



Shot noise limits on binary detection in multiphoton imaging

AARON K. LAVIOLETTE AND CHRIS XU*

School of Applied and Engineering Physics, Cornell University, Ithaca NY 14853, USA

*cx10@cornell.edu

Abstract: Much of fluorescence-based microscopy involves detection of if an object is present or absent (i.e., binary detection). The imaging depth of three-dimensionally resolved imaging, such as multiphoton imaging, is fundamentally limited by out-of-focus background fluorescence, which when compared to the in-focus fluorescence makes detecting objects in the presence of noise difficult. Here, we use detection theory to present a statistical framework and metric to quantify the quality of an image when binary detection is of interest. Our treatment does not require acquired or reference images, and thus allows for a theoretical comparison of different imaging modalities and systems.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Fluorescence based confocal and multiphoton microscopy are powerful tools for biological research and have provided valuable insights into many biological questions. In particular, the use of multiphoton imaging has allowed for deep access into biological tissues and is of utmost importance for biological applications where high spatial resolution imaging deep within intact tissue is required [1–3]. Deep multiphoton imaging has allowed for the visualization and detection of many types of structures [1,2]. Confocal and multiphoton imaging have also been used for dynamic imaging and cell tracking [4,5]. Nearly all confocal and multiphoton imaging experiments used to study biology first rely on the binary detection of an object. For example, in blood vessel imaging, one is tasked with identifying the blood vessels, in recoding neuronal activity one must first find the neurons, and in cell tracking one is tasked with detecting cells in each image frame.

In any such imaging experiment the microscopist must first decide which imaging approach is best. For example, a decision is made about whether to use one-photon (1P, e.g., confocal), two-photon (2P) or three-photon (3P) excitation. This decision is usually made based on the depth of where the imaging needs to take place, because different methods will provide different image quality, relative to each other, at different depths. All three-dimensionally resolved fluorescence microscopy techniques (confocal or multiphoton imaging) are fundamentally limited by out-of-focus background fluorescence, which reduces contrast, degrading quality and inhibiting detection [6–9]. A simple approach to compare the imaging performance in the past is to use the signal-to-background ratio (SBR), which is the ratio of in-focus fluorescence (i.e., signal, S) to out-of-focus fluorescence (i.e., background, B) [6–9]. In multiphoton imaging, the SBR decreases monotonically as a function of imaging depth, and eventually at a certain depth the SBR becomes too small for practical detection of objects [6,7].

One may then be tempted to consider that the SBR is an acceptable metric of image quality. Indeed, in the past a depth limit for 2P and 3P image quality has been arbitrarily defined as the depth at which SBR is reduced to unity [6,7]. However, SBR alone does not consider the noise statistics at the detector, which are described with Poisson statistics (i.e., shot noise) for a typical microscope with high gain photodetectors (e.g., photomultiplier tubes). To make this argument concrete, consider two cases both with $SBR = 1$: in case one $S = 1$ and $B = 1$, and in case two $S = 10$ and $B = 10$. In case one, a pixel containing no object is measured as 1 ± 1 counts

(mean \pm standard deviation) and a pixel containing an object is measured as 2 ± 1.41 counts. In case two these counts are 10 ± 3.16 and 20 ± 4.47 . When one is then tasked with deciding if a measurement contains an object or not, case two allows for better detection of the object when shot noise is considered (i.e., as S increases more faithful measurements of $S + B$ and B are obtained). Therefore, SBR alone cannot be used as a metric for image quality.

Several studies have considered how to evaluate images which can be grouped into reference or non-reference methods. Reference methods need a ground truth image, which is generally not available to the microscopist before setting up the microscope and imaging the sample [10,11]. Non-reference methods do not need a ground truth, however they vary widely from simple measures such as a signal-to-noise ratio (SNR), contrast-to-noise ratio, brightness, etc. [8,11–13], to heuristic combinations of parameters about the image such as resolution, blur, brightness etc. [14], to more complicated mathematical models [10,11,15–17], and ranking methods [10]. Generally, these methods suffer because they (1) need an image to be taken which hinders the microscopist to make predictions about which technique to use before setting up the microscope and imaging the sample, (2) do not relate to the statistics of binary detection, (3) were not developed with microscopy in mind, or (4) any combination of the three reasons listed before. Additionally, most of these techniques do not explicitly show the relationship between image quality and S and B . Two notable exceptions are an SNR used by Sandison et al. in evaluating confocal fluorescence imaging (defined as $\text{SNR} = S/\sqrt{S+B}$) [8,9] and the d' originally introduced in [18] (and modified in [19] to include SBR) for evaluating Ca^{2+} spike detection in the presence of shot noise. Sandison's SNR is a good candidate since it shows the relationship between S and B but is not grounded in binary detection theory. The modified d' is grounded in detection theory but applies to the problem of detection of Ca^{2+} spikes with the assumption that the fluorescence level detected without a spike is much greater than the maximum additional amount of fluorescence generated with a spike. Such an assumption is usually reasonable for in vivo detection of Ca^{2+} spikes with many indicators but cannot be generalized to confocal or multiphoton images for binary detection. Thus, there is a basic, but important, gap in the literature when it comes to quantitative assessment of image quality for binary detection in terms of S and B .

Here we present a statistical framework and metric to compare the quality of images in terms of S and B . We emphasize that this statistical framework and metric applies to binary detection. Our theory relies on making decisions about whether each pixel contains an object or not, which is similar to the case of binary optical telecommunications when one decides if each received bit is a 0 or 1 [20,21]. The utility of our theory is that it allows microscopists to make decisions about which technique is best to use without requiring a reference image or even taking an image. As practical examples, we show that our treatment can be used to compare the performance of 2P and 3P imaging as a function of the imaging depth, excitation wavelength and staining density. Our statistical metric also gives insight to why defining the depth limit of multiphoton microscopy as the depth where $\text{SBR} = 1$ is reasonable, even though it was chosen arbitrarily in the past.

2. Statistical theory of binary detection

We first present a statistical theory to quantify the relative image quality, for binary detection, in terms of S and B . We consider a model where we look to make a decision at every pixel (in isolation) as to whether or not that pixel contains an object. The merit of this pixel wise strategy will be discussed in section 4.

In our binary detection model, we assume that every pixel contains the same level of background fluorescence regardless of if an object is present or not and pixels which contain an object all give off the same level of signal fluorescence in addition to background. This is justified at least in deep 2P and 3P imaging where most of the background is generated away from the focus [6,7].

Additionally, we note that B can include dark counts from the detector. Under these assumptions, on average, any pixel without an object will register B photon counts at the detector, and $S + B$ counts if the pixel contains an object. We note that B and $S + B$ are the mean values of Poisson distributions assuming photon shot-noise limited performance of the imaging system.

We now consider how to use the measurement, y , at a particular pixel (i.e., y measured photon counts at the pixel) to determine if the pixel contains an object or not. Note that y will be distributed according to Poisson statistics, which is used to define two hypotheses: H_0 and H_1 which represent, respectively, the absence and presence of an object. Thus, given a specific measurement, y , the probability of obtaining y under H_0 is,

$$p_0(y) = e^{-B} B^y / y!, \quad (1)$$

and of obtaining y under H_1 is,

$$p_1(y) = e^{-(S+B)} (S+B)^y / y!. \quad (2)$$

In order to decide between the two hypotheses, one can consider a likelihood ratio of the form, $L(y) = p_1(y)/p_0(y)$, and decision rule [18,22,23],

$$\delta(y) = \begin{cases} 1 & L(y) \geq \tau \\ 0 & L(y) < \tau \end{cases} \quad (3)$$

where the 0 and 1 refer to selecting H_0 or H_1 respectively, and τ is a threshold parameter, which is set by minimizing an overall risk [22,23],

$$r(\delta) = \pi_0 \sum_{i=0}^{\infty} C_{i,0} P(i|0) + \pi_1 \sum_{i=0}^{\infty} C_{i,1} P(i|1) \quad (4)$$

where $P(i|j)$ is the probability of choosing H_i when H_j is true, $C_{ij} \geq 0$ is the incurred cost by choosing H_i when H_j is true, and π_i is the prior probability of H_i occurring. Note that π_0 and π_1 are commonly referred to as priors and that $\pi_1 = 1 - \pi_0$ is the staining density in a fluorescence image, which is defined as the fraction of volume containing fluorophore. The staining density is approximately the inverse of the labeling inhomogeneity factor, χ , used in [6] and [7]. Provided that $C_{11} < C_{01}$, τ is optimally chosen as [22,23],

$$\tau = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}. \quad (5)$$

Taking logarithms and doing some manipulation, Eq. (3) can be equivalently written as,

$$\delta(y) = \begin{cases} 1 & y > \gamma \\ 0 & y \leq \gamma \end{cases} \quad (6)$$

where,

$$\gamma = \left\lceil \frac{S}{\ln[(S+B)/B]} + \frac{\ln(\tau)}{\ln[(S+B)/B]} \right\rceil - 1. \quad (7)$$

Note that γ can be written as an integer, as done here, since y is distributed according to a Poisson distribution and can never give a non-integer measurement. Thus γ can be understood as a threshold which discriminates between two Poisson distributions, $p_0(y)$ and $p_1(y)$ (Fig. 1(a)).

Although the priors may be well known (e.g., the mouse brain vasculature has a volume fraction of approximately 2% [2]), the choice of costs is somewhat arbitrary, and a small change

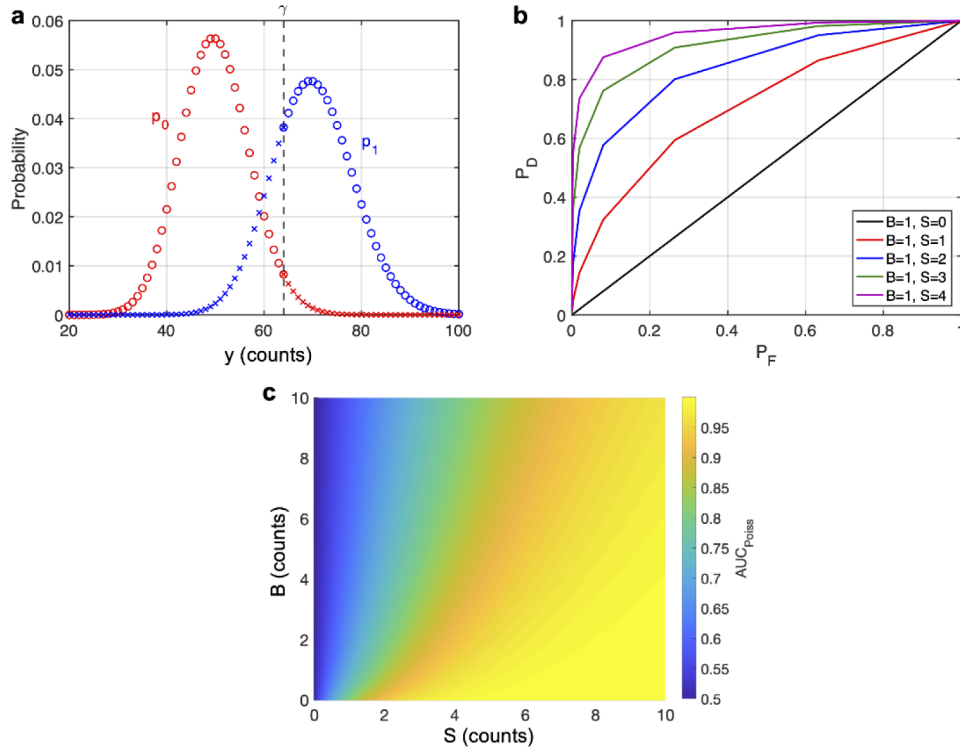


Fig. 1. Statistical Theory. (a) Graph of the probability distributions p_0 (in red) and p_1 (in blue) with arbitrary means. A decision threshold, γ , is shown as a black dashed line at an arbitrary location. Blue circles show true positives and red circles show true negatives, while red x's show false positives and blue x's show false negatives. The points on the cutoff have both symbols to indicate that randomization may be used here. (b) ROC curves using randomization for various values of S and B . (c) Integrating under the ROC curve gives the AUC which is shown for various values of S and B . The infinite sum in Eq. (15) was computed by setting the upper limit to 100. The accuracy of this was checked by computing the sum with the upper limit set to 200, which showed little difference (see Fig. S1).

can potentially give a very different true positive or detection probability, which is defined as [22–24],

$$P_D = \sum_{n=\gamma+1}^{\infty} p_1(n) = 1 - c_1(\gamma), \quad (8)$$

and a very different false positive or false alarm probability, which is defined as [22–24],

$$P_F = \sum_{n=\gamma+1}^{\infty} p_0(n) = 1 - c_0(\gamma). \quad (9)$$

Here $c_j(y) = \sum_{n=0}^y p_j(n)$ is the cumulative probability distribution for $p_j(y)$. Equations (8) and (9) are graphically illustrated in Fig. 1(a). To circumvent the arbitrariness of the costs, we consider a receiver operating characteristic (ROC), which plots the best possible P_D given a P_F [18,23–25]. One can then integrate under the ROC curve to get the area under the curve (AUC), which gives a measure of detection fidelity, and more importantly eliminates the need for choosing an arbitrary decision threshold [18]. The larger the AUC the better the detection fidelity. The concept of ROC analysis and the AUC has been widely applied to diagnostic tests

[26,27], where the ROC and AUC are typically estimated from experimental data. However, this has not been used before for assessing the quality (for binary detection) of a fluorescence image. Probabilistically speaking, AUC is the probability of an ideal observer correctly identifying an object when presented with two measurements, one which contains an object and the other which does not [18]. Additionally, the AUC can be thought of as the average P_D when averaged over all possible P_F . Mathematically speaking, one finds the best possible P_D by specifying an acceptable false alarm rate, α , and maximizing P_D over all possible decision rules such that the actually achievable P_F of the test never exceeds α , i.e. [22,23],

$$\delta = \arg \max_{\delta} P_D \quad \text{s.t.} \quad P_F \leq \alpha. \quad (10)$$

Using the decision rule specified in Eq. (6), this amounts to solving [22–24],

$$\gamma = \arg \max_{\gamma} 1 - c_1(\gamma) \quad \text{s.t.} \quad 1 - c_0(\gamma) \leq \alpha \quad (11)$$

which amounts to finding the smallest integer γ such that $1 - c_0(\gamma) \leq \alpha$ [24].

It is known that Eq. (10) is solved optimally when the specified (α) and achievable (P_F) false alarm probability are the same (provided it is possible to have $P_F \leq \alpha$) [22,23], however, the decision rule specified in Eq. (6) does not allow for arbitrary P_F and P_D to be obtained since the resulting ROC curve is only defined at discrete points due to the Poisson distribution of the photon counts (y) [23,24]. Therefore, one needs to construct continuous ROC curves. In other words, one needs to be able to achieve an arbitrary P_F and P_D . The optimum solution to Eq. (10) is known to be a randomized decision rule [22–25],

$$\delta(y) = \begin{cases} 1 & y > \gamma \\ q & y = \gamma \\ 0 & y < \gamma \end{cases} \quad (12)$$

where q is the probability of choosing H_1 . In this case first γ is found, via Eq. (11), and then one will have that

$$P_F = 1 - c_0(\gamma) + qp_0(\gamma) \quad (13)$$

and

$$P_D = 1 - c_1(\gamma) + qp_1(\gamma), \quad (14)$$

where the value of q is selected so the achievable P_F matches the specified α [22–25]. Thus, it is now possible to choose an arbitrary P_F and P_D . By inspection of Eqs. (13) and (14) one sees that given the same value of γ , P_D will be a linear function of P_F and thus the randomized rule amounts to connecting the discrete P_F and P_D found without randomization with straight lines [24] (Fig. 1(b)). Thus, using the randomized rule, the AUC can be found by summing the areas of the trapezoids,

$$\text{AUC} = \sum_{\gamma=-1}^{\infty} \left\{ \frac{1}{2} [P_D(\gamma) + P_D(\gamma + 1)] [P_F(\gamma) - P_F(\gamma + 1)] \right\}. \quad (15)$$

Note in Eq. (15), that $q = 0$, and that $c_0(-1) = c_1(-1) = 0$. Since the AUC is completely determined by S and B , i.e., $\text{AUC} = \text{AUC}(S, B)$, and the focus of this work is to assess the image quality (in the context of binary detection) in terms of S and B , we have plotted the AUC as a function of S and B in Fig. 1(c). Here we approximated the infinite sum by setting the upper limit to 100 and confirmed the accuracy by comparing this to the values computed with an upper limit

of 200 (Fig. S1). As can be seen from Fig. 1(c), the contour for a constant AUC does not follow a straight line starting from the origin, which means the AUC is not the same for a constant SBR. Indeed, given a higher value of S and the same SBR, the AUC is also higher (Fig. 1(c)).

To confirm that assessing the image quality by the AUC matches reasonably with our intuition (i.e., our visual perception), we present simulated images of beads for various values of S and B as shown in Fig. 2. The images were generated by randomly choosing locations for circles of a desired size, setting pixels within these circles to contain the bead (i.e., they have an average value of $S + B$), and pixels outside to contain background (i.e., they have an average value of B). Then we randomly assigned the pixel values according to shot noise. From Fig. 2 we see that the AUC generally matches our visual perception about image quality, and images with the same SBR, but a higher S do appear to have a better quality (e.g., the quality increases along the lower left to upper right diagonal despite that all these images have $SBR = 1$). Again, this emphasizes the argument that SBR cannot be used as a metric of image quality.

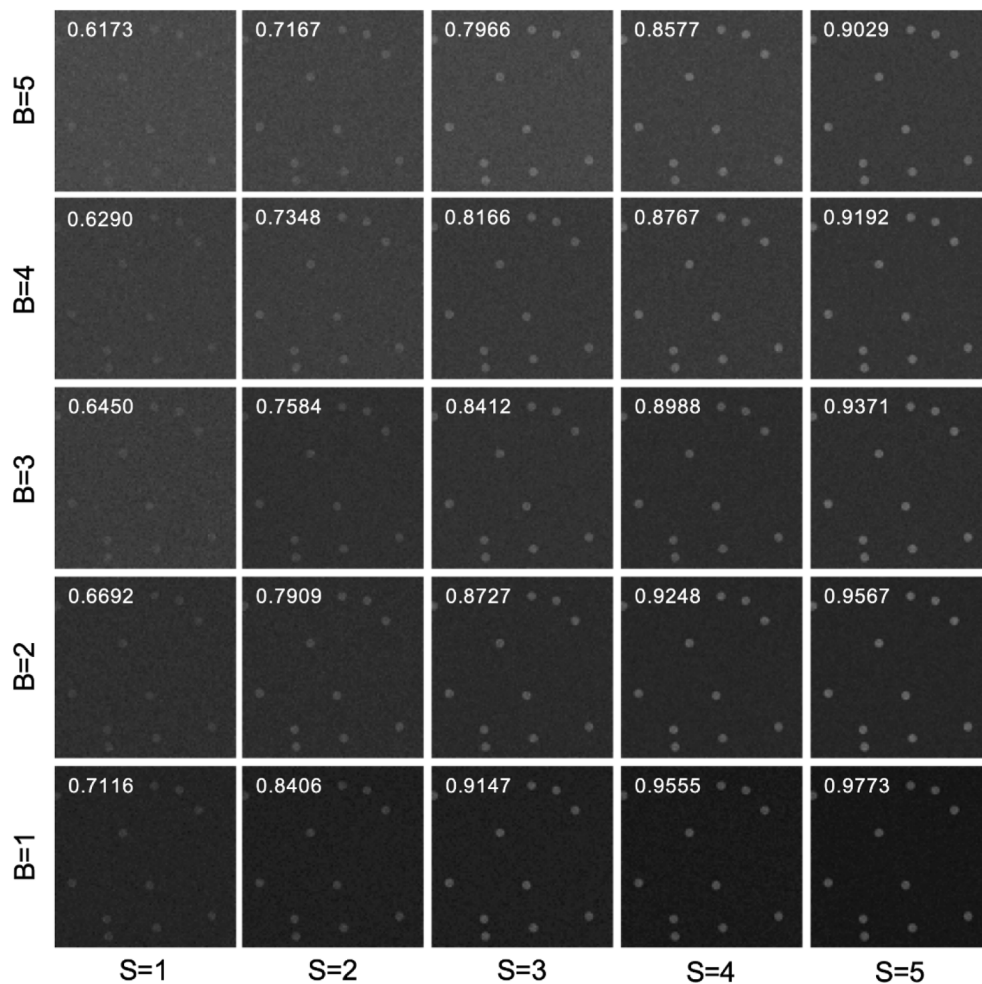


Fig. 2. Simulated Images. Simulated images of beads (diameter of 25 pixels on a 512 by 512 grid; staining density of 2%) are shown with Poisson noise added for various levels of S and B in units of counts/pixel. The white number on each image is the AUC computed using Eq. (15) with setting the upper limit in the infinite sum to 100. The window/level of each image is selected automatically using ImageJ.

In order to better understand the dependence of AUC on S and B , we look for approximate forms to Eq. (15). To do this we approximate p_0 and p_1 as Gaussian distributions (i.e., normal distributions), which are good approximations when the counts are high, and solve for the ROC considering a single threshold, which allows the AUC to be approximated as (see [Supplement 1](#) for details),

$$\text{AUC} \approx \Phi\left(\frac{S}{\sqrt{S+2B}}\right) \equiv \Phi(R), \quad (16)$$

where $\Phi(x) = 0.5\text{erfc}\left(-x/\sqrt{2}\right)$ is the standard normal cumulative distribution function, and we have defined R as,

$$R \equiv \frac{S}{\sqrt{S+2B}} = \frac{\sqrt{S}}{\sqrt{1+2/SBR}}. \quad (17)$$

The AUC under the Gaussian distribution approximation ($\text{AUC}_{\text{Gauss}}$), and the relative error of this approximation, defined as $(\text{AUC}_{\text{Gauss}} - \text{AUC}_{\text{Poiss}})/\text{AUC}_{\text{Poiss}}$, are shown in Fig. 3, where $\text{AUC}_{\text{Poiss}}$ is the AUC calculated using Eq. (15) without the Gaussian approximation. Figure 3(b) shows that the error is minimal provided B is not small (e.g., the relative error of the approximation is approximately less than 3% when S or B are greater than 1). Note that this assumes that S is non-zero, since when S is zero the AUC computed with either method is 0.5 (i.e., the smallest AUC possible), in which case practical detection is not possible. Thus, in all further discussions it will be assumed that S is non-zero.

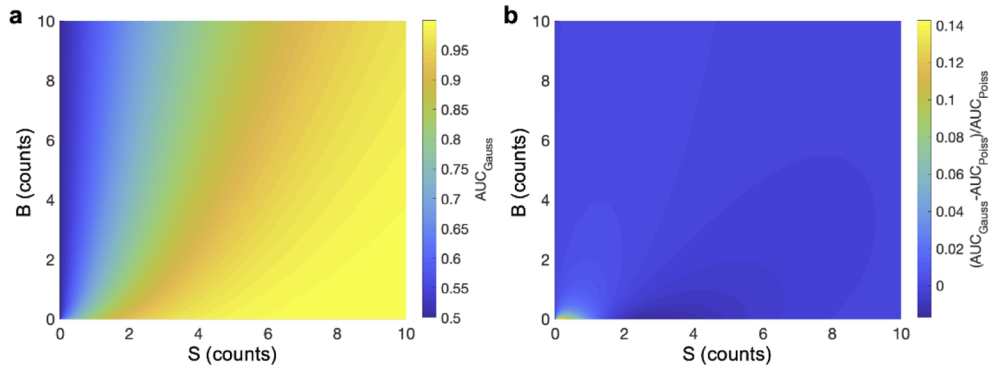


Fig. 3. Approximate AUC with Gaussian distributions. (a) The AUC using Gaussian distributions for the same values of S and B as used in Fig. 1(c). (b) The relative error of this approximation (defined as $(\text{AUC}_{\text{Gauss}} - \text{AUC}_{\text{Poiss}})/\text{AUC}_{\text{Poiss}}$) is shown for various values of S and B . Nearly all errors are small except when S and B are small, assuming S is finite.

We note that since $\Phi(x)$ is monotonically increasing, $R = \Phi^{-1}(\text{AUC}_{\text{Gauss}})$ completely characterizes the AUC in a way that captures the tradeoff between signal and background. When the photon counts are small, e.g., S and $B \ll 1$, the exact AUC (i.e., $\text{AUC}_{\text{Poiss}}$) must be calculated using Eq. (15). In parallel to the parameter R under the Gaussian approximation we introduce the binary detection factor (BDF), defined as $\text{BDF} = \Phi^{-1}(\text{AUC}_{\text{Poiss}})$, as our figure of merit for image quality, which is obtained by first calculating $\text{AUC}_{\text{Poiss}}$. We note that the BDF approaches the value R when the photon counts are large (i.e., when the Gaussian approximation is valid).

3. Utility of the BDF

In this section we will demonstrate the utility of our figure of merit, BDF, by considering some practical examples. We also explore in this section how our metric can justify the arbitrary choice of $\text{SBR} = 1$ for defining the depth limit of multiphoton imaging.

3.1. Quantitative comparison of 2P and 3P imaging for deep tissue imaging

It is qualitatively understood that 3P imaging is advantageous for deep tissue imaging while 2P imaging is preferred in the shallow regions [2,19,28,29]. Some previous attempts to calculate the cross-over depth, defined as the depth where 3P imaging begins to outperform 2P imaging, were based on the signal strength alone, but such calculations establish the upper limit of the depth where 2P imaging is preferred, as taking the SBR into account will shorten this depth [19]. The metric developed here enables us to perform a rigorous and quantitative comparison of 2P and 3P imaging by properly considering the contributions of both S and SBR. To illustrate how this is done we perform this calculation with typical parameters for 2P and 3P imaging, with 2P and 3P cross-sections reflecting those of GCaMP6s [19].

We first calculate the 2P and 3P signal as a function of imaging depth assuming a diffraction-limited focus. Because Fig. 1(c) shows that the AUC (and consequently the BDF) for the same SBR is larger with larger S , we consider the cases where S can be made maximum (likewise one can also note that $R \propto \sqrt{S}$ for the same SBR). For biological imaging, we note that the maximum S will typically depend on the maximum average power allowed in the tissue. We thus constrain this problem to a maximum average power allowed at the surface, $\langle P \rangle$. In our calculations, we choose average powers at the brain surface which are at or less than the maximum power established by previous work for mouse brain imaging: $\langle P \rangle = 100$ mW for 1300 nm 3P imaging [19] and 200 mW for 920 nm 2P Imaging [30]. Another fundamental consideration is the maximum pulse energy at the focus, which is limited by fluorophore saturation (i.e., ground state depletion) and nonlinear damage. We choose to constrain our problem to a maximum allowable saturation level (i.e., the probability of excitation per molecule per pulse), $\alpha_{sat} \equiv g_p^{(n)} \sigma_n \tau I_p^n$, where $g_p^{(n)}$ is a constant related to the temporal pulse shape, σ_n is the n-photon cross-section, τ is the pulse full width at half maximum, and I_p is the peak intensity at the focus [31]. In our calculations we choose $\alpha_{sat} = 20\%$ (of which small changes have minimal effect on calculating the signal; see Supplement 1). The resulting pulse energy at this excitation level is also commonly used for most 2P and 3P deep tissue imaging. Using the maximum average power at the surface and an allowable saturation level, along with knowledge of the concentration of fluorophore, C , the n-photon cross-section, σ_n , pulse full width at half maximum, τ , pulse shape, effective attenuation length, ℓ_e , excitation wavelength, λ , collection efficiency, ϕ , fluorophore quantum efficiency, η , tissue refractive index, n_0 , numerical aperture, NA, and pixel dwell time, T , the n-photon signal, S_n , can be completely determined (see Supplement 1 for details). We chose a constant pulse energy at the focus throughout the imaging depth (i.e., the same level of saturation) to achieve the best imaging performance possible for both 2P and 3P imaging. We note that our calculation is for demonstration purposes, and we have selected the imaging parameters commonly used for 2P and 3P imaging of the mouse brain. S_n as a function of depth is shown in Fig. 4(a).

We then proceed to calculate the 2P and 3P background as a function of depth by using the calculated signal and the 2P and 3P SBR (i.e., $B = S/\text{SBR}$). For simplicity, we consider the bulk background (i.e., fluorescence generation in the light cone above the focus) but not the defocus background (i.e., the fluorescence generated by the side lobes of a distorted point-spread function, See Fig. 3(a) in [2]), which is important when imaging through a highly scattering layer such as the mouse corpus callosum or the intact mouse skull [32]. For the 2P bulk background we note that in [6] it states that the depth limit (defined as the depth where $\text{SBR} = 1$) increases approximately one ℓ_e for a sevenfold increase in χ . Assuming that $\text{SBR} \propto \chi$, and together with the computed values in Fig. 6 of [6], the SBR can be found to be an exponential function of depth. Additionally, since the SBR could be impractically high (which is not practical due to the limited dynamic range of a real detector) at shallow depths, we forced any SBR value above 10^3 to 10^3 . For 3P imaging we note that the SBR is very large for the combinations of imaging depth and staining density considered in this paper [7,33], and so we set the 3P SBR also at

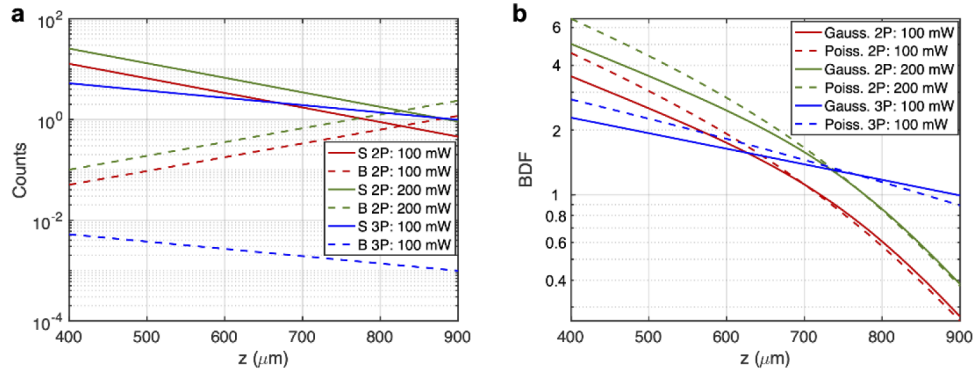


Fig. 4. Comparison of 2P and 3P image quality when imaging the same fluorophores (e.g., GCaMP6). (a) Calculation of the S and B as a function of imaging depth for 2P and 3P imaging. S was calculated under an average power at the surface (100 mW for 3P imaging and 100 mW or 200 mW for 2P imaging) and saturation constraint (20%). Other parameters used are as follows: for 2P imaging: $\sigma_2 = 20$ GM, $\eta = 1$, $\chi = 50$, $\lambda = 920$ nm, $\ell_e = 150$ μm , $g_p^{(2)} = 0.66$, NA=1, $\tau = 60$ fs, $C = 1$ μM , $n_0 = 1.33$, $T = 1$ μs , $\phi = 2\%$; for 3P imaging: $\sigma_3 = 3 \times 10^{-82}$ $\text{cm}^2/(\text{photons}/\text{cm}^2 \text{ s})^{-2}$, $\eta = 1$, $\chi = 50$, $\lambda = 1320$ nm, $\ell_e = 300$ μm , $g_p^{(3)} = 0.51$, NA=1, $\tau = 60$ fs, $C = 1$ μM , $n_0 = 1.33$, $T = 1$ μs , $\phi = 2\%$. B was subsequently calculated from S using the SBR data. SBR values above 10^3 were set to 10^3 to account for the finite dynamic range of a real detector. (b) S and B can then be used to calculate $\text{BDF} = \Phi^{-1}(\text{AUC}_{\text{Poiss}})$. Note that in the Gaussian case $\text{BDF} = R$. The upper limit for the infinite sum in Eq. (15) was set to 2000. The accuracy of this was checked by computing the sum with the upper limit set to 3000, which showed little difference (see Fig. S2).

a constant value of 10^3 . The 2P and 3P SBR values used in our calculation are shown in Fig. S2. The resulting calculated values of B_n are shown in Fig. 4(a). Here we chose to use $\chi = 50$ corresponding to a staining density of 2% (which is typical for mouse brain vasculature [2]). We note that over the depth range shown in Fig. 4(a), B in 3P imaging decreases with depth. This is merely a consequence of the fact that the SBR is held constant at 10^3 over this depth range (Fig. S2) while the 3P signal decreases with depth.

We used the calculated values of S and B to calculate the BDF using both the Poisson and Gaussian statistics. Since S and B are both determined by the imaging depth, we plot the BDF as a function of depth in Fig. 4(b). From our calculations we see that, if the average power at the surface is the same for 2P and 3P imaging (i.e., $\langle P \rangle = 100$ mW), 2P imaging is better at shallower depths due to the higher signal, and 3P imaging outperforms 2P imaging after about 620 μm , which corresponds to about $4.1\ell_e$ at the 2P imaging wavelength.

We then repeated these calculations for 2P imaging using a higher average surface power of 200 mW (which may be of more practical interest), while keeping the 3P imaging average surface power at 100 mW (Fig. 4). We see that the new 2P BDF is about $\sqrt{2}$ times larger, as predicted by theory, and the new cross-over depth at which 3P imaging outperforms 2P imaging is around 730 μm , which corresponds to about $4.9\ell_e$ at the 2P imaging wavelength. These results are consistent with previous experimental investigations which show that 2P imaging generally works better at depths shorter than about 700 μm in the mouse brain [2,19,28,29,34]. Figure 4(b) also shows that the Gaussian approximation (i.e., using R instead of BDF) predicts similar cross-over depths (Fig. 4(b)).

The above calculation uses two different excitation wavelengths for 2P (920 nm) and 3P (1300 nm) imaging. This is practical for comparing 2P and 3P imaging of the same fluorophore (e.g.,

GFP, GCaMP6, fluorescein, etc.), and the longer excitation wavelength and stronger excitation confinement of 3P excitation both contribute to the advantages of 3P imaging when imaging deep. Our metric can also delineate the two advantages of 3P imaging (i.e., longer excitation wavelength and stronger excitation confinement) by comparing 2P and 3P imaging with the same excitation wavelength, e.g., imaging near IR dyes with 2P excitation and green dyes with 3P excitation at around 1300 nm. As an example, we repeat our calculations, except assume the same excitation wavelength of 1300 nm for both 2P and 3P imaging and the same ℓ_e of 300 μm (Fig. 5). Additionally, since we expect the depth at which 3P imaging outperforms 2P imaging to be larger we change the pixel dwell time to 50 μs , which is typical for deep tissue imaging.

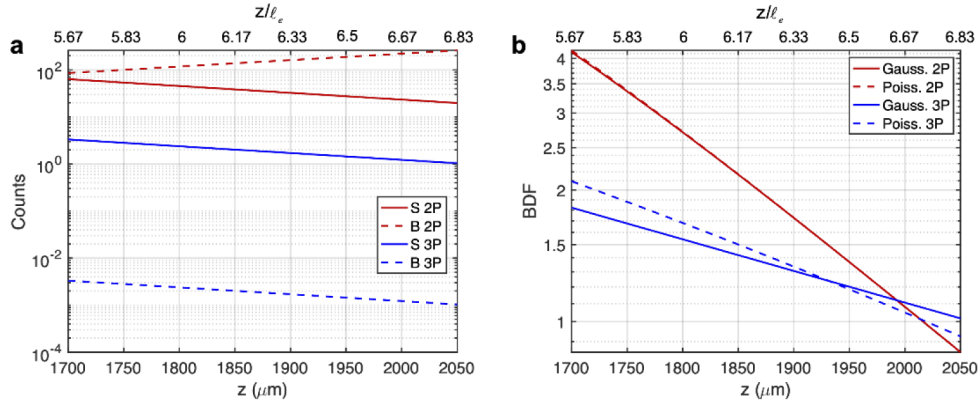


Fig. 5. Comparison of image quality for 2P imaging of near IR fluorophores (e.g., long wavelength Alexa dyes) and 3P imaging of green fluorophores (e.g., GFP). (a) Calculation of the S and B as a function of depth for 2P and 3P. S was calculated using an average power at the surface (100 mW) and saturation constraint (20%). Other parameters used are as follows: 2P imaging: $\sigma_2 = 20$ GM, $\eta = 1$, $\chi = 50$, $\lambda = 1300$ nm, $\ell_e = 300$ μm , $g_p^{(2)} = 0.66$, $\text{NA}=1$, $\tau = 60$ fs, $C = 1$ μM , $n_0 = 1.33$, $T = 50$ μs , $\phi = 2\%$; 3P imaging: $\sigma_3 = 3 \times 10^{-82}$ $\text{cm}^2/(\text{photons}/\text{cm}^2 \text{ s})^{-2}$, $\eta = 1$, $\chi = 50$, $\lambda = 1300$ nm, $\ell_e = 300$ μm , $g_p^{(3)} = 0.51$, $\text{NA}=1$, $\tau = 60$ fs, $C = 1$ μM , $n_0 = 1.33$, $T = 50$ μs , $\phi = 2\%$. B was subsequently calculated from S using SBR data. SBR values above 10^3 were set to 10^3 to account for the finite dynamic range of a real detector. (b) S and B can then be used to calculate $\text{BDF} = \Phi^{-1}(\text{AUC}_{\text{Poiss}})$. Note that in the Gaussian case $\text{BDF} = R$. The upper limit for the infinite sum in Eq. (15) was set to 2000. The accuracy of this was checked by computing the sum with the upper limit set to 3000, which showed little difference (see Fig. S3).

From Fig. 5(b), we see that 3P imaging outperforms 2P imaging after about 2000 μm , which corresponds to about $6.6\ell_e$. We note that the cross-over depth is deeper (even in unit of ℓ_e) than the previous calculations in Fig. 3, simply because the 2P excitation also benefits from the long excitation wavelength and has less attenuation in this case. Figure 5 shows the intrinsic advantage of 3P microscopy for deep tissue imaging due to its stronger excitation confinement.

A recent paper [33] has also shown that it is possible to perform mixed 2P and 3P excitation of the same dye, providing the additional possibility for multicolor 3P imaging. Our theory can also be straightforwardly extended to more complicated cases such as mixed 2P and 3P excitation.

3.2. Deep imaging 3P advantage dependance on staining density

It is well known the excitation confinement advantage of 3P imaging (as compared to 2P imaging) is more pronounced with densely labeled samples, and thus 3P is generally better than 2P for deep imaging in densely labeled samples. We also note that our calculations in section 3.1 are a

function of the staining density (since $SBR \propto \chi$), and thus the 3P advantage for deep imaging as a function of staining density can be quantified. In order to do so, we repeat our calculations from section 3.1, except we vary the staining density (e.g. $1/\chi$) and solve for the depth at which the BDF is the same for 2P and 3P imaging, which we denote as z_{eq} (Fig. 6).

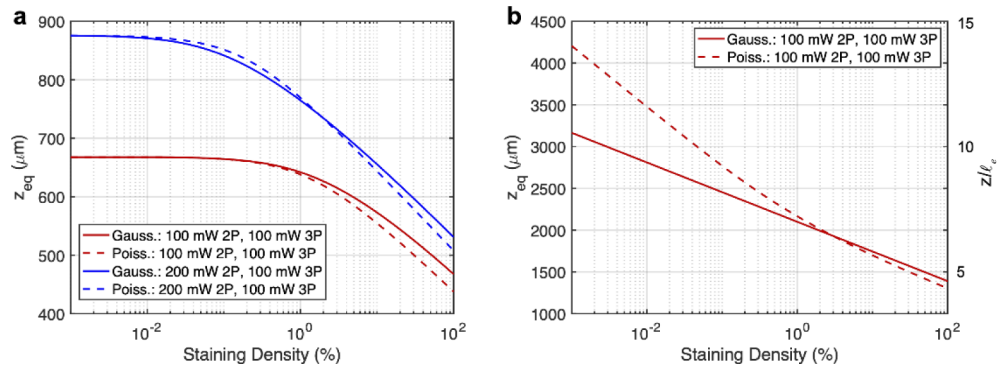


Fig. 6. Calculations of z_{eq} . z_{eq} is calculated as a function of the staining density using the parameters in Fig. 4, and Fig. 5 for panels (a) and (b), respectively. Calculations with the Poisson BDF and the Gaussian approximation are both shown. For the Poisson case the upper limit for the infinite sum in Eq. (15) was set to 2000, and the accuracy of this was checked by computing the sum with the upper limit set to 3000, which showed little difference (see Fig. S4).

From Fig. 6(a) one sees that below a certain value of staining density (about 0.1% for the 100 mW 2P and 100 mW 3P case, and about 0.01% for the 200 mW 2P case and 100 mW 3P case), z_{eq} does not change much. This can be understood because eventually the staining density becomes so small, that very little background is generated. In these cases, one finds that the confinement advantage of 3P excitation no longer matters and only the wavelength advantage plays out. This can be confirmed by noticing that the maximum value of z_{eq} in Fig. 6(a) matches the depth where the 2P and 3P signal cross in Fig. 5(a). In this sense one can consider a ‘densely labeled sample’ to mean a sample where it is necessary to consider the background generation when comparing the performance of 2P and 3P imaging in terms of z_{eq} .

Figure 6(b) does not exhibit the same behavior as Fig. 6(a), because here the 2P signal will always be higher than the 3P signal since the same excitation wavelength is used. In this case the notion of a ‘densely labeled sample’ is less clear because only the excitation confinement advantage of 3P comes into play, and thus the background always needs to be considered in determining z_{eq} despite how small it may actually be. Figure 6(b) also shows that when the staining density becomes very small that BDF and R are no longer predicting a similar z_{eq} . This can be understood since the recorded signal counts become low and so the Gaussian approximation is no longer valid. If the integration time was made longer, then the z_{eq} predicted by BDF would be closer to that predicted by R . We have varied the maximum SBR value between 200 and 5000 and compared the results with Fig. 6 (maximum SBR of 1000), we found negligible difference in z_{eq} .

3.3. Justification of $SBR=1$ for defining the depth limit of multiphoton imaging

Our statistical metric also gives insight to why defining the depth limit of multiphoton microscopy as the depth where $SBR = 1$ is reasonable, even though it was chosen arbitrarily in the past [6,7]. For simplicity we will use our metric under the Gaussian approximation (i.e., using R). We consider two extremes in SBR, where the SBR is much less and much greater than unity and consider how R (i.e., the image quality) behaves when all other parameters are held constant. In

this case we will have that

$$S = K \langle P \rangle^n \exp(-nz/\ell_e), \quad (18)$$

for n -photon excitation, where K is a proportionality constant. Additionally, we assume that the SBR behaves as an exponential function with depth as shown by previous work [6,7]. That is,

$$\text{SBR} = A \exp(-\beta z), \quad (19)$$

where A and β are constants. When $\text{SBR} \gg 1$, then from Eq. (17),

$$R \approx \sqrt{S} = K^{1/2} \langle P \rangle^{n/2} \exp(-nz/2\ell_e). \quad (20)$$

When $\text{SBR} \ll 1$, on the other hand,

$$R \approx \sqrt{S \cdot \text{SBR}/2} = K^{1/2} A^{1/2} \langle P \rangle^{n/2} \exp[-(z/2)(n/\ell_e + \beta)]/\sqrt{2}. \quad (21)$$

In both cases, R decreases exponentially as a functional of imaging depth. However, when $\text{SBR} \ll 1$, a much faster decrease occurs in R as compared to the case when $\text{SBR} \gg 1$. Figure 7 shows R against z on a semi-log plot. For $\text{SBR} \gg 1$ the slope is proportional to n/ℓ_e , and when $\text{SBR} \ll 1$ the slope is proportional to $n/\ell_e + \beta$ (Fig. 7). Such a slope change in the exponential function is identical to the situation of imaging two different layers of tissues where the deeper layer is much more scattering than the superficial layer.

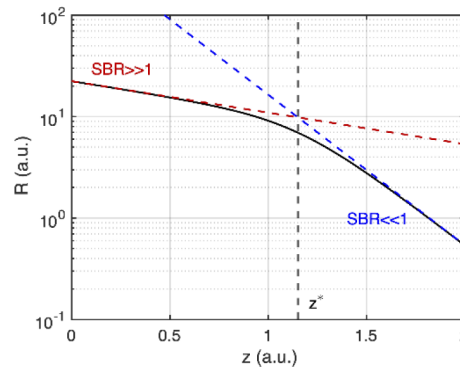


Fig. 7. R as a function of imaging depth on a semi-log plot. Red and blue dashed lines show the approximations made when $\text{SBR} \gg 1$ and $\text{SBR} \ll 1$. The dashed black line shows the approximate depth where the slope changes.

To get an idea of the depth z where this slope change happens, we equate Eqs. (20) and (21). Denoting this depth as z^* , one finds that $z^* = (1/\beta) \ln(A/2)$, and that $\text{SBR}(z^*) = 2$. This suggests that when the SBR falls below 2, R will decrease at a much faster rate (Fig. 7), as if encountering a much more scattering layer of tissue. Thus, one could consider $\text{SBR}=2$ as a reasonable criterion for the depth limit. We note that this is close to the arbitrarily selected $\text{SBR} = 1$ criterion.

4. Discussion

Much of the literature and our understanding of fluorescence imaging have dealt with how imaging parameters, such as NA, pixel dwell time, cross-section, etc., affect the signal strength, resolution, background generation, etc. [1,2,6,31]. However, knowledge of how signal and background photon counts influence the quality of an image, in terms of the binary detection of objects, has remained largely unexplored from a statistical point of view. The metric $\text{BDF} = \Phi^{-1}(\text{AUC}_{\text{Pois}}) \approx R$, thus fills this gap and enables a quantitative treatment of this issue.

Furthermore, as explained in sections 3.1 and 3.2, this metric can be used to evaluate different imaging modalities and instruments, and quantify the dependence of when different modalities outperform each other as a function of experimentally defined parameters such as imaging depth, excitation wavelength, and staining density. Additionally, our metric requires only knowledge of S and B , and so one does not need a reference image, or even to produce an image. This means that one can evaluate and design new imaging systems for biological applications, without the need to first build the instrument and then use it to produce an image.

In section 3.1 we only considered how 2P and 3P imaging compare, but one could also consider confocal imaging too. Given that the confocal point-spread-function is nearly the same as the 2P point-spread-function (neglecting the Stokes shift in the fluorescence emission), one would expect that, using the same excitation wavelength, confocal imaging should always be better than 2P imaging since the one-photon excited signal is stronger. Indeed, long wavelength one-photon confocal imaging can achieve greater than 1 mm imaging depth [35,36]. However, this does not consider the important differences in out-of-focus photobleaching and photodamage, and the significant loss of fluorescence signal in confocal detection. While our metric provides a quantitative assessment of the image quality (i.e., binary detection) in terms of S and B , other factors must also be considered for the best imaging approach, particularly when comparing between linear and nonlinear imaging modalities where the excitation confinement is fundamentally different.

Interestingly, our metric R is similar to the SNR used by Sandison *et al.* [8,9], although their SNR was not derived based on a binary detection problem. Indeed, R is nearly the same as Sandison's SNR with the one difference being the factor of 2 in front of the B in our metric. This factor of 2 can be intuitively understood as follows: to measure S one would need two measurements, one with the object ($S + B$) and one without (B), and then subtract B from $S + B$. Simple uncertainty propagation (assuming Poisson statistics) will show that the result would be $S \pm \sqrt{S + 2B}$ (signal \pm standard deviation), suggesting that our metric represents the effective SNR that could actually be measured.

Our result also validates the arbitrary choice of $SBR = 1$ for defining the depth limit when one considers the decrease in R versus imaging depth with all else held the same. From our analysis we found that a SBR below 2 marks the cutoff for when R will decrease more rapidly with depth. Although not exactly $SBR = 1$, this result is close to the arbitrarily set limit. Interestingly, a similar argument can be carried out for Ca^{2+} spike estimation with the d' modified to include SBR [19], in which case one finds that a SBR below unity marks the cutoff, matching the arbitrary choice. Thus, the SBR limit, as defined by the rate of decrease in detection ability as a function of depth, will depend on the specific detection problem at hand.

It is important to use a metric that is appropriate for the specific detection problem at hand. For example, for a problem involving Ca^{2+} spike detection, one should use d' as this metric was developed specifically for Ca^{2+} imaging. For binary detection of an object, on the other hand, one should use the BDF developed here. The BDF or d' (or other metrics) should not be applied universally, and specific metrics should be used for specific problems which they are designed for. We also note that not only will the depth limit determined by the SBR change with a different metric (such as d' or BDF), but also the cross-over depth will change.

The fact that one universal metric, in terms of S and B , may not adequately describe the detection quality of any detection problem can of course have implications in scenarios where more than one detection problem needs to be applied. For example, in Ca^{2+} imaging one first needs to perform a binary detection problem to detect the neurons, and then once the neurons are identified, needs to be able to detect spikes in time. This two-step process is most obvious in schemes such as adaptive excitation [37]. Because of the multiple metrics in this problem, an image which may be acceptable in one sub-problem may not be acceptable in the other, and the

microscopist should take careful consideration of all sub-problems when evaluating systems in these applications.

We also note that in many applications, and published images, many frames of the same field of view are taken and averaged. Our theory can be extended to this case with some slight modification. In many cases most images are viewed digitally meaning that averaging a certain number of frames and summing the same number of frames look no different when displayed. However, the latter case means that each pixel will behave with poisson statistics, and our theory can be used. We note there will now be an effective signal, $S_{eff} = NS$ and effective background, $B_{eff} = NB$, where N is the number of frames added. Thus, we would have $R_{eff} = \sqrt{NR}$, and so R would still provide a comparison provided that N is held constant.

The careful reader will note that we considered a pixel wise metric, that is to say a decision is made at each pixel, in isolation from its neighbors, as to whether or not it contains an object. We note that this mirrors a similar approach to binary optical communication where a decision is made for every bit [20,21]. In imaging, the detection of object pixels from background pixels using this approach is the same as hard thresholding an image (potentially with randomization). While this is a somewhat simplistic view, it is sufficient for a relative comparison between images. A better image should allow for more pixels to be correctly classified, and since we assume that the only thing that changes between different imaging modalities is S and/or B , the comparison of how many pixels are correctly classified provides a relative comparison of image quality. We note that such a relative comparison is adequate for making decisions about which imaging instruments (e.g., 2P or 3P imaging), dyes, and parameters to use.

Although the AUC is the most common metric for comparing ROC curves, it gives a global metric [26,27]. When not all values of α may be acceptable, one may want to consider instead a partial AUC where only the area under the ROC curve in the acceptable range of α is considered [26,27]. We note that such analysis could be carried out in a similar manner as presented here.

One could consider a more complicated case for detecting a group of pixels containing an object. For example, for a bead of a pre-defined size one would have a circle of pixels that contain an object. We note that this detection problem could also be combined with an estimation problem (e.g., determining where the bead is centered in the image). We did not explore these problems here since it is not needed in our pixel-wise approach and could be considered in future work. We also note that this work did not consider multi-level detections, which could also be examined in the future.

Funding. National Science Foundation (DBI-1707312).

Acknowledgments. The authors acknowledge fruitful discussions with Alejandro Simon, Yi Wang, and Farhan Rana.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," *Nat. Methods* **2**(12), 932–940 (2005).
2. T. Wang and C. Xu, "Three-photon neuronal imaging in deep mouse brain," *Optica* **7**(8), 947–960 (2020).
3. G. E. Stutzmann and I. Parker, "Dynamic multiphoton imaging: a live view from cells to systems," *Physiology* **20**(1), 15–21 (2005).
4. R. N. Germain, M. J. Miller, M. L. Dustin, and M. C. Nussenzweig, "Dynamic imaging of the immune system: progress, pitfalls and promise," *Nat Rev Immunol* **6**(7), 497–507 (2006).
5. J. Wang, M. Hossain, A. Thanabalasuriar, M. Gunzer, C. Meininger, and P. Kubes, "Visualizing the function and fate of neutrophils in sterile injury and repair," *Science* **358**(6359), 111–116 (2017).
6. P. Theer and W. Denk, "On the fundamental imaging-depth limit in two-photon microscopy," *J. Opt. Soc. Am. A* **23**(12), 3139–3149 (2006).

7. N. Akbari and C. Xu, "Theoretical and experimental investigation of the depth limit of three-photon microscopy," *Proc. SPIE* **11648**, 116481A (2021).
8. D. R. Sandison and W. W. Webb, "Background rejection and signal-to-noise optimization in confocal and alternative fluorescence microscopes," *Appl. Opt.* **33**(4), 603–615 (1994).
9. D. R. Sandison, D. W. Piston, and W. W. Webb, "Background rejection and optimization of signal-to-noise in confocal microscopy," in *Three-Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens*, J. K. Stevens, L. R. Mills, and J. E. Trogadis, eds. (Academic Press Inc., 1994).
10. S. Koho, E. Fazeli, J. E. Eriksson, and P. E. Hänninen, "Image quality ranking method for microscopy," *Sci. Rep.* **6**(1), 28962 (2016).
11. S. G. Stanciu, F. J. Ávila, R. Hristu, and J. M. Bueno, "A study on image quality in polarization-resolved second harmonic generation microscopy," *Sci. Rep.* **7**(1), 15476 (2017).
12. U. Kanniyappan, B. Wang, C. Yang, M. Litorja, N. Suresh, Q. Wang, Y. Chen, and T. J. Pfefer, "Performance test methods for near-infrared fluorescence imaging Pejman Ghassemi," *Med. Phys.* **47**(8), 3389–3401 (2020).
13. M.-A. Bray, A. N. Fraser, T. P. Hasaka, and A. E. Carpenter, "Workflow and metrics for image quality control in large-scale high-content screens:," *J. Biomol. Screen* **17**(2), 266–274 (2012).
14. S. G. Stanciu, G. A. Stanciu, and D. Coltuc, "Automated compensation of light attenuation in confocal microscopy by exact histogram specification," *Microsc. Res. Tech.* **73**(3), 165–175 (2010).
15. A. A. Dimia, J. T. Elliott, J. J. Filliben, M. Halter, A. Peskin, J. Bernal, M. Kociolek, M. C. Brady, H. C. Tang, and A. L. Plant, "Comparison of segmentation algorithms for fluorescence microscopy images of cells," *Cytometry* **79**(7), 545–559 (2011).
16. N. B. Nill and B. Bouzas, "Objective image quality measure derived from digital image power spectra," *Opt. Eng.* **31**(4), 813–825 (1992).
17. R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. on Image Process.* **18**(4), 717–728 (2009).
18. B. A. Wilt, J. E. Fitzgerald, and M. J. Schnitzer, "Photon shot noise limits on optical detection of neuronal spikes and estimation of spike timing," *Biophysical Journal* **104**(1), 51–62 (2013).
19. T. Wang, C. Wu, D. G. Ouzounov, W. Gu, F. Xia, M. Kim, X. Yang, M. R. Warden, and C. Xu, "Quantitative analysis of 1300-nm three-photon calcium imaging in the mouse brain," *eLife* **9**, e53205 (2020).
20. G. P. Agrawal, *Fiber-Optic Communication Systems* (John Wiley & Sons, 2002), Chap. 4.
21. G. Keiser, *Optical Fiber Communications* (McGraw Hill, 2000), Chap. 7.
22. H. V. Poor, *An Introduction to Signal Detection and Estimation*, Second (Springer-Verlag Berlin Heidelberg, 1994, Chap. 2).
23. B. C. Levy, *Principles of Signal Detection and Parameter Estimation* (Springer Science and Business Media LLC, 2008), Chap. 2.
24. S. E. Johnson, "Target detection with randomized thresholds for lidar applications," *Appl. Opt.* **51**(18), 4139–4150 (2012).
25. L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* (Addison-Wesley, 1991), Chap. 4.
26. K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Casp. J. Intern. Med.* **4**(2), 627–635 (2013).
27. T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *Journal of Biomedical Informatics* **38**(5), 404–415 (2005).
28. K. Takasaki, R. Abbasi-Asl, and J. Waters, "Superficial bound of the depth limit of two-photon imaging in mouse brain," *eNeuro* **7**(1), ENEURO.0255-19.2019 (2020).
29. D. G. Ouzounov, T. Wang, M. Wang, D. D. Feng, N. G. Horton, J. C. Cruz-Hernandez, Y.-T. Cheng, J. Reimer, A. S. Tolias, N. Nishimura, and C. Xu, "In vivo three-photon imaging of activity of GCaMP6-labeled neurons deep in intact mouse brain," *Nat. Methods* **14**(4), 388–390 (2017).
30. K. Podgorski and G. Ranganathan, "Brain heating induced by near-infrared lasers during multiphoton microscopy," *J. Neurophysiol.* **116**(3), 1012–1023 (2016).
31. C. Xu and W. W. Webb, "Multiphoton excitation of molecular fluorophores and nonlinear laser microscopy," in *Topics in Fluorescence Spectroscopy*, J. R. Lakowicz, ed. (Plenum Press, 2002).
32. T. Wang, D. G. Ouzounov, C. Wu, N. G. Horton, B. Zhang, C.-H. Wu, Y. Zhang, M. J. Schnitzer, and C. Xu, "Three-photon imaging of mouse brain structure and function through the intact skull," *Nat. Methods* **15**(10), 789–792 (2018).
33. Y. Hontani, F. Xia, and C. Xu, "Multicolor three-photon fluorescence imaging with single-wavelength excitation deep in mouse brain," *Sci. Adv.* **7**(12), eabf3531 (2021).
34. M. Yildirim, H. Sugihara, P. T. C. So, and M. Sur, "Functional imaging of visual cortical layers and subplate in awake mice with optimized three-photon microscopy," *Nat. Commun.* **10**(1), 177 (2019).
35. F. Xia, C. Wu, D. Sinefeld, B. Li, Y. Qin, and C. Xu, "In vivo label-free confocal imaging of the deep mouse brain with long-wavelength illumination," *Biomed. Opt. Express* **9**(12), 6545–6555 (2018).
36. F. Xia, M. Gevers, A. Fognini, A. T. Mok, B. Li, N. Akabri, I. E. Zadeh, J. Qin-Dregely, and C. Xu, "Short-wave infrared confocal fluorescence imaging of deep mouse brain with a superconducting nanowire single-photon detector," *ACS Photonics* **8**(9), 2800–2810 (2021).

37. B. Li, C. Wu, M. Wang, K. Charan, and C. Xu, "An adaptive excitation source for high-speed multiphoton microscopy," *Nat. Methods* **17**(2), 163–166 (2020).