

INVITED REVIEW

Made to measure: An introduction to quantifying microscopy data in the life sciences

Siân Culley¹ | Alicia Cuber Caballero¹ | Jemima J Burden² | Virginie Uhlmann³ 

¹Randall Centre for Cell and Molecular Biophysics, King's College London, London, UK

²Laboratory for Molecular Cell Biology, UCL, London, UK

³European Bioinformatics Institute (EMBL-EBI), EMBL, Cambridge, UK

Correspondence

Virginie Uhlmann, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory (EMBL), Cambridge, UK.
Email: uhlmann@ebi.ac.uk

Funding information

Wellcom Trust, Grant/Award Number: 218278/Z/19/Z; Royal Society, Grant/Award Number: URF\R1\211329; Medical Research Council, Grant/Award Number: MC_U12266B; European Molecular Biology Laboratory, Grant/Award Number: Core funding; King's College London, Grant/Award Number: BiPAS CDT

Abstract

Images are at the core of most modern biological experiments and are used as a major source of quantitative information. Numerous algorithms are available to process images and make them more amenable to be measured. Yet the nature of the quantitative output that is useful for a given biological experiment is uniquely dependent upon the question being investigated. Here, we discuss the 3 main types of information that can be extracted from microscopy data: intensity, morphology, and object counts or categorical labels. For each, we describe where they come from, how they can be measured, and what may affect the relevance of these measurements in downstream data analysis. Acknowledging that what makes a measurement 'good' is ultimately down to the biological question being investigated, this review aims at providing readers with a toolkit to challenge how they quantify their own data and be critical of conclusions drawn from quantitative bioimage analysis experiments.

KEYWORDS

bioimage analysis, bioimaging, microscopy, quantification

1 | INTRODUCTION

Quantification is at the heart of modern biological research, and image data is one of the major sources of quantitative information.^{1–4} The past decade has seen an explosion of software tools dedicated to bioimage analysis.^{5–8} These methods most often adapt computer vision algorithms to the specificities of biological image data, such as poor contrast and the presence of complex objects, with the aim to speed up and streamline the quantification process.⁹ Although automated analysis remains a very challenging endeavour even with the best

of images, it opens the possibility to analyse data at scale, in a reproducible and objective manner.¹⁰

While greatly enabling biological research, the democratisation of quantitative image analysis tools also poses some challenges. One of the biggest challenges is identifying which of the dozens of quantifications that can be generated by these tools are most informative and relevant to the biological question being studied. To navigate this, a helpful strategy is to scrutinise all steps from experimental design to image acquisition and ultimately data processing and analysis. This allows each experimental stage to be tailored to best inform the hypothesis

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of Microscopy* published by John Wiley & Sons Ltd on behalf of Royal Microscopical Society.

being investigated but can also highlight aspects that may adversely influence the quantitative output. The first step consists of defining the goal of the imaging experiment (what quantitative property is at the centre of the question being investigated) prior to acquiring data and using this information to guide experimental design.¹¹ Once data start being generated, the focus shifts to the many experimental factors and acquisition parameters that may impact the quality of imaging. These can include sample and labelling properties, such as photobleaching and crosstalk, as well as hardware specifics. Although we briefly touch upon some of these aspects in this article, we invite readers to consult other reviews providing more in-depth discussions on these topics.^{12–14}

During an experiment, appropriately recording imaging parameters as metadata is crucial to ensure that downstream analysis is carried out in a meaningful way.¹⁵ It is crucial that images within an experiment are acquired in similar ways and have broadly similar properties such that they all remain within acceptable tolerances of the analytical methods. This may not always be possible though, as different biological perturbations may inherently affect image quality (for example, in fluorescence microscopy, via a change in the expression level of a fluorescently tagged protein). Similarly, the details of any image processing applied postacquisition, such as methods to remove noise or improve contrast/resolution, should be documented, and recorded. After optimising experimental design, acquisition parameters, postprocessing and recording metadata, one is left with analysing the acquired image data. At this point, the last remaining step is to outline the best way to pull out the desired measurements. This raises questions about which aspects of image data can be measured, which methods are available to do so, and what are their limitations.

Several excellent review papers assess the performance of image analysis methods in a benchmarking setting, meaning that methods are evaluated on the one general task they were designed to solve. General tasks, such as segmentation^{16,17} and denoising,¹⁸ are however generally not the end goal of an experiment. The quantitative output needed to explore a biological question is indeed rarely to improve the quality of an image or partition it. Instead, segmentation and denoising are examples of operations that enable the final quantification task, which is usually to measure a specific phenotype. To complement the existing literature on image analysis algorithms, this review focuses on the quantification problem. We discuss fundamental concepts that are relevant to image quantification, review common categories of quantitative readouts that can be extracted from image data, and introduce the appropriate metrics to do so. We identify three main categories of quantitative information: image intensity, morphology, and counts or labels. We describe the contrast generation,

image formation and image processing methods that are relevant to all these categories, and review how each different quantitative readout can be extracted from the image data. We also discuss aspects impacting the quantification of each type of measurement. We close the paper with a discussion on further processing and analysis of image-based measurements and on quality control and assurance. Considering the breadth of the topic, we choose to limit the scope of this paper to individual 2D images and 3D volumes, acquired with either fluorescence microscopy or electron microscopy (EM). We therefore do not cover measurements that are specific to time-series, multichannel imaging, or to data obtained with specialised imaging modalities such as superresolution microscopy or single particle electron microscopy.

In order to be able to formulate what one *wants* to measure, one must first understand what *can* be measured. By providing an overview of the big categories of quantitative measurements in image analysis, the goal of this review is to provide life science researchers with a framework to appreciate and scrutinise their own image data. We also aim to give insights on aspects of an experiment that impact the relevance of measurements in downstream analysis, and therefore enable readers to be critical about whether the conclusions of studies involving image quantification are meaningful or not.

1.1 | Fundamental concepts

Microscopy images provide a window to the biological world on length scales beyond what the human eye can see. Although capturing a faithful representation of reality, images are indeed a ‘snapshot’ of the real world – not the real world itself. This seemingly philosophical distinction is crucial when images are used as a source of quantitative information. If anything is measured from an image, it becomes necessary to understand the steps involved to make reality visible, to capture it as a digital image, and to make this digital image amenable to measurements – as this is required to gauge the impact of the imaging pipeline on retrieved measurements. This section introduces general concepts that are central to any quantitative endeavour involving biological images.

1.2 | ‘Garbage in, garbage out’

While this article reviews different types of measurements that can be extracted from microscopy data, and how to interpret such measurements, the limiting factor will always be the quality of the original image being analysed. What may be the most important principle to bear in mind is ‘garbage in, garbage out’: flawed input will inevitably

result in flawed output. In other words, quantitative analysis readouts can only ever be as good as the data they were extracted from, and image analysis algorithms cannot compensate for fundamentally bad image data. The general principles discussed hereafter, and the quantitative readouts explored in the rest of the paper are all to be read through the lens of this key concept.

1.3 | Contrast generation

The sample preparation and contrast generation processes fundamentally differ across imaging modalities. Contrast is most often induced by biochemical labels that are artificially introduced in samples specifically for imaging. Contrast can also be obtained in a label-free manner with dedicated optical components, as is the case for phase contrast and Differential Interference Contrast (DIC) imaging.¹⁹

In fluorescence microscopy, molecules of interest are labelled with fluorescent species such as organic dyes and fluorescent proteins. Regardless of the microscope used and downstream analysis performed, the way labelling is performed must be taken into account to provide context to any quantification. For example, if nonendogenous fluorescent protein fusion constructs are being used, how closely do measured intensities reflect endogenous protein distributions?²⁰ If immunofluorescence techniques are being used, what factors are impacting the ratio between number of epitopes of interest and the number of fluorophores (e.g. primary and secondary antibody concentrations, antigen masking affinity, clonality of antibodies)?²¹ The microscope modality also affects how much out-of-focus fluorescence contaminates the in-focus fluorescence measurements (Figure 1A–D).

With EM imaging, the mechanism of contrast generation varies significantly depending upon the type of experiment performed, as well as the type of microscope and detectors used (Figure 1E–H). For most of the different types of EM experiments, contrast is introduced during sample preparation. Sample contrast is enhanced by the introduction of heavy metals (e.g. osmium, lead, uranium, gold, silver etc.), which bind to lipids, proteins, carbohydrates etc. via chemical reactions, whereby the method, sequential order of addition, temperature, time of incubation, etc. can impact the process by which the contrast is incorporated into the sample.²²

1.4 | Image formation

The entire purpose of any microscope is to transmit the biological information contained within the sample to the

detector. The pixel values in the acquired images will depend on several factors in the imaging process, and understanding these factors is important for contextualising quantitative intensity measurements and assessing their accuracy.

In fluorescence microscopy, the intensities in an image represent the number of photons emitted by excited fluorophores. The absolute number of photons emitted by fluorescent molecules in the sample is primarily determined by the intensity of the excitation illumination incident on the sample. Increasing illumination intensity, regardless of the source (typically a lamp, LED, or laser), usually results in an increase in the number of photons emitted by fluorescence. For relatively low excitation intensities, a linear increase in excitation intensity corresponds to a linear increase in the intensity of emitted fluorescence. However, higher excitation intensities can lead to nonlinear saturation of emitted fluorescence²³ (Figure 2A and B), and increase the rate at which permanent photobleaching of fluorophores occurs (Figure 2C). The dependence of emitted fluorescence intensity on illumination intensity means that any local variations in illumination intensity within a field-of-view will affect local fluorescence intensity. The ‘flatness’ of the illumination can be characterised and corrected for further quantitative measurements; this can be done using a homogeneously fluorescent test sample^{12,24} (Figure 2D), or via computational methods^{25,26} without a test sample.

In conventional transmission electron microscopy (TEM), images are formed by detecting electrons that pass through the thin (e.g. <100 nm) sample and reach the detection mechanism (e.g. electron-multiplying charge coupled device (EM-CCD) or direct electron detector or film) (Figure 1E). Alternatively, in scanning electron microscopes (SEM), a focused beam of electrons is scanned across the sample and resulting secondary electrons (SE) and/or backscattered electrons (BSE) and/or x-rays generated are collected to form an image. The most common SEM approach is the collection of SE that have interacted with the surface of a sample, where the resulting image is a view of the surface topology of the sample with a large depth of field (Figure 1F). However, BSE can also be separately collected and mapped onto the sample, providing information about the sample’s elemental composition (Figure 1G). BSE imaging has recently been exploited in a collection of volume EM (vEM) techniques, where either arrayed sections or blocks of ‘resin embedded, fixed, contrasted samples’ are automatically imaged (Figure 1H), generating a large 3D volume of ultrastructural data at nm resolution, across scales of 10s–100s of microns.^{27,28} Regardless of the detection technique, it is important to be aware that working distance, magnification, accelerating voltage, landing

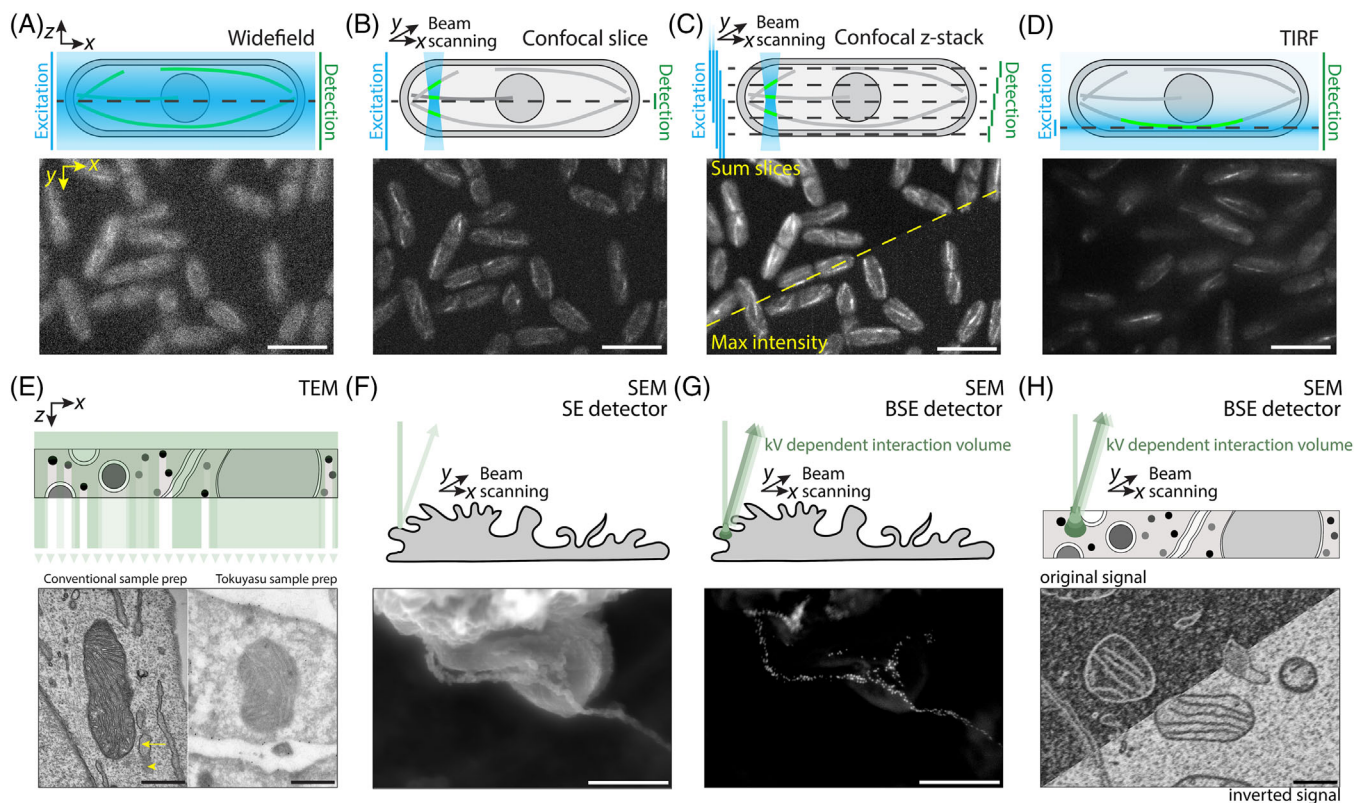


FIGURE 1 Image formation in fluorescence and electron microscopy. (A) Widefield microscopy of live *S. pombe* cells expressing sfGFP-tubulin (curved lines in schematic). The whole sample volume is illuminated simultaneously (cyan shading and ‘excitation’ bar on schematic) and fluorescence is captured on a camera (green ‘detection’ bar on schematic). Fine structure can be swamped by out-of-focus fluorescence. (B) Confocal microscopy of the same field-of-view as (A). Diffraction-limited laser spots are scanned individually (laser-scanning confocal) or swept in an array (spinning disk confocal, shown here) across the field-of-view; a pinhole (or array of pinholes) in the detection path prevents out-of-focus fluorescence from reaching the detector (point detector for laser-scanning confocal, camera for spinning disk) allowing for good contrast in fine structures at a range of sample depths. (C) Confocal slices can be acquired at a range of sample depths to form a ‘z-stack’. This 3D volume can be projected to a single image by adding the images together (‘Sum slices’) or picking the most intense value for each pixel (‘Max intensity’). (D) TIRF microscopy involves the generation of an evanescent field that only illuminates the volume of the sample within a few hundred nanometres of the coverslip. Emitted fluorescence is captured on a camera. Only structures near the coverslip (here, microtubules at the bottom of cells) are present in images. Note, this is a different field-of-view than in (A)–(C). All fluorescence scale bars = 10 μm . (E) TEM showing mitochondria in a thin section of an embedded cell. Different sample preparation protocols are shown: a conventional protocol causes the membrane of the mitochondria to be electron dense (appear dark), while a Tokuyasu protocol causes them to be electron lucent (appear light). Tokuyasu image courtesy of I. J. White. (F) SEM with secondary electron (SE) detection of an exocytosis event on the surface of an endothelial cell, where the topological structure of Von Willebrand factor strings released from the cell can be visualised. (G) SEM with back scattered electron (BSE) detection of sample shown in f, where gold-labelled antibodies highlight Von Willebrand factor strings released from the cell. Images courtesy of K. O’Neill and D. Cutler. (H) SEM with BSE detection of a thin section of a resin embedded sample. In the upper triangle, more heavy metal produces more BSEs and thus a stronger signal reaches the detector (appear light), while regions with less metal incorporation produce less signal (appear white). As this view is an inversion of the more traditional TEM images, researchers often invert image data (lower triangle) to allow for more comparative interpretation. Green ellipses represent increasing interaction volumes of the electron beam with increasing voltage of the primary electron beam. All EM scale bars = 500 nm.

energy, exposure/dwell time, probe size and current are just some of the parameters that impact the resulting images in terms of resolution, depth of field, focus and contrast.

The final acquired images are formed by binning (see Glossary in Table 1) the detected photons (fluorescence microscopy) or electrons (EM) into pixels. The pixel size

in a microscopy image plays a critical role in determining what quantitative information can be retrieved. The physical distance that each pixel represents (the pixel size) is primarily determined by properties of the detection path (for cameras, the physical size of the pixels on the chip, and for point detectors, the scanning parameters) and the total magnification of the system.

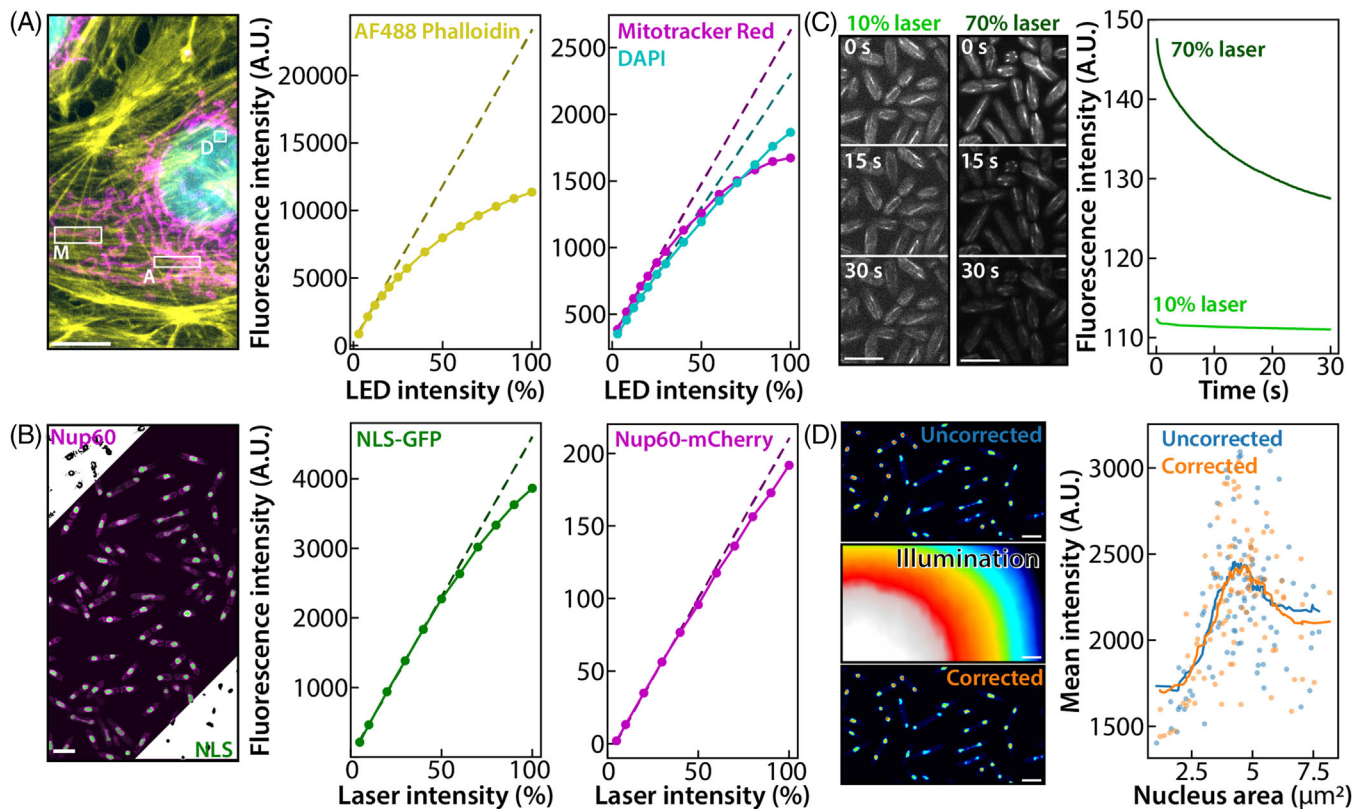


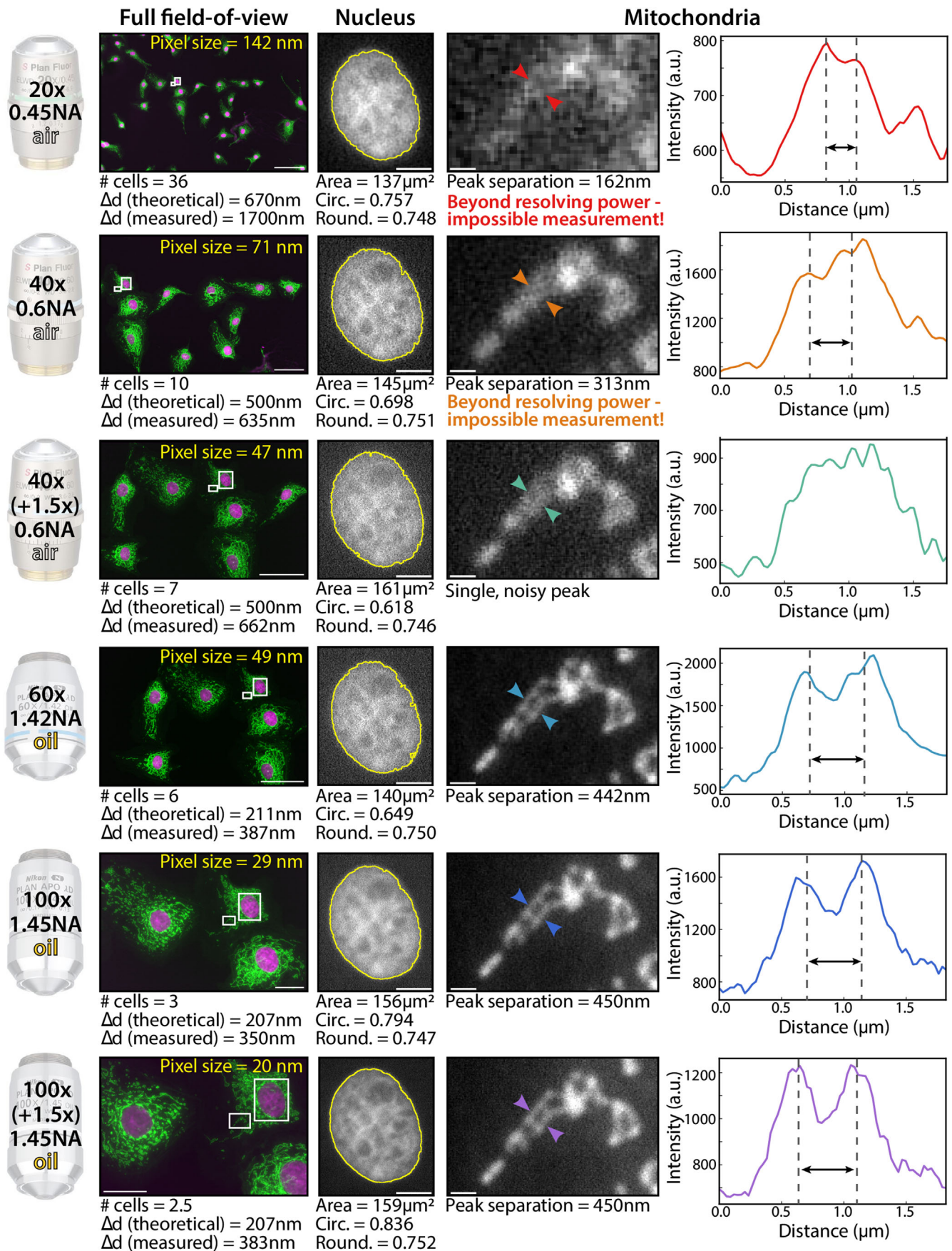
FIGURE 2 Impact of illumination on fluorescence intensity measurements. (A) Measured fluorescence intensities of Alexa Fluor 488 Phalloidin (measured from box 'A'), Mitotracker Red (box 'M') and DAPI (box 'D') in response to increasing LED illumination intensity in fixed BPAAE cells (widefield microscopy, Plan Apo VC 60 \times Oil objective NA = 1.4, 100 ms exposure). Dashed lines indicate what a linear relationship between illumination intensity and fluorescence intensity would be. (B) Measured fluorescence intensities of NLS-GFP and Nup60-mCherry in live *Schizosaccharomyces pombe* cells (strain GD250 as described in Dey et al.⁶⁷) in response to increasing laser illumination intensity (spinning disk confocal microscopy, SR HP Apo TIRF 100 \times AC Oil objective NA = 1.49, 100 ms exposure). Intensities were measured from regions of each channel above the Otsu threshold (portions of masks shown in corners of image). (C) Continuous spinning disk confocal imaging for 30 seconds of *S. pombe* cells expressing sfGFP-tubulin (strain AV2434, as described in Vještica et al.⁸⁸) at either 10% or 70% 488 nm laser intensity results in different photobleaching characteristics. Intensity was measured as the mean intensity above the Otsu threshold for each image. (D) Nuclear fluorescence intensities measured from live *S. pombe* cells expressing NLS-GFP ('Uncorrected') are a product of the true concentration of protein per nucleus and the flatness of the excitation illumination ('Illumination'). Inhomogeneous illumination can be corrected by dividing the acquired image by the illumination image (here, a homogeneously fluorescent slide). Image acquisition as in (B). Scale bars in all panels = 10 μm .

1.5 | Digital image (pre)processing

Per the Nyquist-Shannon sampling theorem, retaining the resolution of a continuous signal (i.e. the spatially varying distribution of photons/electrons incident on the detector) in discrete digital space (i.e. the pixels in the acquired image) requires sampling at least double the frequency of the smallest resolvable feature.²⁹ When the Nyquist-Shannon theorem is applied to the two-dimensional nature of images, the theoretical pixel size for adequate sampling should in fact be ~ 2.8 times smaller than the resolving power of the microscope.³⁰ This sampling should be observed if very fine structures within the sample are to be measured and quantified, as larger pixel sizes will lead to a loss of information due to undersampling. As a result,

the accuracy of any measurement of the same biological structure varies depending on the magnification (which determines the pixel size) and numerical aperture (NA) of the microscope objective (which determines the optical resolution), as shown in Figure 3.

In addition to spatially binning detected photons or electrons into pixels, detectors also convert the measured intensity into an integer number. This value depends on the intensity of emitted fluorescence or scattered electrons, as well as detector settings such as gains and offsets (see Glossary in Table 1). However, it also depends on the bit depth of the acquired images. Bit depth determines the range of values that can be digitally stored within a pixel; most microscopy data is acquired at 8-, 12-, or 16-bit depth. A pixel can only store a number within the range



$0 \rightarrow (2^N - 1)$, where N is the bit depth. During image processing, there are occasions where the bit depth of an image is changed. This is typically when a mathematical operation is performed on the image that generates values that are beyond the range of the bit depth (for example, a negative number) or have a noninteger component. Changing the bit depth will influence downstream measurements, as we shall see later.

Just like contrast generation and image formation, digital preprocessing steps prior to quantification can dramatically affect measurements. One must therefore remain mindful of the impact of these operations on the readout of interest. As an example, whatever gets suppressed by a thresholding operation may be considered as ‘background’ but could in fact contain relevant image information in addition to noise and nonspecific signal.

1.6 | Image segmentation

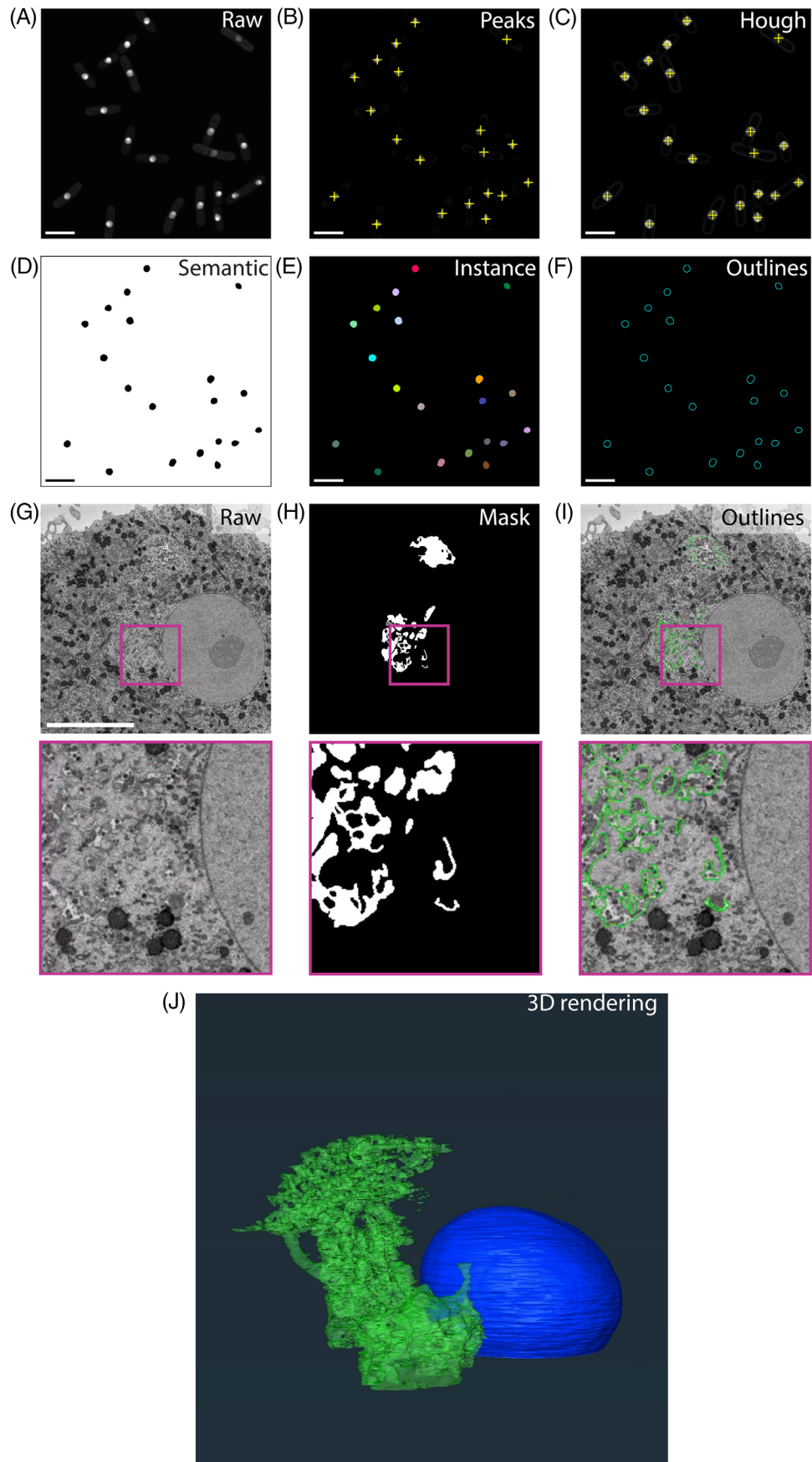
It is rare that every pixel in an image is considered relevant experimental output. Most often, measurements are to be extracted from specific regions of interest, which either correspond to relevant objects or portions thereof.

Segmentation is the process of partitioning an image into different regions, whether background and foreground (semantic, see Glossary in Table 1) or individual objects (instance, see Glossary in Table 1). Segmentation, whether instance or semantic, can be a challenging problem to solve because of the diversity of structures in microscopy data. Some features of interest can be relatively simple to segment (Figure 4A), while others can be highly challenging (Figure 4G–j). The fundamental nature of this challenge has however led to the development of numerous solutions which are available to reuse and adapt, both relying on classical image processing methods and leveraging recent machine learning tools.¹⁶ Segmentation algorithms generally output a ‘mask’ (see Glossary in Table 1), which consists of labels for each pixel in the original image (Figure 4). Such masks can either be binary in the case of semantic segmentation, meaning that pixels (or voxels in the case of 3D volumes) are either labelled

0 (background) or 1 (foreground) (Figure 4D and H), or composed of integer numbers for instance segmentation, whereby all pixels (or voxels) labelled with the same integer value belong to the same object instance (Figure 4E). Alternatively, to masks, instance segmentation algorithms can also output object contours (also sometimes called outlines) or surfaces. In 2D images, each individual object is then identified by the list of 2D coordinates of the pixels composing its contour (Figure 4F and I). In 3D volumes, surfaces can either be represented as a list of 3D voxel coordinates, or as a more structured set of vertices and faces called mesh (Figure 4J) (see Glossary in Table 1). Contour (outlines) or surfaces and mask representations of individual objects can easily be converted into one another by filling the former and finding the boundaries of the latter using classical image processing methods such as connected components or boundary tracing (e.g. the classical marching cubes algorithm). The relevant quantity to be measured for each biological question (e.g. ensemble vs. individual readout, internal vs. membrane readout) informs on the choice of algorithm. If individual objects are not needed, then a binary semantic mask may be sufficient. If only membranes/interfaces are of interest, then individual contours or surfaces may be sufficient.

Relying on computer-based algorithms to automate the process of segmentation has been a topic of central interest since the early days of microscopy image analysis.³¹ In the past decade, advances in deep machine learning have revolutionised bioimage processing and analysis in general, and segmentation algorithms in particular.^{32,33} Larger benchmark datasets of microscopy images and crowdsourced improvements on model architecture have pushed the limits of achievable accuracy and generalisation.³⁴ A large variety of powerful automated segmentation algorithms based on deep learning are now available on open-source software and usable by researchers with little to no computer science expertise.¹⁶ The democratisation of the use of artificial intelligence in image analysis obviously also comes with many challenges, for instance around reproducibility. Although these aspects will not be covered in this paper, other excellent reviews explore them in depth.^{35,36}

FIGURE 3 Impact of acquisition parameters on quantitative measurements. The same field-of-view of fixed BPAE cells stained with Mitotracker Red CMXRos (green) and DAPI (magenta) imaged with a widefield microscope with different air and oil ($n = 1.515$) objectives and additional optical magnifications (‘Full field-of-view’). The number of cells in the FOV, theoretical resolution (Δd) (emission wavelength/2 \times NA), and resolution as measured using Image Decorrelation Analysis⁸⁹ are listed. Scale bars: 20 \times – 100 μm , 40 \times , 60 \times – 50 μm , 100 \times , 150 \times – 20 μm . ‘Nucleus’ column shows a crop of the same nucleus from each magnification (larger white box in full FOV). The nucleus was segmented using Otsu thresholding after applying a 100 nm Gaussian blur to the crop, with the threshold border indicated in yellow. Area, circularity (‘Circ.’) and roundness (‘Round.’) values are indicated below. Nucleus scale bars = 5 μm . ‘Mitochondria’ columns show a crop of mitochondria staining at each magnification (smaller white box in full FOV). A line profile was drawn across the same region (between the arrowheads) and intensity profiles are plotted in the right-hand column (line averaging width of 5 pixels). Distances between prominent adjacent peaks were measured between the dashed lines and are indicated below the images. Mitochondria scale bars = 1 μm .



2 | MEASURING IMAGE INTENSITY

When acquiring microscopy data one of the first things a researcher checks, either by qualitatively inspecting the image or by examining the digital pixel values, is the intensity.

2.1 | What is intensity?

Intensities in a fluorescence microscopy image correspond to photons emitted by excited fluorophores, as described in Section 1.4. Most fluorescence intensity quantifications made from images are of the fluorophores themselves, not directly of the biological molecules of interest. This should be taken into consideration when translating any results from fluorescence intensity quantification into biological conclusions. Depending on the imaging modality used in fluorescence microscopy, intensity information may also arise from fluorescent sources in the sample other than the labelled structures in the focal plane. In techniques capable of optical sectioning, such as confocal microscopy (Figure 1B), two-photon microscopy, and TIRF (see Glossary in Table 1, Figure 1D), out-of-focus fluorescence does not reach the detector, whereas images acquired using widefield microscopy will contain out-of-focus intensities (Figure 1A). All fluorescence microscopy images may also contain intensity contributions from autofluorescence (endogenous fluorescent species present within the sample in the absence of intentional labelling). Confocal z-stacks are frequently projected into a single 2D image for visualisation and analysis (Figure 1C); a 'sum slices' or average intensity projection will retain intensity information, whereas a maximum intensity projection will produce sharper images but with intensities that do not correspond to the total amount of fluorescence below each pixel and thus should not be used for intensity quantification. When fluorescence intensity measurements are important, potential inhomogeneity in the illumination across the field-of-view can also create unwanted variability and should therefore be circumvented (see Section 1.4).

For EM, image intensity corresponds to electrons that reach the detection mechanism having interacted with the sample. For example, in TEM the number of electrons that reach the detector are impacted by microscope parameters (beam kV and intensity) and the sample characteristics (sample thickness and regional electron density of sample throughout its thickness), as illustrated in Figure 1E. Electron dense regions block the electron path and appear dark, while electron lucent regions allow the electrons to reach the detector and appear light. As biological samples often have little inherent capacity for differentially affecting the electron path, heavy metals are usually introduced into the sample to provide differential contrast, as described briefly in Section 1.3. These sample preparation protocols vary widely and can give different views of the sample, as shown in the lower panel of Figure 1E. For BSE imaging in an SEM, the amount of heavy metal incorporated into each part of the sample, as well as the landing energy of the focused electron beam, also influences the intensity captured by the detector as shown in Figure 1G and H. For SE imaging in an SEM, aside from accelerating voltage, the intensity of the signal detected is additionally impacted by the differential angles of the detector and primary electron beam, as well as how the primary and secondary electrons interact with the limiting shape of the sample (Figure 1F). Depending upon the SEM configuration this can result in images that appear 3D with regions of highlights and shadows or comparatively flat images with apparently poorer signal to noise. For all EM imaging modalities, the detector settings including exposure time and whether averaging (line/frame) or automatic gain or scaling is performed can impact acquired image intensity as well.

2.2 | Quantifying intensity

Pixel (or voxel) values are a direct readout of intensity itself. Aggregated measurements summarising collections of pixel values, for instance over the area of an object of interest, are however often more useful than individual

FIGURE 4 Outputs of object counting and segmentation. (A) A spinning disk confocal image of *S. pombe* cells expressing the nuclear marker NLS-sfGFP (strain AV1200, Vještica et al.⁸⁸). Scale bar = 10 μm . If the task is only to count objects in an image, rather than extract morphological information, then this can be performed using peak detection on the image following difference-of-Gaussians filtering (B), or if strong geometric priors are known, a method such as the circular Hough transform on the edge-detected image (C). Segmentation of objects can produce various outputs. (D) Semantic segmentation divides an image into two classes: foreground (i.e. objects of interest, black) and background (white). Such an image is also referred to as a binary mask. (E) Instance segmentation divides an image into background (black) and 'instances' of the object of interest. Each different instance here is randomly assigned a different colour. (F) Segmented objects can alternatively be represented by their boundaries rather than a solid object. (G) BSE image from SEM of HIV infected human monocyte induced macrophage.⁹⁰ Scale bar = 10 μm . (H) Segmentation of intracellular plasma membrane-connected compartment (IPMC) shown as mask. (I) Segmentation of IPMC shown as outlines. (J) 3D rendered mesh of segmentation of IPMC. Magenta boxes highlight locations of insets shown magnified below.

TABLE 1 Glossary of key technical terms in image quantification.

Glossary	
TIRF	(<i>Acquisition</i>) Total Internal Reflection Fluorescence. A fluorescence microscopy technique that uses total internal reflection of excitation light at the interface between the coverslip and the sample to generate a field of light that is most intense at the interface and exponentially decays with increasing depth into the sample (over a range of a few hundred nanometres). This allows for axially restricted excitation of fluorophores close to the sample.
Binning	(<i>Acquisition or analysis</i>) In acquisition, the assimilation of signal into a finite-sized pixel in an image. In analysis, the process of combining the output of adjacent pixels to increase signal, thereby losing resolution.
Gain	(<i>Acquisition</i>) An amplification factor applied to the readout from the photon/electron detector in order to produce the image. It adjusts the sensitivity of the camera, but also amplifies the noise.
Offset	(<i>Acquisition</i>) The minimal intensity captured by the photons/electron detector.
Pipeline	(<i>Analysis</i>) A series of data processing steps that allows extraction of quantitative metrics from raw image data.
Deconvolution	(<i>Analysis</i>) The computational process of enhancing image contrast using knowledge of the way the microscope forms images.
Semantic	(<i>Analysis</i>) Semantic, in the context of segmentation, describes the association of each pixel of an image with a label, typically 'foreground' or 'background'.
Instance	(<i>Analysis</i>) Individual occurrence of an object type. For example, an image with 3 circles has 3 instances of a 'circle' object. In microscopy, 'instances' often correspond to specific biological structures.
Mask	(<i>Analysis</i>) Image in which all pixels/voxels that are part of the foreground are set to an integer value (e.g. 1 or 255), and all pixels/voxels that are part of the background are set to 0.
Mesh	(<i>Analysis</i>) A set of vertices and faces that define polygons (often triangles) and, when taken together, form a surface covering of a 3D object.

pixel intensities. Intensity is therefore usually measured as a distribution – either over the entire image or over a region of interest. That distribution can be analysed with the classical toolkit of statistics: represented as an histogram or characterised by a small number of summary statistics such as the mode, median, standard deviation and higher central moments when appropriate.

Intensity-based measurements extracted from a fluorescence microscopy image, be that from the raw data or after processing, are usually comparative. Standalone measurements of pixel or object intensity in images are indeed often meaningless; they must be reported in the context of some baseline condition such as the background intensity, or the intensity of a comparable structure under a different biological condition. For such comparisons to be made accurately, it is critical that acquisition parameters such as illumination intensity, magnification, pixel dwell time (point detectors) or exposure time (cameras) and detector gains are recorded and ideally kept consistent between different images. Any image processing pipelines should be applied equivalently to each image, including the ones that do not look like they 'need' it. For some biological measurements, it makes sense to work with the absolute fluorescence values in images, such as monitoring the expression of a GFP-tagged protein during successive cell divisions.³⁷ However, when aggregating results from different images, relative or normalised fluorescence intensities are often used so that results can be aggregated. Overall, any absolute measurements of intensity from fluorescence microscopy modalities such as widefield and confocal microscopy are critically dependent on labelling, acquisition settings and postprocessing. If comparisons of fluorescence intensity are to be made between different images, then these parameters should be as identical as possible in each case.

For EM images, electron density or intensity is rarely absolutely quantified, as routinely controlling the factors involved in (a) generating contrast (sample preparation), (b) detecting electron density (microscope configuration, inherent sample characteristics and detector settings) and (c) calibrating these, is fraught with challenges. First, for some tissue samples it is difficult to guarantee homogenous metal incorporation throughout the entire sample.^{38–40} Second, for thin section imaging (TEM or BSE in SEM) one must be aware that the morphology of a structure, and its presentation within the volume of the section, can impact its resulting intensity profile in an image. For example, a limiting membrane of the endoplasmic reticulum cut perpendicular to the electron beam (arrow, Figure 1E) can result in a very different intensity profile as the same membrane cut parallel to the electron beam (arrowhead, Figure 1E). Third, without tilt tomography, it is difficult to rule out that other electron dense

structures, above or below your structure of interest, may be present in the section at that position in x and y , and impact any intensity measurements. Should the research question warrant electron density quantification, great care should be taken to minimise variations in sample preparation and ensure that the same microscope and detector settings are used to ensure homogenous and comparable illumination, and absence of under/over exposure. It is also beneficial to pick an intrinsic, unaffected structure within the samples that can be used as an internal control for electron density calibration. As a notable exception, signal intensity is often quantified in x-ray microanalysis, where the energy spectrum of electron induced x-rays can provide relative quantitative information about the elemental composition of the sample.⁴¹

2.3 | Factors impacting intensity information

A vast range of image processing operations can be applied to raw images following acquisition. If quantitative intensity information is to be extracted following image processing, it is then important to understand how processing affects image intensity (Figures 1C and 5). If intensity measurements are to be made following image processing, then it is important that the processed values are still linearly related to the number of fluorescent molecules present in a given region of the image (Figure 5D). Iterative deconvolution (see Glossary in Table 1) methods have been shown experimentally to be largely linear with respect to intensity, although this can be microscope-dependent⁴² (Figure 5, 'Deconvolution'). An example of a nonlinear image processing operation is the Super-Resolution Radial Fluctuations (SRRF) method⁴³ (Figure 5, 'SRRF'). This is an example of a method which can increase both contrast and resolution of an image dataset but should not be used for quantitative intensity measurements.

The effect of bit depth (see Section 1.5) on intensity information is somewhat analogous to the effect of pixel size on spatial information; higher bit depths provide higher 'sampling' of intensities, which can provide higher precision for quantitative measurements. Critically, measurements should not be made from any pixels having either the minimum (0) or maximum ($2^N - 1$) value as this is likely to represent incomplete or 'clipped' information in the image; unless the image was acquired in a very low fluorescence intensity regime, pixels of value 0 may in fact represent a range of different 'real' fluorescence intensities that are below the range of the detector settings, and pixels of value $2^N - 1$ may represent a range of real fluorescence intensities that are saturating the detector. Intensity quantification is still valid when performed on images after increasing the bit depth, but no intensity quantifications should be made

from images following conversion to a lower bit depth. This is because conversion to a lower bit depth requires a rescaling of pixel values so that they fit within the smaller range, which results in a loss of information from the image.

In addition to bit depth, another important concept when measuring image intensities is that of dynamic range. This can have two slightly different meanings, depending on context. When referring to detectors, the dynamic range refers to the minimum and maximum light intensities incident on the detector that can be measured simultaneously. However, when referring to images, it usually means the minimum and maximum pixel values within a single image. During image acquisition, ideally one wants intensities that span the full dynamic range of the detector. However, this is often impractical for many applications, especially with high bit-depth detectors. Generally, acquisition parameters should be adjusted to maximise image dynamic range as far as is practical without causing saturation or significant bleaching (in the case of fluorescence) to provide a wide range of values for precise extraction of quantitative information.

An emerging field of processing methods for fluorescence microscopy images are deep learning-based methods. Such methods typically require training a neural network with pairs of high-quality and low-quality images of the same field-of-view; the network attempts to 'learn' what series of image processing operations should be applied to reliably convert low-signal images into images closely matching the high-signal equivalent. New low-quality images (without a high-quality equivalent) can then be provided to the trained neural network, and the network will output a high-quality prediction. Example applications of these algorithms are for increasing the signal-to-noise ratio of low-signal images⁴⁴ and increasing resolution of images,⁴⁵ among others.¹⁸ Because these methods impact image intensity in a nonlinear manner (Figure 5, 'CARE'), it is strongly recommended that intensity-dependent quantification is not performed on images processed with deep learning methods.

3 | MEASURING MORPHOLOGY

Most microscopy data, regardless of the modality, hold information that pertains to morphology. Although the concept of 'morphology' is intuitively understood by everyone, it can be challenging to define precisely what it means.

3.1 | What is morphology?

Loosely characterised as the visual appearance in terms of form or structure, morphology is critical in many biological processes because it reflects and influences the

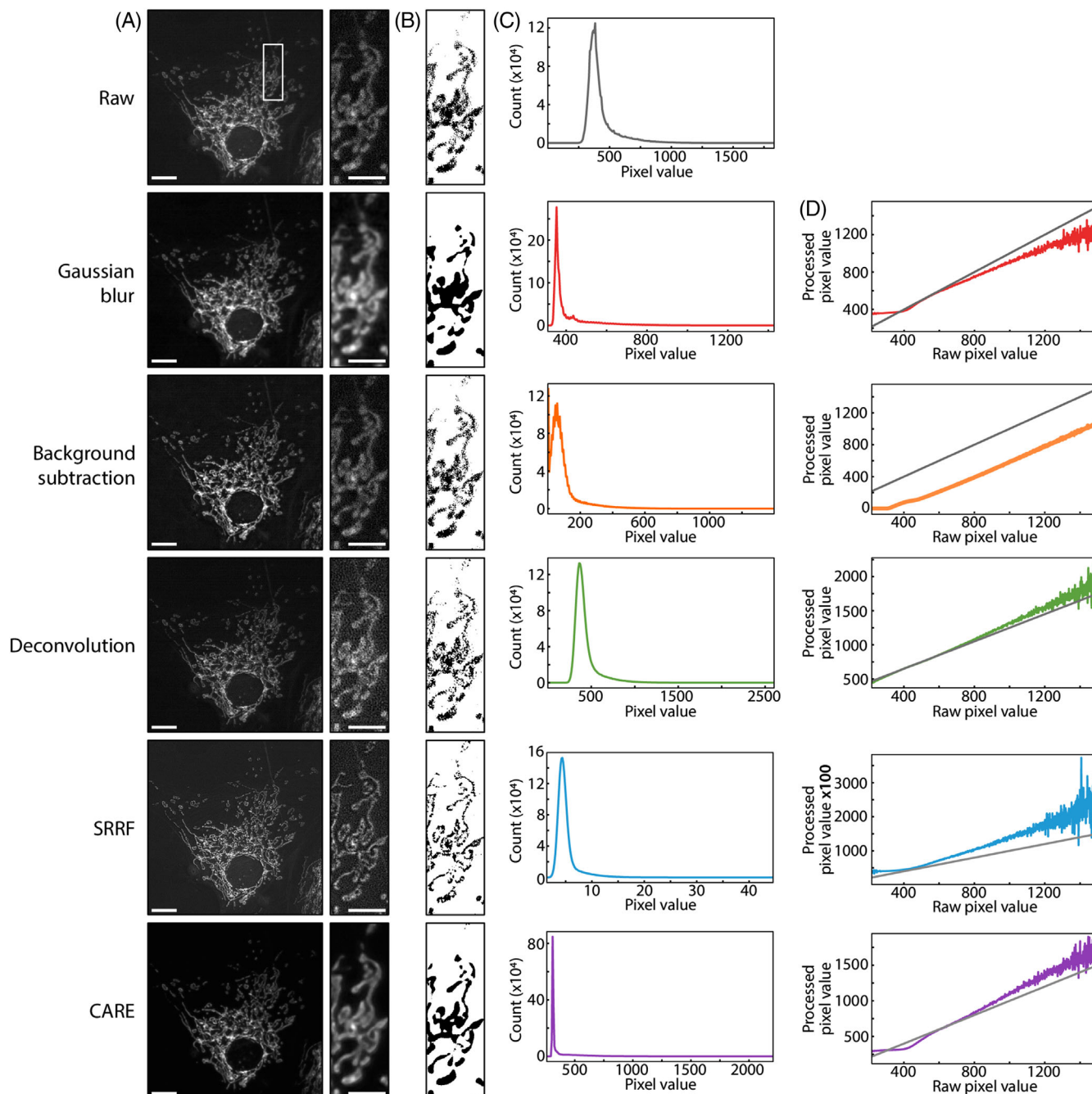


FIGURE 5 Effect of image processing on downstream quantification. (A) Images of a fixed BPAE cell with mitochondria stained. ‘Raw’ image has not been processed following acquisition; other images are processed versions of this image via the methods stated. Inset corresponds to white rectangle in the left image. Scale bars: $10\ \mu\text{m}$ (large image), $5\ \mu\text{m}$ (magnified inset). (B) Mitochondria segmented from the inset images using Otsu thresholding. Note that morphological analysis of these segmentations would yield different results between different image processing methods. (C) Histograms of pixel values within the large images in (A), where count refers to the number of pixels. Histograms cover the full range of pixel values in each image; note that both the shape of the histogram and this value range vary with different processing methods. (D) For each pixel in the processed images, the pixel value is plotted against the value of the corresponding pixel in the raw unprocessed image. The grey line indicates a 1:1 relationship between processed and unprocessed pixel values (i.e. no change following processing).

physiological state of living systems.^{46,47} Though it may be tempting to measure everything that can be measured and ask questions later, it is recommended to identify what kind of morphological information will be relevant to

the question being investigated and how this information may be impacted by the imaging or analysis process. The shape of the objects of interest, for example those labelled with a membrane marker, is the most common readout

of morphology in bioimage analysis.⁴⁸ In addition, texture information is also available in imaging modalities that capture intracellular components such as organelles and cytoskeletal elements.^{49,50} Although the type of image feature informing on morphology may vary (whether edges, textures or a mix of both), most morphology measurements are extracted for individual biologically relevant objects. They therefore share the need for segmentation upstream of the actual quantification step.

3.2 | Quantifying morphology

While segmentation is a necessary step towards the quantification of morphology, it is a means but not the end. The output of segmentation will be used as a basis to quantify morphology. This is worth keeping in mind to assess the level of accuracy needed from segmentation: subtle differences in morphology can be missed if segmentation accuracy is too low (Figure 3, mitochondria). Conversely, large morphological properties such as object size may not require segmentation accuracy to the single pixel or voxel (Figure 3, nucleus). The scale of the morphological readout of interest thus also informs on how precise the segmentation must be for it to be quantitatively captured.

A subset of commonly used handcrafted measurements of morphology in 2D is listed in Table 2. Some of these metrics are adapted from general concepts of geometry and others have been carefully engineered relying on image processing tools. All have been designed to quantify the geometrical (for shape) or visual (for texture) nature of an object in an intuitive and interpretable manner, and several can be directly extended to 3D. Different measurements capture different aspects of morphology, sometimes with very subtle differences (e.g. roundness and circularity).

An alternate route is to let machines learn morphology descriptors directly from the data. This is relevant in many cases, from situations where morphology is too ambiguous to make it possible to craft a relevant set of features, to cases where the biological phenomenon of interest is too poorly understood to allow predicting which aspect of morphology is discriminative. When used well, machine learning strategies, whether supervised or unsupervised, can produce descriptors of morphology that are less biased and that better capture information than manually designed ones, at the expense of interpretability and for a more significant computational cost. Morphology descriptors of individual cell types can be for instance learned by a deep neural network in an unbiased manner relying solely on 3D shape and texture from EM volumes, without specifying a biological question.⁵⁵ Despite the effectiveness of deep neural networks, it is usually not possible to reverse-engineer the exact nature of the morphological features

they rely on, making learned representations potentially difficult to interpret. Efforts to investigate and compare published approaches on benchmark or reference datasets are invaluable to navigate these available options.⁵⁶

3.3 | Factors impacting morphological quantification

It should be noted that when measuring the morphology of objects within images, it is critical to consider both the pixel size and resolution of the image, as these provide information on the lower limits of measurement differences that can be captured by the data (Figure 3). This is especially important if any measurement approaches the resolution limit of the acquired image. For example, any measured sizes should remain above the theoretical resolution limit. If many objects in the image are measured to have sizes comparable to the resolution limit of the system, then this may be a population of objects of varying sizes smaller than what can be resolved. One should also remember that many morphological measurements are computed on 2D projections of structures that are actually tri-dimensional, as for instance in widefield fluorescence (Figure 1A) and projected confocal stacks (Figure 1C). These images do not consider the third dimension and may therefore be misleading when quantifying morphology.


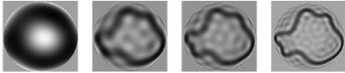

It is important to keep in mind that the morphology we observe in a microscopy image is a product of both the sample preparation and imaging process. Any measurements extracted to describe it are therefore strongly influenced by factors that may not be immediately relevant or obvious to the biological phenomenon of interest. When considering sample preparation factors for instance, some proteins, commonly used as organelle markers, can demonstrate apparently normal organelle morphology while other proteins (also used as organelle markers) can reveal aberrant morphology. For instance, the ER protein Calnexin was shown to reveal ER with apparently normal morphology in cells depleted of GBF1, while Calreticulin, another ER protein, revealed aberrant ER morphology in the same cells. This was later validated by correlative light and electron microscopy (CLEM).⁵⁷ Similarly, if the biological perturbation impacts intracellular trafficking pathways, then the normal intracellular localisation of standard organelle markers, may also be affected. Misinterpretation of imaging data can be avoided by using additional markers, performing trafficking time courses, and using additional experimental approaches such as biochemical assays and CLEM. When considering imaging, refractive index mismatch between the microscope lens and the sample medium can also be a factor impacting morphological

TABLE 2 Common handcrafted morphology features in 2D.

Feature type	Metric	Definition	Interpretation
Shape	Area	A	Size of the region occupied by the object
	Perimeter	L	Length of the contour of the object
	Feret diameters (calliper)	P_H (largest) F_V	Distances between parallel tangents touching two opposite sides of the object
	Convex Hull	C	Smallest convex shape that contains the object
	Circularity	$\frac{4\pi A}{L^2}$	Ratio of the object area to that of a circle with the same perimeter
	Roundness	$\frac{4A}{\pi F_H^2}$	Ratio of the object area to that of a circle with the same width
	Compactness	$\sqrt{\frac{4A}{\pi F_H^2}}$	Square-root of roundness
	Aspect ratio	$\frac{F_H}{F_V}$	Ratio of the object height to its width
	Solidity	$\frac{A}{A_G}$	Ratio of the object area to the area of its convex hull
	Convexity	$\frac{L_G}{L}$	Ratio of the convex hull perimeter to the object perimeter
Texture	Haralick features	Defined in Ref. (51)	Statistics of a matrix counting the cooccurrence of neighbouring intensity values in the image (grey level cooccurrence matrix)
	Gabor filters	Defined in Ref. (52)	Profile of image intensity distributions at different frequencies and orientations

(Continues)

TABLE 2 (Continued)

Feature type	Metric	Definition	Interpretation	
Mixed 	Zernike moments	Defined in Ref. (53)		Decomposition of the shape or texture into a basis of polynomials that are orthogonal on the unit disk
	Fourier descriptors	Defined in Ref. (54)		Decomposition of the shape or texture into the Fourier basis

Descriptors based on an object's shape (object geometry), texture (image intensity), either shape or texture.

quantifications and severely affect downstream analysis, especially for 3D bioimage data.

Beyond acquisition, image processing algorithms also impact morphology measurements, as nonlinear operations can significantly alter the results of automatic thresholding, for example (Figure 5B). Keeping in mind how morphology measurements are computed is therefore crucial to determine whether the considered readouts can be meaningfully compared across different datasets. Whenever absolute image intensity is involved, for instance when relying on texture descriptors that are not only based on relative variations of intensity, one must question whether intensity can reasonably be compared across different images, as discussed in Section 2. Similarly, the observed shape depends on the image resolution in x , y and z and can be strongly affected by how biological samples have been prepared for imaging. Most fluorescence and electron microscopy setups acquiring 3D image volumes produce anisotropic data, meaning that voxels have a different physical size along the x , y and z axis. In such cases, any measurement assuming isotropic voxel size will result in misleading or plainly wrong quantitative readouts. In EM, each TEM will be technically specified to provide resolution in the angstrom range but the ultimate resolution of the acquired images – what can actually be visually resolved – are impacted by the sample, heavy metals introduced, thickness and density of the sample and image acquisition parameters. Introduction of significant amounts of heavy metals may coat ultrastructural features thickly and make it difficult to resolve finer ultrastructural details, thereby limiting the possibility of quantifying morphology. Besides microscope resolution, sample preparation is a notoriously strong factor influencing morphology. Having a good understanding of how different types of preparation distort the morphology of samples therefore provides crucial information on whether measurements can be considered biologically relevant or not.

Electron microscopy has a long-standing history of investigating the effect of sample preparation,⁵⁸ with examples specifically focusing on morphology preservation.⁵⁹ The need for strategies that minimally alter the structure of the imaged sample has inspired several modern fixation techniques.^{60,61} As always, optimisation is required to find a sensible balance of all aspects of the experimental design from sample preparation to quantification, with the ultimate aim to address the research question in mind.

4 | COUNTING AND LABELLING

Intensity and morphology can be considered 'first-order measurements' as they focus on quantifying purely visual information. In contrast, 'second-order measurements' focus on aggregating and combining morphology and intensity metrics to quantify structures that are externally defined.

4.1 | What are counts and labels?

Counts straightforwardly refer to the number of occurrences of a given structure or object in an image. Labels refer to identifiers from a limited and usually fixed set of possibilities (e.g. 'mitochondria' or 'nuclei') assigned to objects in a process referred to as classification. Since labels are used to group objects into categories or classes, they can be referred to as 'categorical labels' or 'class labels'. Labels are most often derived from human-defined categories (e.g. different experimental conditions, different subcellular structure, different cell types), but can also be inferred directly from the result of a clustering algorithm. In that case, the label assigned to an object reflects the identity of the cluster they belong to and may take an arbitrary numerical value.

4.2 | Quantifying and interpreting labels and object counts

Individual object count is immediately obtained whenever objects can be segmented. Labels can be retrieved from morphology and intensity measurements extracted from individual objects and combined in a feature vector through a classification or clustering process, depending on whether annotated examples of categories are available. Labels can also be automatically recovered directly from image crops, without an intermediate measurement extraction step, relying on machine learning.⁶² If no other readouts are required, counting and labelling can therefore simply involve a bounding box detection process and do not necessarily require the definition of precise object boundaries (and therefore segmentation).

Because of its 'second-order' nature, counting exploits known information of the structure or object to be detected. This can be achieved through strong priors, as exploited by blob detectors⁶³ and the Hough transform,⁶⁴ or through a curated example of the object of interest, as is the case in template matching.^{65,66} Blob detectors exploit digital filters with specific shapes such as the Laplacian of Gaussian or the Determinant of Hessian to detect agglomerates of pixels fitting within a circle of predefined radius (Figure 4B). The Hough transform, in contrast, is an algorithm designed to detect occurrences of perfect circles in an image. Both blob detectors and the Hough transform can successfully identify round structures such as nuclei as they appear in fluorescence microscopy (Figure 4C) and serve as basis for the study of more complex cellular processes.⁶⁷ Template matching can be tuned to detect an object of choice (the 'template') and is the preferred method to identify molecular complexes in EM tomograms.^{68,69} Both the Hough transform and template matching are examples of algorithms that have the ability to provide object counts without going through a segmentation step. When visual appearance varies so significantly that a single good object representative is hard to identify, deep learning methods can learn to detect occurrences of complex structures from large collections of visual examples.^{70,71} Although initially designed for the detection of highly structured objects from natural images such as cars and human faces, the same algorithms have shown to generalise enough to provide good enough results in fluorescence microscopy data to allow counting.⁷²

At the extreme, labelling may neither require segmentation nor even object detection. Classification can be successfully carried out from tiles, obtained by splitting an image into a square grid.⁷³ Labels are then assigned to each tile, thus providing a readout of the categories present in the image without relying on the individually defined objects. This approach is successfully exploited in digi-

tal pathology, where object segmentation is particularly challenging.^{74,75}

4.3 | Factors impacting labels and count information

The number of elements present in an image or their category are seemingly absolute measurements, and it is thus reasonable to expect these readouts to be comparable across microscopy data. It is however important to keep in mind that, due to their 'second-order' nature, count and label measurements ultimately rely on morphology and intensity features. As such, when comparing across images, one should carefully consider how the nature of the data may reflect on morphology and intensity measurements as discussed through the other sections of this paper and, in turn, influence the results of counting or labelling quantification pipeline.

5 | BEYOND INDIVIDUAL MEASUREMENTS

The process of quantifying image microscopy data in biology goes beyond understanding individual types of measurements and what they reflect. Once extracted, these measurements are meant to be used to carry out statistical analyses and support conclusions made on the experiments that images captured.

5.1 | Assembling and processing feature vectors

As discussed in previous sections, many different measurements related to any individual object's intensity, morphology, or identity can be extracted, making it challenging to know a priori which ones will be most informative. The best approach is therefore to assemble a large amount of such measurements into a feature vector, which is often simply a list of numbers used to quantitatively represent an object. Being able to measure specific readouts from any input image is therefore of utmost interest, as exemplified by the wealth of available software and libraries providing accessible ways to extract most of the measurements reviewed in previous sections.⁷⁶

A feature vector is said to be handcrafted when it is constructed with measurements that are predefined by the experimenter. The intention behind assembling a large group of measurements is to empirically capture as many aspects of an object as possible, in a quantitative manner, to describe it in the most precise and unambiguous

manner. With this goal in mind, a valid strategy is to create feature vectors by indiscriminately measuring everything one can think of measuring. Although it may make sense to gather more measurements than needed to ensure that no important information is omitted, feature vectors built in that way often end up being highly redundant. This is because different handcrafted measurements may be directly related to one another (for instance roundness and compactness, see Table 2) or may be derived from the same geometrical properties (for instance area, perimeter and circularity). The more numbers a feature vector is composed of, the higher the dimensionality of the space it lives in – and therefore the harder it is to visualise and make sense of. Strong correlation between many elements or large amounts of duplicated elements in a feature vector only make matters worse, as the difficulty of recovering discriminative information increases in higher-dimensional spaces. Feature selection methods such as the Fisher score⁷⁷ can be used to limit redundancy, and dimensionality reduction techniques such as the famous Principal Component Analysis can help prune the collection of measurements and retain only a small number of most informative elements.⁷⁸

Feature vectors can alternatively be inferred directly from image data relying on machine learning.⁷⁹ In this latter case, the numbers composing the feature vector are readouts generated by an automated algorithm through complex combinations of the original image information and cannot be readily expressed with a simple mathematical expression like handcrafted measurements (see Table 2). Feature vectors built in this way have the potential to be more discriminative than handcrafted ones, and to capture properties of objects that human-defined measurements would be missing but lose interpretability as a trade-off. Although automatically generated, learned feature vectors may equally benefit from feature selection and dimensionality reduction as handcrafted ones.

5.2 | Quality control and quality assurance

As already stated in Section 1.1, no matter how well-designed the analysis component of a microscopy experiment, if the images being input have poor quality, or the sample preparation and labelling have been poorly designed or executed, then the results obtained from analysis will have little meaning. In fluorescence microscopy, the most commonly used metrics for assessing image quality are signal-to-noise ratio (SNR) and spatial resolution. SNR values alone are however insufficient to tell whether an image contains important biological information or not, and whether it will be good enough for quantitative anal-

ysis. Spatial resolution measurements are not necessarily an indicator of image quality directly but can be useful for contextualising morphological measurements. For example, large number of measurements clustered at the resolution limit indicate that it may be necessary to use higher resolution data to study the structure of interest. Measuring image properties such as SNR and resolution and recording them alongside other metadata from image acquisition helps to add additional context to results from quantitative analysis.⁸⁰

It is also important to recognise and reduce bias in quantitative image analysis. One major avenue for this is investing time in creating automated analysis pipelines whereby batches of images acquired under different biological conditions can be analysed in the same manner, free from any user input. Where automated analysis is not practical, or manual parameter selection is required, blinding can help reduce user bias.¹⁴ Batch effects, defined as nonbiological experimental variations that confound measurements, are a common source of bias, with possibly dramatic consequences on end results.^{81,82} The influence of batch effects is further demonstrated by Shamir et al. who show that intensity and morphology measurements computed on microscopy images composed only of background signal can allow identifying different organelles.⁸³ As stated throughout this article, sample preparation, acquisition parameters (such as illumination intensity, magnification, exposure time and detector gains) and experimental parameters (such as timestamps and sample id) must be recorded for each image whenever measurements are meant to be compared. Similarly, all parameters that can be kept constant should ideally remain as identical as possible over images. Batch effects can be further mitigated at the level of image data by correcting for intensity variations,^{25,26} or with feature normalisation.⁸⁴ A good summary of strategies to identify and correct batch effects is provided in Caicedo et al.⁹

The laboratory standard for assessing the legitimacy of a scientific analysis is quality control and performance metrics, and image quantification gets no exception from that. Although plenty of established metrics are available to assess the success of algorithms that carry out segmentation, detection, counting and classification among many others, identifying metrics that faithfully reflect performance across datasets and use-cases remains an open challenge.⁸⁵ Although not quantitative, visual inspection remains a robust quality control strategy. This endeavour may however be highly nontrivial when dealing with high-dimensional, dynamic datasets or with rare events, and can be greatly facilitated by dedicated software tools.⁸⁶ Ultimately, the most powerful measure of quality control is reproducibility: the experimental procedures, microscope hardware specifications, image acquisition parameters,

and image quantification algorithms provided in a published study should allow other researchers to recover its quantitative conclusions.⁸⁷

6 | CONCLUSIONS

The most important message of this paper is that any measurements extracted from microscopy image data can only be as good as the image data themselves. Sample preparation and image acquisition therefore significantly impact the quality of quantitative readouts.

Here, we focused on three big families of measurements: image intensity, morphology and object counts or categorical labels. Unless carried out on the entire image at once, most measurements require a first step of instance segmentation. It is generally difficult to accurately compare intensity measurements, whether in electron or light microscopy images. Normalisation to a reference provides a way around this but requires significant care to be done meaningfully. Morphology, unlike intensity, is challenging to define generally as it relates to shape, texture, and complex combinations thereof. Object counts and categorical labels obviously require objects to be identified and assigned but may not necessarily need precise outlines. These types of readout can therefore often be obtained without explicitly segmenting individual objects.

Other key aspects of image quantification are quality control and quality assurance. While it is common practice to account for known distortions and aberrations introduced by sectioning and imaging in specific imaging modalities such as EM, assessing the accuracy or 'success' of these corrections and their impact on downstream measurement is sometimes challenging. Identifying good metrics to assess whether a quantitative readout makes sense can be difficult, and plenty of confounders may adversely affect the extracted measurements. In addition to informing on the type of measurements that can meaningfully be extracted from image data, the essential information about image acquisition, sample preparation, and processing provided by metadata is therefore also crucial to allow randomisation and mitigate batch effects at the analysis stage.

As a summary, the problem of quantifying microscopy data should be approached like any experiment: start by stating a clear definition of the objectives, propose a strategy, study the implications of each step involved in that strategy and challenge the results accordingly. Automation should also be prioritised as much as possible to minimise bias and maximise reproducibility. Such a holistic approach to bioimage quantification is the safest way to ensure that meaningful measurements are extracted and that they are handled in a scientifically rigorous manner.

ACKNOWLEDGEMENTS

VU is supported by EMBL internal funding. JJB was supported by MRC core funding to the MRC Laboratory for Molecular Cell Biology at University College London, award code MC_U12266B and WT funding (218278/Z/19/Z). SC is supported by a Royal Society University Research Fellowship (URF\R1\211329). ACC is funded by the BiPAS CDT at King's College London. The authors thank James Levitt of the Nikon Imaging Centre at King's College London for his support and assistance in this work. The authors thank Gautam Dey and his lab at EMBL, Heidelberg for providing *S. pombe* strains, and David Nkwe, Annegret Pelchen-Matthews, Lucy Collinson, Mark Marsh, Krupa O'Neill, Daniel Cutler and Ian White for permission to use collaborative data for figure panels.

Open access funding enabled and organized by Projekt DEAL.

ORCID

Virginie Uhlmann  <https://orcid.org/0000-0002-2859-9241>

REFERENCES

- Myers, G. (2012). Why bioimage informatics matters. *Nature Methods*, 9, 659–660. <https://doi.org/10.1038/nmeth.2024> PMID: 22743769
- Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A., & Olivo-Marin, J.-C. (2016). Imagining the future of bioimage analysis. *Nature Biotechnology*, 34, 1250–1255. <https://doi.org/10.1038/nbt.3722> PMID: 27926723
- Murphy, R. F. (2014). A new era in bioimage informatics. *Bioinformatics*, 30, 1353–1353. <https://doi.org/10.1093/bioinformatics/btu158> PMID: 24753489
- Jamali, N., Dobson, E. T. A., Eliceiri, K. W., Carpenter, A. E., & Cimini, B. A. (2022). 2020 BioImage Analysis Survey: Community experiences and needs for the future. *Biological Imaging*, 1, e4. <https://doi.org/10.1017/S2633903x21000039> PMID: 35387317
- Eliceiri, K. W., Berthold, M. R., Goldberg, I. G., Ibáñez, L., Manjunath, B. S., Martone, M. E., Murphy, R. F., Peng, H., Plant, A. L., Roysam, B., Stuurman, N., Swedlow, J. R., Tomancak, P., & Carpenter, A. E. (2012). Biological imaging software tools. *Nature Methods*, 9, 697–710. <https://doi.org/10.1038/nmeth.2084> PMID: 22743775
- Guet, R., Burri, O., & Seitz, A. (2019). Open source tools for biological image analysis. In M. Biology, E. Rebollo, & M. Bosch (Eds.), *Computer optimized microscopy methods* (pp. 23–37). New York: Springer. https://doi.org/10.1007/978-1-4939-9686-5_2
- Rubens, U., Mormont, R., Paavolainen, L., Bäcker, V., Pavie, B., Scholz, L. A., Michiels, G., Maška, M., Ünay, D., Ball, G., Hoyoux, R., Vandaele, R., Golani, O., Stanciu, S. G., Sladoje, N., Paul-Gilloteaux, P., Marée, R., & Tosi, S. (2020). BIAFLOWS: A collaborative framework to reproducibly deploy and benchmark bioimage analysis workflows. *Patterns*, 1, 100040. <https://doi.org/10.1016/j.patter.2020.100040> PMID: 33205108
- Levet, F., Carpenter, A. E., Eliceiri, K. W., Kreshuk, A., Bankhead, P., & Haase, R. (2021). Developing open-source

- software for bioimage analysis: Opportunities and challenges. *Fluorescence*, 10, 302. <https://doi.org/10.12688/fl000research.52531.1> PMID: 34249339
9. Caicedo, J. C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A. S., Barry, J. D., Bansal, H. S., Kraus, O., Wawer, M., Paaivolainen, L., Herrmann, M. D., Rohban, M., Hung, J., Hennig, H., Concannon, J., Smith, I., Clemons, P. A., & Carpenter, A. E. (2017). Data-analysis strategies for image-based cell profiling. *Nature Methods*, 14, 849–863. <https://doi.org/10.1038/nmeth.4397> PMID: 28858338
 10. Carpenter, A. E. (2007). Software opens the door to quantitative imaging. *Nature Methods*, 4, 120–121. <https://doi.org/10.1038/nmeth0207-120> PMID: 17264858
 11. Wait, E. C., Reiche, M. A., & Chew, T.-L. (2020). Hypothesis-driven quantitative fluorescence microscopy – The importance of reverse-thinking in experimental design. *Journal of Cell Science*, 133, jcs250027. <https://doi.org/10.1242/jcs.250027> PMID: 33154172
 12. Jonkman, J., Brown, C. M., Wright, G. D., Anderson, K. I., & North, A. J. (2020). Tutorial: Guidance for quantitative confocal microscopy. *Nature Protocols*, 15, 1585–1611. <https://doi.org/10.1038/s41596-020-0313-9> PMID: 32235926
 13. Waters, J. C. (2009). Accuracy and precision in quantitative fluorescence microscopy. *Journal of Cell Biology*, 185, 1135–1148. <https://doi.org/10.1083/jcb.200903097> PMID: 19564400
 14. Jost, A. P.-T., & Waters, J. C. (2019). Designing a rigorous microscopy experiment: Validating methods and avoiding bias. *Journal of Cell Biology*, 218, 1452–1466. <https://doi.org/10.1083/jcb.201812109> PMID: 30894402
 15. Montero Llopis, P., Senft, R. A., Ross-Elliott, T. J., Stephansky, R., Keeley, D. P., Koshar, P., Marqués, G., Gao, Y.-S., Carlson, B. R., Pengo, T., Sanders, M. A., Cameron, L. A., & Itano, M. S. (2021). Best practices and tools for reporting reproducible fluorescence microscopy methods. *Nature Methods*, 18, 1463–1476. <https://doi.org/10.1038/s41592-021-01156-w> PMID: 34099930
 16. Lucas, A. M., Ryder, P. V., Li, B., Cimini, B. A., Eliceiri, K. W., & Carpenter, A. E. (2021). Open-source deep-learning software for bioimage segmentation. *MBoC*, 32, 823–829. <https://doi.org/10.1091/mbc.E20-10-0660> PMID: 33872058
 17. Caicedo, J. C., Roth, J., Goodman, A., Becker, T., Karhohs, K. W., Broisin, M., Molnar, C., McQuin, C., Singh, S., Theis, F. J., & Carpenter, A. E. (2019). Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry*, 95, 952–965. <https://doi.org/10.1002/cyto.a.23863> PMID: 31313519
 18. Laine, R. F., Jacquemet, G., & Krull, A. (2021). Imaging in focus: An introduction to denoising bioimages in the era of deep learning. *The International Journal of Biochemistry & Cell Biology*, 140, 106077. <https://doi.org/10.1016/j.biocel.2021.106077>
 19. Kasprowitz, R., Suman, R., & O'Toole, P. (2017). Characterising live cell behaviour: Traditional label-free and quantitative phase imaging approaches. *The International Journal of Biochemistry & Cell Biology*, 84, 89–95. <https://doi.org/10.1016/j.biocel.2017.01.004>
 20. Gibson, T. J., Seiler, M., & Veitia, R. A. (2013). The transience of transient overexpression. *Nature Methods*, 10, 715–721. <https://doi.org/10.1038/nmeth.2534> PMID: 23900254
 21. Yong, W.H. (Ed.) (2019). *Biobanking: Methods and protocols*. Humana Press.
 22. Hayat, M. A. (1970). *Principles and techniques of electron microscopy* (Vol. 1). Van Nostrand Reinhold.
 23. Visscher, K., Brakenhoff, G. J., & Visser, T. D. (1994). Fluorescence saturation in confocal microscopy. *Journal of Microscopy*, 175, 162–165. <https://doi.org/10.1111/j.1365-2818.1994.tb03479.x>
 24. Model, M. A., & Burkhardt, J. K. (2001). A standard for calibration and shading correction of a fluorescence microscope. *Cytometry*, 44, 309–316. [https://doi.org/10.1002/1097-0320\(20010801\)44:4<309::AID-CYTO1122>3.0.CO;2-3](https://doi.org/10.1002/1097-0320(20010801)44:4<309::AID-CYTO1122>3.0.CO;2-3) PMID: 11500847
 25. Singh, S., Bray, M.-A., Jones, T. R., & Carpenter, A. E. (2014). Pipeline for illumination correction of images for high-throughput microscopy. *Journal of Microscopy*, 256, 231–236. <https://doi.org/10.1111/jmi.12178> PMID: 25228240
 26. Peng, T., Thorn, K., Schroeder, T., Wang, L., Theis, F. J., Marr, C., & Navab, N. (2017). A BaSiC tool for background and shading correction of optical microscopy images. *Nature Communications*, 8, 14836. <https://doi.org/10.1038/ncomms14836> PMID: 28594001
 27. Peddie, C. J., Genoud, C., Kreshuk, A., Meechan, K., Micheva, K. D., Narayan, K., Pape, C., Parton, R. G., Schieber, N. L., Schwab, Y., Titze, B., Verkade, P., Weigel, A., & Collinson, L. M. (2022). Volume electron microscopy. *Nature Reviews Methods Primers*, 2, 51. <https://doi.org/10.1038/s43586-022-00131-9>
 28. Collinson, L. M., Bosch, C., Bullen, A., Burden, J. J., Carzaniga, R., Cheng, C., Darrow, M. C., Fletcher, G., Johnson, E., Narayan, K., Peddie, C. J., Winn, M., Wood, C., Patwardhan, A., Kleywegt, G. J., & Verkade, P. (2023). Volume EM: A quiet revolution takes shape. *Nature Methods*, <https://doi.org/10.1038/s41592-023-01861-8>
 29. Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the Ire*, 37, 10–21. <https://doi.org/10.1109/JRPROC.1949.232969>
 30. Pawley, J. B. (2006). *Points, pixels, and gray levels: Digitizing image data*. In J. B. Pawley (Ed.), *Handbook of biological confocal microscopy* (pp. 59–79). US: Springer. https://doi.org/10.1007/978-0-387-45524-2_4
 31. Meijering, E. (2012). Cell segmentation: 50 Years down the road [life sciences]. *IEEE Signal Processing Magazine*, 29, 140–145. <https://doi.org/10.1109/MSP.2012.2204190>
 32. Meijering, E. (2020). A bird's-eye view of deep learning in bioimage analysis. *Computational and Structural Biotechnology Journal*, 18, 2312–2325. <https://doi.org/10.1016/j.csbj.2020.08.003> PMID: 32994890
 33. Uhlmann, V., Donati, L., & Sage, D. (2022). A practical guide to supervised deep learning for bioimage analysis: Challenges and good practices. *IEEE Signal Processing Magazine*, 39, 73–86. <https://doi.org/10.1109/MSP.2021.3123589>
 34. Spiers, H., Songhurst, H., Nightingale, L., Folter, J., Hutchings, R., Peddie, C. J., Weston, A., Strange, A., Hindmarsh, S., Lintott, C., Collinson, L. M., & Jones, M. L. (2021). Deep learning for automatic segmentation of the nuclear envelope in electron microscopy data, trained with volunteer segmentations. *Traffic (Copenhagen, Denmark)*, 22, 240–253. <https://doi.org/10.1111/tra.12789> PMID: 33914396
 35. Laine, R. F., Arganda-Carreras, I., Henriques, R., & Jacquemet, G. (2021). Avoiding a replication crisis in deep-learning-based bioimage analysis. *Nature Methods*, 18, 1136–1144. <https://doi.org/10.1038/s41592-021-01284-3>

36. Hoffman, D. P., Slavitt, I., & Fitzpatrick, C. A. (2021). The promise and peril of deep learning in microscopy. *Nature Methods*, *18*, 131–132. <https://doi.org/10.1038/s41592-020-01035-w> PMID: 33479523
37. Chaigne, A., Labouesse, C., White, I. J., Agnew, M., Hannezo, E., Chalut, K. J., & Paluch, E. K. (2020). Abscission couples cell division to embryonic stem cell fate. *Developmental Cell*, *55*, 195–208. e5 <https://doi.org/10.1016/j.devcel.2020.09.001>
38. Hua, Y., Laserstein, P., & Helmstaedter, M. (2015). Large-volume en-bloc staining for electron microscopy-based connectomics. *Nature Communications*, *6*, 7923. <https://doi.org/10.1038/ncomms8923> PMID: 26235643
39. Song, K., Feng, Z., & Helmstaedter, M. (2022). High-contrast en-bloc staining of mouse whole-brain samples for EM-based connectomics. *Neuroscience*, <https://doi.org/10.1101/2022.03.30.486341> bioRxiv
40. Genoud, C., Titze, B., Graff-Meyer, A., & Friedrich, R. W. (2018). Fast homogeneous en bloc staining of large tissue samples for volume electron microscopy. *Frontiers in Neuroanatomy*, *12*, 76. <https://doi.org/10.3389/fnana.2018.00076> PMID: 30323746
41. Fernandez-Segura, E., & Warley, A. (2008). Electron probe X-ray microanalysis for the study of cell physiology. *Methods in cell Biology*, *88*, 19–43. [https://doi.org/10.1016/S0091-679X\(08\)00402-0](https://doi.org/10.1016/S0091-679X(08)00402-0)
42. Swedlow, J. R., Hu, K., Andrews, P. D., Roos, D. S., & Murray, J. M. (2002). Measuring tubulin content in *Toxoplasma gondii*: A comparison of laser-scanning confocal and wide-field fluorescence microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 2014–2019. <https://doi.org/10.1073/pnas.022554999> PMID: 11830634
43. Culley, S., Tosheva, K. L., Matos Pereira, P., & Henriques, R. (2018). SRRF: Universal live-cell super-resolution microscopy. *The International Journal of Biochemistry & Cell Biology*, *101*, 74–79. <https://doi.org/10.1016/j.biocel.2018.05.014>
44. Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., Rocha-Martins, M., Segovia-Miranda, F., Norden, C., Henriques, R., Zerial, M., Solimena, M., Rink, J., Tomancak, P., Royer, L., ..., Myers, E. W. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, *15*, 1090–1097. <https://doi.org/10.1038/s41592-018-0216-7> PMID: 30478326
45. Speiser, A., Müller, L.-R., Matti, U., Obara, C. J., Legant, W. R., Kreshuk, A., Macke, J. H., Ries, J., & Turaga, S. C. (2020). Deep learning enables fast and dense single-molecule localization with high accuracy. *Biophysics*, <https://doi.org/10.1101/2020.10.26.355164>. bioRxiv
46. Bertossa, R. C. (2011). Morphology and behaviour: Functional links in development and evolution. *Philosophical Transactions of the Royal Society B*, *366*, 2056–2068. <https://doi.org/10.1098/rstb.2011.0035>
47. Bettinger, C. J., Langer, R., & Borenstein, J. T. (2009). Engineering substrate topography at the micro- and nanoscale to control cell function. *Angewandte Chemie International Edition*, *48*, 5406–5415. <https://doi.org/10.1002/anie.200805179>
48. Andrews, T. G. R., Pönisch, W., Paluch, E. K., Steventon, B. J., & Benito-Gutierrez, E. (2021). Single-cell morphometrics reveals ancestral principles of notochord development. *Development* (Cambridge, England), *148*, dev199430. <https://doi.org/10.1242/dev.199430>
49. Vergara, H. M., Pape, C., Meechan, K. I., Zinchenko, V., Genoud, C., Wanner, A. A., Mutemi, K. N., Titze, B., Templin, R. M., Bertucci, P. Y., Simakov, O., Dürichen, W., Machado, P., Savage, E. L., Schermelleh, L., Schwab, Y., Friedrich, R. W., Kreshuk, A., Tischer, C., & Arendt, D. (2021). Whole-body integration of gene expression and single-cell morphology. *Cell*, *184*, 4819–4837. e22 <https://doi.org/10.1016/j.cell.2021.07.017>
50. Diamond, J., Anderson, N. H., Bartels, P. H., Montironi, R., & Hamilton, P. W. (2004). The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. *Human Pathology*, *35*, 1121–1131. <https://doi.org/10.1016/j.humpath.2004.05.010> PMID: 15343515
51. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*, 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
52. Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, *61*, 103–113. <https://doi.org/10.1007/BF00204594>
53. Bhatia, A. B., & Wolf, E. (1954). On the circle polynomials of Zernike and related orthogonal sets. *Mathematical Proceedings of the Cambridge Philosophical Society*, *50*, 40–48. <https://doi.org/10.1017/S0305004100029066>
54. Burger, W., & Burge, M. J. (2013). Fourier shape descriptors. In I. Mackie (Ed.), *Principles of digital image processing undergraduate topics in computer science* (pp. 169–227). London: Springer. https://doi.org/10.1007/978-1-84882-919-0_6
55. Zinchenko, V., Hugger, J., Uhlmann, V., Arendt, D., & Kreshuk, A. (2022). MorphoFeatures: Unsupervised exploration of cell types, tissues and organs in volume electron microscopy. *Developmental Biology*, *12*, e80918. <https://doi.org/10.1101/2022.05.07.490949>. bioRxiv
56. Dmitrenko, A., Masiero, M. M., & Zamboni, N. (2022). Comparing representations of biological data learned with different AI paradigms, augmenting and cropping strategies. arXiv:2203.04107. <https://doi.org/10.48550/ARXIV.2203.04107>
57. Lopes-Da-Silva, M., McCormack, J. J., Burden, J. J., Harrison-Lavoie, K. J., Ferraro, F., & Cutler, D. F. (2019). A GBF1-dependent mechanism for environmentally responsive regulation of ER-Golgi transport. *Developmental Cell*, *49*, 786–801. e6 <https://doi.org/10.1016/j.devcel.2019.04.006>
58. Shah, F. A., Johansson, B. R., Thomsen, P., & Palmquist, A. (2015). Ultrastructural evaluation of shrinkage artefacts induced by fixatives and embedding resins on osteocyte processes and pericellular space dimensions: Ultrastructural evaluation of shrinkage artefacts. *Journal of Biomedical Materials Research*, *103*, 1565–1576. <https://doi.org/10.1002/jbm.a.35287>
59. Gendelman, H. E., Moench, T. R., Narayan, O., & Griffin, D. E. (1983). Selection of a fixative for identifying T cell subsets, B cells, and macrophages in paraffin-embedded mouse spleen. *Journal of Immunological Methods*, *65*, 137–145. [https://doi.org/10.1016/0022-1759\(83\)90310-1](https://doi.org/10.1016/0022-1759(83)90310-1) PMID: 6317754
60. Bullen, A., Taylor, R. R., Kachar, B., Moores, C., Fleck, R. A., & Forge, A. (2014). Inner ear tissue preservation by rapid freezing: Improving fixation by high-pressure freezing and hybrid methods. *Hearing Research*, *315*, 49–60. <https://doi.org/10.1016/j.heares.2014.06.006> PMID: 25016142

61. Zenner, H. L., Collinson, L. M., Michaux, G., & Cutler, D. F. (2007). High-pressure freezing provides insights into Weibel-Palade body biogenesis. *Journal of Cell Science*, *120*, 2117–2125. <https://doi.org/10.1242/jcs.007781> PMID: 17535847
62. Gupta, A., Harrison, P. J., Wieslander, H., Pielawski, N., Kartasalo, K., Partel, G., Solorzano, L., Suveer, A., Klemm, A. H., Spjuth, O., Sintorn, I.-M., & Wählby, C. (2019). Deep learning in image cytometry: A review. *Cytometry*, *95*, 366–380. <https://doi.org/10.1002/cyto.a.23701> PMID: 30565841
63. Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, *30*, 79–116. <https://doi.org/10.1023/A:1008045108935>
64. Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, *13*, 111–122. [https://doi.org/10.1016/0031-3203\(81\)90009-1](https://doi.org/10.1016/0031-3203(81)90009-1)
65. Goshtasby, A. (1985). Template matching in rotated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *7*, 338–344. <https://doi.org/10.1109/TPAMI.1985.4767663>
66. Uhlmann, V., Püspöki, Z., Depeursinge, A., Unser, M., Sage, D., & Fageot, J. (2022). Steer'n'Detect: Fast 2D template detection with accurate orientation estimation. *Bioinformatics*, *38*, 3146–3148. <https://doi.org/10.1093/bioinformatics/btac270> PMID: 35435214
67. Dey, G., Culley, S., Curran, S., Schmidt, U., Henriques, R., Kukulski, W., & Baum, B. (2020). Closed mitosis requires local disassembly of the nuclear envelope. *Nature*, *585*, 119–123. <https://doi.org/10.1038/s41586-020-2648-3> PMID: 32848252
68. Lucas, B. A., Himes, B. A., Xue, L., Grant, T., Mahamid, J., & Grigorieff, N. (2021). Locating macromolecular assemblies in cells by 2D template matching with cisTEM. *eLife*, *10*, e68946. <https://doi.org/10.7554/eLife.68946> PMID: 34114559
69. Zhou, N., Wang, H., & Wang, J. (2017). EMBuilder: A template matching-based automatic model-building program for high-resolution cryo-electron microscopy maps. *Scientific Reports*, *7*, 2664. <https://doi.org/10.1038/s41598-017-02725-w> PMID: 28572576
70. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. arXiv:1506.02640 [cs].
71. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In 2017 IEEE International Conference on Computer Vision (ICCV) (IEEE), Venice, Italy. pp. 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>
72. Waithe, D., Brown, J., Reglinski, K., Diez-Sevilla, I., Roberts, D., & Eggeling, C. (2019). Object detection networks and augmented reality for cellular detection in fluorescence microscopy acquisition and analysis. *Bioinformatics*, <https://doi.org/10.1101/544833>. bioRxiv
73. Uhlmann, V., Singh, S., & Carpenter, A. E. (2016). CP-CHARM: Segmentation-free image classification made accessible. *BMC Bioinformatics [Electronic Resource]*, *17*, 51. <https://doi.org/10.1186/s12859-016-0895-y> PMID: 26817459
74. Fu, Y., Jung, A. W., Torne, R. V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L. R., Jimenez-Linan, M., Moore, L., & Gerstung, M. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nature Cancer*, *1*, 800–810. <https://doi.org/10.1038/s43018-020-0085-8> PMID: 35122049
75. Wang, Y., Acs, B., Robertson, S., Liu, B., Solorzano, L., Wählby, C., Hartman, J., & Rantalainen, M. (2022). Improved breast cancer histological grading using deep learning. *Annals of Oncology*, *33*, 89–98. <https://doi.org/10.1016/j.annonc.2021.09.007> PMID: 34756513
76. Haase, R., Fazeli, E., Legland, D., Doube, M., Culley, S., Belevich, I., Jokitalo, E., Schorb, M., Klemm, A., & Tischer, C. (2022). A Hitchhiker's guide through the bio-image analysis software universe. *FEBS Letters*, *596*, 2472–2485. <https://doi.org/10.1002/1873-3468.14451> PMID: 35833863
77. Orlov, N. V., Chen, W. W., Eckley, D. M., Macura, T. J., Shamir, L., Jaffe, E. S., & Goldberg, I. G. (2010). Automatic classification of lymphoma images with transform-based global features. *IEEE Transactions on Information Technology in Biomedicine*, *14*, 1003–1013. <https://doi.org/10.1109/TITB.2010.2050695>
78. Saeyns, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*, 2507–2517. <https://doi.org/10.1093/bioinformatics/btm344> PMID: 17720704
79. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., & Van Valen, D. (2019). Deep learning for cellular image analysis. *Nature Methods*, *16*, 1233–1246. <https://doi.org/10.1038/s41592-019-0403-1> PMID: 31133758
80. Schmied, C., Nelson, M., Avilov, S., Bakker, G.-J., Bertocchi, C., Bischof, J., Boehm, U., Brocher, J., Carvalho, M., & Chiritescu, C., Christopher, J., Cimini, B., Conde-Sousa, E., Ebner, M., Ecker, R., Eliceiri, K., Fernandez-Rodriguez, J., Gaudreault, N., Gelman, L., Grunwald, D., ... Jambor, H. K. (2023). Community-developed checklists for publishing images and image analysis. ArXiv, arXiv:2302.07005v1.
81. Howard, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O. I., Kather, J. N., Cipriani, N., Grossman, R. L., & Pearson, A. T. (2020). The impact of digital histopathology batch effect on deep learning model accuracy and bias. *Bioinformatics*, *12*, 4423. <https://doi.org/10.1101/2020.12.03.410845> bioRxiv
82. Sonesson, C., Gerster, S., & Delorenzi, M. (2014). Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. *PLoS ONE*, *9*, e100335. <https://doi.org/10.1371/journal.pone.0100335> PMID: 24967636
83. Shamir, L. (2011). Assessing the efficacy of low-level image content descriptors for computer-based fluorescence microscopy image analysis: Assessing pattern recognition in microscopy image analysis. *Journal of Microscopy*, *243*, 284–292. <https://doi.org/10.1111/j.1365-2818.2011.03502.x> PMID: 21605118
84. Kothari, S., Phan, J. H., Stokes, T. H., Osunkoya, A. O., Young, A. N., & Wang, M. D. (2014). Removing batch effects from histopathological images for enhanced cancer diagnosis. *IEEE Journal of Biomedical and Health Informatics*, *18*, 765–772. <https://doi.org/10.1109/JBHI.2013.2276766> PMID: 24808220
85. Reinke, A., Tizabi, M. D., Sudre, C. H., Eisenmann, M., Rädtsch, T., Baumgartner, M., Acion, L., Antonelli, M., Arbel, T., Bakas, S., Bankhead, P., Benis, A., Cardoso, M. J., Cheplygina, V., Christodoulou, E., Cimini, B., Collins, G. S., Farahani, K., van Ginneken, B., ... Maier-Hein, L. (2021). Common limitations of image processing metrics: A picture story. 10.48550/ARXIV.2104.05642.
86. Bray, M.-A., & Carpenter, A. E. (2015). CellProfiler Tracer: Exploring and validating high-throughput, time-lapse

- microscopy image data. *BMC Bioinformatics [Electronic Resource]*, *16*, 369. <https://doi.org/10.1186/s12859-015-0759-x>
87. Huisman, M., Hammer, M., Rigano, A., Boehm, U., Chambers, J. J., Gaudreault, N., North, A. J., Pimentel, J. A., Sudar, D., Bajcsy, P., Brown, C.M., Corbett, A.D., Faklaris, O., Lacoste, J., Laude, A., Nelson, G., Nitschke, R., Grunwald, D., & Strambio-De-Castilla, C. (2019). A perspective on Microscopy Metadata: Data provenance and quality control. <https://doi.org/10.48550/ARXIV.1910.11370>
88. Vještica, A., Marek, M., Nkosi, P. J., Merlini, L., Liu, G., Bérard, M., Billault-Chaumartin, I., & Martin, S. G. (2019). A toolbox of Stable Integration Vectors (SIV) in the fission yeast *Schizosaccharomyces pombe*. *Journal of Cell Science*, *133*, jcs240754 <https://doi.org/10.1242/jcs.240754>
89. Descloux, A., Großmayer, K. S., & Radenovic, A. (2019). Parameter-free image resolution estimation based on decorrelation analysis. *Nature Methods*, *16*, 918–924. <https://doi.org/10.1038/s41592-019-0515-7>
90. Nkwe, D. O., Pelchen-Matthews, A., Burden, J. J., Collinson, L. M., & Marsh, M. (2016). The intracellular plasma membrane-connected compartment in the assembly of HIV-1 in human macrophages. *BMC Biology*, *14*, 50. <https://doi.org/10.1186/s12915-016-0272-3> PMID: 27338237

How to cite this article: Culley, S., Caballero, A. C., Burden, J. J., & Uhlmann, V. (2023). Made to measure: An introduction to quantifying microscopy data in the life sciences. *Journal of Microscopy*, 1–22. <https://doi.org/10.1111/jmi.13208>