. Bu, Y. Wang, Y. Tang, C. Bi, T. Kwok, A. A. Chubykin, P. Yin, S. Calve, G. E. Landreth, and F. Huang, ''Three-dimensional nanoscopy of whole cells and tissues with *in situ* point spread function retrieval,'' *Nature Methods*, vol. 17, no. 5, pp. 531–540, May 2020.

[155] L. Alvarenga. (Jun. 5, 2018). *The Power of Modern Super-Resolution Microscopy*. Labcompare. Accessed: Feb. 22, 2020. [Online]. Available: https://www.labcompare.com/10-Featured-Articles/350672-The-Power-of-Modern-Super-Resolution-Microscopy/

[156] Z. Wang, W. Guo, L. Li, B. Luk'yanchuk, A. Khan, Z. Liu, Z. Chen, and M. Hong, ''Optical virtual imaging at 50 nm lateral resolution with a white-light nanoscope,'' *Nature Commun.*, vol. 2, no. 1, pp. 1–6, Sep. 2011.

[157] D. Matthews, ''Virtual-reality applications give science a new dimension,'' *Nature*, vol. 557, no. 7703, pp. 127–128, May 2018.

[158] M. Roelandse and K. Wallerius. (Oct. 30, 2018). *How Augmented Reality and Virtual Reality Can Revolutionize Science*. Scientific American. Accessed: Feb. 17, 2020. [Online]. Available: https://blogs.scientificamerican.com/observations/how-augmented-reality-and-virtual-reality-can-revolutionize-science/

[159] C. L. Chen, A. Mahjoubfar, L.-C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, and B. Jalali, ''Deep learning in label-free cell classification,'' *Sci. Rep.*, vol. 6, no. 1, Mar. 2016, Art. no. 21471.

[160] M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, and B. Fei, ''Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning,'' *Cancers*, vol. 11, no. 9, p. 1367, Sep. 2019.

[161] R. Brent and L. Boucheron, ''Deep learning to predict microscope images,'' *Nature Methods*, vol. 15, no. 11, pp. 868–870, Nov. 2018.

[162] drbyos. (Apr. 2, 2020). *CPU Manufacturing: Intel Has a Netburst déjávu*. Accessed: Apr. 21, 2020. [Online]. Available: https://world-today-news.com/cpu-manufacturing-intel-has-a-netburst-deja-vu/

[163] C. Belthangady and L. A. Royer, ''Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction,'' *Nature Methods*, vol. 16, no. 12, pp. 1215–1225, Dec. 2019.

[164] S. Souverville, J. A. Rosales, F. J. Gallegos, M. Dehesa, I. V. Hernández, and L. V. Lozano, ''Fuzzy logic applied to improvement of image resolution using Gaussian membership functions,'' *Res. Comput. Sci.*, vol. 102, no. 1, pp. 77–88, Dec. 2015.

[165] P. Purkait, N. Ranjan Pal, and B. Chanda, ''A fuzzy-rule-based approach for single frame super resolution,'' *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2277–2290, May 2014.

[166] G. Jeon, M. Anisetti, E. Damiani, and O. Monga, ''Real-time image processing systems using fuzzy and rough sets techniques,'' *Soft Comput.*, vol. 22, no. 5, pp. 1381–1384, Mar. 2018.

[167] (Jul. 29, 2020). *Meet Silq—The First Intuitive High-Level Language for Quantum Computers*. Artibatm. Accessed: Aug. 14, 2020. [Online]. Available: https://www.artiba.org/blog/meet-silq-the-first-intuitive-high-level-language-for-quantum-computers

SHIRAZ S. KADERUPPAN received the B.Sc. degree in life sciences (concentrations in biomedical science and molecular and cell biology) from the National University of Singapore, in 2007, and the M.Sc. degree in information systems from Nanyang Technological University, in 2017. He is currently pursuing the Ph.D. degree in biomedical signal processing with Newcastle University in conjunction with NewRIIS. His research interests include optical microscopical imaging, the IoT, and machine learning.

EUGENE WAI LEONG WONG (Member, IEEE) received the B.Eng. and Ph.D. degrees in mechanical engineering from the National University of Singapore, in 2003 and 2008, respectively. He is currently an Associate Professor and the Director of Undergraduate Studies with Newcastle University in Singapore. His research interests include additive manufacturing, development of composite materials, and application of machine learning in materials prediction. He is also a member of the American Society of Mechanical Engineers (ASME).

ANURAG SHARMA (Member, IEEE) received the bachelor's degree in electrical engineering from the Malaviya National Institute of Technology (MNIT) Jaipur, India, in 2012, and the Ph.D. degree from the National University of Singapore, in 2017. He is currently working as an Assistant Professor with Newcastle University in Singapore. His research interests include applications of computational intelligence and data analytic techniques in electrical distribution systems, especially for energy management, integration of distributed energy resources, and service restoration.

WAI LOK WOO (Senior Member, IEEE) received the B.Eng. degree in electrical and electronics engineering and the M.Sc. and Ph.D. degrees in statistical machine learning from Newcastle University, U.K., in 1993, 1995, and 1998, respectively. He is currently a Professor of machine learning with Northumbria University, U.K. His research interests include the mathematical theory and algorithms for artificial intelligence, machine learning, and computational imaging. He has published more than 400 articles on these topics and won several best IEEE paper awards. He also serves as an Associate Editor to several international machine learning journals.

● ● ●