

A guide to accurate reporting in digital image processing – can anyone reproduce your quantitative analysis?

Jesse Aaron and Teng-Leong Chew*

ABSTRACT

Considerable attention has been recently paid to improving replicability and reproducibility in life science research. This has resulted in commendable efforts to standardize a variety of reagents, assays, cell lines and other resources. However, given that microscopy is a dominant tool for biologists, comparatively little discussion has been offered regarding how the proper reporting and documentation of microscopy relevant details should be handled. Image processing is a critical step of almost any microscopy-based experiment; however, improper, or incomplete reporting of its use in the literature is pervasive. The chosen details of an image processing workflow can dramatically determine the outcome of subsequent analyses, and indeed, the overall conclusions of a study. This Review aims to illustrate how proper reporting of image processing methodology improves scientific reproducibility and strengthens the biological conclusions derived from the results.

KEY WORDS: Accurate reporting, Data reproducibility, Image analysis, Image processing, Microscopy

Introduction

Image processing and analysis have become an indispensable component of the biology toolbox. There are several important forces that drive this trend, including (1) the increasing complexity and volume of modern imaging data (Ouyang and Zimmer, 2017); (2) the emphasis placed by journals and funding agencies on quantitative, rather than qualitative, data interpretation; and (3) the efforts by the biological community to reliably share data [e.g. the brain connectome (Dance, 2015), or the OMER0 project (Allan et al., 2012)]. Collectively, the ubiquitous reliance on quantitative imaging as a biological tool highlights an important problem that has remained inadequately spotlighted until recently – the lack of accurate and sufficient reporting of the processing steps required for image analysis (Marques et al., 2020).

An image file contains information that can yield important insights into a wide range of biological structures and processes. Unfortunately, any image is intrinsically an imperfect representation of the object(s) being studied (Aaron et al., 2019; Sluder and Wolf, 2013). The complexity of a biological specimen and its interaction with light often further degrades the fidelity of how the biological reality is represented in the acquired images (Booth and Patton, 2014; Ji et al., 2010) by distorting the image data, resulting in suppressed image signal and resolution.

Image processing is the vital intermediary step that aims to isolate and/or emphasize the desired signals in a raw acquired

image before eventual analysis (see Fig. 1A). In this discussion, we make a distinction between image processing and image analysis, whereby the latter comprises a vast set of diverse approaches to extract biologically meaningful, quantitative measurements from a dataset. Image processing, on the other hand, serves to digitally transform an acquired dataset by enhancing or isolating signals of interest and/or suppressing other signals and noise that will otherwise prevent accurate analysis. Furthermore, as indicated in Fig. 1A, there are often steps that are required to be taken before image data can be properly processed, termed pre-processing. Such steps generally do not serve to enhance particular features in an image per se, but rather correct for imperfections commonly encountered in imaging systems. This may include, for example, corrections for proper color channel registration (Zhang and Carter, 1999) or microscope stage drift (Lee et al., 2012).

The myriad of available image processing and analysis software packages enable biologists to perform highly complex digital operations with ease. However, herein lies the ‘double-edge sword’. While easy access to such algorithms is empowering, their foundations are grounded in complex mathematical concepts that may be unfamiliar to many researchers. An ill-advised application of these algorithms can drastically change the underlying image data in unanticipated and counterproductive ways. Yet, many image processing software programs, especially commercial solutions, provide an easy ‘one-click’ route to unwittingly alter images with little regard to understanding the resulting effects on the data, as long as the end justifies the means – the background is silenced, the image is sharpened, and the desired objects are segmented. This conceptual opacity, combined with the lack of safeguards in implementation, has perpetuated the notion that image processing requires little thought, and that the underlying details of digital operations can be ignored and go unreported.

It is important to point out that digitally processing raw images for later analyses is not intrinsically unethical; indeed, good use of processing techniques can be integral to achieving the experimental goals (Miura and Sladoje, 2020; Miura and Tosi, 2017). What becomes unacceptable, usually unintentionally, is the failure to document the processing steps taken. Proper reporting would allow possible contradictory findings to be reconciled by retracing the processing steps, even if a published conclusion is erroneous due to an ill-advised processing workflow. Furthermore, many image processing techniques, while entirely justifiable as a means to accurate quantitative analysis, are not appropriate to generate ‘improved’ images for display. Indeed, the displayed images themselves should be presented in a format that is as close as possible to the original data, with any alterations consistently applied and carefully disclosed. Finally, there do exist certain manipulations, such as image ‘cutting and pasting’, the use of lossy compression, or any alteration that is applied to a user-selected

Advanced Imaging Center, Howard Hughes Medical Institute Janelia Research Campus, Ashburn, VA 20147, USA.

*Author for correspondence (chewt@janelia.hhmi.org)

id J.A., 0000-0002-7543-9590; T.-L.C., 0000-0002-3139-7560

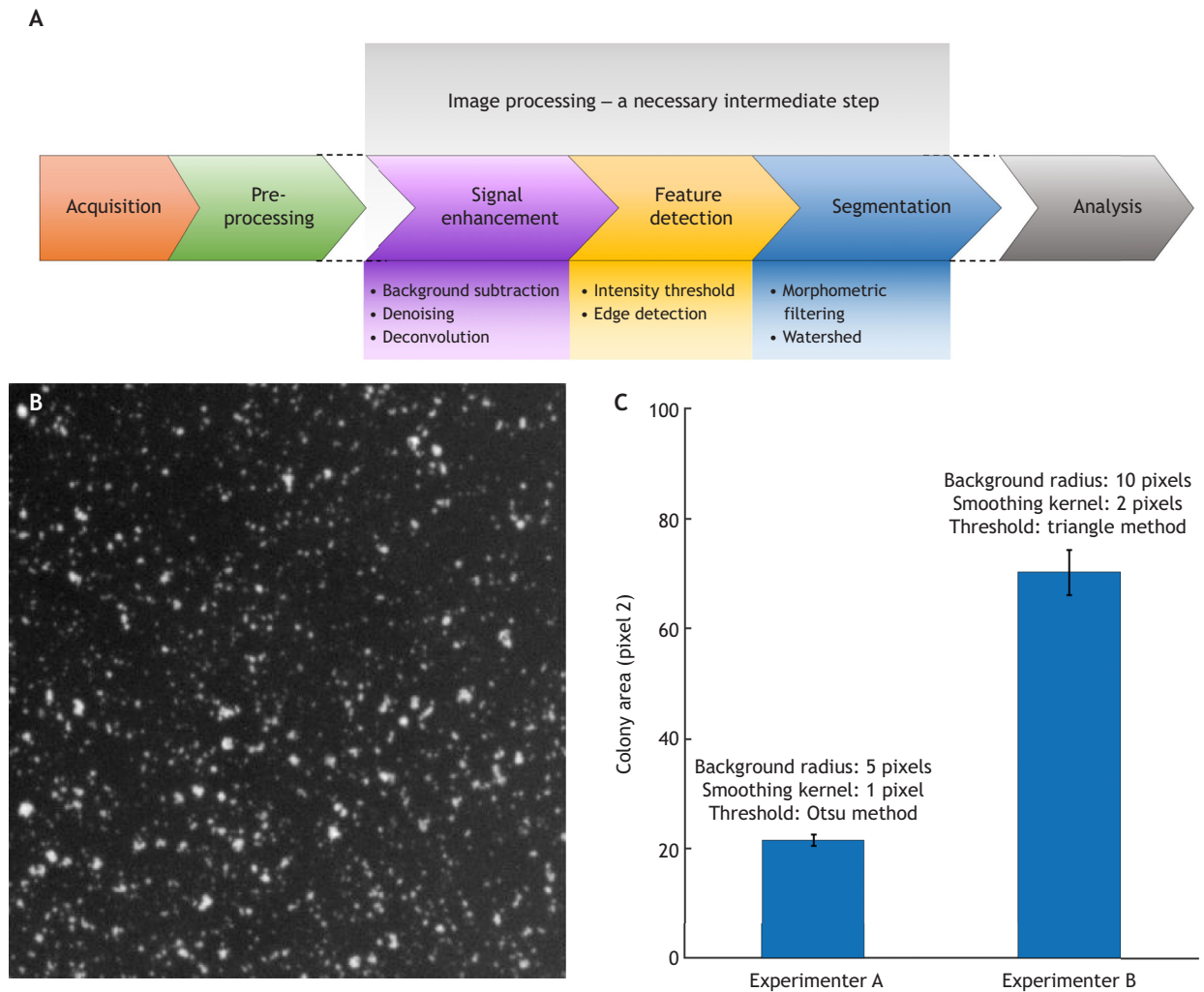


Fig. 1. The importance of image processing in imaging experiments and its impact on final results. (A) Image processing is necessary in nearly any microscopy experiment, occurring between the image acquisition stage and any subsequent analysis steps. Image processing can include feature enhancement/detection and/or segmentation operations. Each of these image processing steps can alter the final results and skew data interpretation. (B) Example fluorescence image of bacterial colonies is shown (taken from freely available FIJI sample data, see <https://imagej.net/Samples>). (C) Two researchers subjecting the image in B to different background subtraction, Gaussian de-noising, and automatic thresholding operations arrive at considerably different estimates of bacterial colony size (error bars represent s.e.m.).

sub-region of an image that will nearly always be inappropriate regardless of their application (Cromey, 2010). In any case, proper documentation of all digital alterations serves as a traceable means to identify inappropriate manipulations.

Our goal here is not to exhaustively discuss the many image processing methods available, as this is a topic far too vast for the scope of this paper. Many excellent and comprehensive treatments on the principles behind image processing are available for review (Bhabatosh and Dutta, 2011; Burger and Burge, 2013; Ekstrom, 2012; Gonzalez, 2002; Nixon and Aguado, 2019; Pitas, 2000; Szeliski, 2010; Wu et al., 2010). Likewise, it is not our goal to discourage experimenters from performing any particular technique, so long as such use is well-understood and scientifically justified. We aim to provide instructive examples that illustrate how various image processing operations, as outlined in Fig. 1A, can impact the resulting image data, and how their inaccurate and insufficient documentation impede or even prevent reproducibility. Additionally, we will provide the readers guidance and examples for proper reporting.

Reproducibility in image processing – a simple example

Reporting image-processing methodology inaccurately and/or incompletely can easily result in findings that cannot be reproduced or reconciled – often because of a seemingly trivial omission in detail. This can lead to contradictory conclusions from multiple researchers, when in fact, the true inconsistency is one of data processing. As a simple theoretical example, consider Fig. 1B, which shows a fluorescence image of bacterial colonies expressing a reporter of interest. Two different experimenters were each interested in reporting the size distribution of these colonies from the image. However, high background signal and noise were deemed to likely bias this measurement. Therefore, both experimenters decide to first employ a processing method, and report it as, ‘Prior to analysis, images were background-subtracted, followed by Gaussian smoothing, and automatic intensity threshold calculation in FIJI’.

While seemingly a comprehensive description of a reasonable approach, it is in fact insufficient to reproduce the true workflow. To see why, note the bar graph in Fig. 1C that reports the results of

average colony size (with s.e.m.) as determined by each experimenter. While both followed the same general steps as reported, the results from each researcher differ by more than three-fold.

As may be surmised, the source of this inconsistency lies in the details. While Experimenter A used a background subtraction radius (Stenberg, 1983) and Gaussian smoothing (Nixon and Aguado, 2019) kernel size of 5 and 1 pixels, respectively, experimenter B used 10 and 2 pixels. And in one case Otsu's automatic threshold (Otsu, 1979) was applied, while the triangle method (Zack et al., 1977) was used in the other. Otsu's method serves to minimize the variance in intensity between above- and below-threshold pixels, while the triangle method finds a suitable threshold based on how pixel intensities are distributed between the most common and highest values, respectively. Either approach may be technically justified; but unfortunately, such specifics can often be deemed too nuanced or unimportant to report in a manuscript with a tightly controlled word limit. As a result, even relatively small unreported differences in processing methodology can translate into apparently inconsistent results with multi-fold differences. The two experimenters may draw quite different conclusions based on these competing results, each claiming, for example, a significantly different growth rate for the bacteria. However, proper documentation and reporting will allow these two disparate results to be easily reconciled.

Image enhancement

The above example describes two common image enhancement steps – background subtraction, and image de-noising. Both techniques are designed to isolate the signal of interest from the rest of the image. To better understand why image processing is often necessary prior to analysis, it is important to understand what factors contribute to image degradation. These factors include: (1) optical diffraction that 'blurs' the image of an observed object; (2) extraneous or uninformative signals not associated with the structure of interest, referred to as 'background'; and (3) random intensity fluctuations, collectively described as noise (Lambert and Waters, 2014). Overall, biological, optical and electronic factors will always work in concert to degrade an image. The choice of strategy for recovering or amplifying the signal of interest, and how it is applied, can substantially affect the outcome of any subsequent analysis. There are image processing techniques designed to address each of these factors, which will be briefly surveyed in the following sections. We will also articulate why complete reporting of their implementation is critical.

Background subtraction

Background can refer to any detected but non-informative signal, whose presence may render subsequent analysis inaccurate or otherwise misleading. There are many sources that can contribute to background signals: (1) fluorescence generated from molecules that are not the reporter of interest (e.g. riboflavin in cell culture medium) (Aubin, 1979); (2) out-of-focus signals that may emanate from the fluorescent reporter molecules, but are too far away from the microscope focal plane to form a suitable image; and/or (3) non-specific signals, which emanate from the reporter of interest, and may be located in the focal plane, but are not associated with the biological structure of interest (e.g. non-specific binding of a secondary antibody). Regardless of their source, background signals will often skew the results of many image analysis techniques.

Fortunately, there is a retinue of methods that can estimate and remove some unwanted background signals from an image (Gonzalez, 2002). These estimates, however, almost always rely on one or more limiting assumptions. Thus, it is paramount to

clearly describe any background removal technique when reporting results, as illustrated with an image of fluorescently labeled viral particles in a cell (Fig. 2A).

To accurately measure the fluorescence intensity of individual viral particles, it is almost certainly necessary to employ some background subtraction method. There are many viable options for background subtraction, including (1) Fourier domain filtering (Nixon and Aguado, 2019), (2) Gaussian smoothing (Nixon and Aguado, 2019), or (3) 'rolling ball' subtraction (Stenberg, 1983).

Fig. 2C–E show the estimated image background signal (as surface plots) calculated from each of these methods, respectively. Note that the Gaussian smoothing and rolling ball filters each had a 5-pixel radius, and the Fourier frequency filter removed the first five frequency components (including the offset). Then, the total pixel intensity within the box indicated in Fig. 2A was summed for each background subtraction scenario as a measure of the total fluorescence signal associated with the encompassed single viral particle. Strikingly, a 75% variation in total intensity among the differently processed images is observed (Fig. 2B). Therefore, simply indicating that an image was 'background-subtracted' is not sufficiently specific to render a workflow reproducible; rather, explicitly stating the chosen method with the accompanying parameter values is required.

Image denoising

As outlined previously, noise is inevitable in any imaging system. While every effort should be made to maximize the signal-to-noise ratio (SNR) of an image, practical considerations, such as fluorophore choice, imaging speed, duration and phototoxicity, can force experimenters to compromise (Galdeen and North, 2011; Jonkman et al., 2020; North, 2006). This can render subsequent measurements more difficult. Fortunately, image analysis can be aided by denoising methods. While such algorithms never remove noise entirely from an image, they can often substantially reduce its contribution (Fan et al., 2019), albeit often with a loss in image detail. How, and to what extent this is done can vary widely from one technique to another – potentially producing widely variable results, thus reinforcing the importance of proper documentation.

The noise component of an image is often assumed to vary more from one pixel to the next than does the structure itself. Thus, many denoising algorithms force each pixel within an image to be more similar in intensity to its neighbors. As an example, a noisy image of a neuron is shown in Fig. 3A; it was subjected to Gaussian smoothing (Nixon and Aguado, 2019), median filtering (Huang et al., 1979) and non-local means denoising (NLMD) (Buades et al., 2011), with end-results shown in Fig. 3B–D, respectively. The Gaussian smoothing and median filtering each used a filter radius of 6 pixels. NLMD was used as implemented in FIJI, with a smoothing factor of 1 and automatic estimation of sigma parameters. In Fig. 3E, an intensity profile plot is shown for the same region in each image, indicated by white lines, for the original data and for the indicated denoising method. It is useful to compare areas primarily with and without the signal of interest, indicated by the yellow and light blue areas, respectively.

Note that Gaussian smoothing results in less noise than the median filter in areas of low signal (light blue areas). However, it also tends to reduce the intensity and broaden the apparent axon widths by blurring their edges (yellow areas), as denoted by the arrows. The median filter results in greater pixel variation in the dimmer areas of the image, but with better intensity preservation and sharper axon edges.

NLMD (shown in Fig. 3D) can often circumvent the limitations of the Gaussian and median filtering methods, as it does not assume

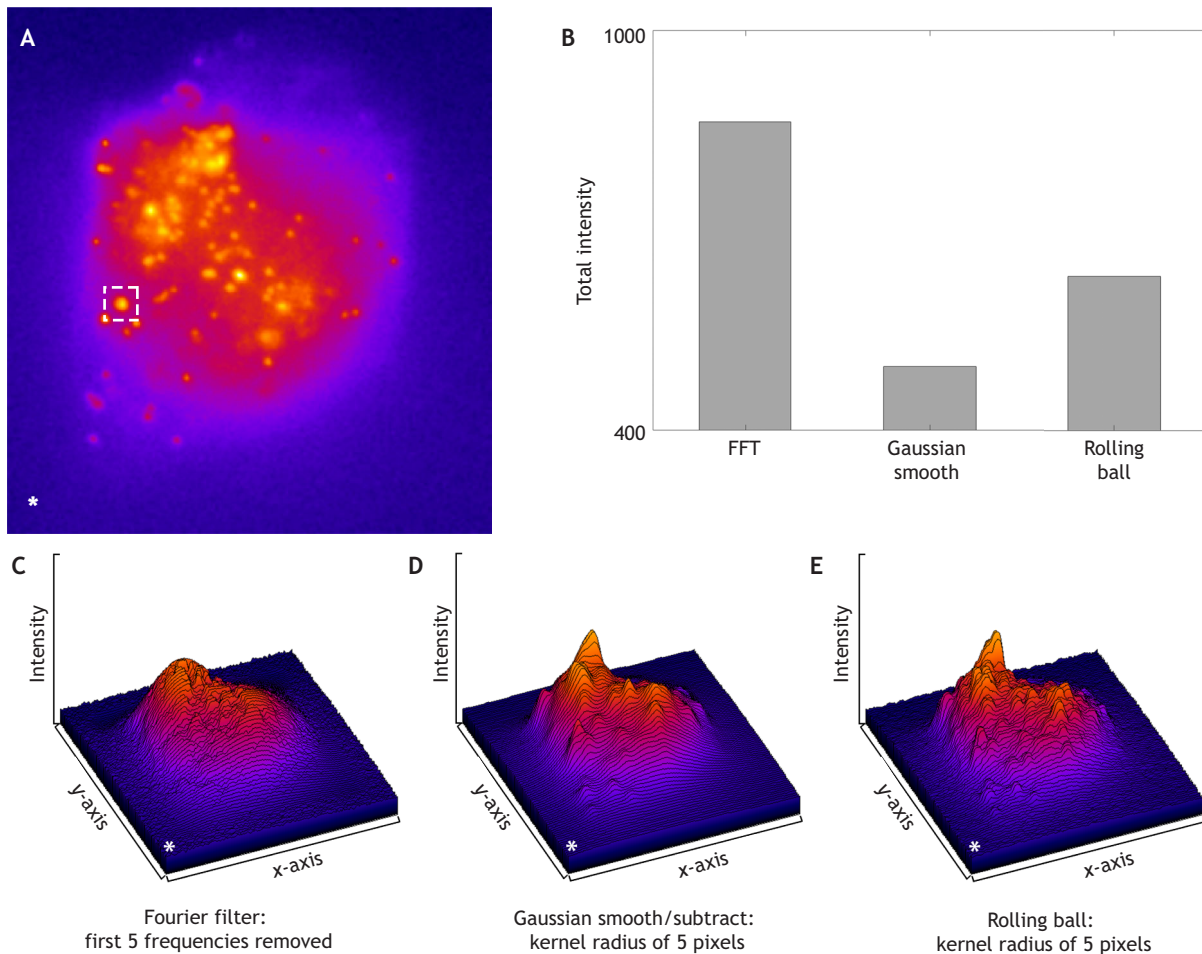


Fig. 2. Background subtraction. (A) Image of a cell infected with fluorescently labeled viral particles, taken from FIJI sample data (<https://imagej.net/Samples>). (B–E) The total summed pixel intensity of the single virus encompassed by the white box in A is plotted (B) after application of three background-removal techniques, specifically, Fourier-based filtering – with the first five frequency components removed (C), as well as Gaussian smoothing (D) and rolling minimum methods (E), both of the latter using a 5-pixel radius kernel size. Note that the orientation surface plots in C–E is rotated relative to the original data. An asterisk (*) indicates the corresponding lower left corner in each image C–E relative to the same indicator in A. The apparent inconsistency in integrated intensity shown in B can be attributed to the different estimated image backgrounds calculated from each method shown in C–E.

that neighboring pixels need to have similar intensity. But because of its ‘non-local’ nature, NLMD can give different results depending on whether an image is cropped – another potential source of irreproducibility.

Deconvolution

Background signal and noise can represent formidable foes in image analysis. But diffraction has historically been the ultimate limiter of image quality. Any acquired image will be the convolution of the structure being observed and the point spread function (PSF) of the microscope (Sibarita, 2005). Deconvolution techniques therefore attempt to reverse the effects of diffraction by approximating a real structure, given (1) a captured image and (2) a known or estimated PSF. The desired result is a crisper image whose details may better facilitate image analysis. The sharpened image, however, can only be an estimate. The accuracy of this estimate is further complicated by the extent of image noise and the accuracy of the assumed or measured PSF. In practice, deconvolution algorithms must balance image deblurring and noise – increasing the former will unfortunately increase the latter. Many algorithms, therefore, incorporate one or more user-defined parameters that attempt to optimize that balance.

Deconvolution techniques can be grouped into ‘direct’ or ‘iterative’ algorithms. A common example of a direct method is based on a Wiener filter (Gonzalez, 2002; Sekko et al., 1999). The Richardson–Lucy algorithm, on the other hand, is a common iterative scheme (Lucy, 1974; Richardson, 1972), whereby an experimenter specifies the number of cycles to perform. In any study that makes use of deconvolution, the specific algorithm employed should be cited. In addition, reporting the user-defined parameters is equally vital to ensure reproducibility. To illustrate this point, consider the image in Fig. 4A, which shows mitochondria in several HeLa cells. The images in Fig. 4B–E show an enlargement of the area denoted by the white box in Fig. 4A and indicate the results of four deconvolution attempts. Each image was obtained using the same Richardson–Lucy algorithm. The only variation between each image is the number of user-defined iterations, with Fig. 4B–E corresponding to 5, 10, 25 and 75 iterations, respectively. This example illustrates the marked effect of iteration number on a deconvolved image. Note that, particularly in Fig. 4E, the chosen number of iterations is too large, as unintended structures have become apparent in dim regions of the image. But a complete reporting of this processing step can be vital in identifying such errors. More generally, the chosen number of deconvolution

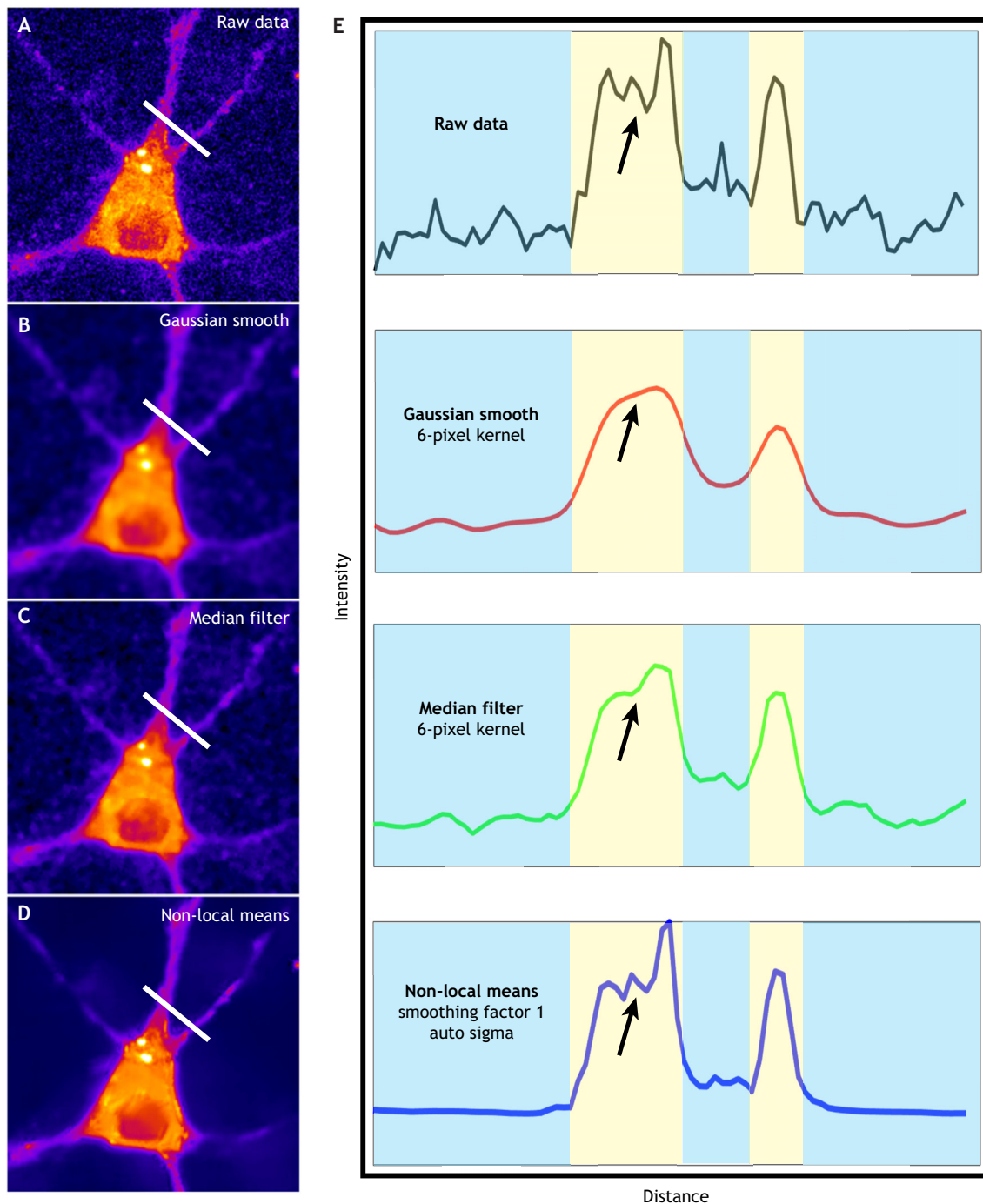


Fig. 3. Image de-noising. (A) Image of a fluorescently labeled neuron (taken from FIJI sample data; <https://imagej.net/Samples>) with relatively low signal-to-noise ratio (SNR). (B-D) The image in A is subjected to three denoising techniques – Gaussian smoothing (B), median filtering (C), and non-local means (D). (E) Pixel intensity profiles across the areas indicated by the white line in each image are shown, with dark and bright regions of the image denoted by light blue and yellow areas in the graphs, respectively. Parameters for each denoising method are listed in the graph. Arrows indicate the apparent variation in structure due to different denoising methods.

iterations can have significant effects on downstream analyses; in this example, measurements of mitochondrial size, shape, and number can all be expected to vary depending on precisely how the preceding deconvolution was implemented. However, this image processing detail is often omitted in published studies.

Deconvolution can be a conceptually difficult image-processing task for many researchers, in part due to its mathematical complexity. Additionally, commercial implementations may not

disclose a rigorous description of the exact algorithm used for proprietary reasons, or a clear explanation of the user-defined parameters. This is compounded in the case of iterative algorithms, where the optimal number of deconvolution iterations may not be clear. Some implementations make use of a particular image quality criterion that, when met, would terminate the operation (Laasmaa et al., 2010). In any case, researchers should obtain as much information about the chosen deconvolution algorithm as possible

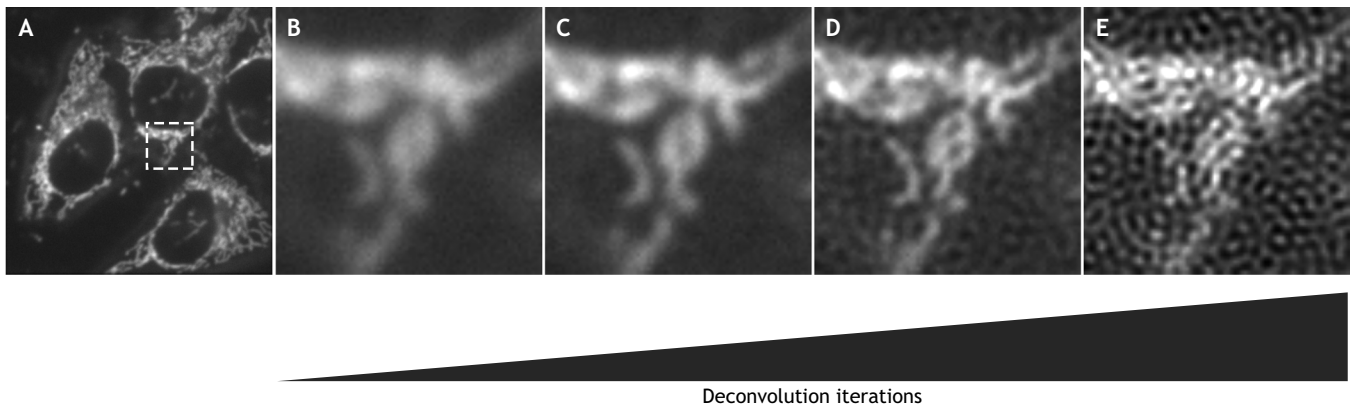


Fig. 4. Image deconvolution. (A) An image of fluorescently labeled mitochondria (taken from FIJI sample data; <https://imagej.net/Samples>) was subjected to Richardson–Lucy deconvolution using an assumed point spread function (PSF). (B–E) Effects of deconvolution of the area indicated by the white box in A after 5 (B), 10 (C), 25 (D), and 75 (E) iterations of the same Richardson–Lucy deconvolution algorithm are shown.

from their microscopy facility. At a minimum, any utilized commercial software tools should be clearly listed (with version number), along with all the requisite user-defined parameters. While details on commercial implementation of deconvolution may be difficult to obtain, there are other open source options available (Sage et al., 2017; Sun et al., 2009) with more transparent descriptions.

Feature detection and object segmentation

The image enhancement techniques described above can be powerful techniques to reduce unwanted signals and noise, and to amplify signals of interest. However, this may only be an initial step in an image processing workflow. Distinguishing pixels that primarily contain signal of interest – termed foreground – from those that do not is often the next logical processing step toward eventual analyses.

Intensity threshold

The simplest way to identify foreground pixels is to select a minimum threshold intensity value, although other strategies exist (Maini and Aggarwal, 2009). Manual selection of a threshold value is often used, based on a visual inspection of the resulting foreground pixels, but this approach can be prone to user-bias. To minimize such bias, automated threshold methods, which do not rely on human perception, are common alternatives.

Few topics have seen as wide a variety of implementations as automated intensity threshold calculation. For example, FIJI (Rueden et al., 2017; Schindelin et al., 2015; Schneider et al., 2012) currently offers no fewer than 17 different methods to automatically calculate an image intensity threshold with a diversity in results (Fig. 5). Fig. 5A features an image of fluorescently labeled intracellular vesicles after background subtraction (with rolling ball subtraction, 5-pixel radius) and denoising (using median filtering, 3-pixel radius). Three different automatic thresholding techniques (Kittler and Illingworth, 1986; Otsu, 1979; Prewitt and Mendelsohn, 1966), each included in ImageJ/FIJI, were applied with results displayed in Fig. 5B–D, respectively. White pixels denote foreground, while black pixels correspond to those pixels deemed not to contain signal of interest. Clearly, different thresholding techniques return very different sets of foreground pixels. This is primarily due to the unique assumptions of each method about the underlying pixel intensity distribution.

As suggested before, microscopy facility staff can often be the best initial resource in selecting an appropriate threshold application

technique. However, describing the method being used and its associated parameters is of even greater importance. Even if a threshold method is poorly chosen, documenting its use can prompt a manuscript reviewer to suggest a different method prior to publication, and/or resolve any issues of reproducibility after publication.

Object segmentation

Object segmentation, in many instances, can be thought of as a final processing step that groups adjacent foreground pixels into discrete objects for analysis. Despite the apparent simplicity of this task, there are important details that will affect the outcome. Most fundamentally, there are multiple ways to define ‘adjacent’ in this context. For example, pixels can be joined into an object if they share a common edge (4-connectivity rule), or if they share either an edge or a corner (8-connectivity rule) (He et al., 2017). While seemingly inconsequential, this difference can have a large impact on the measured number, size, and shape of objects – particularly if those objects are comprised of relatively few pixels.

Once pixels have been grouped, further refinements are often made. Spuriously bright pixels are often misclassified as objects but tend to be of different size or shape from the objects of interest. Thus, it is common to simply not consider any objects that fall outside acceptable bounds of shape and size. Such operations are not inherently unethical, so long as they are well-justified based on what is known about the biological structure(s), and so long as they are applied in a consistent fashion. But most importantly, such morphological filtering should always be noted with a complete description of all parameters that were considered.

Binary operations can also refine segmented objects. These operations represent image processing techniques that are reserved only for binary images – that is, images with only two possible intensity values, which typically denote whether a pixel is above or below a specified threshold, respectively. For example, binary erosion followed by dilation – called ‘image opening’ – can be useful to remove small unwanted objects produced by spurious signals. Similarly, binary dilation followed by erosion – or ‘image closing’ – can be useful to fill holes within objects that happen to have areas of relatively dim signal (Gonzalez, 2002). Watershed operations can separate objects that share a common border (Roerdink and Meijster, 2000). Although each of these processing steps aim to improve upon the initially segmented objects, many results will depend heavily on their precise implementation.

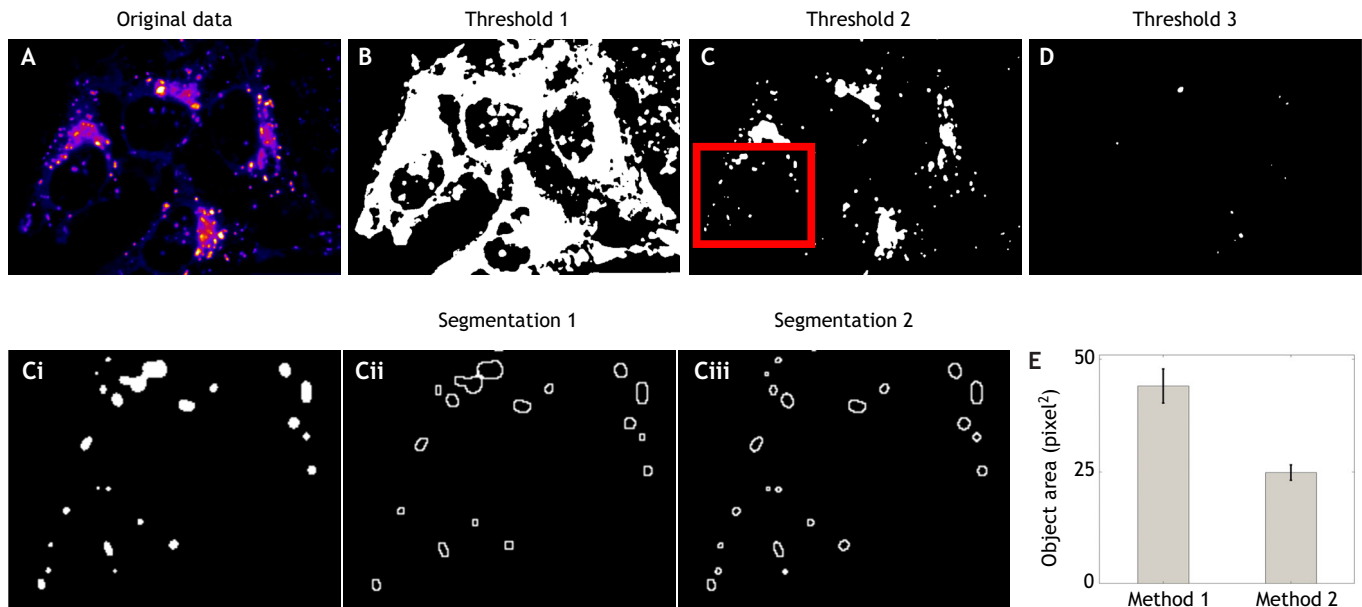


Fig. 5. Feature extraction and object segmentation. (A) Fluorescence image of labeled intracellular vesicles (taken from FIJI sample data; <https://imagej.net/Samples>) after background removal (with rolling ball subtraction, 5-pixel radius) and denoising (using median filtering, 3-pixel radius). The results of three different automatic intensity-threshold algorithms applied to A, are shown in B–D, corresponding to the methods described in Kittler and Illingworth (1986), Otsu (1979) and Prewitt and Mendelsohn (1966), respectively. The image in C (enlarged views of area denoted by the red box, as shown in Ci), was then analyzed to measure the average vesicle size using two segmentation methods to refine the objects to be considered (Cii and Ciii). Method 1 used a single iteration of image opening, followed by a binary watershed operation (Cii), while Method 2 applied a maximum object size (<100 pixel²) and minimum circularity (>0.5) for any object to be considered (Ciii). (E) There is an approximate 75% difference in the vesicle area obtained with Methods 1 and 2. Error bars are s.e.m.

This can be seen by continuing to consider the images in Fig. 5. Assuming the threshold technique represented in Fig. 5C, with the enlarged view shown in Fig. 5Ci (Otsu, 1979), is deemed optimal, measuring the average area of each vesicle may seem to be a straightforward task and not subject to hazards of irreproducibility. However, the removal of specious small objects and apparent vesicle aggregates may be critical to avoid skewing these results. One experimenter may choose to use image opening and a watershed operation to accomplish this task, which we term Segmentation Method 1. Another experimenter may simply apply a maximum object area and minimum object circularity to select putative vesicles, termed Segmentation Method 2. While either approach may be appropriate, the segmented objects derived from each method are notably different, as illustrated in Fig. 5Cii and Ciii. It should then come as little surprise that when computing the average vesicle area, a 75% difference in outcome is found between these two methods (Fig. 5E). Importantly, differing results would be expected even if individual parameters within each method are varied. For example, varying the acceptable bounds of object circularity in Method 2 is likely to change the quantitative outcome. Thus, even when a pair of images is subject to identical background subtraction, image denoising and automated threshold application, the analytical result can still vary by a surprising amount due to differences in object segmentation and refinement.

An often overlooked, yet vital, aspect of proper reporting of image processing is the sequence in which multiple tasks are applied to an image. For example, applying a rolling ball background subtraction, followed by median-filter denoising will produce a different image from the one obtained with these same steps performed in the reverse order. As such, differences in the processed image will almost certainly affect any later segmentation and, ultimately, an analysis outcome. Thus, when describing any image processing workflow, careful attention should be paid to denote

both the details of all individual steps, as well as the precise order in which they were applied. We have included a user-fillable form (see Table S1) aimed at aiding readers to properly document their image processing workflows.

The various examples shown here serve to illustrate how seemingly inconsequential details in the implementation of image processing can have dramatic effects on a final image analysis outcome. A common reason these details are deemed unimportant is largely due to the fact that many imaging experiments are comparative in nature. As such, it can be tempting to presume that the overall difference in outcome between experimental and control samples will be preserved, regardless of exactly how each image is processed and analyzed, so long as such workflows are applied consistently. It must be stressed, however, that this assertion cannot be assumed to be true. The preceding examples indicate that varying even one parameter in a single processing step can produce unintuitive and non-linear effects in the final outcome. Thus, extending such variability over several steps in a workflow would therefore render any changes in final results wholly unpredictable.

Perspectives and conclusions

As described here and elsewhere, digital image processing is an integral step toward quantitative image analyses. In fact, judicious utilization of these methods – not avoidance of them out of indiscriminate ethical concern – is often required for the success of accurate quantitative image analysis (Miura and Sladoje, 2020). A more insidious problem, however, is the lack of proper reporting of these essential digital operations in the literature (Marques et al., 2020). The preceding examples demonstrate the often-significant effect on the analysis outcome as a result of inconsistently applying seemingly insignificant image processing factors. Promoting accurate and sufficient documentation of image processing is not novel. In fact, it has been advocated by many (Lee and Kitaoka,

2018; Limare and Morel, 2011; National Academies of Sciences and Medicine, 2019). Furthermore, there are a number of organized efforts by the imaging science community dedicated to providing resources and guidelines aimed at improving microscopy data quality assurance and transparency, as exemplified by the OME (Allan et al., 2012) and QUAREP-LiMi (Nelson et al., 2021 preprint) initiatives. Furthermore, in a companion Review, we follow a similar approach to discuss how to properly report image acquisition parameters (Heddleston et al., 2021). Indeed, the increasing utilization of public data repositories such as OMERO and Zenodo (Sicilia et al., 2017) can only be truly effective with proper accompanying reporting of both image acquisition and processing/analysis metadata. Unfortunately, advocacy alone has thus far not garnered a sufficient response. In this Review and its companion (Heddleston et al., 2021), we take a different approach to this issue than mere advocacy.

Here, we assume that the widespread problem of under- and mis-reporting of image processing steps in published literature is not primarily driven by ill-intention to conceal pertinent information, but due to the general lack of understanding and appreciation of how these methods could affect the data, often in unintuitive ways. In other words, we presume many end-users simply do not recognize what matters, and therefore do not know precisely what to report. Shaw and Hinchcliffe (2013) observed that the propagation of computer-aided image analysis among life scientists often follows an 'oral' tradition, facilitated mainly by colleagues or collaborators. Another alarming occurrence is to see early-career scientists, such as postdocs and graduate students, receiving little or no guidance at all when performing image processing and analysis tasks. To further exacerbate the challenge, many imaging facilities – where such support should be readily available – do not receive sufficient institutional and federal funding to offer image processing and analyses as a core service, or training of their user base (Ferrando-May et al., 2016). This problem frequently leads to a situation wherein the biologist arbitrarily tries to process the image without a logically designed strategy, resulting in either suboptimal results, or in many cases, failure to achieve the experimental goals. Even more consequentially, it may also inadvertently allow unethical image manipulations to go undetected.

It is impractical and unreasonable to expect a full understanding of the theories behind every image processing step. Here, we aim to provide readers with an appreciation of how image data can be affected by these digital operations, and to increase the awareness of why it matters to report them. To help readers toward this goal, we have summarized several major image-processing tasks, examples of common algorithms used for each task, and the corresponding parameters that should be reported (Table 1). While not comprehensive, this table can aid readers in critically approaching a broad range of image-processing algorithms with an eye toward documenting and reporting the necessary elements to maintain reproducibility. More importantly, this is expected to raise awareness of the issues, so that researchers seek expert assistance when in doubt, thereby potentially avoiding costly mistakes.

A myriad of strategies can aid researchers in tracking, properly documenting and, ultimately, standardizing their image processing workflows. While the simplest approaches can merely rely on diligent notetaking and screenshot documentation, there are built-in features in many software packages that can further aid in this regard. For example, the OMERO platform (Allan et al., 2012) can allow storage and tracking of not just imaging data, but extensive annotations and metadata associated with both image acquisition and digital transformations. ImageJ/FIJI (Rueden et al., 2017; Schindelin et al., 2015; Schneider et al., 2012) features the ability to record individual processing steps, together with the associated parameters. This capability can be leveraged further to easily create customized image processing macros that can be applied to large data collections, thus increasing both the transparency, as well as efficiency of an image processing workflow. Similar functions exist in commercial software packages that allow researchers to store and recall function histories for later reporting. In general, workflow macros and custom software code should always be shared whenever practical. We encourage readers to download a user-fillable form (available as Table S1), based on Table 1, in order to aid in summarizing their image processing workflows for later reporting.

The need for accurate reporting will only become more acute as machine learning techniques become more commonly applied to perform a broad spectrum of image processing tasks (Arganda-

Table 1. Summary of the discussed image processing tasks, examples of common algorithms for each, and the requisite parameters to report for each example

| Processing task | Common algorithms | Parameter(s) to report |
|------------------------|--|---|
| Background subtraction | Rolling ball/rolling minimum Gaussian smoothing and subtraction FFT high pass filter | Kernel size and shape Kernel size and shape Frequency cut-off values |
| Denoising | Gaussian smoothing Median filtering Non-local means | Kernel size and shape Kernel size and shape Noise sigma and smoothing value |
| Deconvolution | Wiener deconvolution Richardson–Lucy Blind | PSF and Wiener factor (1/SNR) PSF and number of iterations Estimated PSF (or none) |
| Intensity threshold | Manual Automatic Local | Intensity value or percentile None Varies |
| Segmentation | Pixel connectivity Morphometric filtering Binary operations | 4 or 8 (2D images); 6, 18 or 26 (3D images) Shape/size parameter(s) with upper and lower bounds Operations type (e.g. erosion, dilation), number and order of implementations |

Carreras et al., 2017; Smal et al., 2010; Sommer and Gerlich, 2013; Waller and Tian, 2015). This computational field, while often extraordinarily powerful (Ounkomol et al., 2018), presents formidable challenges regarding experimental reproducibility – not least of which is due to their reliance on training datasets. This is further complicated by the fact that the bases upon which machine learning algorithms arrive at decisions are generally not easily conveyed in non-specialist terms.

The end-user community, however, is not our sole target audience. This systemic problem in reporting is, disappointingly, further abetted by the poor implementation of the editorial guidelines set by the journals themselves (Marques et al., 2020). There is unfortunately no fool-proof mechanism that would serve as a perfect remedy to poor reporting. However, it is helpful to identify a series of ‘checkpoints’ that would have to be breached to lead to an irreproducible quantitative analysis – (1) lack of documentation of the image processing workflow by the end-users during the experimentation phase, (2) lack of oversight by someone knowledgeable to identify the mis- or under-reporting during the manuscript preparation phase, and (3) failure of the journals to assign the manuscript to at least one reviewer who could guide the inclusion of this pertinent information during the peer-review phase.

The root problem of under-documentation and under-reporting by the end-users during the experimental phase can only be remedied by better training and is therefore a loftier, longer term goal. However, there are immediate actionable solutions that could be taken by the journals to blunt the perpetuation of the problem. Firstly, inquire during manuscript submission if the authors had altered any of the images for analysis and whether all the image processing workflow and parameters had been documented. While this may not completely stem the problem, it would convey the gravity of the matter. More importantly, the emphasis by journals would encourage end-users to keep the record of image processing workflow as detailed as they would a biochemical or molecular biology assay. Secondly, the journal editors should assign the manuscript to at least one reviewer who has the related expertise when image data analysis either makes up a considerable proportion of the data presented in the manuscript, or the key findings are generated through image quantification. Importantly, ensuring proper reporting of image processing workflows not only helps to guard against unintended non-reproducibility, but can also serve as a valuable means in identifying and deterring deliberate malfeasance before publication of a study. Such instances, while comparatively rare, arguably damage public confidence in the scientific method far more than honest mistakes. A digital image represents an information-rich data source, from which valuable quantitative biological insights can be derived. However, simply demanding that modern biological data be more quantitative rather than descriptive is inadequate without the ability for independent validation. Scientific data are only as credible and valuable as they can be verified. To overlook the importance of documenting accurate and sufficient image processing details subsequently renders the published results unverifiable. This oversight, therefore, undermines all the time, effort and resources that go into generating the data for a publication, and severely diminishes the value of the scientific discovery.

Acknowledgements

We thank Wendye Quayle for her design assistance in creating the downloadable supplemental forms.

Competing interests

The authors declare no competing or financial interests.

Funding

The Advanced Imaging Center at Janelia Research Campus is generously supported by the Gordon and Betty Moore Foundation and Howard Hughes Medical Institute.

Supplementary information

Supplementary information available online at <https://jcs.biologists.org/lookup/doi/10.1242/jcs.254151.supplemental>

References

- Aaron, J., Wait, E., DeSantis, M. and Chew, T.-L. (2019). Practical considerations in particle and object tracking and analysis. *Curr. Protoc. Cell Biol.* **83**, e88. doi:10.1002/cpcb.88
- Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., MacDonald, D., Moore, W. J., Neves, C., Patterson, A. et al. (2012). OMERO: flexible, model-driven data management for experimental biology. *Nat. Methods* **9**, 245–253. doi:10.1038/nmeth.1896
- Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A. and Sebastian Seung, H. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* **33**, 2424–2426. doi:10.1093/bioinformatics/btx180
- Aubin, J. E. (1979). Autofluorescence of viable cultured mammalian cells. *J. Histochem. Cytochem.* **27**, 36–43. doi:10.1177/27.1.220325
- Bhabatosh, C. and Dutta, D. (2011). *Digital Image Processing and Analysis*. PHI Learning Pvt. Ltd.
- Booth, M. J. and Patton, B. R. (2014). Adaptive optics for fluorescence microscopy. In *Fluorescence Microscopy* (ed. A. Cornea), pp. 15–33. Academic Press.
- Buades, A., Coll, B. and Morel, J.-M. (2011). Non-local means denoising. *Image Process. Line* **1**, 208–212. doi:10.5201/ijpol.2011.bcm_nlm
- Burger, W. and Burge, M. J. (2013). *Principles of Digital Image Processing: Advanced Methods*. London: Springer.
- Cromey, D. W. (2010). Avoiding twisted pixels: ethical guidelines for the appropriate use and manipulation of scientific digital images. *Sci. Eng. Ethics* **16**, 639–667. doi:10.1007/s11948-010-9201-y
- Dance, A. (2015). Connectomes make the map. *Nature* **526**, 147–149. doi:10.1038/526147a
- Ekstrom, M. P. (2012). *Digital Image Processing Techniques*. Elsevier Science.
- Fan, L., Zhang, F., Fan, H. and Zhang, C. (2019). Brief review of image denoising techniques. *Vis. Comput. Ind. Biomed. Art* **2**, 7. doi:10.1186/s42492-019-0016-7
- Ferrando-May, E., Hartmann, H., Reymann, J., Ansari, N., Utz, N., Fried, H. U., Kukat, C., Peychl, J., Liebig, C., Terjung, S. et al. (2016). Advanced light microscopy core facilities: Balancing service, science and career. *Microscopy Res. Technique* **79**, 463–479. doi:10.1002/jemt.22648
- Galdeen, S. A. and North, A. J. (2011). Live cell fluorescence microscopy techniques. In: *Cell Migration. Methods in Molecular Biology* (Methods and Protocols) (eds Wells C., Parsons M.), vol. 769. Humana Press. doi:10.1007/978-1-61779-207-6_14
- Gonzalez, R. C. (2002). *Digital Image Processing*, 2nd edn. Prentice-Hall Of India Pvt. Limited.
- He, L., Ren, X., Gao, Q., Zhao, X., Yao, B. and Chao, Y. (2017). The connected-component labeling problem: a review of state-of-the-art algorithms. *Pattern Recognit.* **70**, 25–43. doi:10.1016/j.patcog.2017.04.018
- Huang, T., Yang, G. and Tang, G. (1979). A fast two-dimensional median filtering algorithm. *IEEE Trans. Acoust.* **27**, 13–18. doi:10.1109/TASSP.1979.1163188
- Heddlleston, J. M., Aaron, J. S., Khoun, S. and Chew, T.-L. (2021). A guide to accurate reporting in digital image acquisition – can anyone replicate your microscopy data? *J. Cell Sci.* **134**, jcs254144. doi:10.1242/jcs.254144
- Ji, N., Milkie, D. E. and Betzig, E. (2010). Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues. *Nat. Methods* **7**, 141–147. doi:10.1038/nmeth.1411
- Jonkman, J., Brown, C. M., Wright, G. D., Anderson, K. I. and North, A. J. (2020). Tutorial: guidance for quantitative confocal microscopy. *Nat. Protoc.* **15**, 1585–1611. doi:10.1038/s41596-020-0313-9
- Kittler, J. and Illingworth, J. (1986). Minimum error thresholding. *Pattern Recognit.* **19**, 41–47. doi:10.1016/0031-3203(86)90030-0
- Laasmaa, M., Vendelin, M. and Peterson, P. (2010). 3D confocal microscope image enhancement by Richardson-Lucy deconvolution algorithm with total variation regularization: parameters estimation. *Biophys. J.* **98**, 178a. doi:10.1016/j.bpj.2009.12.954
- Lambert, T. J. and Waters, J. C. (2014). Chapter 3 - assessing camera performance for quantitative microscopy. In *Quantitative Imaging in Cell Biology* (ed. J. C. Waters and T. Wittman), pp. 35–53. Academic Press.
- Lee, J.-Y. and Kitaoka, M. (2018). A beginner's guide to rigor and reproducibility in fluorescence imaging experiments. *Mol. Biol. Cell* **29**, 1519–1525. doi:10.1091/mbc.E17-05-0276
- Lee, S. H., Baday, M., Tjioe, M., Simonson, P. D., Zhang, R., Cai, E. and Selvin, P. R. (2012). Using fixed fiducial markers for stage drift correction. *Opt. Exp.* **20**, 12177–12183. doi:10.1364/OE.20.012177

- Limare, N. and Morel, J.-M.** (2011). The IPOL initiative: publishing and testing algorithms on line for reproducible research in image processing. *Proc. Comput. Sci.* **4**, 716-725. doi:10.1016/j.procs.2011.04.075
- Lucy, L. B.** (1974). An iterative technique for the rectification of observed distributions. *Astron. J.* **79**, 745. doi:10.1086/111605
- Maini, R. and Aggarwal, H.** (2009). Study and comparison of various image edge detection techniques. *Int. J. Image Process.* **3**, 1-11.
- Marques, G., Pengo, T. and Sanders, M. A.** (2020). Imaging methods are vastly underreported in biomedical research. *eLife* **9**, e55133. doi:10.7554/eLife.55133
- Miura, K. and Sladoje, N.** (2020). *Bioimage Data Analysis Workflows*. Cham: Springer Nature PP.
- Miura, K. and Tosi, S.** (2017). Epilogue. In *Standard and Super-Resolution Bioimaging Data Analysis* (eds A. Wheeler and R. Henriques), Wiley. doi:10.1002/9781119096948.ch11
- National Academies of Sciences, Engineering and Medicine.** (2019). *Reproducibility and Replicability in Science*. National Academies Press.
- Nelson, G., Boehm, U., Bagley, S., Bajcsy, P., Bischof, J., Brown, C. M., Dauphin, A., Dobbie, I. M., Eriksson, J. E., Faklaris, O. et al.** (2021). QUAREP-LiMi: A community-driven initiative to establish guidelines for quality assessment and reproducibility for instruments and images in light microscopy. arxiv.org/abs/2101.09153
- Nixon, M. and Aguado, A.** (2019). *Feature Extraction and Image Processing for Computer Vision*. Elsevier Science.
- North, A. J.** (2006). Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. *J. Cell Biol.* **172**, 9-18. doi:10.1083/jcb.200507103
- Otsu, N.** (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man, Cybern.* **9**, 62-66. doi:10.1109/TSMC.1979.4310076
- Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. and Johnson, G. R.** (2018). Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917-920. doi:10.1038/s41592-018-0111-2
- Ouyang, W. and Zimmer, C.** (2017). The imaging tsunami: computational opportunities and challenges. *Curr. Opin. Syst. Biol.* **4**, 105-113. doi:10.1016/j.coisb.2017.07.011
- Pitas, I.** (2000). *Digital Image Processing Algorithms and Applications*. Wiley.
- Prewitt, J. M. S. and Mendelsohn, M. L.** (1966). The analysis of cell images. *Ann. NY Acad. Sci.* **128**, 1035-1053. doi:10.1111/j.1749-6632.1965.tb11715.x
- Richardson, W. H.** (1972). Bayesian-based iterative method of image restoration. *J. Optical Soc. Am.* **62**, 55-59. doi:10.1364/JOSA.62.000055
- Roerdink, J. B. T. M. and Meijster, A.** (2000). The watershed transform: definitions, algorithms and parallelization strategies. *Fundam. informaticae* **41**, 187-228. doi:10.3233/FI-2000-411207
- Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T. and Eliceiri, K. W.** (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529. doi:10.1186/s12859-017-1934-z
- Sage, D., Donati, L., Soulez, F., Fortun, D., Schmit, G., Seitz, A., Guiet, R., Vonesch, C. and Unser, M.** (2017). DeconvolutionLab2: an open-source software for deconvolution microscopy. *Methods* **115**, 28-41. doi:10.1016/j.ymeth.2016.12.015
- Schindelin, J., Rueden, C. T., Hiner, M. C. and Eliceiri, K. W.** (2015). The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518-529. doi:10.1002/mrd.22489
- Schneider, C. A., Rasband, W. S. and Eliceiri, K. W.** (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675. doi:10.1038/nmeth.2089
- Sekko, E., Thomas, G. and Boukrouche, A.** (1999). A deconvolution technique using optimal Wiener filtering and regularization. *Signal Process.* **72**, 23-32. doi:10.1016/S0165-1684(98)00161-3
- Shaw, S. L. and Hinchcliffe, E. H.** (2013). 65,000 shades of grey: use of digital image files in light microscopy. *Methods Cell Biol.* **114**, 317-336. doi:10.1016/B978-0-12-407761-4.00013-0
- Sibarita, J.-B.** (2005). Deconvolution microscopy. In *Microscopy Techniques* (ed. J. Rietdorf), pp. 201-243. Springer.
- Sicilia, M.-A., García-Barriocanal, E. and Sánchez-Alonso, S.** (2017). Community curation in open dataset repositories: insights from Zenodo. *Proc. Comput. Sci.* **106**, 54-60. doi:10.1016/j.procs.2017.03.009
- Sluder, G. and Wolf, D. E.** (2013). *Digital Microscopy*. Academic Press.
- Smal, I., Loog, M., Niessen, W. and Meijering, E.** (2010). Quantitative Comparison of Spot Detection Methods in Fluorescence Microscopy. *IEEE Trans. Med. Imaging* **29**, 282-301. doi:10.1109/TMI.2009.2025127
- Sommer, C. and Gerlich, D. W.** (2013). Machine learning in cell biology—teaching computers to recognize phenotypes. *J. Cell Sci.* **126**, 5529-5539. doi:10.1242/jcs.123604
- Sternberg, S.** (1983). Biomedical image processing. *Computer.* **16**, 22-34. doi:10.1109/MC.1983.1654163
- Sun, Y., Davis, P., Kosmacek, E. A., Ianzini, F. and Mackey, M. A.** (2009). An open-source deconvolution software package for 3-D quantitative fluorescence microscopy imaging. *J. Microsc.* **236**, 180-193. doi:10.1111/j.1365-2818.2009.03205.x
- Szeliski, R.** (2010). *Computer Vision: Algorithms and Applications*. London: Springer.
- Waller, L. and Tian, L.** (2015). Computational imaging: machine learning for 3D microscopy. *Nature* **523**, 416-417. doi:10.1038/523416a
- Wu, Q., Merchant, F. and Castleman, K.** (2010). *Microscope Image Processing*. Elsevier Science.
- Zack, G. W., Rogers, W. E. and Latt, S. A.** (1977). Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **25**, 741-753. doi:10.1177/25.7.70454
- Zhang, Y.-Z. and Carter, D.** (1999). Multicolor fluorescent microspheres as calibration standards for confocal laser scanning microscopy. *Appl. Immunohistochem. Mol. Morphol.* **7**, 156-163. doi:10.1097/00129039-199906000-00009