

Local retrodiction models for photon-noise-limited images

Matthias Sonnleitner^a, John Jeffers^b, and Stephen M. Barnett^a

^aSchool of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom

^bSUPA, Department of Physics, University of Strathclyde, Glasgow G4 0NG, United Kingdom

ABSTRACT

Imaging technologies working at very low light levels acquire data by attempting to count 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 and a host of successful algorithms trying to reconstruct the original image from this noisy data have been developed. Here we review a recently proposed scheme that complements these algorithms by calculating the full probability distribution for the local intensity distribution behind the noisy photocount measurements. Such a probabilistic treatment opens the way to hypothesis testing and confidence levels for conclusions drawn from image analysis.

Keywords: Poisson noise, image analysis, image optimisation, Bayesian inference, photon-limited imaging, image retrodiction

1. INTRODUCTION

Several modern imaging technologies work at the lowest possible light level in that they image an object by detecting individual photocounts at each pixel.¹⁻⁴ The resulting images are corrupted by Poisson noise such that even after processing with advanced denoising algorithms one might not obtain a clear distinction between real features of an object and a random cluster of noise, especially for an average of less than one photocount per pixel. To assess the reliability of conclusions drawn from image analysis, one therefore needs to define confidence levels, which in turn require a probabilistic treatment of the noisy image data.

In a recent work⁵ we presented a Bayesian model inspired by quantum state retrodiction⁶⁻⁹ to produce a probability distribution for the original light intensity, given the measured number of photocounts at each pixel. Such a probabilistic approach then gives not only the most likely value for the intensity at each pixel, but allows one to derive uncertainty levels or test certain hypotheses on the image. This retrodictive approach therefore opens the way to a more cautious and quantitative analysis of measured images, especially in situations with high noise levels.

Here we will start by reviewing the ideas behind image retrodiction in more detail. We start with the simple approach where only data from a single pixel is used, then go on to derive a more involved analysis where also the information from neighbouring pixels is taken into account in section 3. In section 4 we discuss the performance of these models in a more quantitative way by comparing the results of image retrodiction with the known original image shown in figure 1.

2. SINGLE-PIXEL RETRODICTION

Let us assume that an object is illuminated by an attenuated single-mode light source such as a weak laser beam; alternatively, the object itself could be weakly fluorescent. Then the number of photons n coming from the object towards the detector follows a Poissonian distribution, $p(n|\lambda) = \text{Pois}(n; \lambda)$ with

$$\text{Pois}(n; \lambda) := \frac{e^{-\lambda} \lambda^n}{n!}. \quad (1)$$

Corresponding author: matthias.sonnleitner@glasgow.ac.uk

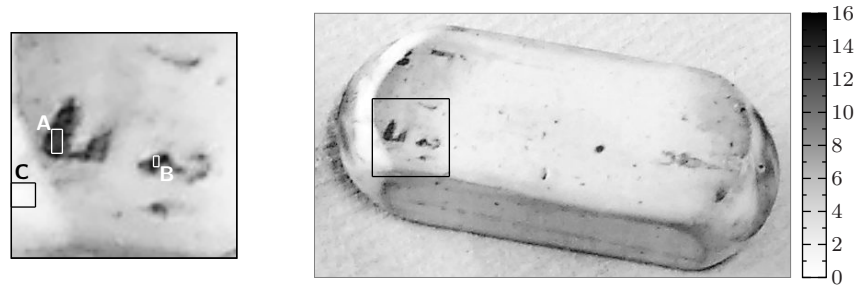


Figure 1. Images used for retrodiction in this work. Right: original image of an eraser used for the examples in figures 2, 3 and 5 ($N = 520 \times 289$ pixels). Left: detail image ($N = 85 \times 85$ pixels) used for quantitative measurements in section 4. The areas in the boxes are selected regions of relatively high (A and B) and low intensity (C) used for the example in section 4.2. Note that the intensity map is inverted here so dark regions correspond to high intensities.

The value λ is proportional to the intensity of the light field emitted by the object towards the detector within the detection time. If we have an array of photodetectors then each of them will image a different region of the object, corresponding to a different value of λ . The set $\hat{\lambda}_1, \dots, \hat{\lambda}_N$ thus constitutes the true image for N pixels.

However, each detector only has a finite detection efficiency $\eta \in (0, 1]$ such that the number of photocounts actually is given by

$$p(m|\lambda) = \text{Pois}(m; \eta\lambda). \quad (2)$$

An example for an artificial measurement of the original image from figure 1 using this Poisson process is given in figure 2. Note that any realistic detector will also suffer from external noise which results in a certain level of dark counts.⁹ These are not included in this work, but are discussed in the additional material of a previous work.⁵

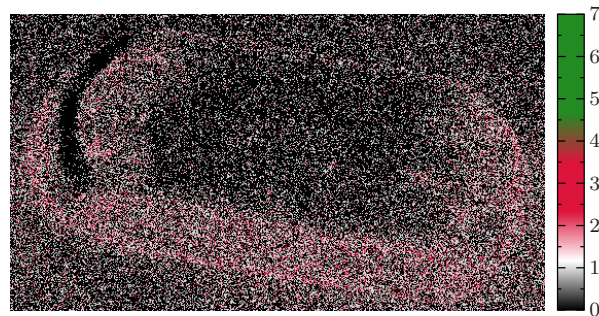


Figure 2. “Measured” image obtained from the original in figure 1 using artificial Poisson noise corresponding to a detection efficiency $\eta = 0.1$, see also equation (2). Note that most of the pixels have $m = 0$ such that the global average number of photocounts is $\bar{m} \simeq 0.5$.

As we know the measurement value $m \geq 0$ we seek the probability distribution for the possible intensities λ , given a measurement m . This can be obtained using Bayes’ theorem

$$p(\lambda|m) = \frac{p(m|\lambda)p(\lambda)}{p(m)}, \quad (3)$$

with $p(m|\lambda)$ as given above and a normalisation factor $p(m) = \int_0^\infty p(m|\lambda)p(\lambda)d\lambda$. Choice of a correct prior probability $p(\lambda)$ is a crucial step in Bayesian probability theory.¹⁰

As discussed in an earlier work⁵ useful prior information can be obtained from a first estimate of the average detected photon number \bar{m} . This can be either found by a global average across the whole image, $\bar{m} = \sum_{i=1}^N m_i/N$, where m_i is the number of photocounts measured at the i th of N pixels; or one derives a local prior using, for instance, Gaussian weights to find an estimate \bar{m}_i for the average photon number in the

region of the i th pixel, see also equation (10). If we have such an estimate of the average \bar{m} , the least biased discrete probability distribution is $p(m|\bar{m}) = \text{Geom}(m; \bar{m})$,¹¹ with the geometric distribution*

$$\text{Geom}(m; x) = \frac{x^m}{(x + 1)^{m+1}}. \quad (4)$$

By using $p(m|\bar{m}) = \int_0^\infty p(m|\lambda)p(\lambda|\bar{m})d\lambda$ we then find that this requires an exponential prior for the intensities $p(\lambda|\bar{m}) = \frac{\eta}{\bar{m}}e^{-\lambda\eta/\bar{m}}$ such that

$$p(\lambda|m, \bar{m}) = \lambda^m \frac{(\eta(1 + 1/\bar{m}))^{m+1}}{m!} e^{-\lambda\eta(1+1/\bar{m})}. \quad (5)$$

The retrodicted probability distribution for the intensity λ given a measurement m and some estimate \bar{m} thus follows a gamma distribution, $p(\lambda|m, \bar{m}) = \text{Gam}(\lambda; m + 1, 1/\bar{\eta})$ with $\bar{\eta} = \eta(1 + 1/\bar{m})$ and

$$\text{Gam}(x; \alpha, \beta) := \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}. \quad (6)$$

Note that setting $\bar{m} \rightarrow 0$ gives the result for a flat prior $p(\lambda)$ which is constant up to some arbitrarily large value of λ . Such a flat prior thus gives reasonable results if we expect a high number of photocounts, but is a poor estimate for the situations discussed here where we usually have $\bar{m} < 1$.

As $p(\lambda|m, \bar{m})$ follows a gamma distribution we immediately find the expectation value

$$E(\lambda|m, \bar{m}) = \frac{m + 1}{\eta(1 + 1/\bar{m})}, \quad (7)$$

while the cumulative distribution function, which contains all the information about the intensity distribution given the counts at each pixel, $P(\lambda|m, \bar{m}) := \int_0^\lambda p(\lambda'|m, \bar{m})d\lambda'$ reads

$$P(\lambda|m, \bar{m}) = \gamma(m + 1, \lambda\eta(1 + 1/\bar{m})), \quad (8)$$

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

$$\gamma(n, x) := \frac{1}{\Gamma(n)} \int_0^x e^{-t} t^{n-1} dt. \quad (9)$$

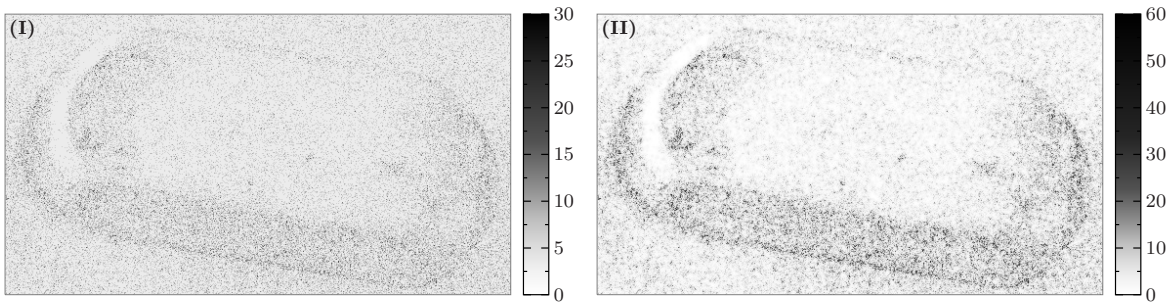


Figure 3. Expectation values of single pixel retrodiction as given by equation (7) using measurements from figure 2, i. e. $\eta = 0.1$. The left image used a global value for $\bar{m} \simeq 0.5$ while the right image used a local version for \bar{m}_i as given by equation (10) with $\sigma = 1.3$. These models are labelled as I and II in table 1. Although the intensity map had to be rescaled to account for some outliers, the right hand side image appears to be closer to the original from figure 1.

*Note that $\lim_{x \rightarrow 0} \text{Geom}(m; x) = 1$ for $m = 0$ and $\lim_{x \rightarrow 0} \text{Geom}(m; x) = 0$ else.

Two examples for the expectation values $E(\lambda|m, \bar{m})$ obtained from an artificial “measurement” given in figure 2 are displayed in figure 3. In the first panel we used a global estimate $\bar{m} = \sum_{i=1}^N m_i/N$ and in the second a local version where the prior is slightly different at each pixel,

$$\bar{m}_i = \frac{\sum_{j=1}^N m_j \exp[-d_{ij}^2/(2\sigma^2)]}{\sum_{j=1}^N \exp[-d_{ij}^2/(2\sigma^2)]} \quad (10)$$

with d_{ij} being the Euclidean distance between the i th and the j th pixel. As discussed in more detail in section 3.1 such a local prior \bar{m}_i is probably the most obvious way to improve the quality of the retrodiction.

3. MIXED RETRODICTION

Single pixel retrodiction as introduced above is valid for any type of photodetection setup and thus does not make use of the additional information that we are trying to retrodict an image. The essential feature of an interesting image is that the individual pixels are not uncorrelated and in most cases one can expect that pixels within a certain neighbourhood have the same or similar true values $\hat{\lambda}$. This is the key to most image denoising algorithms.

3.1 Retrodiction from a constant source

First, let us assume we know that all pixels within a certain neighbourhood $\mathcal{N} = \{m_1, m_2, \dots, m_N\}$ should ideally show the same true intensity value, i.e. $\hat{\lambda}_1 = \dots = \hat{\lambda}_N = \hat{\lambda}$. The inherently random nature of the measurement will still result in varying integer photocounts across this neighbourhood, but we know that the measurement at each pixel is just one out of many individual and independent experiments with the same mean value $\hat{\lambda}$.

If we start with $m_1 \in \mathcal{N}$ and some guess for \bar{m} we will find $p(\lambda|m_1, \bar{m}) = \text{Gam}(\lambda; m_1 + 1, \bar{\eta})$ as described in equation (5). But as we know that measurements m_2, \dots, m_N come from a source with the same intensity we can use a Bayesian update, $p(\lambda|m_1, m_2, \bar{m}) \sim p(m_2|\lambda)p(\lambda|m_1, \bar{m})$, such that

$$p(\lambda|m_1, m_2, \bar{m}) = \text{Gam}(\lambda; m_1 + m_2 + 1, 1/[\eta(2 + 1/\bar{m})]), \quad (11)$$

where we used $p(m_2|\lambda) = \text{Pois}(m_2, \eta\lambda)$ and

$$\text{Pois}(k; \alpha\lambda) \text{Gam}(x; l + 1, 1/\beta) = \binom{k+l}{l} \frac{\alpha^k \beta^{l+1}}{(\alpha + \beta)^{k+l+1}} \text{Gam}(x; k + l + 1, 1/(\alpha + \beta)). \quad (12)$$

Knowing that all N measurements in this neighbourhood are independent we find

$$p(\lambda|\mathcal{N}, \bar{m}) \sim \prod_{j=2}^N p(m_j|\lambda)p(\lambda|m_1, \bar{m}), \quad (13)$$

$$p(\lambda|\mathcal{N}, \bar{m}) = \text{Gam}(\lambda; m_1 + m_2 + \dots + m_N + 1, 1/[\eta(N + 1/\bar{m})]), \quad (14)$$

where $p(\lambda|\mathcal{N}, \bar{m}) \equiv p(\lambda|m_1, m_2, \dots, m_N, \bar{m})$. As this is again a Gamma distribution we know its expectation value,

$$E(\lambda|\mathcal{N}, m_N, \bar{m}) = \frac{1}{\eta(N + 1/\bar{m})} \left(\sum_{j=1}^N m_j + 1 \right). \quad (15)$$

Obviously we would expect that having N independent experiments instead of just one will give a much better estimate on the true intensity $\hat{\lambda}$. To quantify just how much better N experiments are, we calculate the expected difference between the true $\hat{\lambda}$ and $E(\lambda|\mathcal{N}, \bar{m})$, i.e.

$$E(E(\lambda|\mathcal{N}, \bar{m}) - \hat{\lambda}) = \sum_{m_1=0}^{\infty} \dots \sum_{m_N=0}^{\infty} p(m_1, m_2, \dots, m_N|\hat{\lambda}) (E(\lambda|\mathcal{N}, \bar{m}) - \hat{\lambda}), \quad (16)$$

where $p(m_1, m_2, \dots, m_N | \hat{\lambda}) = \prod_{i=1}^N \text{Pois}(m_i | \eta \hat{\lambda})$. For the example with $N = 2$ we find

$$\begin{aligned} & \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \left[\text{Pois}(m_1 | \eta \hat{\lambda}) \text{Pois}(m_2 | \eta \hat{\lambda}) \left(\frac{m_1 + m_2 + 1}{\eta(2 + 1/\bar{m})} - \hat{\lambda} \right) \right] = \\ & = \frac{1}{\eta(2 + 1/\bar{m})} \left[\sum_{m_1=0}^{\infty} m_1 \text{Pois}(m_1 | \eta \hat{\lambda}) \sum_{m_2=0}^{\infty} \text{Pois}(m_2 | \eta \hat{\lambda}) + \sum_{m_2=0}^{\infty} \text{Pois}(m_2 | \eta \hat{\lambda}) \sum_{m_1=0}^{\infty} m_1 \text{Pois}(m_1 | \eta \hat{\lambda}) \right. \\ & \quad \left. + \left(1 - \eta \hat{\lambda} (2 + 1/\bar{m}) \right) \sum_{m_1=0}^{\infty} \text{Pois}(m_1 | \eta \hat{\lambda}) \sum_{m_2=0}^{\infty} \text{Pois}(m_2 | \eta \hat{\lambda}) \right], \quad (17) \end{aligned}$$

such that $E(E(\lambda | m_1, m_2, \bar{m}) - \hat{\lambda}) = (1 - \eta \hat{\lambda} / \bar{m}) / (\eta(2 + 1/\bar{m}))$. Here where we only used that $\sum_{m=0}^{\infty} \text{Pois}(m | x) = 1$ and $\sum_{m=0}^{\infty} m \text{Pois}(m | x) = x$. The procedure for N elements is exactly the same and we find

$$E(E(\lambda | \mathcal{N}, \bar{m}) - \hat{\lambda}) = \frac{1 - \eta \hat{\lambda} / \bar{m}}{\eta(N + 1/\bar{m})}. \quad (18)$$

This shows that the expected error of a retrodiction using N pixels (from the same source) is reduced by a factor $\sim 1/N$. It also shows the importance of having $\bar{m} \sim \eta \hat{\lambda}$, which is a good reason to work with a local prior \bar{m}_i such as, for example, the one given in equation (10).

By using the fact that $\max_{\lambda} p(\lambda | \mathcal{N}, \bar{m}) = \sum_{j=1}^N m_j / (\eta(N + 1/\bar{m}))$ is the maximum of the Gamma distribution we find that the expected difference between the maximum likelihood and the true value is

$$E(\max_{\lambda} p(\lambda | \mathcal{N}, \bar{m}) - \hat{\lambda}) = -\frac{\eta \hat{\lambda} / \bar{m}}{\eta(N + 1/\bar{m})}. \quad (19)$$

One might therefore be tempted to set $\bar{m} \rightarrow \infty$ as this will obviously ensure that the maximum of the probability distribution is close to the true value $\hat{\lambda}$, even for a single measurement $\hat{\lambda}$. Calculation of the variance however shows that

$$\text{Var}(\max_{\lambda} p(\lambda | \mathcal{N}, \bar{m}) - \hat{\lambda}) = \text{Var}(E(\lambda | \mathcal{N}, \bar{m}) - \hat{\lambda}) = \frac{N \hat{\lambda}}{\eta(N + 1/\bar{m})^2}, \quad (20)$$

and we see that especially $\bar{m} \ll 1$ helps to reduce the variance. A similar result for single-pixel retrodiction has been shown in an earlier work.⁵

For this derivation we used the assumption that we know that all measurements $m_i \in \mathcal{N}$ come from a source with the same true intensity $\hat{\lambda}$. This might be the case in some over-sampled setups where, for instance, the point spread function of an imaged object covers several pixels. But in a more general situation this assumption of a constant true intensity has to be weakened.

Figure 4 shows the convergence of mixed retrodiction towards a true value $\hat{\lambda}$ for a constant image with increasing number of measurements N . One sees that starting from a prior with $\bar{m} \approx \eta \hat{\lambda}$ helps to reduce the difference between $E(\lambda | \mathcal{N}, \bar{m})$ and $\hat{\lambda}$, especially if one doesn't know that all measurements come from a source of equal intensity. One also sees that the variation remains high.

3.2 Mixed retrodiction with weights

In the example above we *knew* that the measurements at two different pixels, m_i and m_j , came from a source of equal intensity. Without this knowledge we might still *assume* that this is the case, for instance because m_i and m_j are next to each other and we expect that our image contains several constant regions. This assumption is quantified by introducing a weight $W_{ij} \in [0, 1]$ which reflects the assumed probability for $\hat{\lambda}_i = \hat{\lambda}_j$.

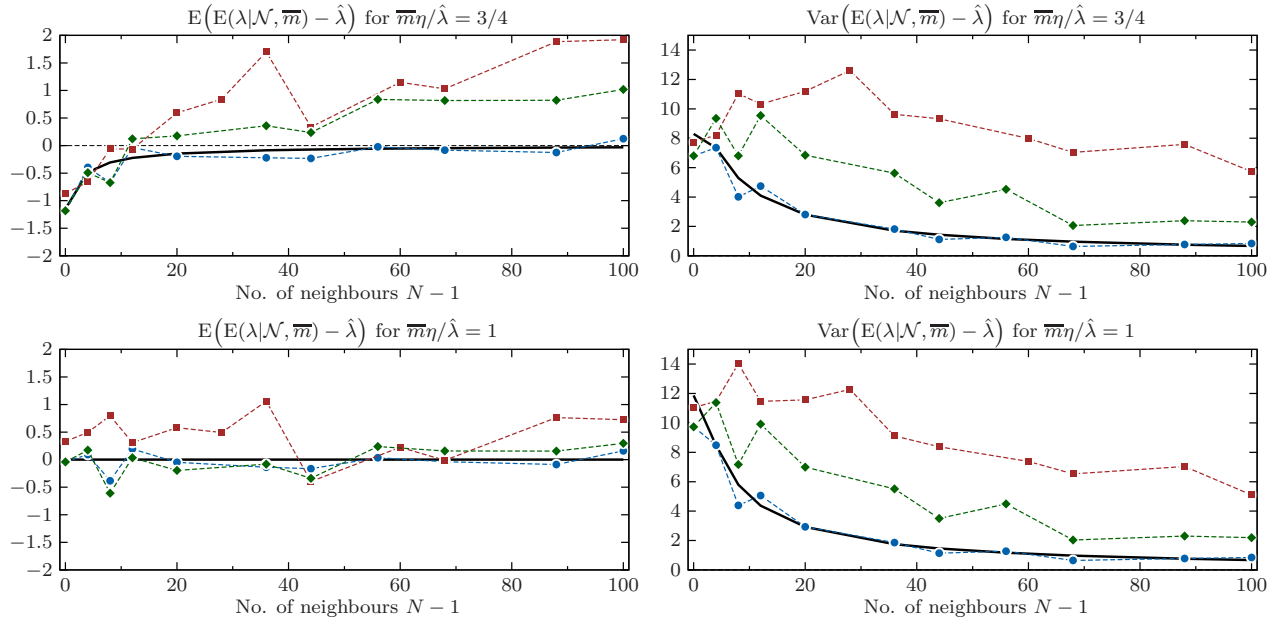


Figure 4. Difference between a true image value $\hat{\lambda}$ and the expectation values obtained from mixed retrodiction using different models and an increasing number of measurements $\mathcal{N} = \{m_1, \dots, m_N\}$. The left column shows the expected difference $E(E(\lambda|\mathcal{N}, \bar{m}) - \hat{\lambda})$ for different values of \bar{m} , the right column shows the corresponding variances. The theoretical curves from equations (18) and (20) are given in black.

The blue dots show the results for mixed retrodiction when we know that all data points come from the same source as discussed in section 3.1 and agree well with the theoretical curves. The red squares [green diamonds] are for mixed retrodiction using Gaussian [polynomial] weights as described in table 1 (models III and IV, but with increasing width σ). Each data point is the average of $E(E(\lambda|\mathcal{N}, \bar{m}) - \hat{\lambda})$ and $\text{Var}(E(\lambda|\mathcal{N}, \bar{m}) - \hat{\lambda})$ after 100 simulations with different sets of measurements m_1, \dots, m_N . Here we chose $\hat{\lambda} = 7$ and $\eta = 0.1$.

Generally, if we think a measurement m could have come from a source of intensity λ or a different source of intensity λ' we can write $p(m|\lambda, \lambda', W) = \text{Pois}(m; \eta W \lambda + \eta(1 - W)\lambda')$ such that

$$\begin{aligned}
 p(m|\lambda, W, \bar{m}) &= \int_0^\infty d\lambda' p(m|\lambda, \lambda', W) p(\lambda'|\bar{m}) \\
 &= \sum_{k=0}^m \text{Geom}(m - k; (1 - W)\bar{m}) \text{Pois}(k; \eta W \lambda),
 \end{aligned} \tag{21}$$

where we used an exponential prior $p(\lambda'|\bar{m}) = \frac{\eta}{\bar{m}} e^{-\eta\lambda'/\bar{m}}$ and the geometric distribution as defined in equation (4). Hence an update on the probability of λ using a measurement m_j and a weight W_{ij} reads (up to some normalisation factor)

$$\begin{aligned}
 p(\lambda|m_i, m_j, W_{ij}, \bar{m}) &\sim p(m_j|\lambda, W_{ij}, \bar{m}) p(\lambda|m_i, \bar{m}) \\
 &\sim \sum_{k=0}^{m_j} \text{Gam}(\lambda; m_i + k + 1, 1/[\eta(1 + W_{ij} + 1/\bar{m})]) \times \\
 &\quad \binom{m_i + k}{k} \frac{(W_{ij})^k}{(1 + W_{ij} + 1/\bar{m})^k} \text{Geom}(m_j - k; (1 - W_{ij})\bar{m}).
 \end{aligned} \tag{22}$$

Here we used $p(\lambda|m_i, \bar{m})$ from single-pixel retrodiction and relation (12).

We see that this updated probability distribution $p(\lambda|m_i, m_j, W_{ij}, \bar{m})$ is a sum of $m_j + 1$ gamma distributions with some coefficients and a normalisation, which is simply the sum of these coefficients. To use more data points

we repeat this procedure using equations (21) and (12) to find

$$\begin{aligned}
 p(\lambda|m_i, m_j, m_{j'}, W_{ij}, W_{ij'}, \bar{m}) &\sim p(m_{j'}|\lambda, W_{ij'}, \bar{m})p(\lambda|m_i, m_j, W_{ij}, \bar{m}) \\
 &\sim \sum_{k=0}^{m_j} \sum_{k'=0}^{m_{j'}} \text{Gam}(\lambda; m_i + k + k' + 1, 1/[\eta(1 + W_{ij} + W_{ij'} + 1/\bar{m})]) \frac{(W_{ij})^k (W_{ij'})^{k'}}{(1 + W_{ij} + W_{ij'} + 1/\bar{m})^{k+k'}} \times \\
 &\quad \binom{m_i + k + k'}{m_i, k, k'} \text{Geom}(m_j - k; (1 - W_{ij})\bar{m}) \text{Geom}(m_{j'} - k'; (1 - W_{ij'})\bar{m}),
 \end{aligned} \tag{23}$$

where $\binom{m_i+k+k'}{m_i, k, k'} := (m_i + k + k')! / (m_i! k! k')$ is a multinomial coefficient. Comparing this to what we found for $p(\lambda|m_i, m_j, W_{ij}, \bar{m})$, we identify a pattern which allows one to calculate the mixed retrodiction using all other pixels

$$p(\lambda|m_i, \{m_j, W_{ij}\}_{j \neq i}, \bar{m}) \sim \prod_{j \neq i} p(m_j|\lambda, W_{ij}, \bar{m})p(\lambda|m_i, \bar{m}). \tag{24}$$

As mentioned above, $p(\lambda|m_i, \{m_j, W_{ij}\}_{j \neq i}, \bar{m})$ is a sum of gamma distributions with coefficients given by the measurement values and the set of weights W_{ij} . Once these coefficients are computed, it is straightforward to calculate expectation values $E(\lambda|m_i, \{m_j, W_{ij}\}_{j \neq i}, \bar{m})$ or cumulative distribution functions $P(\lambda|m_i, \{m_j, W_{ij}\}_{j \neq i}, \bar{m})$ using equations (7) or (8), respectively. Thus, the mixed retrodiction framework allows one to calculate probability distributions for the intensity at a pixel with measurement m_i , using other measurements $m_j, j \neq i$, even if one cannot be sure that the true intensities are the same $\hat{\lambda}$, as we assumed in section 3.1.

Of course, the quality of such a mixed retrodiction crucially depends on the choice of weights W_{ij} . But finding a suitable model for these weights is the same type of problem as discussed when designing image denoising algorithms using local or non-local averaging methods of varying complexity.¹²

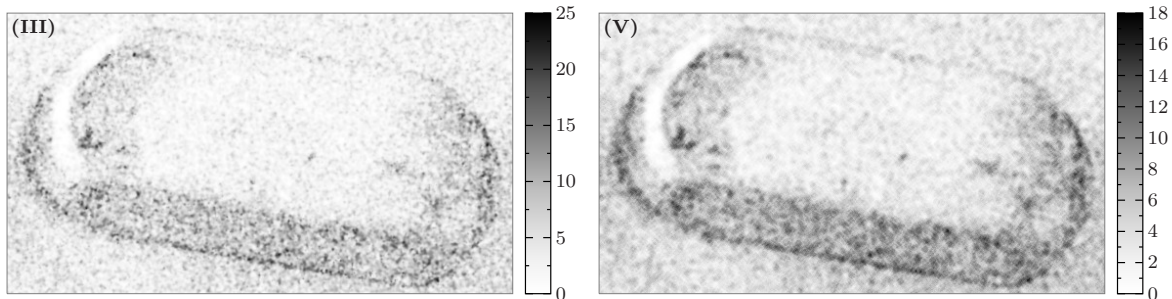


Figure 5. The expectation values from mixed retrodiction using Gaussian (left, $\sigma = 1.3$) and step-function (right, $\sigma = 3.5$) weights W_{ij} as given in equations (25) and (27) or table 1 (models III and V). The measurements are shown in figure 2.

To discuss some of the features of this mixed retrodiction we continue with three simple models where the weights are proportional to the distance d_{ij} between the i th and the j th pixel and thus resemble local averaging algorithms. The first is a Gaussian model where

$$W_{ij} = e^{-d_{ij}^2/(2\sigma^2)}, \tag{25}$$

while the second follows a polynomial

$$W_{ij} = \begin{cases} 1 - (d_{ij}/\sigma)^2 & \text{for } d_{ij} \leq \sigma, \\ 0 & \text{else.} \end{cases} \tag{26}$$

The Gaussian mixed retrodiction is by design very similar to usual local averaging.¹² Even with a cutoff where weights with $W_{ij} < 1/100$ are ignored, the number of neighbouring measurements included in the calculation grows rapidly, although most weights are rather small. The polynomial weights by comparison include fewer

neighbours for the same value of σ , but have a higher fraction of weights close to unity. As seen from table 1 or figure 5, we use $\sigma = 1.3$ [$\sigma = 4.5$] for the examples with Gaussian [polynomial] weights. This typically corresponds to $N = 45$ [$N = 48$] measurements included in retrodicting a single pixel.

The third is a much simpler, but surprisingly effective model and uses a step function for the weights,

$$W_{ij} = \begin{cases} 1 & \text{for } d_{ij} \leq \sigma, \\ 0 & \text{else.} \end{cases} \quad (27)$$

This model is thus an implementation of mixed retrodiction from a locally constant source as described in section 3.1. As seen in figure 5 (right panel) and described in table 1 we here use an example with $\sigma = 3.5$ which corresponds to a typical number of $N = 37$ measurements used to retrodict a single pixel.

4. PERFORMANCE OF DIFFERENT IMAGE RETRODICTION MODELS

In figure 4 we saw that the expectation values obtained from mixed retrodiction converge towards the true intensity value $\hat{\lambda}$ in regions of constant intensity. This justifies some optimism on the validity of this approach also for more interesting, non-constant images. However we also saw the significant variation in the quality of this retrodiction.

This unsettling uncertainty comes from the fact that the measurement values discussed here usually range in the low-digit numbers, with the large majority being either 0 or 1. For instance, having $\eta = 0.1$ we can expect that $\hat{\lambda} = 4$ or $\hat{\lambda} = 6$ will result in an almost identical set of Poissonian measurements. This is why fluctuations in the distribution of the number of photocounts used to retrodict the intensity of a specific pixel have such an impact on the quality of the retrodiction.

In this section we therefore present some results of an ongoing project to quantify the reliability of different retrodiction models. For this we take the subset of the image shown on the left in figure 1 and generate several artificial set of “measured” images. Each of these “measurements” is then processed with different retrodiction methods and compared to the original. In the following sections we show the results of some benchmark tests on the correctness and usefulness of the retrodicted probability distributions.

Table 1. The models tested using different artificial “measurements” of the image given in fig 1 and their labels as used in figures 3 to 8. Here d_{ij} denotes the Euclidean distance between the i th and the j th pixel.

Model	Label
single-pixel retrodiction with global $\bar{m} \simeq 0.5$	I
single-pixel retrodiction with local \bar{m}_i as given in equ. (10) with $\sigma = 1.3$.	II
mixed retrodiction with Gaussian weights, $W_{ij} = \exp(-d_{ij}^2/(2\sigma^2))$ with $\sigma = 1.3$ and local $\bar{m}_i \sim \sum_{j=1}^N W_{ij}m_j$	III
mixed retrodiction with polynomial weights, $W_{ij} = 1 - (d_{ij}/\sigma)^2$ (for $d_{ij} \leq \sigma$) with $\sigma = 4.5$ and local $\bar{m}_i \sim \sum_{j=1}^N W_{ij}m_j$	IV
mixed retrodiction with weights from a step function, $W_{ij} = 1$ for $d_{ij} \leq \sigma$ with $\sigma = 3.5$ and local $\bar{m}_i \sim \sum_{j=1}^N W_{ij}m_j$	V

4.1 Global performance compared to a true image

In order to learn about the correctness of a given retrodiction model we use two quantities similar to the common signal-to-noise ratio (SNR) and normalised root-mean-square error (nRMSE), but adapted to the situation where neither the original nor the estimated picture have a defined maximum value, i. e.

$$\text{SNR} := \frac{\sum_{i=1}^N \hat{\lambda}_i}{\sum_{i=1}^N \sqrt{(\hat{\lambda}_i - \text{E}(\lambda|m_i, \dots))^2}}, \quad (28a)$$

$$\text{nRMSE} := \frac{\sqrt{\sum_{i=1}^N (\hat{\lambda}_i - \text{E}(\lambda|m_i, \dots))^2 / N}}{\sum_{i=1}^N \hat{\lambda}_i / N}. \quad (28b)$$

Here $\hat{\lambda}_i$ is the true image value at the i th pixel while $\text{E}(\lambda|m_i, \dots)$ is the expectation value of the corresponding probability distribution. The notation $\text{E}(\lambda|m_i, \dots)$ will be used to indicate that this expectation value can be obtained using different retrodiction models.

As retrodiction gives not just a single number per pixel, but a full probability distribution we can also ask for the average probability to find λ close to the true image value $\hat{\lambda}_i$, given the corresponding measurements, i. e.

$$\text{Prob}(\lambda \in [(1-x)\hat{\lambda}, (1+x)\hat{\lambda}]) := \frac{1}{N} \sum_{i=1}^N \left(P((1+x)\hat{\lambda}_i|m_i, \dots) - P((1-x)\hat{\lambda}_i|m_i, \dots) \right), \quad (29)$$

with the cumulative distribution function $P(\lambda|m_i, \dots)$ as obtained using equations (24) and (8). In section 4.2 we will discuss how to use the probabilistic information to distinguish brighter or darker features.

In figure 6 we see the values for nRMSE, SNR and $\text{Prob}(\lambda \in [0.75\hat{\lambda}, 1.25\hat{\lambda}])$ for different artificial measurements of the sub-image shown in figure 1. The models for retrodiction used in figures 6 and 8 are given in table 1. We find that the polynomial retrodiction model IV and, surprisingly, the locally constant model V are on average far closer to the true image than the other models. The poor performance of the single-pixel retrodiction with local prior \bar{m}_i (model II) is surprising and appears to be due to a comparatively high noise level. Note that the Gaussian retrodiction (model III) uses the same local \bar{m}_i , but the retrodiction process adds further reliability to the predictions.

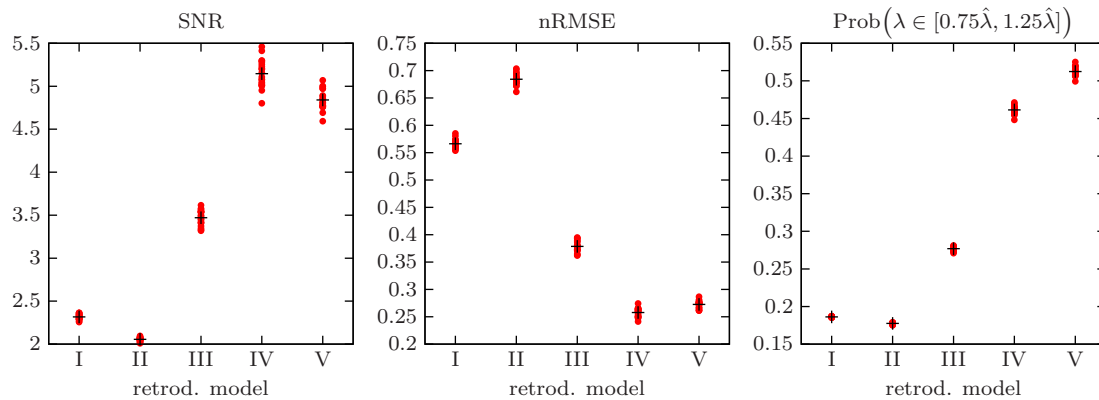


Figure 6. Results of benchmark tests of the adapted signal-to-noise ratio and normalised root-mean-square error as defined in equation (28) and the average probability of being close to the true value given in equation (29) for the retrodiction models as given in table 1. Each red dot corresponds to the results from one of 20 “measurements” of the detail image given in figure 1, the black crosses are the corresponding averages. It is quite surprising that model II (single-pixel retrodiction with local prior \bar{m}_i) is worse than model I with the global \bar{m} , but both are clearly outperformed by the mixed retrodiction models III to V.

4.2 Hypothesis testing using different retrodiction models

The distinguishing feature between image retrodiction and the evolved field of image denoising is that image retrodiction provides the full probability distribution $p(\lambda|m_i, \dots)$ for each pixel $i = 1, \dots, N$. In most of the previous examples we only discussed the corresponding expectation values, although the probability of λ being

close to the true value as given in equation (29) and tested in figure 6 was a first example making use of the probability distributions.

In this section we discuss the performance of the retrodiction models given in table 1 in distinguishing bright and dark areas. The framed areas of the detail image in figure 1 are three regions that are singled out because they are relatively bright (region A and B) or dark (region C).

In order to establish whether the intensity at a pixel is larger [smaller] than the average, we define the background as the median of the retrodicted expectation value across the whole image, i. e. $bg := \text{median}_i E(\lambda|m_i, \dots)$. Figures 7 and 8 show the evidence $ev(\lambda \geq bg)$ [or $ev(\lambda \leq bg)$] for the whole detail image or averaged across the regions of interest, respectively. The evidence that the intensity at a point is larger than a certain value x is given by¹⁰

$$ev(\lambda \geq x|m_i, \dots) = 10 \log_{10} \frac{1 - P(x|m_i, \dots)}{P(x|m_i, \dots)}, \quad (30)$$

with the cumulative distribution function P as used in equation (29).

The evidence as defined here is a good measure to see how well the hypothesis “ $\lambda \geq x$ ” is supported by a given probability distribution. Values of $ev \gg 0$ indicate that the statement “ $\lambda \geq x$ ” is strongly supported by this model, while negative values for the evidence indicate that this assumption is wrong. If we find $ev \approx 0$, then this model cannot help us to decide whether the hypothesis “ $\lambda \geq x$ ” is true or false.

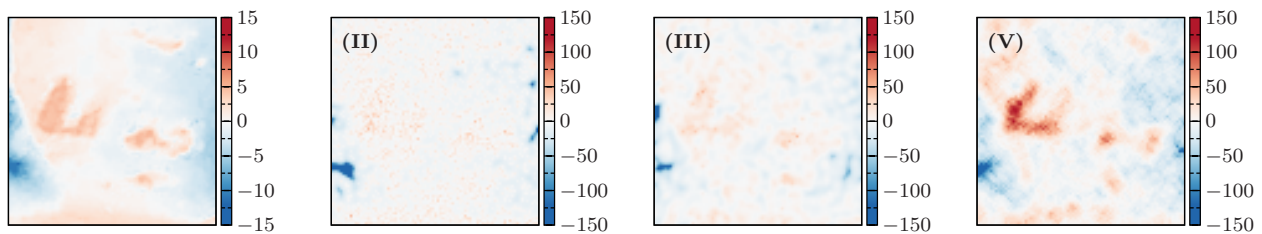


Figure 7. From left to right: values of $10 \log_{10}[\hat{\lambda}/\text{median}\hat{\lambda}]$ to show areas of relatively high or low intensity of the detail image from figure 1. Evidence for high/low intensity (compared to background) for the models described in table 1. We see that most models are relatively good in identifying low intensity regions, but only locally-constant mixed retrodiction (model V) gives an overall pleasant result. See figure 8 for the averaged results of multiple different “measurements”.

Figures 7 and 8 show that mixed retrodiction models (and there especially the locally constant model V, followed by the polynomial model IV) can reliably help to identify exceptionally bright or dark regions even in noisy images. Here we also see a key difference between single-pixel and mixed retrodiction: Choice of a good local prior \bar{m}_i might ensure that an image of the expectation values looks pleasantly like a “proper” image. However, the retrodicted probability distributions for single-pixel retrodiction remain very broad such that a test related to the probabilities will always give rather vague answers (that is, a low absolute value for the evidence).

By having a good prior and model for the weights W_{ij} one can produce retrodiction models that are correct when compared to a true image (see figure 6) and at the same time give more narrow probability distributions such that proper hypothesis testing is possible.

5. CONCLUSIONS

Usual (Poissonian) image denoising^{1,2,4,13-15} is highly successful in replacing a noisy image with a smoothed, more correct version. But especially if one tries to analyse data with on average less than one photocount per pixel, having a new, nicer image without detailed information on the reliability of the reconstruction might not be enough.

Designed for noisy images recorded at low light levels, image retrodiction⁵ is a framework that allows one to calculate the Bayesian probability distributions for the intensity at each pixel. In this work we have performed

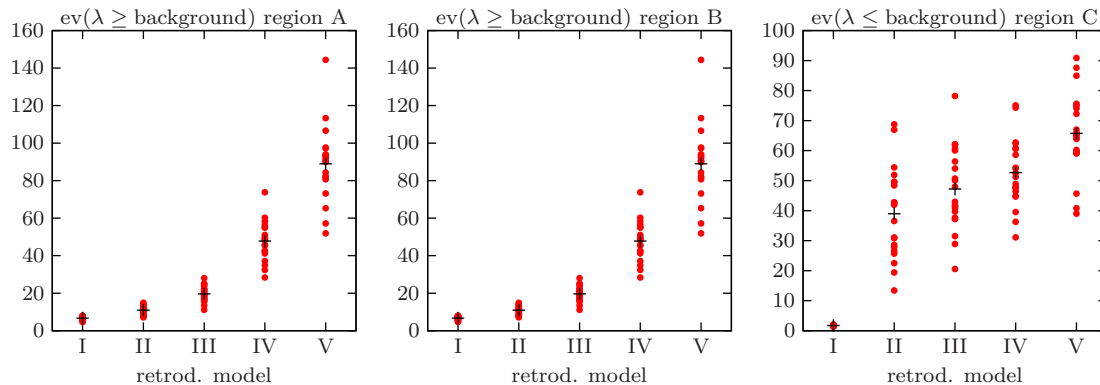


Figure 8. Evidence that the regions A, B and C defined in figure 1 have on average a higher (lower for region C) intensity than the background of the detail image, calculated for different retrodiction models as defined in table 1. The background is defined as the median of the retrodicted expectation value across the whole image, $bg := \text{median}_i E(\lambda|m_i, \dots)$, and is thus independent from any knowledge about the true image. Each red dot corresponds to the results from one of 20 “measurements” of the detail image given in figure 1, the black crosses are the corresponding averages. The evidence as defined in equation (30) is a logarithmic quantity which is here limited to a maximum value of ≈ 160 due to numerical reasons.

several benchmark tests on different retrodiction models to assess their reliability and usefulness for hypothesis testing.

The models used here are very simple yet surprisingly effective and we believe that combining this approach with for more evolved algorithms from image denoising can be of significant use for image-analysis based experiments and technologies in physics, astrophysics, medical imaging or fluorescence microscopy.

ACKNOWLEDGMENTS

This work was supported by the EPSRC UK Quantum Technology Programme under grant no. EP/M01326X/1 (QuantIC), the Challenges in Orbital Angular Momentum Programme, grant no. EP/I012451/1 and the Austrian Science Fund (FWF), J 3703-N27.

REFERENCES

- [1] Molina, R., Núñez, J., Cortijo, F. J., and Mateos, J., “Image restoration in astronomy: a Bayesian perspective,” *IEEE Signal Process Mag.* **18**(2), 11–29 (2001).
- [2] Bertero, M., Boccacci, P., Desiderà, G., and Vicidomini, G., “Image deblurring with Poisson data: from cells to galaxies,” *Inverse Prob.* **25**(12), 123006 (2009).
- [3] Kirmani, A., Venkatraman, D., Shin, D., Colaço, A., Wong, F. N., Shapiro, J. H., and Goyal, V. K., “First-photon imaging,” *Science* **343**(6166), 58–61 (2014).
- [4] Morris, P. A., Aspden, R. S., Bell, J. E., Boyd, R. W., and Padgett, M. J., “Imaging with a small number of photons,” *Nat. Commun.* **6**, 5913 (2015).
- [5] Sonnleitner, M., Jeffers, J., and Barnett, S. M., “Image retrodiction at low light levels,” *Optica* **2**(11), 950–957 (2015).
- [6] Watanabe, S., “Symmetry of physical laws. Part III. Prediction and retrodiction,” *Rev. Mod. Phys.* **27**(2), 179 (1955).
- [7] Aharonov, Y., Bergmann, P. G., and Lebowitz, J. L., “Time symmetry in the quantum process of measurement,” *Phys. Rev.* **134**(6B), B1410 (1964).
- [8] Barnett, S. M., Phillips, L. S., and Pegg, D. T., “Imperfect photodetection as projection onto mixed states,” *Opt. Commun.* **158**(1), 45–49 (1998).
- [9] Jedrkiewicz, O., Loudon, R., and Jeffers, J., “Retrodiction for optical attenuators, amplifiers, and detectors,” *Phys. Rev. A* **70**(3), 033805 (2004).

- [10] Jaynes, E. T., [*Probability Theory, The Logic of Science*], Cambridge University Press (2003).
- [11] Louisell, W. H., [*Quantum statistical properties of radiation*], vol. 2, Wiley (1973).
- [12] Buades, A., Coll, B., and Morel, J.-M., “A review of image denoising algorithms, with a new one,” *Multiscale Model. Simul.* **4**(2), 490–530 (2005).
- [13] Le, T., Chartrand, R., and Asaki, T. J., “A variational approach to reconstructing images corrupted by Poisson noise,” *J. Math. Imaging Vision* **27**(3), 257–263 (2007).
- [14] Lefkimiatis, S., Maragos, P., and Papandreou, G., “Bayesian inference on multiscale models for Poisson intensity estimation: Applications to photon-limited image denoising,” *IEEE Trans. Image Process.* **18**(8), 1724–1741 (2009).
- [15] Shin, D., Kirmani, A., Goyal, V. K., and Shapiro, J. H., “Photon-efficient computational 3d and reflectivity imaging with single-photon detectors,” *arXiv:1406.1761* (2014).