Research Article Vol. 2, No. 11 / November 2015 / Optica 950

Image retrodiction at low light levels

Matthias Sonnleitner,[1,*] John Jeffers,[2] and Stephen M. Barnett[1]

[1]School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
[2]SUPA, Department of Physics, University of Strathclyde, Glasgow G4 0NG, UK
*Corresponding author: matthias.sonnleitner@glasgow.ac.uk

Received 15 July 2015; revised 28 September 2015; accepted 28 September 2015 (Doc. ID 246037); published 4 November 2015

Imaging technologies working at very low light levels acquire data by counting the number of photons impinging on each pixel. Especially in cases with, on average, less than one photocount per pixel, the resulting images are heavily corrupted by Poissonian noise. To tackle this problem, we use methods from Bayesian statistics to retrodict the spatial intensity distribution responsible for the photocount measurements. Unlike the usual photon-limited image denoising algorithms, we calculate the full probability distributions for the intensities at each pixel. The knowledge of these probability distributions helps to assess the validity of results from image analysis using data corrupted by Poisson noise with low photon-count numbers and dark counts.

OCIS codes: (100.3008) Image recognition, algorithms and filters; (110.3010) Image reconstruction techniques; (040.5160) Photodetectors.

http://dx.doi.org/10.1364/OPTICA.2.000950

1. INTRODUCTION

Numerous existing and proposed imaging technologies work in very low light intensities and gather data using arrays of photodetectors [1–4]. The resulting measurements are tables of natural numbers where each entry is the number of photocounts on a specific pixel, a single realization of a Poisson process whose mean value is proportional to the time-integrated intensity of the light field imaged by the respective pixel.

At very low photocount rates the resulting images suffer heavily from quantum or Poisson noise. This noise is inherently different from the independent and identically distributed (Gaussian) noise tackled by most image optimization algorithms, but the importance of imaging technologies based on photon counts led to a range of image optimization algorithms specifically dealing with Poisson noise.

The basic principle of image denoising is to replace the noisy measured data with an assumed intensity distribution, and of course there exist several publications on photon-limited image denoising that show spectacular results [1,2,4–7]. In particular, maximum likelihood denoising models search for the intensity distribution that satisfies certain smoothness criteria or other known features of "typical images" while maximizing the likelihood of reproducing the measured data.

Once we consider measurements with on average less than one count per pixel, however, noise tends to obscure many structures in an image, or, worse, it might produce deceptive features. Even elaborate denoising algorithms might then produce smoothed versions of misleading images, and without the full probability distribution, it is impossible to assess the validity of conclusions based on analysis of such a denoised image. Only a probabilistic treatment allows us to provide qualified statements about noisy images, just as an error bar is necessary to interpret the quality of a smooth fitting curve in noisy 1D data. Thus the present work uses simple but accessible methods to calculate the (Bayesian) probability that a certain intensity level λ_i was responsible for a measurement m_i at the ith pixel.

Our approach to this problem is inspired by quantum retrodiction [8–10], in which a premeasurement state is assigned on the basis of a later measurement and used to obtain statistical information on earlier events. In particular, the detection of a given number of photocounts by a photon counter of known efficiency leads directly to a premeasurement probability distribution for the number of photons present [11].

This work is organized as follows. In Section 2 we introduce the basic notation and discuss how the intensity of the light field incident on a single pixel can be retrodicted from the measurement value at this pixel. In Section 3 we refine this retrodiction both by using measurements from neighboring pixels and with methods inspired by averaging algorithms [12]. After assessing the general validity of the retrodiction schemes in Section 4, we go on in Section 5 to describe how to use the probabilities obtained from various retrodiction models to test actual hypotheses about an image.

The different retrodiction methods discussed in this work are illustrated using a picture of a statue of Lord Kelvin [13] with added artificial Poisson noise, as shown in Fig. 2. The image data is gray-scale, but colors are used in Figs. 2, 3, 5, and 8 for better visibility. For the computer-generated noisy image in Fig. 2(b), we assume a very low detection efficiency $\eta = 0.05$, as experienced in some quantum optics experiments [4]. In Figs. 3 and 5 we show the expectation values of the probability distributions retrodicted from this noisy image, but we emphasize that these

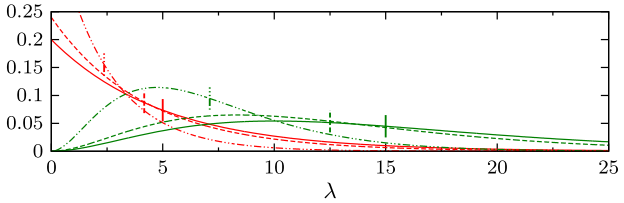

Fig. 1. Comparison of the probabilities for the intensity retrodictions with and without prior information on the expected mean photocount number $\bar{m}$. Broken lines show $p(\lambda|m,\bar{m})$ for $\bar{m}=0.9$ (dashed–dotted) and $\bar{m}=5$ (dashed). The solid line shows $p(\lambda|m)$ using the flat prior $p(\lambda)$, which is equivalent to $\bar{m}\to\infty$. The red lines are for measurement $m=0$, and green for $m=2$; the respective expectation values are indicated by the short vertical lines in the respective style; $\eta=0.2$. We see how information about $\bar{m}$ shifts the retrodictive distributions of λ to lower values.

are not necessarily less noisy than the original. This work focuses on the probabilities and not on typical image denoising.

The paper is intended to be self-contained, but some of the more technical calculations and derivations were moved into Supplement 1. There one can also find some of the results recalculated for the more general case in which measurements are distorted by dark counts.

2. SINGLE-PIXEL RETRODICTION

First, let us attempt to understand the sequence of random processes that determine the number of photocounts for a single pixel. If we have a detector with a finite detection efficiency η, then the probability of getting m counts given that n photons were incident upon the detector is given by a binomial distribution [11,14]

$$p(m|n)=\binom{n}{m}\eta^m(1-\eta)^{n-m}. \tag{1}$$

In this work we shall assume an attenuated single-mode light source such that the number of photons coming from the illuminated object during a certain time interval follows a Poisson distribution of mean λ, $p(n|\lambda)=\mathrm{Pois}(n;\lambda)$, with

$$\mathrm{Pois}(n;\lambda):=\frac{e^{-\lambda}\lambda^n}{n!}. \tag{2}$$

This means, in turn, that the number of photocounts is also given by a Poisson process of mean $\eta\lambda$, i.e., $p(m|\lambda)=\sum_{n=m}^{\infty}p(m|n)p(n|\lambda)=\mathrm{Pois}(m;\eta\lambda)$. The value λ hence determines the amount of light coming from the section of the object imaged by our pixel. A larger value of λ results in, on average, a higher number of photons traveling toward the pixel, n, and being detected later, m.

A. Intensity Retrodiction

We use the term *intensity retrodiction* to describe the process of finding the probability distribution for λ given m photocounts. Using Bayes' theorem, we have

$$p(\lambda|m)=\frac{p(m|\lambda)p(\lambda)}{p(m)}. \tag{3}$$

While $p(m|\lambda)=\mathrm{Pois}(m;\eta\lambda)$ has been calculated above and $p(m)=\int_0^\infty p(m|\lambda)p(\lambda)\mathrm{d}\lambda$, we lack an expression for the prior probability $p(\lambda)$.

A common approach for choosing a prior in Bayesian statistics would be that $p(\lambda)$ is constant (up to some arbitrarily large value). (That is, we set $p(\lambda)=1/\lambda_{\max}$ for $\lambda\in[0,\lambda_{\max}]$ and $p(\lambda)=0$ otherwise. The term $1/\lambda_{\max}$ thus cancels from Eq. (3), and taking the limit $\lambda_{\max}\to\infty$ then gives a gamma distribution.) This gives a gamma distribution for λ as $p(\lambda|m)=\mathrm{Gam}(\lambda;m+1,\eta^{-1})$, with

$$\mathrm{Gam}(x;\alpha,\beta):=\frac{x^{\alpha-1}}{\beta^\alpha}\frac{e^{-x/\beta}}{\Gamma(\alpha)}, \tag{4}$$

and a mean value $E(\lambda|m)=(m+1)/\eta$.

Such a flat prior suggests that we assume that the brightness λ could reach any absurdly high value. In reality, however, we can check the data from all N pixels to get an average detected photon number $\bar{m}=\sum_{i=1}^{N}m_i/N$, where m_i denotes the measurement in the ith out of N pixels. Given only this information, it is more reasonable to search for a prior $p(\lambda|\bar{m})$ that ensures that $p(m|\bar{m})$ has an expectation value $E(m|\bar{m})=\bar{m}$. The least biased—i.e., entropy maximizing—discrete distribution with a given mean value $\bar{m}$ is a geometric distribution [15]

$$p(m|\bar{m})=\frac{\bar{m}^m}{(\bar{m}+1)^{m+1}}. \tag{5}$$

Considering that $p(m|\bar{m})=\int_0^\infty p(m|\lambda)p(\lambda|\bar{m})\mathrm{d}\lambda$, this requires an exponential prior distribution for the brightness

$$p(\lambda|\bar{m})=\frac{\eta}{\bar{m}}e^{-\eta\lambda/\bar{m}}. \tag{6}$$

Using this prior in Eq. (3) results again in a gamma distribution

$$p(\lambda|m,\bar{m})=\mathrm{Gam}(\lambda;m+1,(\eta(1+1/\bar{m}))^{-1}), \tag{7}$$

with an expectation value

$$E(\lambda|m,\bar{m})=\frac{m+1}{\eta(1+1/\bar{m})}. \tag{8}$$

If we set $\bar{\eta}:=\eta(1+1/\bar{m})$, we see that the distribution obtained from the flat prior in Eq. (4) is recovered for $\bar{m}\to\infty$. The effect of this more informed prior can be seen in Fig. 1, where $p(\lambda|m)$ is compared against $p(\lambda|m,\bar{m})$.

In Section 4 we shall show that the retrodiction gives the best results if $\bar{m}/\eta$ is close to the true intensity value at the respective pixel. As we calculate only one $\bar{m}$ for the whole image, however, this will usually not be the case for all pixels. It is therefore useful to compute locally expected mean values $\bar{m}_i$ for the ith pixel, for instance, using successful denoising algorithms from the literature [1,2]. The retrodicted distributions for the intensity then still follow gamma distributions, but with different mean values. To simplify the discussion here, we will continue to use only the global average $\bar{m}$ in this work.

Based on the measurements m_i shown in Fig. 2(b), the mean retrodicted intensities $E(\lambda|m_i,\bar{m})$ from Eq. (8) are shown in Fig. 3(a). This replacement corresponds to a linear shift in each pixel value, so there is no enhancement regarding the perceived image quality. It gives, however, a better estimate of the actual amount of light that was sent from the object.

B. Transmission Retrodiction

In some applications one not only has information about the number of detected photons, but one also knows the amount of light used to illuminate the object. Let us assume the number of photons emitted from the source toward the detector within a certain time follows a Poisson distribution of mean value ν,

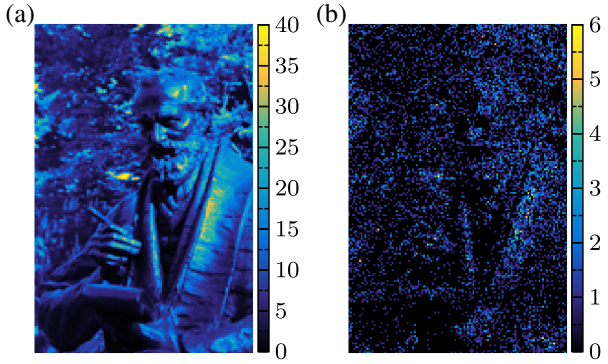

Fig. 2. (a) Original image [13] and (b) image distorted by artificial Poisson noise serving as the measured image for the upcoming examples. The colors indicate (a) the true intensity values λ_i and (b) the measurement values m_i, $i = 1, \ldots, N$. Detection efficiency $\eta = 0.05$; average number of photocounts $\bar{m} \approx 0.4$; no. of pixels $N = 120 \times 180$.

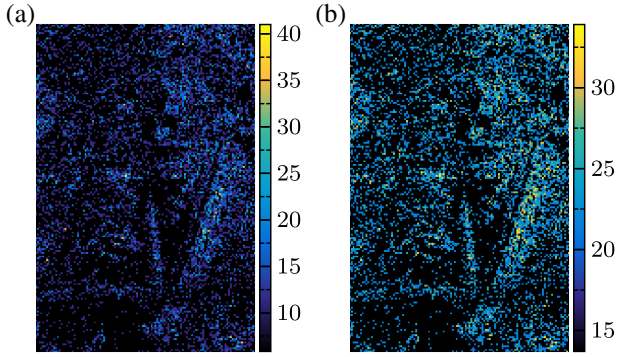

Fig. 3. Mean values of single-pixel retrodictions of the data from Fig. 2(b). (a) Mean value for intensity retrodiction $E(\lambda_i|m_i, \bar{m})$ as given in Eq. (8); (b) corresponding image $\nu_i E(\tau|m_i, \nu_i)$ from transmission retrodiction, cf. Eq. (12). The illumination is assumed to be constant with $\nu_i = 40 \quad \forall\, i = 1, \ldots, N$. Note that these images are generated using only values from the individual pixels and hence cannot be less noisy than the raw data.

i.e., $p(n_{\text{inc}}|\nu) = \text{Pois}(n_{\text{inc}}; \nu)$. In a simplified model we assume that the number of photons passing through the object, n, is determined by the transmission rate $\tau \in [0, 1]$ such that

$$p(n|\tau, \nu) = \sum_{n_{\text{inc}}=n}^{\infty} \binom{n_{\text{inc}}}{n} \tau^n (1-\tau)^{n_{\text{inc}}-n} p(n_{\text{inc}}|\nu) = \text{Pois}(n; \tau\nu). \tag{9}$$

Including the single photon detection efficiency η, we find $p(m|\tau, \nu) = \text{Pois}(m; \eta\tau\nu)$. Using Bayes' theorem and a uniform prior $p(\tau) = 1$ for $\tau \in [0, 1]$, we find that the transmission rate τ follows a distribution similar to a gamma distribution, but with an incomplete gamma function, i.e., $p(\tau|m, \nu) = \text{Gami}(\tau; m+1, 1/(\eta\nu))$, with

$$\text{Gami}(x; \alpha, \beta) := \frac{x^{\alpha-1}}{\beta^\alpha} \frac{e^{-x/\beta}}{\Gamma(\alpha, 1/\beta)}, \tag{10}$$

where $\Gamma(n, x)$ denotes the incomplete gamma function

$$\Gamma(n, x) := \int_0^x e^{-t} t^{n-1} \mathrm{d}t. \tag{11}$$

The corresponding expectation value is

$$E(\tau|m, \nu) = \frac{1}{\eta\nu} \frac{\Gamma(m+2, \eta\nu)}{\Gamma(m+1, \eta\nu)}. \tag{12}$$

(Using this method to retrodict the intensities gives $p(\lambda|m, \nu) = \text{Gam}(\lambda; m+1, 1/\eta) m!/\Gamma(m+1, \eta\nu)$ and $E(\lambda|m, \nu) = \Gamma(m+2, \eta\nu)/(\eta\Gamma(m+1, \eta\nu))$. Note that in this case $0 \le \lambda \le \nu$.)

From a basic physical point of view, it is equivalent to retrodict the apparent brightness λ or the transmission rate τ. From a logical perspective, however, adding the knowledge about the initial illumination level ν makes quite a difference: an outlying measurement m can easily result in a mean retrodicted intensity larger than the actual source, $E(\lambda|m) > \nu$. This cannot happen with the equivalent value from transmission retrodiction, $\nu E(\tau|m, \nu)$.

This is visible in Fig. 3(b), where we show the expected intensity using transmission retrodiction, $\nu_i E(\tau|m_i, \nu_i)$, given the measurements from Fig. 2(b). There [and in further examples in Figs. 7(b) and 8(b)] we see that the flat prior used here tends to bias the retrodicted probabilities toward values where $\tau \sim 1/2$ or $\lambda = \nu/2$. Obviously this is least suitable for images where we have uniform illumination, such as is assumed for the example depicted in Fig. (2). Transmission retrodiction as discussed here is therefore only useful in connection with a known, inhomogeneous illumination pattern.

Note also that the assumption $0 \le \tau \le 1$ leads to errors if light traveling through the object shows (positive) interference effects or is cross scattered such that some pixels actually see more light than without the object.

To simplify the remaining discussion, we will use the terminology of intensity retrodiction and sometimes give equivalent results for transmission retrodiction in brief remarks.

Let us briefly note that this transmission retrodiction model can be adapted for images using light reflected off an object, though there one often has to deal with (unknown) background illumination as well [3,7,16].

3. MIXED RETRODICTION

The methods discussed in Section 2 attempt to retrodict the light intensity incident on a single pixel using only data from this one pixel (and the average measurement for all pixels, $\bar{m}$). These methods thus attempt to find a probability distribution using only a single measurement value.

But from experience we know that typical images have regions of constant brightness such that we might assume that measurements m and m' at two neighboring pixels are the results of Poisson processes with the same mean, $\lambda = \lambda'$. In this case a Bayesian update of the probability distribution for λ using the second data point m' is possible, i.e., $p(\lambda|m, m', \bar{m}) = p(m'|\lambda)\, p(\lambda|m, \bar{m})/p(m'|m, \bar{m})$:

$$p(\lambda|m, m', \bar{m}) = \text{Gam}(\lambda; m + m' + 1, (\eta(2 + 1/\bar{m}))^{-1}). \tag{13}$$

Figure 4 illustrates the line of thought from single-pixel retrodiction to mixed retrodiction as outlined below. In Section 2 we used individual measurements m and m' to retrodict their respective λ, λ'. Assuming $\lambda = \lambda'$ allows us to calculate an updated distribution using both m and m' in Eq. (13). After an intermediate step in Section 3.A, we will discuss in Section 3.B how the strong claim $\lambda = \lambda'$ can be toned down by introducing a weight W' representing an *ad hoc* estimate that there is a chance W' for $\lambda = \lambda'$ and

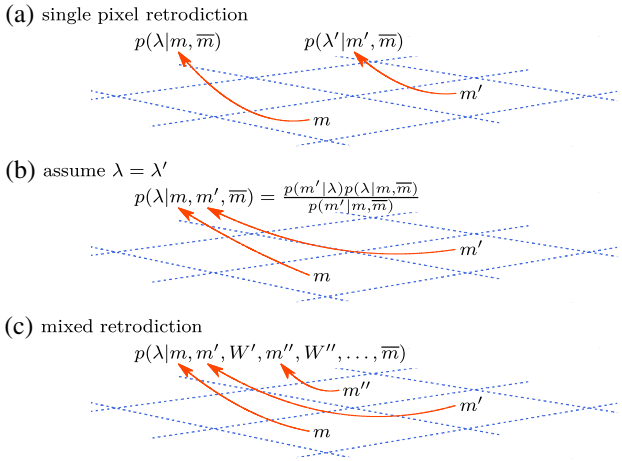

Fig. 4. Illustration of the mixed retrodiction model. (a) In single-pixel intensity retrodiction, described in Section 2, the number of photocounts at each pixel is used to calculate a probability distribution for λ at this pixel; (b) the probability distribution obtained after assuming that measurements at two different pixels, m and m', came from the same intensities, cf. Eq. (13); (c) mixed retrodiction as outlined in Section 3.B then uses predefined weights W', W'', … reflecting the assumed probabilities for the cases $\lambda'=\lambda;\lambda''=\lambda;\ldots$. Hence, assumed correlations between the neighboring pixels are used to update the probability distribution $p(\lambda|\ldots)$.

$1-W'$ for $\lambda\neq\lambda'$. Sections 3.C and 3.D will give two examples for possible definitions of these weights.

A. Retrodiction from a Mixed Source

Let us assume a situation where each photon hitting a certain pixel has a chance W of coming from a source with intensity λ and a chance $1-W$ of coming from some other source with λ', i.e.,

$$p(m|\lambda,\lambda',W)=\mathrm{Pois}(m;\eta W\lambda+\eta(1-W)\lambda'). \quad (14)$$

Such a situation might occur for an image that is known to have simple bright and dark sectors.

If we lack further knowledge about λ', we marginalize using $p(\lambda'|\bar{m})\sim\exp(\eta\lambda/\bar{m})$,

$$p(m|\lambda,W,\bar{m})=\int_0^\infty \mathrm{d}\lambda' p(m|\lambda,\lambda',W)p(\lambda'|\bar{m})$$
$$=\sum_{k=0}^{m}\mathrm{Geom}(m-k;(1-W)\bar{m})\mathrm{Pois}(k;\eta W\lambda), \quad (15)$$

where we set $\mathrm{Geom}(n;x):=x^n/(x+1)^{n+1}$ as the geometric distribution of mean value x as used in Eq. (5).

To retrodict λ from this measurement of m with the assumed weight W, we use $p(\lambda|m,W,\bar{m})\sim p(m|\lambda,W,\bar{m})p(\lambda|\bar{m})$,

$$p(\lambda|m,W,\bar{m})\sim\sum_{k=0}^{m}\mathrm{Gam}(\lambda;k+1,[\eta(W+1/\bar{m})]^{-1})$$
$$\times\mathrm{Geom}(m-k;(1-W)\bar{m})\mathrm{Geom}(k;W\bar{m}). \quad (16)$$

The symbol $\sim$ indicates that we dropped the normalization factor $p(m|W,\bar{m})=\int_0^\infty \mathrm{d}\lambda p(m|\lambda,W,\bar{m})p(\lambda|\bar{m})$. But calculating $p(m|W,\bar{m})$, expectation values $E(\lambda|m,W,\bar{m})$ or higher moments is a simple task, as it only requires integration over the gamma distribution. [To recover the result from Eq. (7), we set $W=1$ and use $\lim_{x\to0}\mathrm{Geom}(n;x)=\begin{cases}1 & \text{for } n=0,\\ 0 & \text{for } n>0.\end{cases}$]

B. Retrodiction Updates Assuming Mixed Sources

Equation (13) gave the Bayesian update for the case where it is known that both measurement values, m and m', come from the same source of value λ. In line with the results given above, we go on with the updated retrodiction for the case in which we assume some weights W and W' for the case where m, m' come from the same λ. Using Eqs. (15) and (16) gives

$$p(\lambda|m,m',W,W',\bar{m})\sim p(m'|\lambda,W',\bar{m})p(\lambda|m,W,\bar{m})$$
$$=\sum_{k=0}^{m}\sum_{k'=0}^{m'}\mathrm{Gam}(\lambda;k+k'+1,[\eta(W+W'+1/\bar{m})]^{-1})$$
$$\times\binom{k+k'}{k}\frac{(\bar{m}W)^k(\bar{m}W')^{k'}}{(\bar{m}(W+W')+1)^{k+k'+1}}$$
$$\times\mathrm{Geom}(m-k;(1-W)\bar{m})\mathrm{Geom}(m'-k';(1-W')\bar{m}). \quad (17)$$

Again, calculation of the normalization and the mean value are easy, as λ appears only in the gamma distribution terms. Aside from a normalization factor, the mean value has the form

$$E(\lambda|m,m',W,W',\bar{m})\sim\sum_{k=0}^{m}\sum_{k'=0}^{m'}\frac{k+k'+1}{\eta(W+W'+1/\bar{m})}$$
$$\times\binom{k+k'}{k}\frac{(\bar{m}W)^k(\bar{m}W')^{k'}}{(\bar{m}(W+W')+1)^{k+k'+1}}$$
$$\times\mathrm{Geom}(m-k;(1-W)\bar{m})\mathrm{Geom}(m'-k';(1-W')\bar{m}). \quad (18)$$

Although these expressions look complicated, they follow a simple pattern that is best explained by discussing what happens after a further update using a measurement m'' and a weight W''. First, this gives another sum $\sum_{k''=0}^{m''}$ and a geometric distribution $\mathrm{Geom}(m''-k'';(1-W'')\bar{m})$. The gamma distribution in Eq. (17) or its mean value in Eq. (18) changes as $k+k'+1\to k+k'+k''+1$ and $(W+W'+1/\bar{m})\to(W+W'+W''+1/\bar{m})$. The binomial is replaced by a multinomial coefficient

$$\binom{k+k'+k''}{k,k',k''}:=\frac{(k+k'+k'')!}{k!k'!k''!}, \quad (19)$$

and the fraction in the third line is changed by an additional term $(\bar{m}W'')^{k''}$ in the numerator, while the denominator changes to $(\bar{m}(W+W'+W'')+1)^{k+k'+k''+1}$. This pattern is kept for all further updates $m''',\ldots$ with corresponding weights $W''',\ldots\in[0,1]$. The result from Eq. (13) is recovered from Eq. (17) by setting $W=W'=1$.

In summary, this framework allows us to update the probability distribution for the intensity at the ith pixel using m_i and all other m_j, $j\neq i$, provided we have a model for the weights W_{ij} describing the assumed probability for $\lambda_i=\lambda_j$. By extending the first line of Eq. (17) to more than two pixels, we find

$$p(\lambda|\{m_j,W_{ij}\}_{j=1,\ldots,N},\bar{m})\sim\prod_{j\neq i}p(m_j|\lambda,W_{ij},\bar{m})p(\lambda|m_i,W_{ii},\bar{m}). \quad (20)$$

There is, of course, a multitude of options that can be used to define such weights. The following sections will present two very basic examples inspired by common averaging techniques.

C. Local Averaging Using Mixed Retrodiction

As defined above, the weights W_{ij} represent our *ad hoc* ansatz on the probability that the true intensities at the ith and jth pixels are equal. Without further information about the imaged object, one has to use general experience such as the assumption that typical images have smooth regions. In this case it is reasonable to assume that pixels in the same neighborhood share similar values.

Defining d_{ij} as the Euclidean distance between the ith and the jth pixel, we thus assume weights $W_{ij} = D_\sigma(i,j)$,

$$D_\sigma(i,j) = \begin{cases} 1 - (d_{ij}/\sigma)^2 & \text{for } d_{ij} \le \sigma \\ 0 & \text{else} \end{cases}, \tag{21}$$

with some width $\sigma > 0$ to update the retrodiction for the ith pixel ($W_{ii} = 1$). This means that the measurements m_j of the pixels surrounding the ith pixel are combined with weights W_{ij} such that the probability for λ_i follows a distribution $p(\lambda|\{m_j, W_{ij}\}_{j=1,\ldots,N}, \tilde{m})$ as outlined in Eq. (20).

The weights $W_{ij} = 1 - (d_{ij}/\sigma)^2$ can be seen as an approximation to a Gaussian, but with a well-defined boundary so as to reduce computational cost. Note that these weights are not normalized, as they do not describe the usual averaging procedures, but represent our assumption of the probability that the jth pixel has the same intensity value as the ith, i.e., that the measurement m_j is a suitable choice to update the probability for λ_i.

D. Nonlocal Averaging

Basic local averaging methods are easy to implement and optimal for smooth images, but they fail at the edges. This is why we present an adaptation of the nonlocal means algorithm to add to the mixed retrodiction scheme introduced above.

The nonlocal means denoising algorithm was introduced by Buades *et al.* in 2005 [12]. In this algorithm each measured pixel value is replaced by a weighted average of all other pixels in the image. Contrary to the local average mentioned above, however, the weights do not depend on the distance between the pixels, but rather on the average similarity between the local neighborhoods of the respective pixels. If, for instance, pixel i and its surrounding pixels are similar to pixel j and its surrounding pixels, then the jth pixel is weighted higher when retrodicting the ith pixel, and vice versa. For Gaussian noise, it has been shown that this approach reliably reduces noise while preserving structures as well [12,17].

In the original work [12], the similarity between two pixels is evaluated using the difference between the measured pixel values, but different measures targeted at Poisson noise have been proposed [18]. We use a more Bayesian approach by expressing the (dis-)similarity between two pixels with measurement values m_l and m_k using the single-pixel retrodicted intensity difference $\Delta_{lk} := \lambda_l - \lambda_k$. In Supplement 1 we derive the probability distribution for Δ_{lk} using intensity retrodiction from Eq. (7) and the corresponding mean value

$$E(\Delta|m_l, m_k, \tilde{m}) = \sum_{n=0}^{m_k} \binom{m_l + m_k - n}{m_l} \frac{n+1}{2^{m_l+m_k-n+1}\tilde{\eta}} - \sum_{n=0}^{m_l} \binom{m_l + m_k - n}{m_k} \frac{m_l - n + 1}{2^{m_l+m_k-n+1}\tilde{\eta}}, \tag{22}$$

where we again use $\tilde{\eta} = \eta(1 + 1/\tilde{m})$.

The average similarity between a neighborhood of width σ around the ith pixel and the respective region around the jth pixel is then defined as

$$\hat{\Delta}_{i,j} := \frac{1}{\sum_{k'=1}^{N} D_\sigma(i,k')} \sum_{k=1-i}^{N-i} E(\Delta|m_{i+k}, m_{j+k}, \tilde{m}) D_\sigma(i, i+k), \tag{23}$$

with a local averaging factor $D_\sigma(i, i+k)$ that is proportional to the Euclidean distance from the central pixel i as used in Eq. (21). The weight factor W_{ij} is then defined by the average similarity and a tolerance parameter t:

$$W_{ij} = \exp(-\hat{\Delta}_{i,j}^2/(2\,t^2)). \tag{24}$$

The weight W_{ij} is thus proportional to the average similarity of the respective neighborhoods at a distance σ from i and j as well as a tolerance t. As $E(\Delta|m_i, m_i, \tilde{m}) = 0$ it is ensured that $W_{i,i} = 1$.

Figure 5(b) shows the expectation values of a retrodiction using data from Fig. 2 and this approach inspired by the nonlocal means algorithm. For this example we chose parameters $\sigma = 2.5$ and $t = 0.5$. For computational reasons we restricted the search for similar neighborhoods to a 5 pixel × 5 pixel window centered at the ith pixel ($W_{ij} = 0$ for j outside this window). The same parameters are also used in Figs. 6–8.

The measure of similarity here is based on the distance, given in Eq. (22), which is calculated using a single-pixel retrodiction scheme. Additionally, one could consider an iterative scheme where the probabilities retrodicted using the weights from Eq. (24) are used to compute an updated expectation value for Δ. A different iterative Bayesian scheme for the nonlocal means algorithm has been shown to be successful for Gaussian noise [17].

4. COMPARING DIFFERENT RETRODICTION MODELS

To assess the quality of different retrodiction methods, we assume that we know the true intensity for a pixel λ_t and calculate the expected error for our retrodiction scheme, i.e.,

$$E_{\text{err}}(\lambda_t, \tilde{m}) := E(E(\lambda|m, \tilde{m}) - \lambda_t|\lambda_t) = \sum_{m=0}^{\infty} \left(\frac{m+1}{\tilde{\eta}} - \lambda_t\right) \text{Pois}(m; \eta\lambda_t) = \frac{1}{\eta(1 + 1/\tilde{m})} \left(1 - \frac{\eta\lambda_t}{\tilde{m}}\right). \tag{25}$$

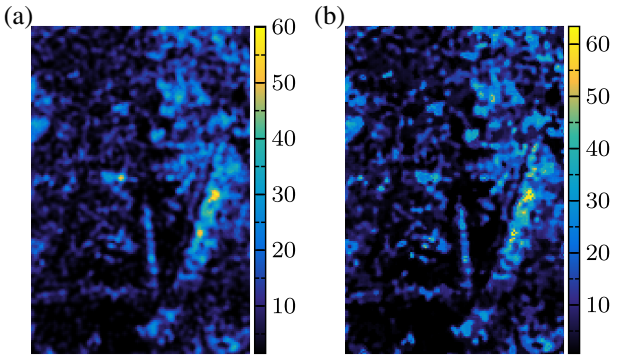

Fig. 5. Expectation values from mixed retrodiction $E(\lambda|\{m_j, W_{ij}\}_{j=1,\ldots,N}, \tilde{m})$ using the data from Fig. 2(b) and weights from (a) local averaging [cf. Eq. (21)] and (b) nonlocal averaging [Eq. (24)]. The chosen width $\sigma = 2.5$ in both cases, and tolerance $t = 0.5$ for the nonlocal-means-inspired retrodiction.

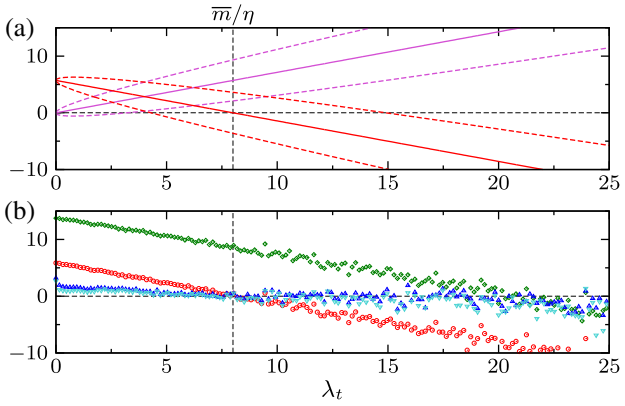

Fig. 6. (a) Expected error for intensity retrodiction using the mean value $E_{\mathrm{err}}(\lambda_t, \bar{m})$ [cf. Eq. (25)] (red solid line) or the maximum likelihood $E^L_{\mathrm{err}}(\lambda_t, \bar{m})$ given in Eq. (27) (magenta solid). The dashed lines give the expected margin of error $\pm\sqrt{V_{\mathrm{err}}(\lambda_t, \bar{m})}$ for both cases. Note that for $\bar{m} = \eta\lambda_t$ the expected error E_{err} crosses zero, though the variance keeps growing $\sim\lambda_t$. (b) Average experimental error obtained by comparing retrodicted images with the original from Fig. 2(a). Error from intensity retrodiction (red circles) behaves as expected; error from transmission retrodiction (green diamonds) shows that this method is optimal only for $\tau \sim 1/2$, i.e., $\lambda_t \sim \nu/2$ (see Fig. 3). The mixed retrodictions described in Section 3 (blue triangles, local weights; upside-down cyan triangles, nonlocal weights) give better average results; see also Fig. 5 and Table 1.

Thus, as we would hope, if our guess for $\bar{m}$ is correct and $\bar{m} \approx \eta\lambda_t$, we can expect that the replacement $m \to E(\lambda|m, \bar{m})$ gives the true intensity λ_t. The variance for the retrodiction error is

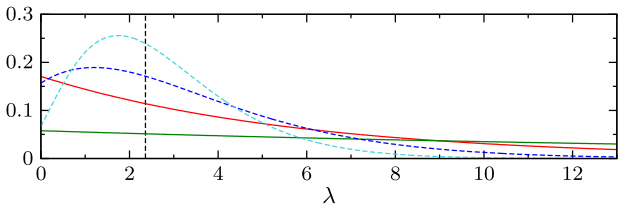

Fig. 7. Retrodicted probability $p(\lambda|\ldots)$ for a specific pixel from Fig. 2(b) with $m = 0$, $\bar{m} \approx 0.4$, $\eta = 0.05$, for different retrodiction models: single-pixel intensity retrodiction (red, expectation value $E \approx 5.9$), single-pixel transmission retrodiction (green, $E \approx 13.7$, $\nu = 40$), and the multipixel method as given in Section 3 with local averaging (blue dashed, $E \approx 3.4$) and nonlocal means (cyan dashed, $E \approx 2.7$). The true value $\lambda_t \approx 2.4$ is indicated by the black dashed line.

$$V_{\mathrm{err}}(\lambda_t, \bar{m}) := E((E(\lambda|m, \bar{m}) - \lambda_t)^2|\lambda_t) - E^2_{\mathrm{err}}(\lambda_t, \bar{m})$$
$$= \frac{\lambda_t}{\eta(1 + 1/\bar{m})^2}. \tag{26}$$

When we replace the measurements m_i by a single value to obtain a single example for the retrodiction, we could also use the maximum likelihood instead of the expectation value. The gamma distributions for intensity retrodiction have their maximum at $L(\lambda|m, \bar{m}) = m/\bar{\eta}$ such that the expected error for the replacement $m \to L(\lambda|m, \bar{m})$ is

$$E^L_{\mathrm{err}}(\lambda_t, \bar{m}) := E(L(\lambda|m, \bar{m}) - \lambda_t|\lambda_t)$$
$$= \sum_{m=0}^{\infty}\left(\frac{m}{\bar{\eta}} - \lambda_t\right)\mathrm{Pois}(m; \eta\lambda_t) = \lambda_t(1 - \eta/\bar{\eta}). \tag{27}$$

Therefore, if we use the most uninformed flat prior where $\eta = \bar{\eta}$ (i.e., $\bar{m} \to \infty$), we can expect that a retrodiction using maximum likelihood is, on average, correct. This is a surprising result that would suggest that all efforts to refine our prior were in vain. However, once we calculate the corresponding variance, we find that $V^L_{\mathrm{err}}(\lambda_t, \bar{m}) = V_{\mathrm{err}}(\lambda_t, \bar{m})$, which grows $\sim\lambda_t/\eta$ for the flat prior $\eta = \bar{\eta}$.

We thus conclude that if we have a reasonably good prior for intensity retrodiction, where $\bar{m} \approx \eta\lambda_t$, then an image where $m \to E(\lambda|m, \bar{m})$ can be expected to be correct, especially for a small $\bar{m}$. Using the maximum likelihood and an uninformed prior will be correct on average, yet terribly wrong most of the time.

As noted at the end of Section 2.A, one can use (Poisson) image denoising algorithms to calculate an optimal local prior $\bar{m}_i$ at each pixel. This will obviously enhance the reliability of the retrodiction, as $\eta\lambda_{t,i} \approx \bar{m}_i$ will be true for a larger set of pixels, $i = 1, \ldots, N$.

The situation is similar if we use transmission retrodiction, as introduced in Section 2.B. There we used a flat prior for the transmission τ, and the expected error for retrodiction using the mean value became zero at $\tau \sim 1/2$ regardless of other choices.

Figure 6 shows the distribution of error if the expectation values from retrodictions as given in Figs. 3 and 5 are subtracted from the original given in Fig. 2(a).

To quantify the difference between the retrodicted mean value for various retrodiction methods and the true value λ_t, we use the peak signal-to-noise ratio (PSNR) [7],

$$\mathrm{PSNR}(\lambda_t, x) = 10\log_{10}\left(\frac{\max_i(\lambda_{t,i})^2}{\sum_{i=1}^N(\lambda_{t,i} - x_i)^2/N}\right), \tag{28}$$

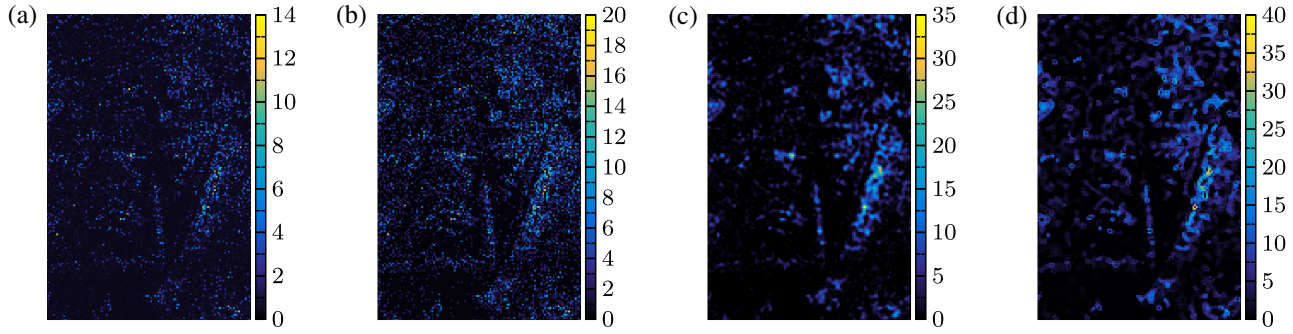

Fig. 8. Given the measurement values from Fig. 2, one can expect that the local intensities are larger than the values for (a) single-pixel intensity retrodiction, (b) transmission retrodiction, and mixed retrodiction inspired by (c) local or (d) nonlocal averaging. The chosen evidence level is $\mathrm{ev}(\lambda_i \geq \xi) = 20$ [cf. Eq. (30)]; other parameters are as in Figs. 3 and 5.

where x has to be replaced by the respective mean value, e.g., $x_i = E(\lambda|m_i, \bar{m})$, for single-pixel intensity retrodiction. The results for the examples illustrated in Figs. 3 and 5 are given in Table 1, where we see that the local averaging model as described in Section 3.C performs best. This would also be expected from Fig. 7(b).

An example of the retrodicted probabilities for a pixel from the example in Fig. 2 is shown in Fig. 7. There we see clearly how the mixed retrodiction models produce narrower probability distributions. This is due to the Bayesian updates using neighboring pixels, as outlined in Eq. (20). Keep in mind that this individual example represents only a single pixel; the average performance is shown in Fig. 6.

5. HYPOTHESIS TESTING

Simply replacing measured photocounts with the expectation values of some retrodicted probabilities, as we did in Figs. 3 and 5, does not take advantage of the full potential of retrodiction. If we condense the probabilistic information we have about the intensities to a single (possibly denoised) image, then we conceal the uncertainty behind that image.

Consider a measurement showing a small bright spot. In order to determine whether this bright spot shows the location of, say, a fluorescent molecule or is simply noise, we need to ask questions such as how likely it is that the intensity in this region is larger than a certain value. More importantly, we can easily infer the reliability of a statement by means of Bayesian hypothesis testing.

For example, using single-pixel intensity retrodiction, as given in Eq. (7), the probability that the intensity at the ith pixel is less than a certain value $\xi > 0$ is

$$\Pr(\lambda_i < \xi|m_i, \bar{m}) = \int_0^\xi p(\lambda|m_i, \bar{m})\mathrm{d}\lambda = \Gamma\frac{(m_i + 1; \bar{\eta}\xi)}{(m + 1)!}. \quad (29)$$

The probability of the opposite situation is $\Pr(\lambda_i \geq \xi|m_i, \bar{m}) = 1 - \Pr(\lambda_i < \xi|m_i, \bar{m})$. The evidence [19] for a statement "$\lambda_i \geq \xi$" reading

$$\mathrm{ev}(\lambda_i \geq \xi|m_i, \bar{m}) = 10\log_{10}\frac{\Pr(\lambda_i \geq \xi|m_i, \bar{m})}{\Pr(\lambda_i < \xi|m_i, \bar{m})} \quad (30)$$

is then a reliable measure of whether the brightness of a pixel exceeds a certain level ξ. Similarly, one may ask whether the brightness lies within a certain interval, $\lambda_i \in [\xi, \xi']$, or test hypotheses regarding correlations between different pixels. Of course, similar expressions can also be derived for the transmission retrodiction scheme described in Section 2.B or the retrodiction schemes from Section 3.

The ability to ask and reliably test such questions shows the benefit of full Bayesian image retrodiction in addition to more common image denoising. Advanced denoising algorithms certainly give more pleasant images and help the observer to identify previously hidden features. But in the end only proper hypothesis testing can provide the level of confidence needed for qualified scientific statements.

Table 1. Peak Signal-to-Noise Ratio for Different Retrodiction Models as Compared to the True Image Given in Fig. 2(a)

Model	PSNR(λ_t,·) (dB)
Single-pixel inten. retrod., Fig. 3(a)	16.5
Single-pixel transm. retrod., Fig. 3(b)	11.7
Mixed retrod., local average, Fig. 5(a)	18.7
Mixed retrod., nonlocal average, Fig. 5(b)	17.5

Figure 8 shows a possible application of this hypothesis testing. Here every measurement m_i from the raw data given in Fig. 3 is replaced with a value $\xi_{>20}$, which is defined such that $\mathrm{ev}(\lambda_i \geq \xi_{>20}|m_i, \bar{m}) = 20$. This means that the probability that $\lambda_i < \xi_{>20}$ is 100 times smaller than the probability that $\lambda_i \geq \xi_{>20}$. Similarly, we can define $\xi_{<20}$ as the value for which we can be very certain that $\lambda_i \leq \xi_{<20}$.

A comparison with the real image from Fig. 2(a) shows that the claim "$\lambda_i > \xi_{>20}$" is true 98.2% for single-pixel intensity retrodiction, 96.7% for single-pixel transmission retrodiction, 98.4% for multipixel retrodiction using local averaging, and 97.7% for multipixel retrodiction using nonlocal means. Similar ratios hold for the claim "$\lambda_i < \xi_{<20}$" (99.6%, 99.99%, 98.9%, and 96.3%, respectively). The poor performance of the local and, in particular, the nonlocal algorithm is a result of the narrow probability distributions obtained after multiple updates, which give a false sense of confidence. This can be tackled, for instance, by reducing the value of the weights W_{ij}.

Further information can be obtained if, for instance, questions on the probability that the average of several neighboring pixels' intensity values lies in a certain region are asked.

6. CONCLUSIONS

In this work we provided some basic ideas to calculate the Bayesian probability distributions for the intensity at each pixel of an image obtained using photodetector arrays. These images are usually very distorted by Poisson noise, which is why we argue that a full probabilistic treatment is an essential extension to the host of algorithms focusing on image denoising.

The single-pixel intensity retrodiction scheme presented in Section 2 can be easily applied to the usual measurement data. In particular, the possibility of estimating local mean values $\bar{m}_i$ using other denoising techniques might significantly enhance the quality of this method. The mixed retrodiction schemes from Section 3 try to include the measurements of other pixels and some *ad hoc* correlations across the image. As illustrated in Section 4, all these retrodiction methods give slightly different estimates, as one would expect for Bayesian inference using different priors.

The difference between image denoising and image retrodiction as described here can be illustrated by the following analogy: image denoising algorithms search for the optimal compromise between the noisy data and general considerations about the image so they can be seen as a fit to two-dimensional experimental data. As image retrodiction provides the full probability distribution for the intensity at each pixel, it plays a role analogous to error bars. A fitting curve is certainly more pleasant and often sufficient for high-quality data, yet every scientist is aware that it is essential to provide information about the margin of error as well, especially for noisy data.

We thus think that image retrodiction and hypothesis testing can serve as an important supplement to image-analysis-based experiments and technologies in physics, astronomy, bio-chemistry, and medicine [1–4,20]. Especially fast processes that cannot be easily reproduced are likely to be accompanied by severe noise.

Conclusions drawn from image analysis using (denoised) data can be quantified using the techniques provided in this work.

Funding. Engineering and Physical Sciences Research Council (EPSRC) (EP/I012451/1, EP/M01326X/1).

Acknowledgment. We thank Miles J. Padgett for drawing our attention to the problem; Reuben Aspden, Peter Morris, and Lena Mertens for data from their ghost imaging setup and discussions on their image denoising technique; Adrain Bowman and Ludger Evers for an encouraging discussion; Vašek Potoček for disturbing questions; and many more colleagues from the QuantIC network for stimulating ideas and interest.

See Supplement 1 for supporting content.

REFERENCES

1. R. Molina, J. Núñez, F. J. Cortijo, and J. Mateos, "Image restoration in astronomy: a Bayesian perspective," IEEE Signal Process. Mag. **18**(2), 11–29 (2001).
2. M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini, "Image deblurring with Poisson data: from cells to galaxies," Inverse Probl. **25**, 123006 (2009).
3. A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. Wong, J. H. Shapiro, and V. K. Goyal, "First-photon imaging," Science **343**, 58–61 (2014).
4. P. A. Morris, R. S. Aspden, J. E. Bell, R. W. Boyd, and M. J. Padgett, "Imaging with a small number of photons," Nat. Commun. **6**, 5913 (2015).
5. T. Le, R. Chartrand, and T. J. Asaki, "A variational approach to reconstructing images corrupted by Poisson noise," J. Math. Imaging Vis. **27**, 257–263 (2007).
6. S. Lefkimmiatis, P. Maragos, and G. Papandreou, "Bayesian inference on multiscale models for Poisson intensity estimation: applications to photon-limited image denoising," IEEE Trans. Image Process. **18**, 1724–1741 (2009).
7. D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3D and reflectivity imaging with single-photon detectors," arXiv:1406.1761 (2014).
8. S. Watanabe, "Symmetry of physical laws. Part III. Prediction and retrodiction," Rev. Mod. Phys. **27**, 179–186 (1955).
9. Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, "Time symmetry in the quantum process of measurement," Phys. Rev. **134**, B1410 (1964).
10. D. T. Pegg and S. M. Barnett, "Retrodiction in quantum optics," J. Opt. B **1**, 442–445 (1999).
11. S. M. Barnett, L. S. Phillips, and D. T. Pegg, "Imperfect photodetection as projection onto mixed states," Opt. Commun. **158**, 45–49 (1998).
12. A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," Multiscale Model. Simul. **4**, 490–530 (2005).
13. M. Sonnleitner, "Statue of Lord Kelvin, Kelvingrove Park, Glasgow" (2015).
14. O. Jedrkiewicz, R. Loudon, and J. Jeffers, "Retrodiction for optical attenuators, amplifiers, and detectors," Phys. Rev. A **70**, 033805 (2004).
15. W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, 1973), Vol. **2**.
16. Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform based analysis of depth images constructed using sparse single-photon data," arXiv:1507.02511 (2015).
17. C. Kervrann, J. Boulanger, and P. Coupé, "Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal," in *Scale Space and Variational Methods in Computer Vision* (Springer, 2007), pp. 520–532.
18. C.-A. Deledalle, F. Tupin, and L. Denis, "Poisson NL means: unsupervised non local means for Poisson noise," in *17th IEEE International Conference on Image Processing* (IEEE, 2010), pp. 801–804.
19. E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University, 2003).
20. L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imaging **1**, 113–122 (1982).