Contents

Supplementary Note 1: PG-PEM algortihm

1.1 DAR imaging model

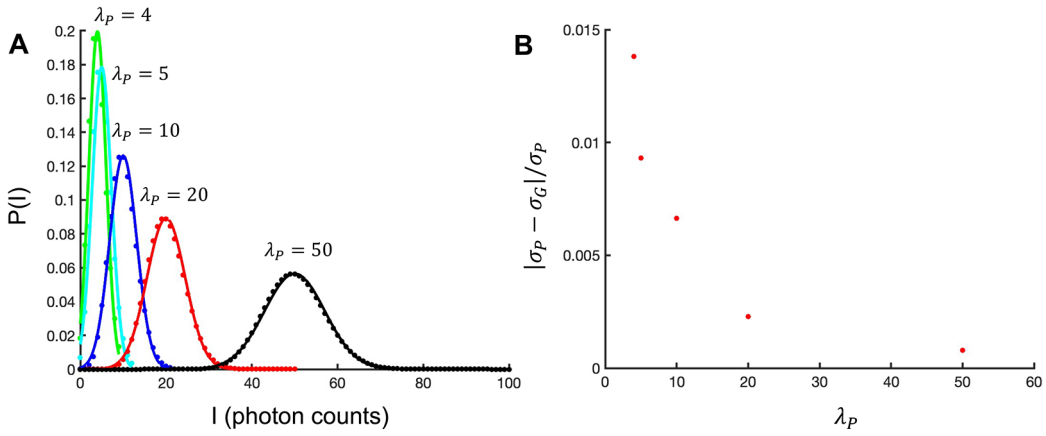
According to the DAR imaging process, its physical model can be expressed as:

$$R_p = \alpha Q_p + N_p, Q_p \sim \mathcal{P}[(X * h)_p + b_p], N_p \sim \mathcal{N}(0, \sigma_G^2), \tag{S1}$$

where p is the pixel index ($p \in \{1, 2, ..., P\}$), R is the raw DAR image of a tissue, α is a scaling factor corresponding to the gain of the imaging system, X is the "clean" image, h is the PSF, b is the mean of background, $\mathcal{P}[x]$ refers to the Poisson noise with mean x, $\mathcal{N}(0, \sigma_G^2)$ represents Gaussian noise and σ_G is its standard deviation. Because Q and N are both random fields in Eq. (S1), R is also a random field. Thus, we define r as the available observations of R.

According to the definition of Poisson process,

$$\mathcal{P}[(X * h)_p + b_p] = \mathcal{P}[(X * h)_p] + \mathcal{P}[b_p]. \tag{S2}$$



Supplementary Figure 1. Possion distributions approximated as Gaussian distsributions. (A) Poisson-distributed data (dots) for different values of photon flux λ_P and the corresponding Gaussian fitting functions $\mathcal{N}(\lambda_P, \lambda_P)$ (solid lines). **(B)** The relative error between the two distributions, in which σ_P and σ_G are the standard deviation of the Possion- and Gaussian-distributed data, respectively. Based on these results, Poisson distribution can be feasibly approximated as a Gaussian distribution when $\lambda_P > 3$.

When $b > 3$, $\mathcal{P}[b] \approx \mathcal{N}(b, b)$ [1] (Suppmentary Fig. 1). In the DAR images, b is normally assumed to be spatially invariant around the tissue and larger than 3. Therefore, Eq. (S3) is derived from Eqs. (S1) and (S2),

$$\begin{aligned} R &= \alpha\mathcal{P}[X * h] + \alpha\mathcal{N}(b, b) + \mathcal{N}(0, \sigma_G^2) \\ &= \alpha\mathcal{P}[X * h] + \mathcal{N}(\mu_N, \sigma_N^2). \end{aligned} \tag{S3}$$

where $\mu_N = \alpha b$ and $\sigma_N = \sqrt{\alpha^2 b + \sigma_G^2}$. Notably, the raw image can be splitted into Poisson-distributed signal and Gaussian-distributed noise. To estimate the noise parameters μ_N and σ_N, the background part needs to be extracted.

1.2 Patch-based estimation of noise parameters

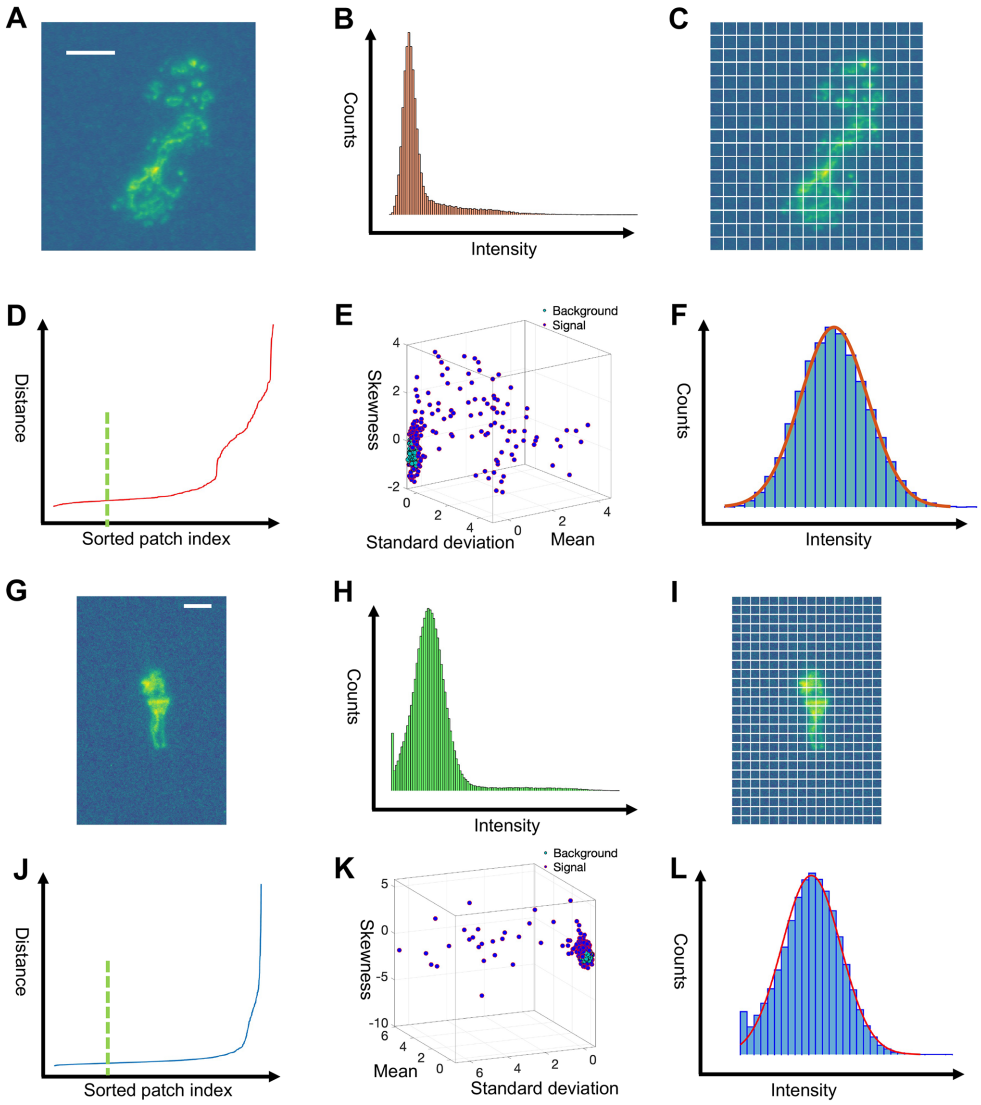
As Eq. (S3), the raw image can be splitted into Poisson-distributed signal $\alpha\mathcal{P}[X_p * h]$ and Gaussian-distributed noise $\mathcal{N}(\mu_N, \sigma_N^2)$. Because of the unknown true PSF and the non-white noise, the frequency domain-based noise parameters estimation method in [2] cannot be used. To estimate the noise parameters, the background part needs to be extracted. Due to the continuity of the histogram of DAR images (**Supplementary Fig. 2B, H**), simple thresholding method based on intensity values will result in signal pixels aberrantly being classified as background, known as false negative in detection theory. In fact, most of the areas without tissues in DAR images normally should not have radioactive signal from the tissues. Thus, these areas should only have background and noise and be highly similar to each other. Based on this assumption, we propose a patch-based estimation algorithm with Density-Based Spatial Clustering of Applications with Noise [3] (DBSCAN) by searching patches with similar features to robustly estimate the noise parameters. The process is as **Supplementary Algorithm 1** shows.

Supplementary Algorithm 1 Patch-based estimation of noise parameters

Input: Observation of the raw image, r;
Patch size, (M, N);
Output: Mean, αb;
Standard deviation, $\sqrt{\alpha^2 b + \sigma_G^2}$;

1: Split r (**Supplementary Fig. 2A, G**) into multiple patches with size of M rows and N columns (**Supplementary Fig. 2C, I**);
2: Calculate the mean, standard deviation, skewness and kurtosis of every patch;
3: Use the Z-score means method to normalize the data and cluster them by DBSCAN (**Supplementary Fig. 2D, J**);
4: Select the background patches based on the cluster result and form a new dataset: $\{X_p | p = 1, 2, ..., T\}$, where T is the total pixel number of the extracted background;
5: The final values are estimated by MLEM algorithm based on the histograms of the extracted background (**Supplementary Fig. 2E, K**).
6: **return** αb, $\alpha^2 + \sigma_G$.

In this algorithm, M and N are usually set as 10 for our dataset. Besides, it should be noted that when the noise level of the raw image is very high (**Supplementary Fig. 2G**), the corresponding histogram may be cut off at 0 (**Supplementary Fig. 2H**). Under this circumstance, a truncated Gaussian distribution parameters estimation method based on maximum likelihood expectation maximization (MLEM) algorithm is needed [4].



Supplementary Figure 2. Noise parameters estimation. **(A,G)** Raw DAR images with low and high noise levels; **(B,H)** The histogram of the raw images; **(C,I)** Split the raw images into multiple patches and calculate the mean, standard deviation, skewness and kurtosis of each patch; **(E,K)** Plot of the sorted (minPts-1)-th nearest distance of every patch; **(E,K)** DBSCAN results for the patches; **(F,L)** The histograms of the extracted background field which can be fitted by a Gaussian distribution; while **(B)** and **(H)** cannot be. The green dotted lines in **(D)** and **(J)** correspond to the ϵ used in **(E)** and **(K)**. The histogram in **(H)** is truncated at 0 while that in **(B)** is not. To estimate the parameters of the truncated histogram, we apply the truncated MLEM method. Scale bar: **(A)**: 2.3 mm, **(G)**: 4.65 mm.

When implementing DBSCAN, its two parameters ϵ and minPts needs to be manually set, in which ϵ is a parameter specifying the radius of a neighborhood with respect to the same point and minPts is the minimum number of points required to form a dense region. As a rule of thumb, minPts is usually twice as the dimension number of the features. Therefore, we set minPts as 8 in PG-PEM. Meanwhile, we use k-nearest neighbour algorithm to determine ϵ. In detail, a k-distance graph is first built based on the data, and then we find every point's (minPts-1)-th nearest distance, sort them in the order from low to high and plot them (**Supplementary Fig. 2D and J**). Next, unlike the approach [3] which select the "elbow" point as ϵ, we experimentally select the distance before the first 25% position in the range as ϵ, which is far from the "elbow" point for our DAR images. This approach could ensure enough points to form the Gaussian-shape histograms in **Supplementary Fig. 2F and L**. Simultenously, it could also avoid classify signal as background by mistake.

1.3 MLEM algorithm for the mixed Poisson-Gaussian model

By simple variable substitutions using Eq. (S4), Eq. (S3) can be simplified as Eq. (S5), where R' and X are both 2D matrices with P pixels and h is a 2D kernel with S pixels. For our DAR images, normally $S \ll P$.

$$R' = \frac{R}{\alpha}, \mu'_N = \frac{\mu_N}{\alpha} = b, \sigma'_N = \frac{\sigma_N}{\alpha} = \sqrt{b + \frac{\sigma_G^2}{\alpha^2}}, \tag{S4}$$

$$R' = \mathcal{P}[X * h] + \mathcal{N}(b, b + \frac{\sigma_G^2}{\alpha^2}). \tag{S5}$$

Here we donate Q as $\mathcal{P}[X * h]$ and U as $\mathcal{N}(b, b + \frac{\sigma_G^2}{\alpha^2})$, whose elements follow Poisson and Gaussian distributions, respectively. Our goal is to estimate the unknown parameters $\theta = (X, h)$ with the given parameters r' (the available observations of R'), b and $b + \frac{\sigma_G^2}{\alpha^2}$ by following the assumptions for the Poisson-distributed signal and Gaussian-distributed noise in Eq. (S5): 1) they are mutually independent; 2) their components are independent. Under these assumptions, Eq. (S6) is obtained by applying Bayes rule, which is the mixed continuous-discrete probability distribution of (R'_p, Q_p) for every $p \in \{1, 2, ..., P\}$.

$$\begin{aligned} Pr_{R'_p, Q_p}(r'_p, q_p|\theta) &= Pr(Q_p = q_p|\theta) f_{R'_p|Q_p=q_p}(r'_p|b, b + \frac{\sigma^2}{\alpha^2}) \\ &= Pr(Q_p = q_p|\theta) f_{U_p}(r'_p - q_p|b, b + \frac{\sigma^2}{\alpha^2}). \end{aligned} \tag{S6}$$

where $f_{R'_p|Q_p=q_p}(\cdot|b, b+\frac{\sigma^2}{\alpha^2})$ is the conditional probability density function (PDF) of R'_p knowing that $Q_p = q_p$ and $f_{U_p}(\cdot|b, b + \frac{\sigma^2}{\alpha^2})$ is the PDF of U_p. In detail,

$$Pr(Q_p = q_p|\theta) = exp[(-X * h)_p]\frac{(X * h)_p^{q_p}}{q_p!}, \tag{S7}$$

$$f_{U_p}(r'_p - q_p|b, b + \frac{\sigma^2}{\alpha^2}) = \frac{1}{(2\pi)^{1/2}(b + \frac{\sigma_G^2}{\alpha^2})^{1/2}} exp\Big(-\frac{(r'_p - q_p - b)^2}{2(b + \frac{\sigma_G^2}{\alpha^2})}\Big). \tag{S8}$$

Therefore, the likelihood of Eq. (S5) takes the form as Eq. (S9):

$$\begin{aligned} f_{R'}(r'|\theta) &= \prod_{p=1}^{P}\sum_{q_p=0}^{+\infty} Pr_{R'_p,Q_p}(r'_p, q_p|\theta) \\ &= \frac{1}{(2\pi)^{P/2}(b + \frac{\sigma_G^2}{\alpha^2})^{P/2}}\prod_{p=1}^{P} exp[(-X * h)_p] \\ &\sum_{q_p=0}^{+\infty} exp\Big(-\frac{(r'_p - q_p - b)^2}{2(b + \frac{\sigma_G^2}{\alpha^2})}\Big)\frac{(X * h)_p^{q_p}}{q_p!}. \end{aligned} \tag{S9}$$

To solve the parameters θ in Eq. (S9), an iterative MLEM approach is utilized:

$$\theta^{(n+1)} = argmax_\theta J(\theta|\theta^{(n)}), \tag{S10}$$

where $J(\theta|\theta^{(n)}) = E_{Q|R'=r',\theta^{(n)}}[lnPr_{R',Q}(R', Q|\theta)]$ and $Pr_{R',Q}(R', Q|\theta) = \prod_{p=1}^{P} Pr_{R'_p,Q_p}(R'_p, Q_p|\theta)$.

According to Eq. (S6),

$$\begin{aligned} lnPr_{R',Q}(R', Q|\theta) = &-\frac{1}{2(b + \frac{\sigma_G^2}{\alpha^2})^{P/2}}\sum_{p=1}^{P}(R'_p - Q_p - b)^2 - \frac{P}{2}ln(2\pi(b + \frac{\sigma_G^2}{\alpha^2})) \\ &-\sum_{p=1}^{P}(X * h)_p + \sum_{p=1}^{P} ln(X * h)_p Q_p - \sum_{p=1}^{P} ln(Q_p!) \end{aligned} \tag{S11}$$

By dropping the terms that are independent of θ, Eq. (S10) is simplified as:

$$\theta^{(n+1)} = argmin_\theta \tilde{J}(\theta|\theta^{(n)}), \tag{S12}$$

where

$$\tilde{J}(\theta|\theta^{(n)}) = \sum_{p=1}^{N}(X * h)_p - \sum_{p=1}^{N} ln(X * h)_p E_{Q_p|R'_p=r'_p,\theta^{(n)}}(Q_p). \tag{S13}$$

For every $p \in \{1, 2, ..., P\}$, we have

$$\begin{aligned} E_{Q_p|R'_p=r'_p,\theta^{(n)}}(Q_p) &= \sum_{q_p=0}^{+\infty} q_p Pr(Q_p = q_p | R'_p = r'_p, \theta^{(n)}) \\ &= \frac{\sum_{q_p=0}^{+\infty} q_p Pr_{R'_p Q_p}(r'_p, q_p|\theta^{(n)})}{f_{R'_p}(r'_p|\theta^{(n)})} \\ &= \frac{\sum_{q_p=0}^{+\infty} q_p Pr_{R'_p Q_p}(r'_p, q_p|\theta^{(n)})}{\sum_{q_p=0}^{+\infty} Pr_{R'_p, Q_p}(r'_p, q_p|\theta^{(n)})}. \end{aligned} \tag{S14}$$

By combining Eqs. (S6)-(S8), Eq. (S14) simplifies to:

$$E_{Q_p|R'_p=r'_p,\theta^{(n)}}(Q_p) = \frac{\sum_{q_p=1}^{+\infty} exp\Big(- \frac{(r'_p - q_p - b)^2}{2(b+\sigma_G^2/\alpha^2)} \Big) \frac{(X*h)_p^{q_p}}{(q_p-1)!}}{\sum_{q_p=0}^{+\infty} exp\Big(- \frac{(r'_p - q_p - b)^2}{2(b+\sigma_G^2/\alpha^2)} \Big) \frac{(X*h)_p^{q_p}}{q_p!}}. \tag{S15}$$

Therefore, a 2D matrix $E_{Q|R'=r',\theta^{(n)}}(Q)$, sharing the same shape with R' and X, is acquired. When implementing Eq. (S15), there is an issue calculating the infinite terms of q_p on numerator and denominator. Nonetheless, the infinite terms are bounded and the Lambert W function is used to estimate the terms of summarizations, as described previously [5]. The iteration process is done by differentiating Eq. (S13) with respect to X and h and setting the derivative to zero [6–8]. The resulting iterative scheme is given by alternating:

$$\hat{h}^{(n+1)} = \hat{h}^{(n)} \cdot \left[\left(\frac{E_{Q|R'=r',\theta^{(n)}}(Q)}{\hat{h}^{(n)} * \hat{X}^{(n)}} \right) * \hat{X}^{(n),m} \right], \tag{S16}$$

$$\hat{X}^{(n+1)} = \hat{X}^{(n)} \cdot \left[\left(\frac{E_{Q|R'=r',\theta^{(n)}}(Q)}{\hat{h}^{(n)} * \hat{X}^{(n)}} \right) * \hat{h}^{(n),m} \right], \tag{S17}$$

where $\hat{h}^{(n),m}$ and $\hat{X}^{(n),m}$ are the mirrored results of $\hat{h}^{(n)}$ and $\hat{X}^{(n)}$, respectively. In addition, $\hat{h}$ has three constraints: circularly symmetric constraint, non-negativity ($\forall s$, $\hat{h}_s > 0$) and and norm to 1 ($\sum_{s=1}^{S} \hat{h}_s = 1$). To ensure the circularly symmetric constraint, $\hat{h}$ is averaged in the angular direction as Eq. (S18), where r is the magnitude and ϕ is the phase in the polar coordinate system; To ensure $\sum_{s=1}^{S} \hat{h}_s = 1$, Eq. (S19) is conducted; To ensure $\forall s$, $\hat{h}_s > 0$, Eq. (S20) is conducted.

$$\hat{h}'(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{h}(rcos\phi, rsin\phi) d\phi. \tag{S18}$$

$$\hat{h} = \frac{\hat{h}'}{\sum_{s=1}^{S} \hat{h}'_s}. \tag{S19}$$

$$\exists \hat{h}_s < 0, \text{set } \hat{h}_s = 0. \tag{S20}$$

1.4 Regularization for h and X

In practice, the blind deconvolution problem is highly ill-posed, so regularizations for both h and X are needed. Through the iteration process h tends to converge as a delta function. To avoid this trivial solution and considering the smooth characteristics of h, h is regularized by L2 norm. The noise of X may amplify in the iteration process, so total variation (TV) is normally added for X [8, 9]. Nevertheless, TV penalty always oversharpens the boundaries between different regions, generating "staircase" effect. To avoid this, we adopted a Hessian Frobenius penalty for the estimated X to enable smoother transitions between different regions and to suppress noise simultaneously [10–12]. As a result, this leads to a penalized MLEM algorithm as Eq. (S21), where λ_h and λ_X are the regularization parameters for h and X, $P_h = exp(-\frac{1}{2}\sum_{s=1}^{S} h_s^2)$ and $P_X = exp(-\sum_{p=1}^{P} |\mathcal{H}X|_p)$ are the prior probability functions for h and X, respectively. Here, $\mathcal{H}$ is a Hessian operator and defined as $[\partial_{xx}, \partial_{xy}; \partial_{xy}, \partial_{yy}]$, where $\partial_{xx} = \partial^2/\partial x^2$, $\partial_{xy} = \partial^2/\partial x \partial y$ and $\partial_{yy} = \partial^2/\partial y^2$. $|\mathcal{H}X|$ is the Hessian Frobenius (HF) norm for X and defined as $\sqrt{(\partial_{xx}X)^2 + (\partial_{yy}X)^2 + 2(\partial_{xy}X)^2}$.

$$\theta^{(n+1)} = argmin_\theta[\tilde{J}(\theta|\theta^{(n)}) + \lambda_h \sum_{s=1}^{S} h_s^2 + 2\lambda_X \sum_{p=1}^{P} |\mathcal{H}X|_p]. \tag{S21}$$

The penalized optimization process for h and X can both be implemented by a forward-backward splitting algorithm similar to [13]. In [13], this algorithm was originally designed for TV regularization. Nevertheless, since the regularization terms for h and X are all convex and the data fidelity term can be extended to a Kullback-Leibler functional without affecting the stationary points [14], which is the same as the condition of TV regularization, the algorithm framework still works for our problem. Therefore, the original Eq. (S16) is modified as EM step and L2 norm regularization step:

$$\begin{cases} \hat{h}^{(n+\frac{1}{2})} = \hat{h}^{(n)} \cdot \left[\left(\frac{E_{Q|R'=r',\theta^{(n)}}(Q)}{\hat{h}^{(n)} * \hat{X}^{(n)}} \right) * \hat{X}^{(n),m} \right] & \text{(EM step)} \\ \hat{h}^{(n+1)} = argmin_h \left\{ \sum_{s=1}^{S} \frac{(\hat{h}_s - \hat{h}_s^{(n+\frac{1}{2})})^2}{\hat{h}_s^{(n)}} + \lambda_h \sum_{s=1}^{S} \hat{h}_s^2 \right\} & \text{(L2 norm regularization step).} \end{cases} \tag{S22}$$

Expand the L2 norm regularization step:

$$\hat{h}^{(n+1)} = argmin_h \left\{ \sum_{s=1}^{S} \frac{1}{\hat{h}_s^{(n)}} \left[\left(1 + \lambda_h \hat{h}_s^{(n)} \right) \hat{h}_s^2 - 2\hat{h}_s^{(n+\frac{1}{2})} \hat{h}_s + const \right] \right\}. \tag{S23}$$

By setting the derivative of every $\hat{h}_s$ to be zero, we get

$$\hat{h}_s^{(n+1)} = \frac{\hat{h}_s^{(n+\frac{1}{2})}}{1+\lambda_h \hat{h}_s^{(n)}}. \tag{S24}$$

Therefore, the solution of Eq. (S22) is

$$\hat{h}^{(n+1)} = \hat{h}^{(n)} \cdot \left[\left(\frac{E_{Q|R'=r',\theta^{(n)}}(Q)}{\hat{h}^{(n)} * \hat{X}_p^{(n)}} \right) * \hat{X}_p^{(n),m} \right] \Big/ \left(1+\lambda_h \hat{h}^{(n)}\right). \tag{S25}$$

Likewise, the original Eq. (S17) is modified as EM step and HF norm regularization step:

$$\begin{cases} \hat{X}^{(n+1)} = \hat{X}^{(n)} \cdot \left[\left(\frac{E_{Q|R'=r',\theta^{(n)}}(Q)}{\hat{h}^{(n)} * \hat{X}^{(n)}} \right) * \hat{h}^{(n),m} \right] & \text{(EM step)} \\ \hat{X}^{(n+1)} = argmin_X \left\{ \sum_{p=1}^{P} \frac{(\hat{X}_p - \hat{X}_p^{(n+\frac{1}{2})})^2}{\hat{X}_p^{(n)}} + 2\lambda_X \sum_{p=1}^{P} |\mathcal{H}\hat{X}|_p \right\} & \text{(HF norm regularization step).} \end{cases} \tag{S26}$$

The HF norm regularization step can be solved by the $majorization-minimization$ (MM) framework [10]. Based on this method, the point-wise regularization term can be transformed as Eq. (S27). Its equality holds if and only if $|\mathcal{H}\hat{X}|_p = |\mathcal{H}\hat{X}^{(n)}|_p$, which can be achieved when the iteration converges.

$$|\mathcal{H}\hat{X}|_p \leq \frac{|\mathcal{H}\hat{X}^{(n)}|_p}{2} + \frac{|\mathcal{H}\hat{X}|_p^2}{2|\mathcal{H}\hat{X}^{(n)}|_p}. \tag{S27}$$

Thus, minimizing the HF norm regularization step in Eq. (S26) can be conducted by minimizing a surrogate function as Eq. (S28), where $\mathcal{W} = [\partial_{xx}, \partial_{yy}, \sqrt{2}\partial_{xy}]^T$.

$$\hat{X}^{(n+1)} = argmin_X \left\{ \sum_{p=1}^{P} \frac{(\hat{X}_p - \hat{X}_p^{(n+\frac{1}{2})})^2}{\hat{X}_p^{(n)}} + \lambda_X \sum_{p=1}^{P} \frac{|\mathcal{H}\hat{X}|_p^2}{|\mathcal{H}\hat{X}^{(n)}|_p} + const \right\}. \tag{S28}$$

Expanding Eq. (S28), we have:

$$\hat{X}^{(n+1)} = argmin_X \left\{ \sum_{p=1}^{P} \frac{1}{\hat{X}_p^{(n)}} \left[\left(1 + 2\lambda_X \left(\mathcal{W}^T \frac{\mathcal{W}\hat{X}^{(n)}}{|\mathcal{H}\hat{X}^{(n)}|} \right)_p \right) \hat{X}_p^2 - 2\hat{X}_p \hat{X}_p^{n+\frac{1}{2}} + const \right] \right\}. \tag{S29}$$

By setting the derivative of $\hat{X}_p$ to be zero, Eq. (S28) is solved as Eq. (S30), where $\Xi = \frac{\partial}{\partial_{xx}} + \frac{\partial}{\partial_{yy}} + \sqrt{2}\frac{\partial}{\partial_{xy}}$.

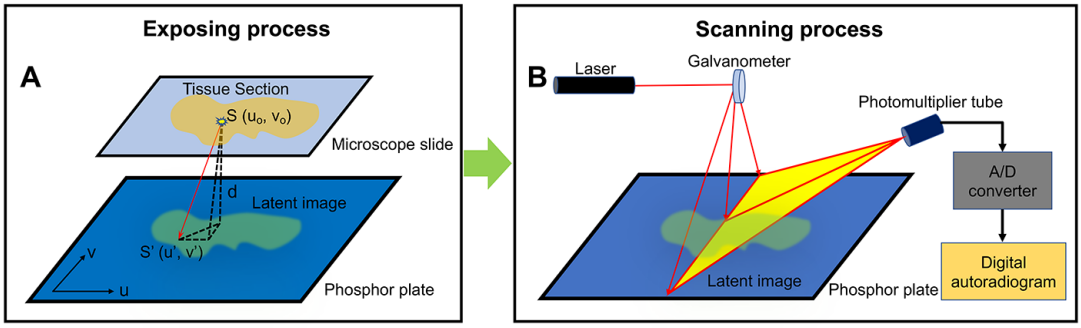
$$\hat{X}_p^{(n+1)} = \frac{\hat{X}_p^{(n+1/2)}}{1 + \lambda_X \left(\Xi \frac{\mathcal{W}\hat{X}^{(n)}}{|\mathcal{H}\hat{X}^{(n)}|} \right)_p}. \tag{S30}$$

Therefore, the solution of Eq. (S26) is derived as:

$$\hat{X}^{(n+1)} = \hat{X}^{(n)} \cdot \left[\left(\frac{E_{Q|R'=r',\theta^{(n)}}(Q)}{\hat{h}^{(n)} * \hat{X}^{(n)}} \right) * \hat{h}^{(n),m} \right] \Big/ \left[1 + \lambda_X \left(\Xi \frac{\mathcal{W}\hat{X}^{(n)}}{|\mathcal{H}\hat{X}^{(n)}|} \right) \right]. \tag{S31}$$

1.5 PSF model

The DAR imaging system involves exposing the plate to the radioactive section which produces a latent image through a trapping of electron-hole (**Supplementary Fig. 3A**), and then scanning a laser over the latent image on the phosphor plate to stimulate release of photons (**Supplementary Fig. 3B**). As discussed in the paper, h is circularly symmetric, also affected by the finite focal point of the laser scanner and the modulation transfer function (MTF) of the phospher plate. However, in blind restoration it is not needed to consider all these points to initialize a PSF. Instead, we model it based on the scattering effect using the inverse square law [15]. In **Supplementary Fig. 3A**, the energy of one point in the latent image can be expressed as Eq. (S32), where E_S is the total energy of its source S, E_C is the cut-off energy of the phosphor plate, $d\Omega$ is the unit solid angle corresponding to each pixel. Here, we assume $d\Omega$ is the same for each pixel under the blind restoration framework. As a result, h is approximated as Eq. (S33) when initializing, in which a controls its size. Then its sum is normalized to 1. In fact, because of the regularization for h, the initial value of a does not have a good impact on the result. In our application, we set a as 1.



Supplementary Figure 3. DAR imaging process and PSF modelling. (A) Expose the plate to the radioactive section to produce a latent image; **(B)** Scan a laser over the latent image on the phosphor plate to generate the corresponding digital autoradiographic image. d is the distance between the tissue section and the phosphor plate. In uv-plane, the coordinate of S and S' are (u_o,v_o) and (u',v') seperately.

$$E_{S'} = \begin{cases} \frac{E_S}{4\pi[(u'-u_o)^2+(v'-v_o)^2+d^2]} d\Omega & \text{when } E_{S'} > E_C \\ 0 & \text{else} \end{cases}. \tag{S32}$$

$$h = \begin{cases} \frac{1}{a[(u-u_o)^2+(v-v_o)^2]+1} & \text{when } E_{S'} > E_C \\ 0 & \text{else} \end{cases}. \tag{S33}$$

1.6 The impact of the scaling factor α

According to Eqs. (S1), (S4) and (S5), the scaling factor α needs to be pre-calibrated. However, this process is very time-consuming and may not be robust as discussed in the paper. Here, we investigate the impact of α on the restoration result. Assuming α is not calibrated correctly and the calibrated scaling factor is $\alpha^* = \beta\alpha$. Then Eq. (S5) is transformed as Eq. (S34), where X' is the estimated image under this condition.

$$\frac{R'}{\beta} = \mathcal{P}[X' * h] + \frac{\mathcal{N}(b, b + \frac{\sigma_G^2}{\alpha^2})}{\beta}. \tag{S34}$$

By using Poisson distribution to estimate Gaussian distribution, Eq. (S34) is estimated as a shift-Poisson format [12] as Eq. (S35), in which b/β and $(b + \frac{\sigma_G^2}{\alpha^2})/\beta^2$ can both be estimated from R'/β with Algorithm 1.

$$\frac{R' - b}{\beta} + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} = \mathcal{P}[X' * h + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2}]. \tag{S35}$$

Based on Eq. (S35), the likelihood of the Poisson statistics is as Eq. (S36), where $k_p = \frac{r'_p - b}{\beta} + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2}$ $(p \in \{1, 2, ..., P\})$.

$$f_K(k|\theta) = \prod_{p=1}^{P} exp\left[-(X' * h)_p - \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2}\right] \frac{\left[(X' * h)_p + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2}\right]^{k_p}}{k_p!}. \tag{S36}$$

Therefore, the negative log-likelihood of Eq. (S36) is:

$$-lnf_K(k|\theta) = \sum_{p=1}^{P} \left\{ (X' * h)_p + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} - k_p ln\left[(X' * h)_p + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2}\right] + ln(k_p!) \right\}. \tag{S37}$$

By setting the deriative of every $(X' * h)_p$ to be 0, we have:

$$1-\left[(X'*h)_p+\frac{b+\frac{\sigma_G^2}{\alpha^2}}{\beta^2}\right]\Bigg/\left(\frac{r'_p-b}{\beta}+\frac{b+\frac{\sigma_G^2}{\alpha^2}}{\beta^2}\right)=0. \tag{S38}$$

Under the MLEM iterative framework, the result of Eq. (S38) can be expressed as Eq. (S39), where $E(r'_p)$ is the expectation of r'_p.

$$(X'*h)_p=\frac{E(r'_p)-b}{\beta}. \tag{S39}$$

Thus, when the estimated PSF h is the same, the relationship between the results using α^* and α can be approximated as Eq. (S40),

$$X=\beta X'. \tag{S40}$$

Based on these results, the scaling factor α does not have an appreciable impact on the restoration result except as a multiplier parameter β. The reason is that the Gaussian-distributed noise parameters μ_N and σ_N are estimated directly from the background, and this condition is quite similar to that of pure Poisson noise, whose scaling parameter is not required to be calibrated. In fact, digital light unit (DLU) itself is meaningless. When implementing dosimetry calculation, several phantoms should be built for calibrating the mapping relationship between DLU and dosimetry values. In this sense, α will not affect the result of Eq. (S15) and is not needed to be pre-calibrated for DAR image restoration if Eqs. (S34) and (S35) are approximately equal to each other. However, when α^* is too large, the computation process in Eq. (S15) will generate sampling errors, especially for the low intensity regions of the images. To maintain this accurate Gaussian-Poisson (continuous to discrete) transformation, we should make sure $\frac{b+\frac{\sigma_G^2}{\alpha^2}}{\beta^2}=\frac{b}{\beta^2}+(\frac{\sigma_G}{\alpha^*})^2\gg 1$. Therefore, we empirically set $\frac{\sigma_G}{\alpha^*}\geq 10$. In real applications, we can directly use α^* to replace α. Before the restoration process, the raw image could be divided by a large α to decrease the number of summations in Eq. (S15) and then multiply the same α after estimation to maintain the final result with the same scale with the raw image. Then, we set $\alpha\leq\frac{\sigma_G}{10}$. In practice, we set α as $\frac{\sigma_G}{10}$ considering both the computation efficiency and the accuracy of PG-PEM.

1.7 PG-PEM algorithm summary, software availability and runtime analysis

The algorithm is summarized as **Supplmentary Algorithm 2**.

Supplementary Algorithm 2 PG-PEM algorithm

Initialization:

Estimate αb and $\alpha^2 b + \sigma_G^2$ using Algorithm 1.
Rescale r, αb and $\alpha^2 b + \sigma_G^2$ to r', b and $b + \sigma_G^2/\alpha^2$ using the pre-set α and (S4).
Set the raw image observation r' as the initial X.
Initialize the PSF h using Eq. (S33).
Set the regularization parameters λ_X and λ_h.

Iteration:

for each $i = 1, 2, ..., N$ **do**
  E step: Use Eq. (S15) to estimate $E_{Q|R'=r',\theta^{(n)}}(Q)$.
  M step: Use Eq. (S31) to estimate X.
    Use Eq. (S25) to estimate h.
    Use Eqs. (S18), (S19) and (S20) to normalize h.
  if $\sqrt{\sum_{p=1}^{P}(\hat{X}_p^{(n+1)} - \hat{X}_p^{(n)})^2} \Big/ \sqrt{\sum_{p=1}^{P}(\hat{X}_p^{(n)})^2} < threshold$ **then**
    break
  end if
end for

All the codes for PG-PEM were written in Matlab 2019a (MathWorks). The software supporting all proposed methods and example data are available upon reasonable request from the corresponding author.

The runtime of PG-PEM mainly depends on the image size, the number of summation terms in Eq. (S15) and the value of the threshold in Algorithm 2. In our experiments, we typically set the value of the threshold from 0.0005 to 0.001. In such a range, the iteration numbers are usually around 100. Here we report the performance of our software under various image sizes both with and without graphics processing unit (GPU, NVIDIA Quadro RTX 6000) on a workstation using 12 cores Intel(R) Xeon(R) W-2133, 3.60GHz central processing unit (CPU). Specifically, we set the iteration numbers as 100 for all the groups. Besides, the number of summation terms in Eq. (S15) is set as around 700, which can satisfy the criteria in **Supplementary Note 1.6**. The results are presented in **Supplementary Table 1**.

From the table, when the image size is small, CPU-based restoration runs fast. However, as the image sizes grow larger, GPU-based restoration performs better than CPU. Fortunately, the size of all the DAR images utilized in research practice are smaller than 500x500 pixels (pixel size: 0.042x0.042mm^2). Therefore, CPU can fulfill most of the requirements. Notably, PG-PEM could be slower than the algorithm based on the shifted-Poisson model [12] due to the computation of Eq. (S15). Nevertheless, PG-PEM achieves much better denoising

Supplementary Table 1. Computational speed of PG-PEM in different conditions

Image Size (px)	CPU	GPU
250x250	12.3s	42.6s
500x500	85.5s	62.6s
750x750	193.6s	90.7s

performance especially for the images with high noise level. Meanwhile, compared to the very long exposure time of the DAR imaging process (several hours to several days), our PG-PEM algorithm still runs very fast (from several seconds to less than 2 minutes).

Supplementary Note 2: Quality metrics

The root mean squared error (RMSE) is an pixel-wise difference between two input images, where the ideal value is zero. It is computed as Eq. (S41), where Y^{est} is the estimated image, Y^{true} is the ground truth, p is the pixel index and P is the total pixel number for every image.

$$\mathrm{RMSE}(Y^{est}, Y^{true}) = \sqrt{\frac{1}{P}\sum_{p=1}^{P}(Y_p^{est} - Y_p^{true})^2} \tag{S41}$$

The signal power to noise power ratio (SNR) indicates the ratio of the power of a signal to the power of background noise. It is defined as Eq. (S42), where Y^{est} is the estimated image from restoration algorithms and Y^{true} is the ground truth.

$$\mathrm{SNR}(Y^{est}, Y^{true}) = 10log\frac{\sum_{p=1}^{P}(Y_p^{true})^2}{\sum_{p=1}^{P}(Y_p^{est} - Y_p^{true})^2} \tag{S42}$$

The structure similarity [16] (SSIM) is a perception-based model that considers image degradation as perceived change in structural information, while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms. Compared to RMSE and SNR, it is supposed to give more information about image distortion by the computation of local image structure, luminance and contrast into a single local quality score. In this paper, the luminance and contrast are normalized and SSIM is defined as Eq. (S43),

$$\mathrm{SSIM}(Y^{est}, Y^{true}) = \frac{2\mu_{Y^{est}}\mu_{Y^{true}} + C_1}{\mu_{Y^{est}}^2 + \mu_{Y^{true}}^2 + C_1} \cdot \frac{2\sigma_{Y^{est}Y^{true}} + C_2}{\sigma_{Y^{est}}^2 + \sigma_{Y^{true}}^2 + C_2} \tag{S43}$$

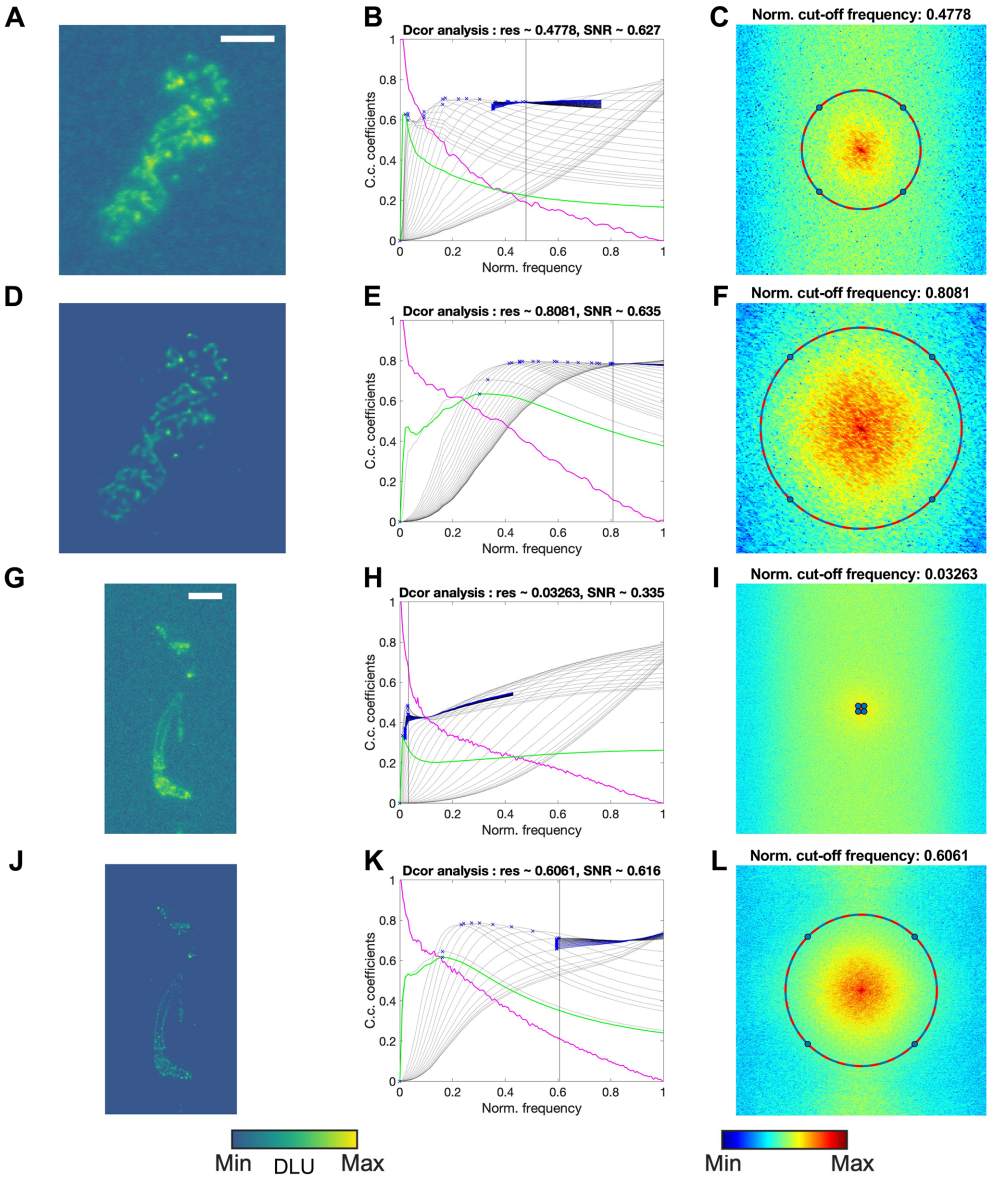
where Y^{est} is the estimated image from restoration algorithms, Y^{true} is the ground truth, $\mu_{Y^{est}}$, $\mu_{Y^{true}}$, $\sigma_{Y^{est}}$, $\sigma_{Y^{true}}$ and $\sigma_{Y^{est}Y^{true}}$ are the local means, standard deviations and cross-covariance for images Y^{est} and Y^{true}, C_1 and C_2 are the regularization constants to avoid instability for image regions where the local mean or standard deviation is close to zero.

Contrast to noise ratio (CNR) is defined as Eq. (S44),

$$\mathrm{CNR} = (C_{sig} - C_{bg})/\sigma_{bg} \tag{S44}$$

where C_{sig} and C_{bg} are the mean of the signal and background and σ_{bg} is the standard deviation of the background. In this metric, the signal area is defined as the high activity region in the DAR images, and the background is extracted by our patch-based estimation method.

Effective resolution is estimated based on the recently published decorrelation-based method [17]. This method estimates the highest frequency with sufficiently high signal in relation to noise (**Supplementary Fig. 4**).



Supplementary Figure 4. Effective resolution estimation by decorrelation analysis. **(A,G)** Two Raw DAR images with low and high noise levels, respectively. **(D,J)** The corresponding PG-PEM restored images of **A** and **G**. **(B,E,H,K)** The corresponding decorrelation analysis of **(A)**,**(D)**,**(G)** and **(J)**. **(C,F,I,L)** The log-scale frequency map of **(A)**,**(D)**,**(G)** and **(J)** labeled with cut-off frequency estimated from **(B)**,**(E)**,**(H)** and **(K)**, respectively. Scale bar: **(A)**: 1.8 mm. **(B)**: 5.3 mm.

Supplementary Note 3: Reference methods

We modified five restoration algorithms as referenced to blind restore DAR images, namely Richardson-Lucy (RL) [6], RL with wavelet-based residue denoising (RD) [18], Shift-Poisson (SP) [12], PG-PEM with no regularization for X (NP) and PG-PEM with TV regularization (TV).

3.1 RL algorithm

RL algorithm is fit for Poisson-distributed data. For DAR image restoration problem, it assumes the imaging model is:

$$R = \alpha \mathcal{P}[X * h + b]. \tag{S45}$$

Based on this model, Eq. (S46) demonstrates the iterative deconvolution scheme without regularization.

$$\begin{aligned} \hat{h}^{(n+1)} &= \hat{h}^{(n)} \cdot \left[\left(\frac{r'}{\hat{h}^{(n)} * \hat{X}^{(n)} + b} \right) * \hat{X}^{(n),m} \right], \\ \hat{X}^{(n+1)} &= \hat{X}^{(n)} \cdot \left[\left(\frac{r'}{\hat{h}^{(n)} * \hat{X}^{(n)} + b} \right) * \hat{h}^{(n),m} \right]. \end{aligned} \tag{S46}$$

3.2 RD algorithm

Different from RL, RD algorithm performs residual denoising during each iteration. In [7, 8], the authors utlized median filter as the denoising algorithm. Not the same as them, wavelet denoising algorithm [18] is applied here. The basic scheme of RD is as Eq. (S47), where $Denoise()$ represents wavelet denoising algorithm.

$$\begin{aligned} r'^{(n)} &= r' - \hat{h} * \hat{X}^{(n)}, \\ \overline{r'^{(n)}} &= Denoise(r'^{(n)}), \\ \hat{h}^{(n+1)} &= \hat{h}^{(n)} \cdot \left[\left(\frac{\hat{h}^{(n)} * \hat{X}^{(n)} + \overline{r'^{(n)}}}{\hat{h}^{(n)} * \hat{X}^{(n)} + b} \right) * \hat{X}^{(n),m} \right], \\ \hat{X}^{(n+1)} &= \hat{X}^{(n)} \cdot \left[\left(\frac{\hat{h}^{(n)} * \hat{X}^{(n)} + \overline{r'^{(n)}}}{\hat{h}^{(n)} * \hat{X}^{(n)} + b} \right) * \hat{h}^{(n),m} \right]. \end{aligned} \tag{S47}$$

In the wavelet denoising algorithm, the input image is decomposed for 7 levels, and Stein's Unbiased Risk Estimate (SURE) and a soft-thresholding approach are conducted.

3.3 SP algorithm

As discussed in Section 1.6, Eq. (S5) can be transformed to a shifted-Poisson distribution:

$$R' + \frac{\sigma_G^2}{\alpha^2} = \mathcal{P}[X * h + b + \frac{\sigma_G^2}{\alpha^2}]. \tag{S48}$$

Based on this equation, the iteration scheme can be conducted as Eq. (S49).

$$\begin{aligned} \hat{h}^{(n+1)} &= \hat{h}^{(n)} \cdot \left[\left(\frac{r' + \frac{\sigma_G^2}{\alpha^2}}{\hat{h}^{(n)} * \hat{X}^{(n)} + b + \frac{\sigma_G^2}{\alpha^2}} \right) * \hat{X}^{(n),m} \right], \\ \hat{X}^{(n+1)} &= \hat{X}^{(n)} \cdot \left[\left(\frac{r' + \frac{\sigma_G^2}{\alpha^2}}{\hat{h}^{(n)} * \hat{X}^{(n)} + b + \frac{\sigma_G^2}{\alpha^2}} \right) * \hat{h}^{(n),m} \right]. \end{aligned} \tag{S49}$$

3.4 NP and TV algorithms

Note that NP and TV have almost the same framework with PG-PEM except the regularization for X. We aim to show the competitive performance of Hessian Frobenius norm regularization in DAR images by comparing it with NP and TV. NP does not have regularization for X while TV algorithm utilizes TV norm as its regularization for X as Eq. (S50), in which $\bigtriangledown$ represents $[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ and $div = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$.

$$\hat{X}^{(n+1)} = \hat{X}^{(n)} \cdot \left[\left(\frac{E_{Q|R'=r',\theta^{(n)}}(Q)}{\hat{h}^{(n)} * \hat{X}_p^{(n)}} \right) * \hat{h}^{(n),m} \right] \Big/ \left[1 - \lambda_X div \left(\frac{\bigtriangledown \hat{X}^{(n)}}{| \bigtriangledown \hat{X}^{(n)} |} \right) \right]. \tag{S50}$$

3.5 Regularization strategies for the reference methods

RL, RD and SP have the same regularization strategies for both h and X with PG-PEM, while NP and TV have the same regularization stragegy for h. Further, similar to PG-PEM, the scaling parameter α here does not impact the estimation result except a multiplier parameter.

Supplementary Note 4: Simulations

4.1 Simulated data generation

We use Eq. (S51) to generate simulated data, in which p is the pixel index, X the ground truth, h the pre-set PSF, b the background, α the scaling factor and $\mathcal{N}(0, \sigma_G^2)$ the Gaussian noise with mean of 0 and standard deviation of σ_G. Here, α and σ_G control the level of Poisson and Gaussian noises, respectively. Note that we divide $X * h + b$ with α to ensure the generated images with the same range with the ground truth.

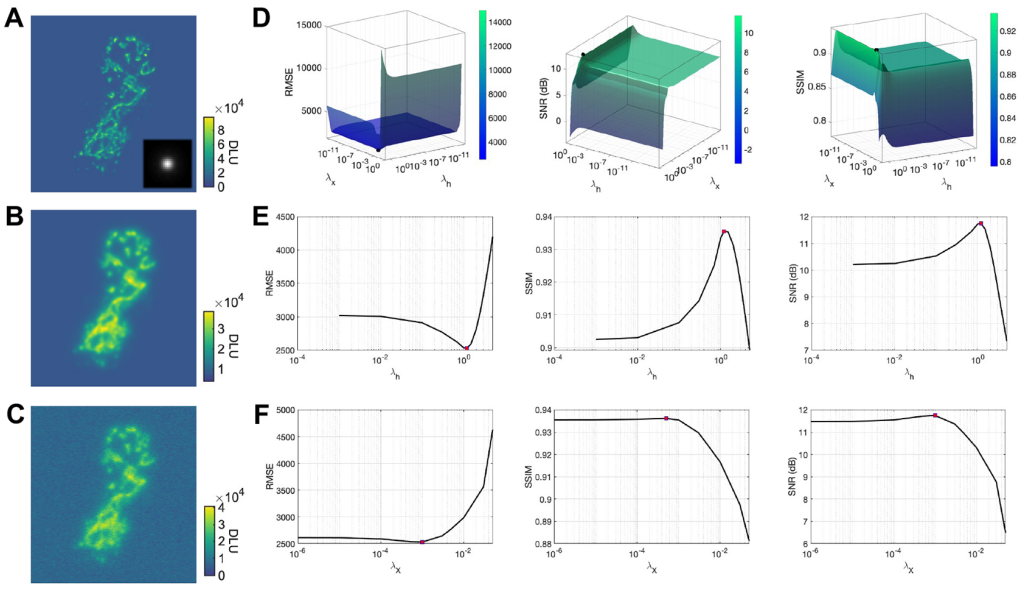
$$R = \alpha \mathcal{P}\left[\frac{X * h + b}{\alpha}\right] + \mathcal{N}(0, \sigma_G^2) \tag{S51}$$

Because it is impossible to acquire a "clean" raw DAR image without noise and blurring effect, the ground truth image needs to be generated. To do so, we selected a raw DAR image with low noise level and blurring effect. Then, it was blindly restored by PG-PEM. The regularization parameters were carefully tuned so that the restored image could achieve its best quality. After restoration, the background of the image was cleared to further remove background noise. The pre-set PSF was generated using a more blurred raw DAR image so that the kernel size of the PSF is larger. To do this, we aim to better test the deblurring ability of the restoration algorithms. The generated ground truth image and PSF are both shown in **Supplementary Fig. 5A**.

4.2 Characterization of the regularization parameters of PG-PEM

To test the impact of regularization parameters, different λ_X (0 to 0.1) and λ_h (0 to 10) were selected to test the performance of the algorithm using a simulated image. To generate the image, the ground truth image was convoluted with the pre-set PSF, to which a constant background was added (b=4000) and then corrupted with Poisson noise (α=20) and Gaussian noise with standard deviation (σ_G) of 1500. The generated image (**Supplementary Fig. 5C**) was restored using the PG-PEM algorithm with the different regularization parameters. The results are shown in **Supplementary Fig. 5D-L**. Meanwhile, several restoration results and their corresponding PSFs with different regularization parameters are shown in **Supplementary Fig. 6**. Here, RMSE, SSIM and SNR are set as the accuracy metrics.

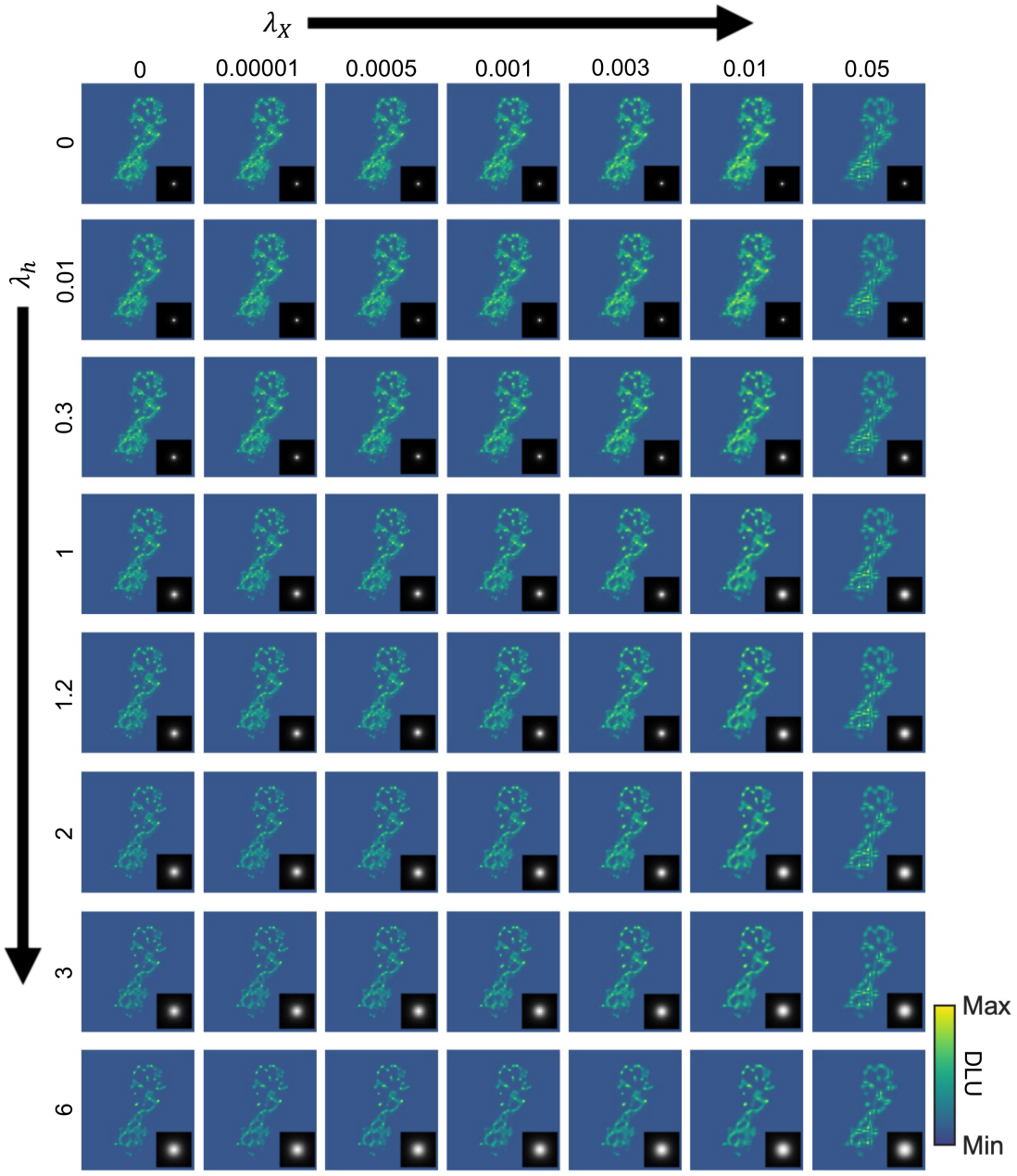
From **Supplementary Fig. 5**, the regularization parameters indeed increases the restoration accuracy compared to the condition without regularizations. The results are very similar across the range of λ_h. For example, in **Supplementary Fig. 5E**, the optimal value for λ_h is approximately 1.2, but the results are very close to each other when λ_h is in [0.7, 1.7]. On the contrary, in **Supplementary Fig. 5F**, the optimal value for λ_X is in a narrow range of approximately 0.0005 to 0.0015. In fact, λ_X should be a small value so as to keep the



Supplementary Figure 5. Data generation for simulation and accuracy metrics for different parameters. **(A)** Ground truth image and corresponding PSF; **(B)** Blurred image with background; **(C)** Noisy image; **(D)** RMSE, SSIM and SNR for all the parameters; **(E)** RMSE, SSIM and SNR with different λ_h when $\lambda_X = 0.001$; **(F)** RMSE, SSIM and SNR with different λ_X when $\lambda_h = 1.2$. Note that in the figures the dots indicate the optimal values.

estimated X from being over-smoothed.

From **Supplementary Fig. 6**, the parameters λ_h and λ_X control the strength of regularization on the PSF h and specimen image X, respectively. With larger λ_X, the restored images tend to become less noisy. But when it is large enough, the restored images will lose fine details and become blurry. Therefore, it is essential to set λ_X properly in order to restore the raw images with a good quality. Combined with the conclusion from **Supplementary Fig. 6**, λ_X is experimentally set as a small value around 0.001 to suppress the noise while preserving the resolution simultaneously. When λ_h=0, the kernel size of h is very small, indicating the estimated h tends to converge as a delta function in this condition. As λ_h increases, the kernel size of the estimated h becomes larger, and the restored image becomes less noisy and blurry. This validates that the regularization is able to ensure its smoothness and to prevent h from converging to a delta function. However, when λ_h is large enough, parts of the restored images become over-smoothed. In this case, the restored images lose some details due to the over-estimated h. This phenomenon can be interpreted in frequency domain (k-space). When λ_h is smaller, the corresponding optical transfer function (OTF) in k-space is broader. Therefore, there will be more high frequency noises and the restoration results are not ideal. On the contrary, when λ_h is larger, the OTF is narrower. As a result, more high frequency signal is filtered out and X looks more blurry. In this sense, the regularization strategy for h is very similar to a Wiener filter. Because λ_h impacts the size of the OTF, it should be positively



Supplementary Figure 6. Restoration images and corresponding PSFs using PG-PEM algorithm with λ_h from 0 to 10 and λ_X from 0 to 0.1. With λ_h increasing, the kernel size of the estimated PSF become larger and larger, and the restored images tend to become less noisy. But when the PSF kernel is large enough, the restored images become blurry. With λ_X increasing, the image noise is also suppressed. But with too large λ_X, the images lose fine details and become blurry.

correlated to the noise level of the input images. For our DAR images, the parameter λ_h is normally set between 0 and 4, determined by the noise level of the input images. In our work, we use the reciprocal of the CNR of the images to reflect their noise levels.

$$\text{Noise level} = \sigma_{bg}/(C_{sig} - C_{bg}), \tag{S52}$$

where C_{sig} and C_{bg} are the mean of the signal and background and σ_{bg} is the standard deviation of the background. The same is done for CNR; the signal area is defined as the high activity region in the DAR images, and the background is extracted by our patch-based estimation method. Reasonably, this can reflect the normalized noise after considering both the noise power and the averaged signal energy. We also found that λ_h grows slower with the noise level increasing because the E step in Eq. (S15) can also suppress parts of the noise. We thus emperically set λ_h as:

$$\lambda_h = 4\sqrt{\text{Noise level}}. \tag{S53}$$

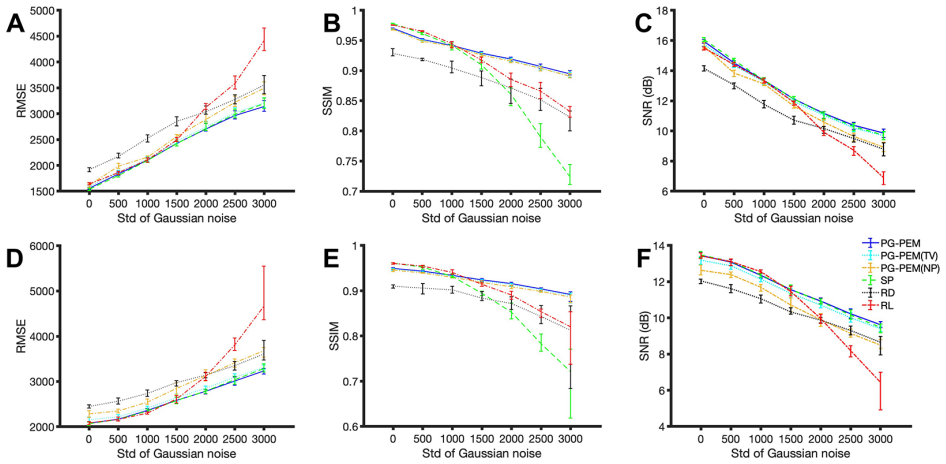
In fact, according to the results in **Supplementary Fig. 5 and 6**, the restored data are quite similar along a range of λ_h, which brings more flexibility to our empirical setting of λ_h in Eq. (S53). We can further tune the λ_h manually, for example, by visual assessment or with the help of the decorrelation analysis method [17]. The most optimal λ_h should correspond to the point with the best signal-noise trade-off.

4.3 Comparison with the reference algorithms

Next, to evaluate the denoising performance of PG-PEM algorithm, the simulated image had added to it a constant background (b=4000), blurred by the same PSF again and corrupted with two different levels of Poisson noise (α=20 and 100) and Gaussian noise with different standard deviations (σ_G) from 0 to 3000 with interval of 500. We generated 10 groups of data for each noise level. In this way, by generating simulated images with different Poisson and Gaussian noise levels, we aim to have a thorough comparison of the algorithms under different conditions.

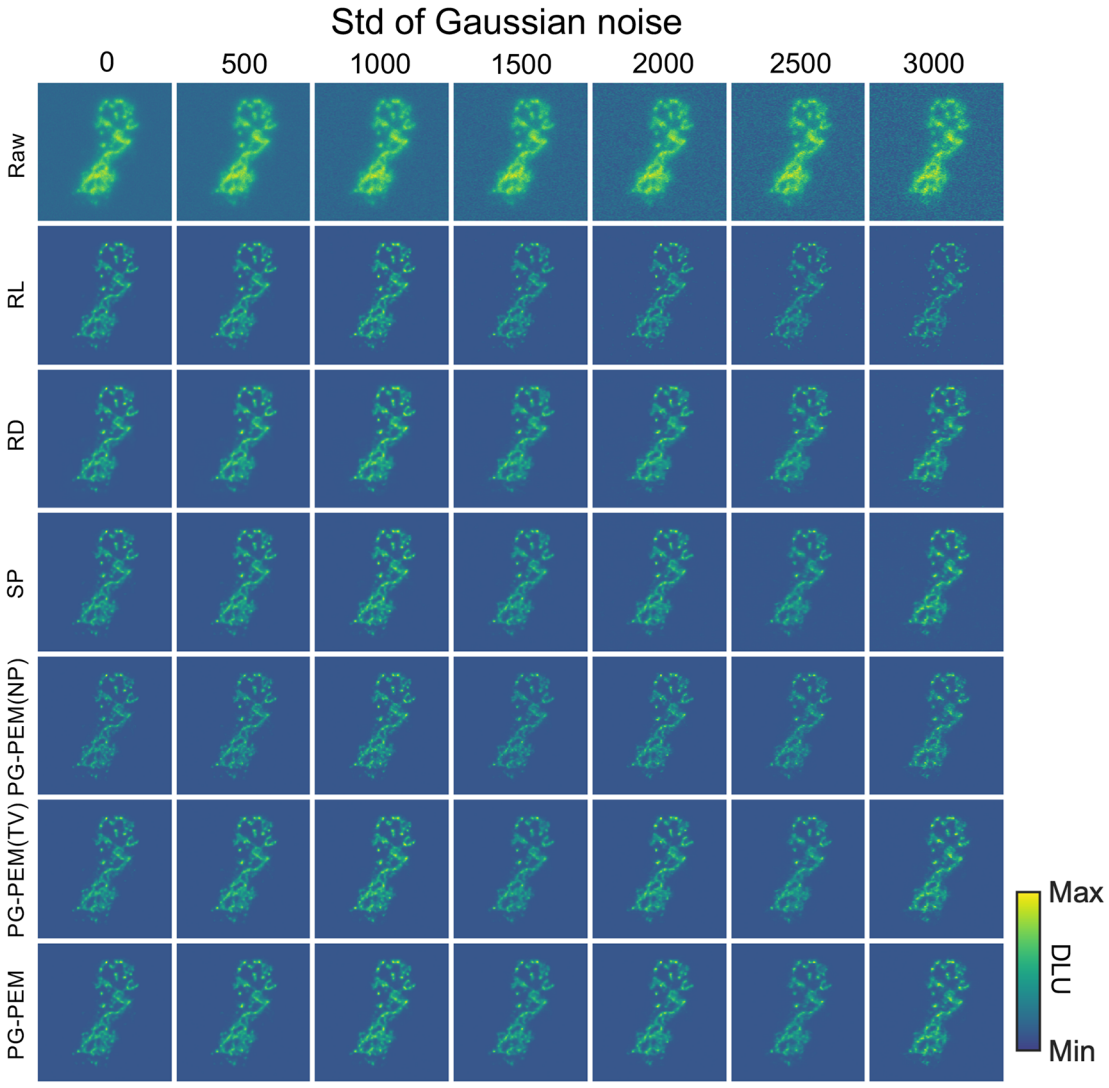
The PG-PEM algorithm was compared with RL, RD, SP, NP and TV. In this simulation, λ_X is set as 0.001. RL, RD and SP shared the same λ_X but different λ_h to make sure the PSFs of these algorithms have the same shape. We did this to eliminate the impact of PSF and only focused on the different frameworks of these models. RMSE, SSIM and SNR were used to compare the restoration performance between different algorithms. The results are shown in **Supplementary Fig. 7**.

$$R' = \mathcal{P}[X * h] + \gamma\mathcal{P}(b/\gamma) \tag{S54}$$

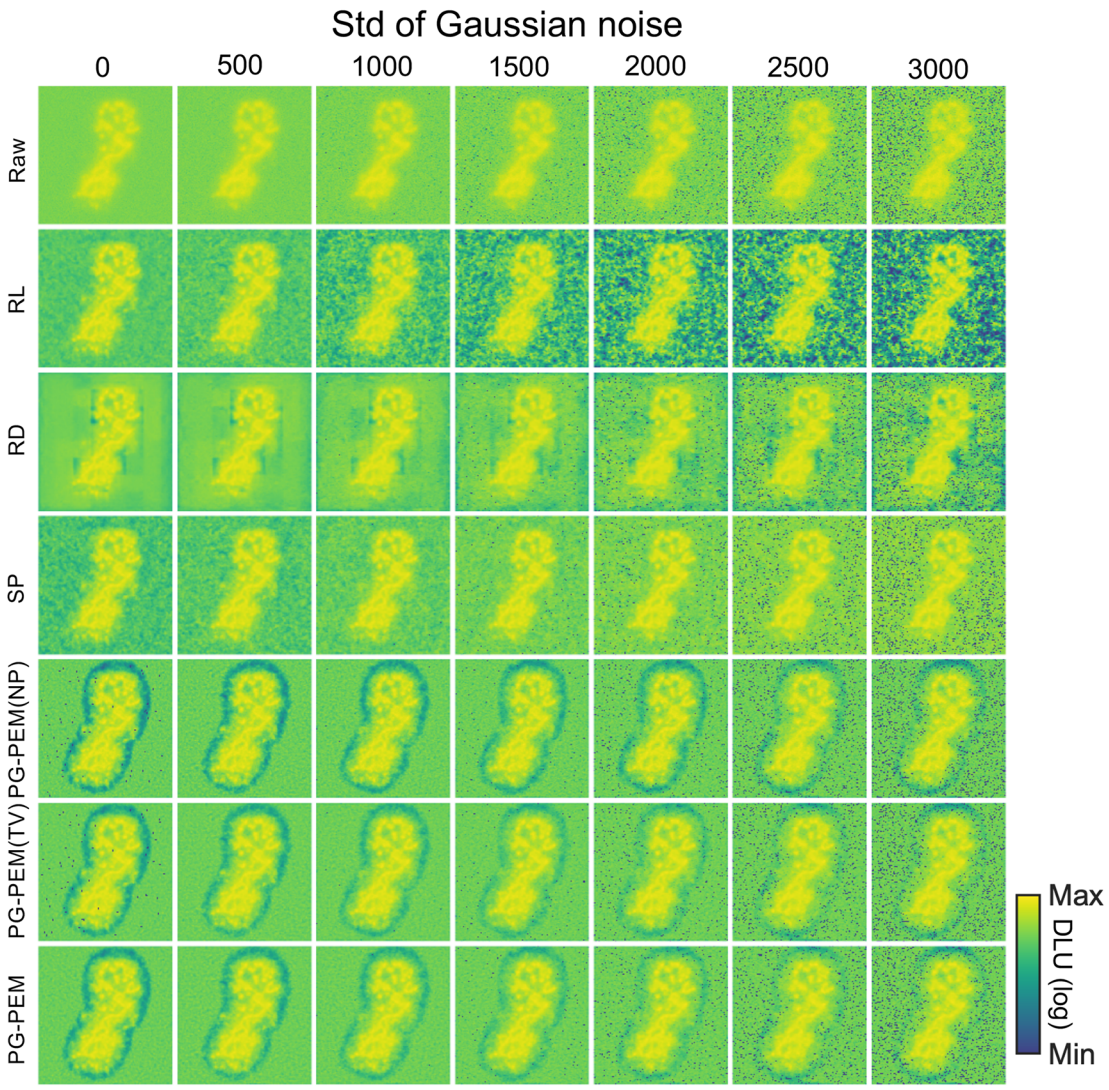


Supplementary Figure 7. Evaluation results for different parameters and methods when α=20 and 100. **(A)** RMSE evaluation when α=20; **(B)** SSIM evaluation when α=20; **(C)** SNR evaluation when α=20; **(D)** RMSE evaluation when α=100; **(E)** SSIM evaluation when α=100; **(F)** SNR evaluation when α=100. From the figures, PG-PEM is the best performer among all the conditions regarding RMSE, SSIM and SNR.

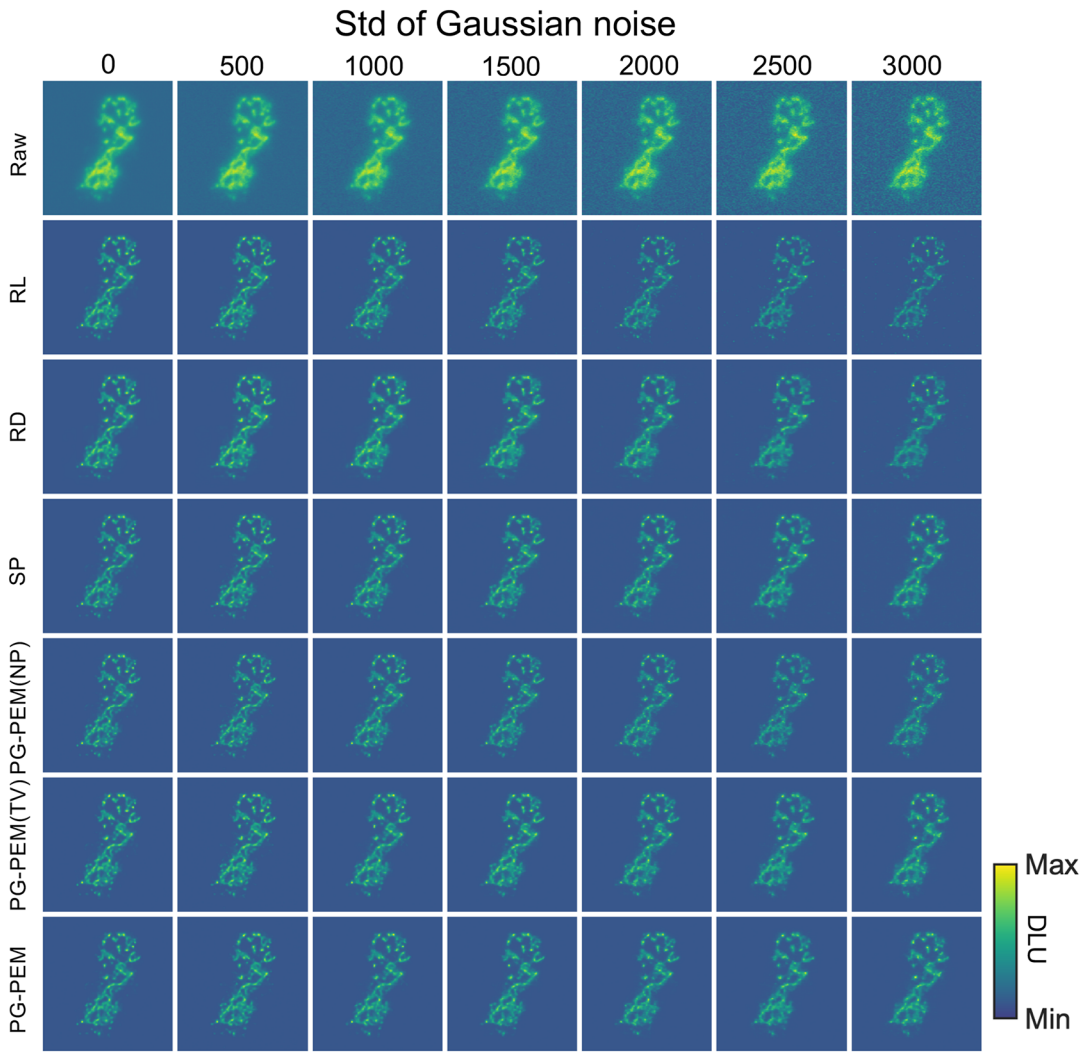
From **Supplementary Fig. 7**, in the parametric space where the standard deviation of the Gaussian noise is small, RL outperforms RD and is similar to SP and PG-PEM. However, with additional noise, RD algorithm performs better than RL. This can be interpreted using Eq. (S54) estimated from Eq. (S5), where $(\gamma-1)b = \sigma_G^2/\alpha^2$. When σ^2 is small, the Gaussian noise part can be approximated as a Poisson distribution; while σ^2 is large enough, the Gaussian distribution will be truncated. Therefore, the approximation will cause large errors. This is why as Gaussian noise increases above a level, the performance of RL decreases dramatically. In aspects of RMSE and SNR, the performances of SP and PG-PEM are comparable, both of which are not worse than those of RL and RD. When the level of Gaussian noise is higher, the SSIM of PG-PEM is much better than that of SP. Next, NP, TV and PG-PEM were compared to evaluate the different regularization strategies for X. For all the conditions with different noise levels, PG-PEM always outperforms TV and NP is the worst of all. In summary, PG-PEM is the best performer in aspects of RMSE, SSIM and SNR for simulated image data restoration. Parts of the restored images with their log-scale versions are shown in **Supplementary Figs. 8-11**. The same as the results from the accuracy metrics, PG-PEM outperforms alternative methods by means of visual inspection.



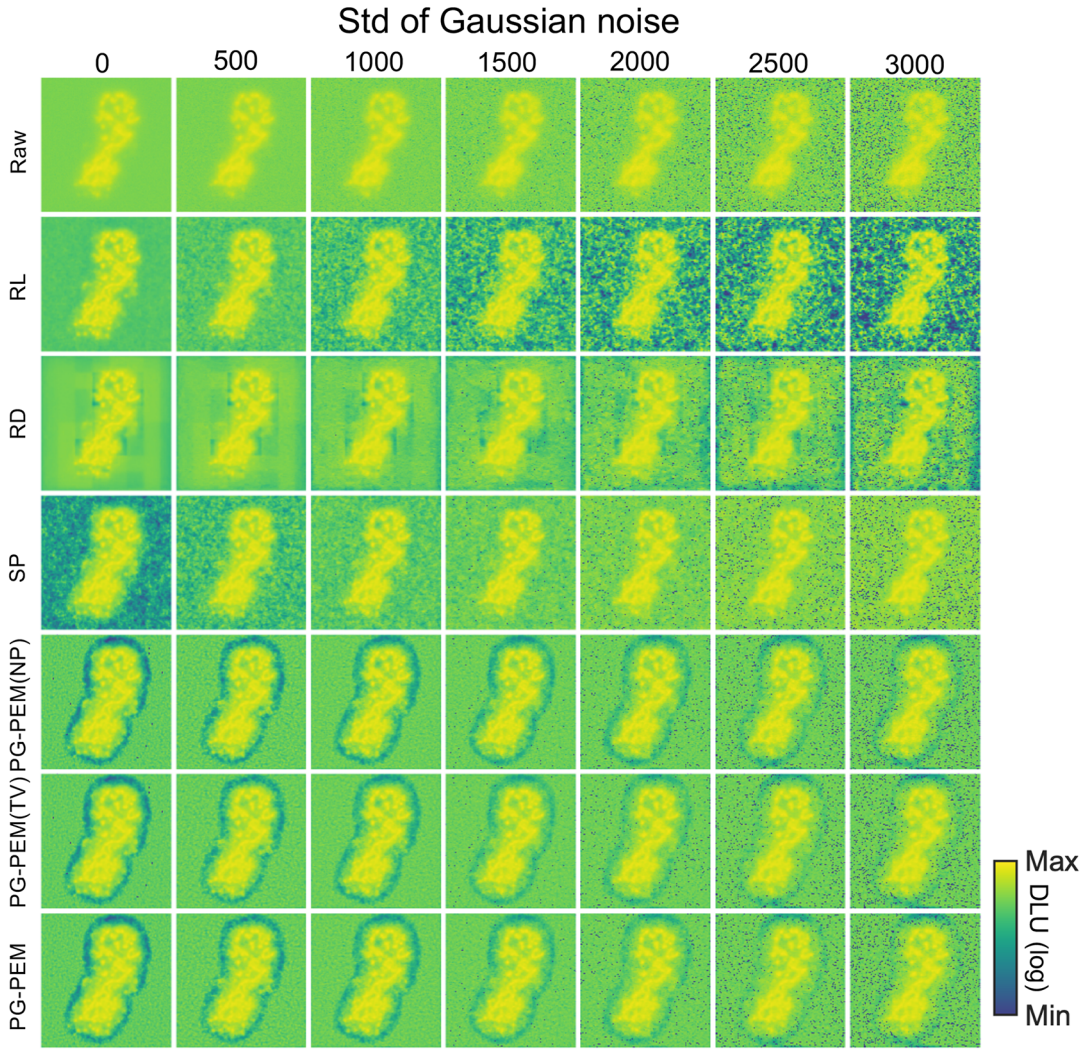
Supplementary Figure 8. Restoration results from different noise levels and methods when α=20.



Supplementary Figure 9. Restoration results from different noise levels and methods when α=20 (log scale). It is easier to compare the denoising ability between different methods using the log-scale images.

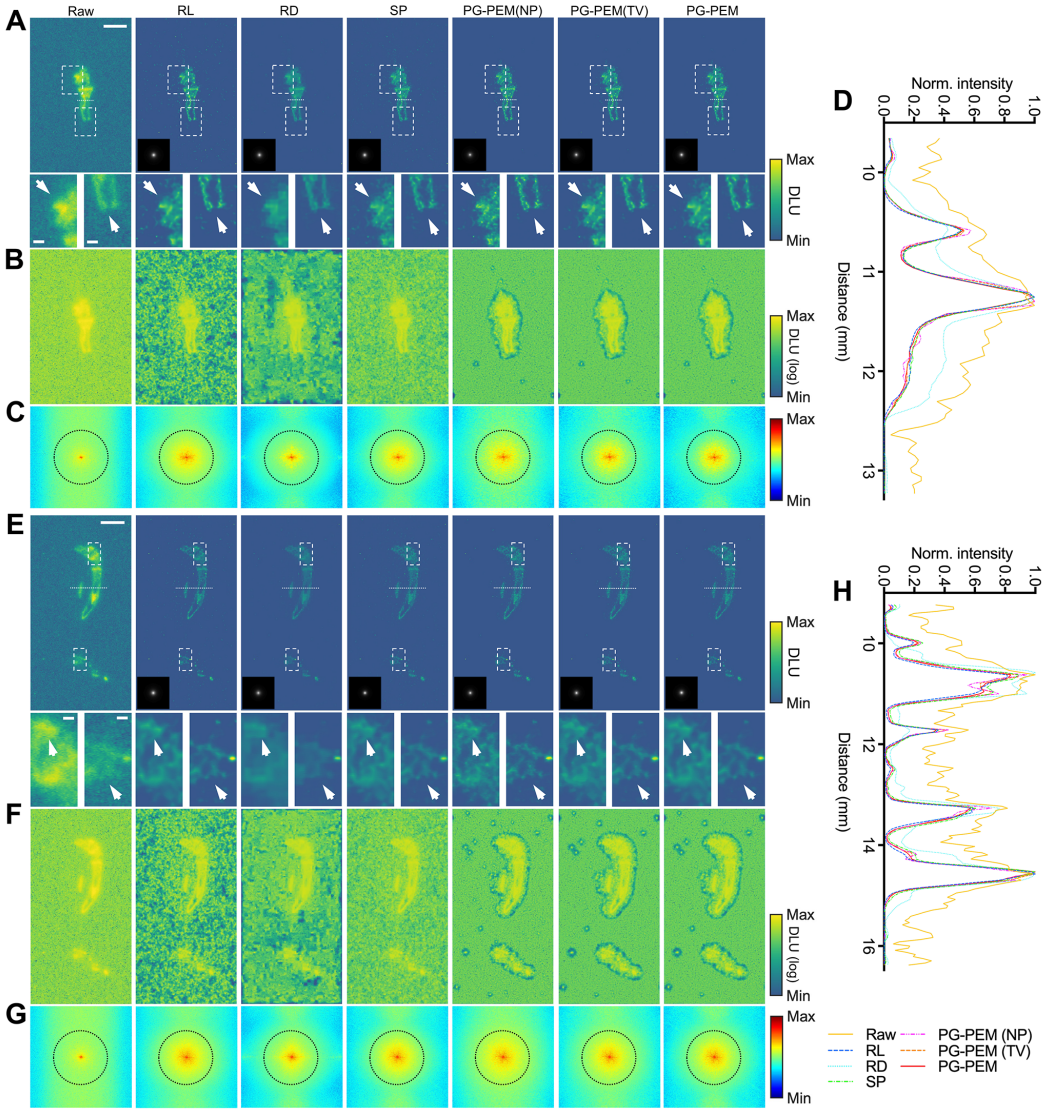


Supplementary Figure 10. Restoration results from different noise levels and methods when α=100.

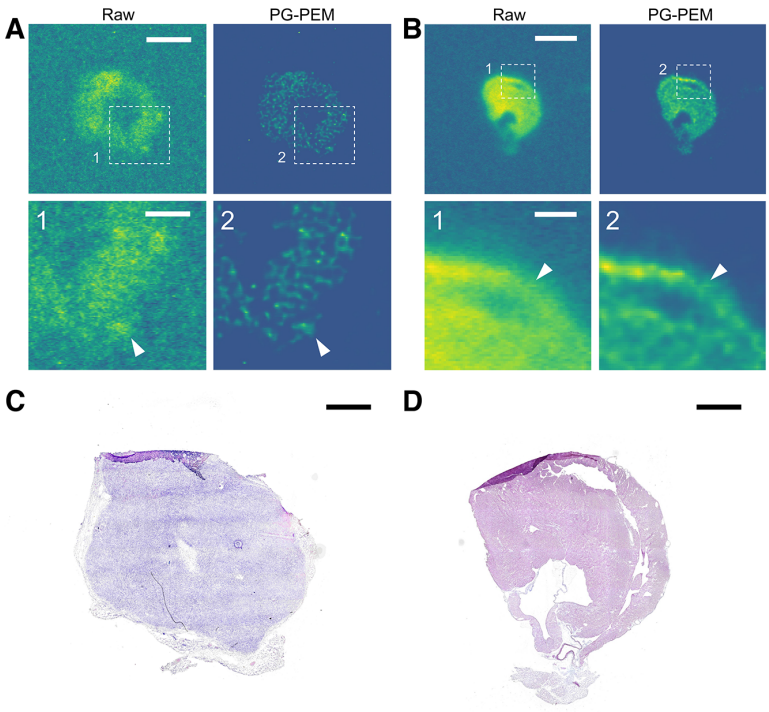


Supplementary Figure 11. Restoration results from different noise levels and methods when α=100 (log scale). It is easier to compare the denoising ability between different methods using the log-scale images.

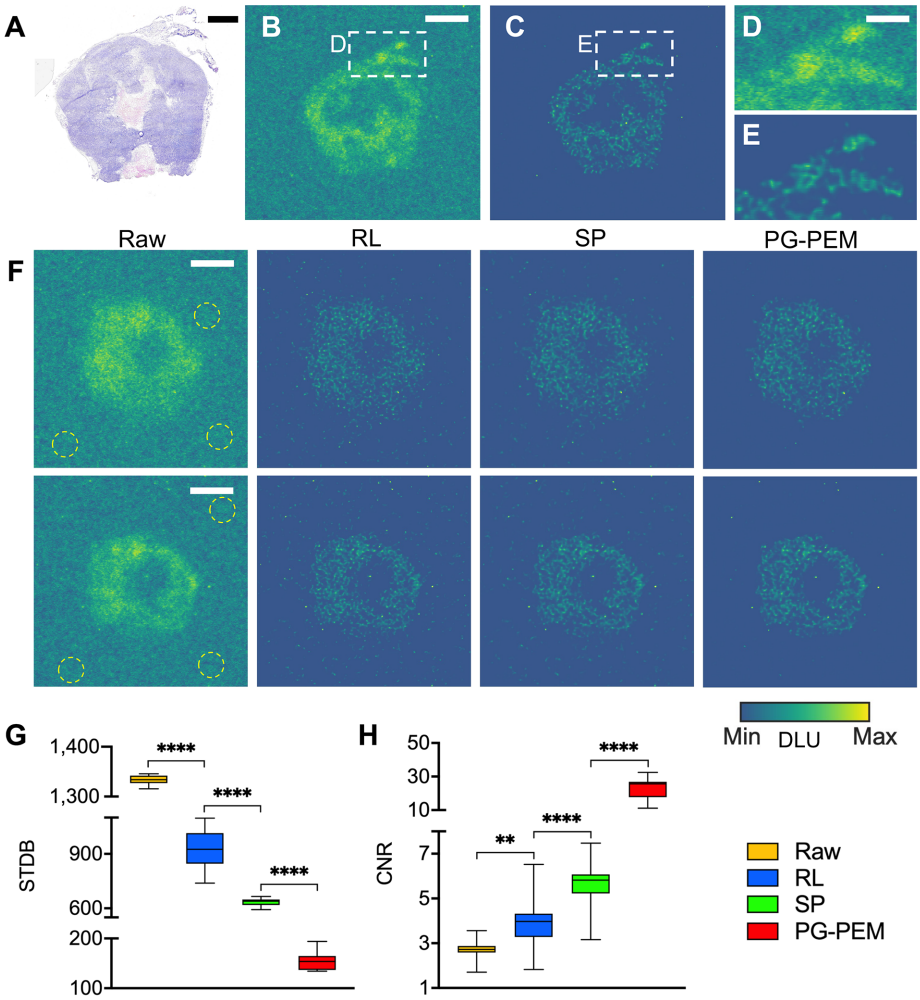
Supplementary Figures 12-21



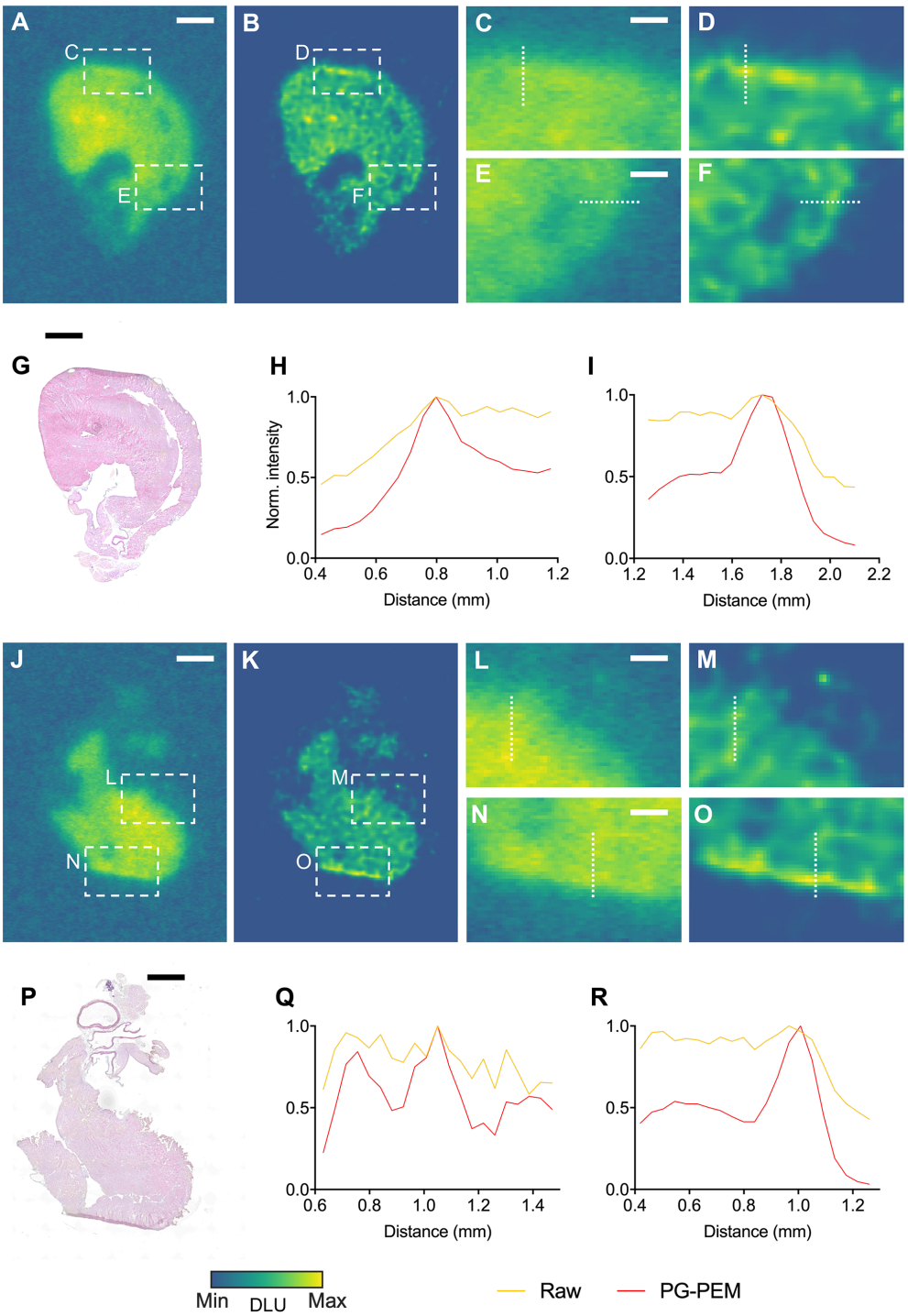
Supplementary Figure 12. Assessment of PG-PEM in comparison with alternative blind restoration algorithms. **(A,E)** Raw DAR images from the mouse hindlimb following [^{18}F]NaF PET imaging and the restoration results with the corresponding PSFs using modified restoration algorithms, PSF estimation and patch-based segmentation. The corresponding estimated PSFs are inset by grayscale. The lower images are the magnificiation of dash-outlined regions from the raw and restored images in **(A,E)**. **(B,F)** Log-scale transformed images from **(A,E)** for background appraisal. **(C,G)** Log-scale magnitude of the Fourier transform of the raw and restored images in **(A,E)**, respectively. **(D,H)** Line profiles of the images in **(A,E)** along the corresponding dashed lines. Scale bar: **(A)**: upper images: 4.7 mm, lower images: 1.4 mm. **(E)**: upper images: 4.85 mm, lower images: 0.59 mm.



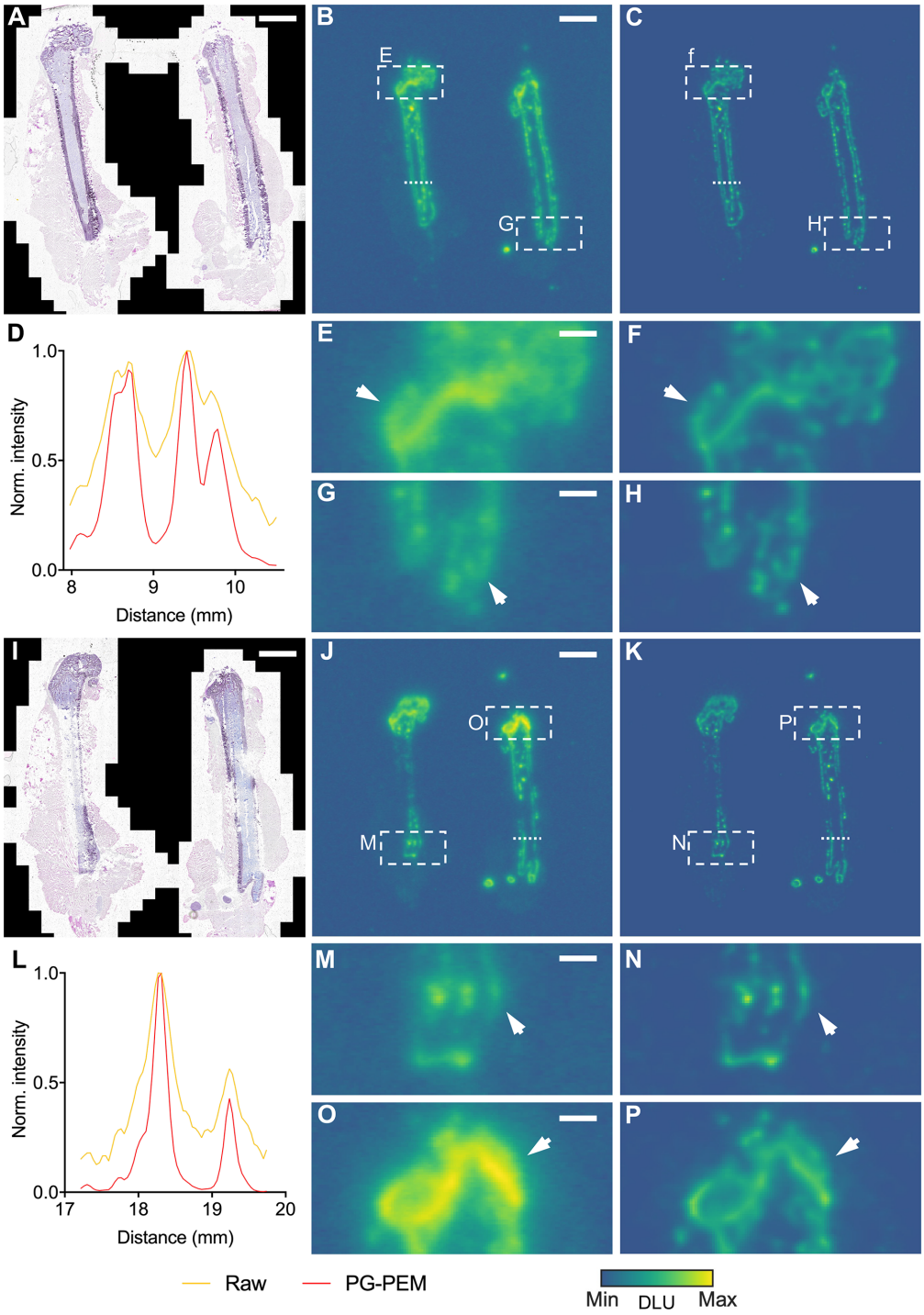
Supplementary Figure 13. Content-adaptive blind restoration improves preclinical DAR. **(A)**, **(B)** Raw and PG-PEM restored DAR images of mice tumor and heart treated with ^{18}F-FDG. **(C)**, **(D)** The corresponding H&E stained tumor and heart tissues. Scale bar: **(A)**: 4.2 mm, **(A1)**: 1.4 mm; **(B)**: 3.7 mm, **(B1)**: 0.75 mm; **(C)**: 1.5 mm; **(D)**: 1.25 mm.



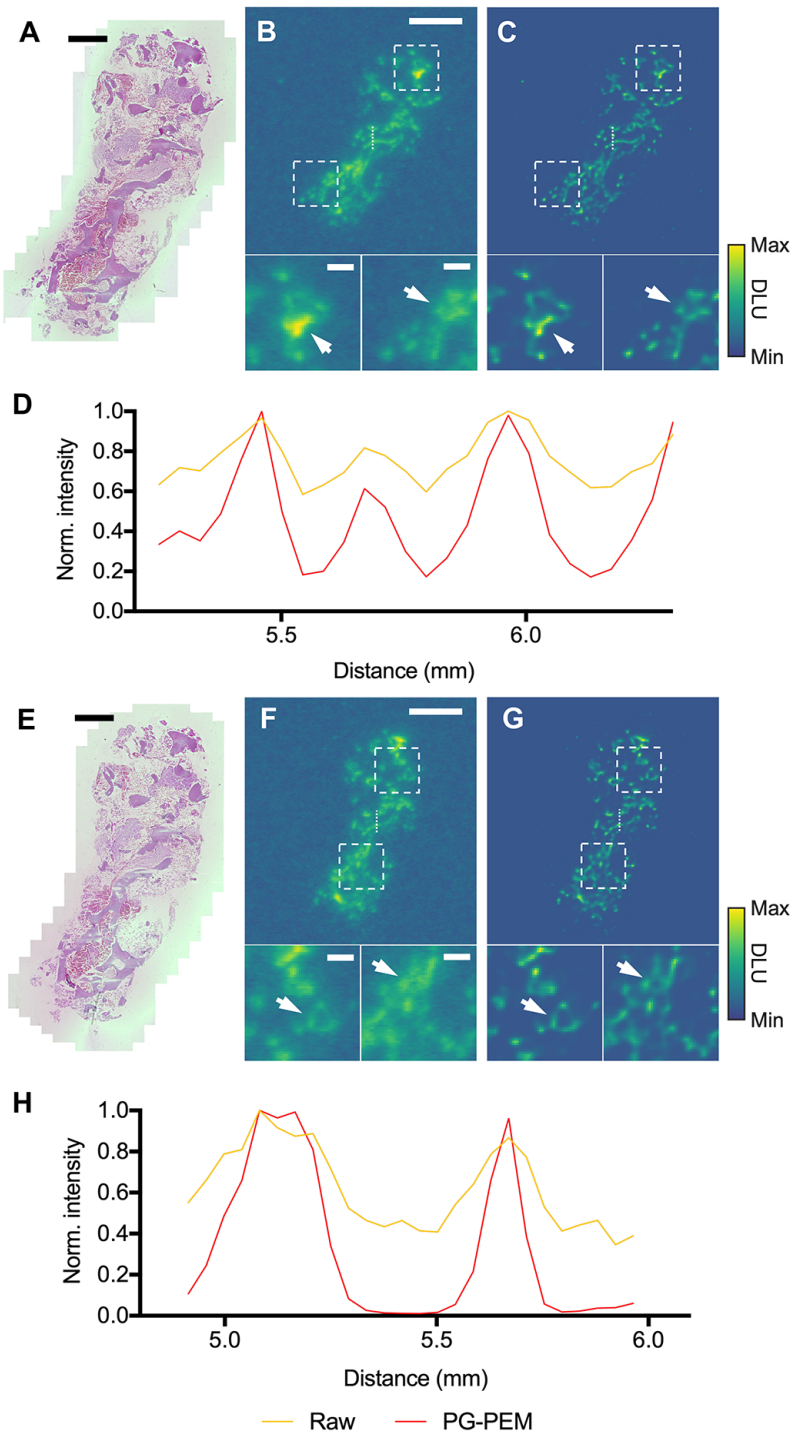
Supplementary Figure 14. PG-PEM outperforms RL and SP in the restoration of raw DAR images from [^{18}F]FDG uptake by xenograft tumors. **(A)** A H&E stained tumor tissue image. **(B)** The corresponding raw DAR image of **(A)**. **(C)** The restored image of **(B)** using PG-PEM. **(D,E)** Magnified regions of the two dashed areas in **(B,C)**. **(F)** Another two groups of raw DAR images with the corresponding restored results using RL, SP and PG-PEM. By observing the dashed yellow circles in the raw images and the corresponding regions in the restored results of RL, SP and PG-PEM, PG-PEM can effectively suppress the noise in the circles while RL and SP cannot. **(G,H)** STDB and CNR assessment of totally 10 groups of DAR images. These results demonstrate that PG-PEM performs significantly better than RL and SP in denoising. Scale bar: **(A)**: 2 mm. **(B)**: 3.6 mm. **(D)**: 1.5 mm. **(F)**: first row: 3.3 mm, second row: 3.28 mm.



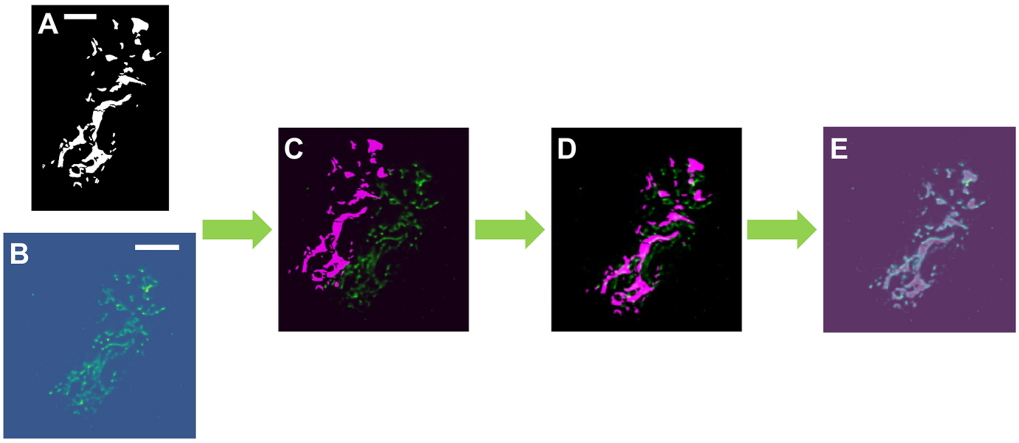
Supplementary Figure 15. PG-PEM improves the DAR image quality from [^{18}F]FDG administered murine heart. **(A,J)** Raw DAR images of mice heart sections. **(B,K)** Corresponding restored images of **(A,J)**. **(C–F)** Zoomed-in regions of the four dashed boxes in **(A,B)**. **(L–O)** Magnified Regions of Dashed Box Regions in **(J,K)**. **(G,P)** H&E stained tissue images corresponding to **(A,J)**. **(H,I,Q and R)** Line profile comparison along the corresponding dashed lines in **(C,E,L and N)**. Scale bar: **(A)**: 2.4 mm. **(C,E)**: 0.75 mm. **(G)**: 1.2 mm. **(J)**: 2.7 mm. **(L,N)**: 0.84 mm. **(P)**: 1 mm.



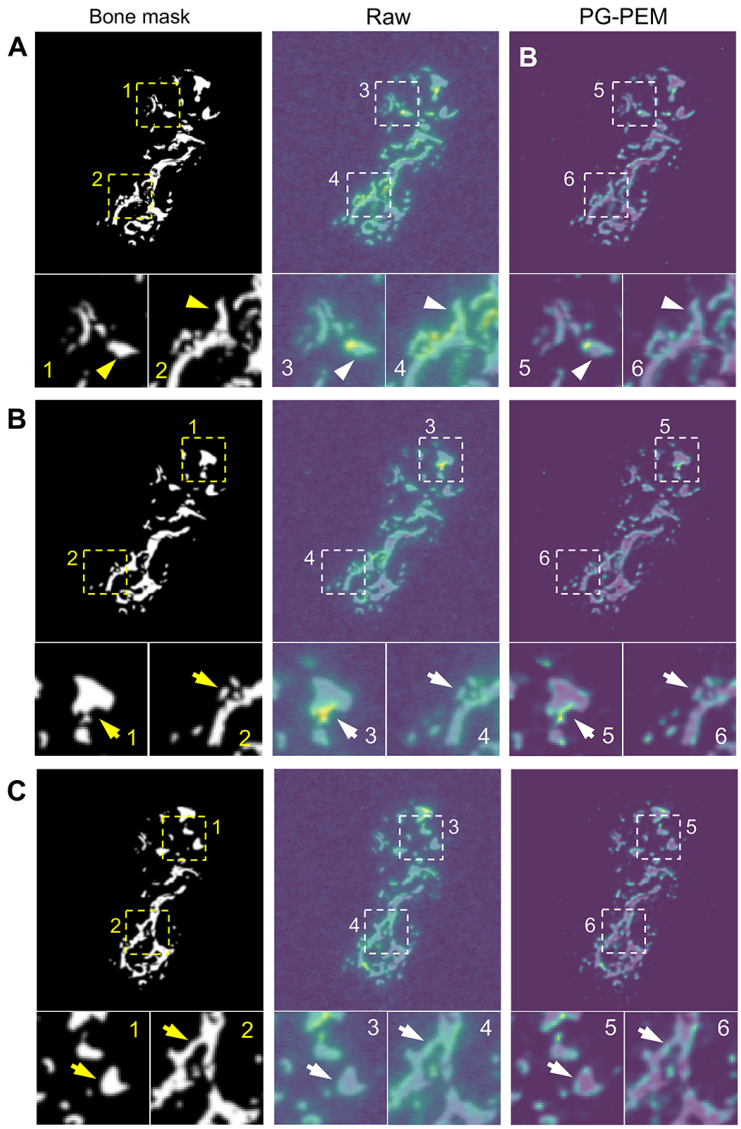
Supplementary Figure 16. PG-PEM improves DAR image quality from [^{18}F]NaF administered mouse femurs. **(A,I)** H&E stained mice femur images. **(B,J)** Corresponding raw DAR images of **(A,I)**. **(C,K)** The restored DAR images of **(B,J)** by PG-PEM. **(D,L)** Line profile comparison along the corresponding dashed lines in **(B,J)**. **(E–H)** Zoomed-in regions of the four dashed boxes in **(B,C)**. **(M–P)** Zoomed-in regions of the four dashed boxes in **(J,K)**. Scale bar: **(A)**: 2.25 mm. **(B)**: 3.15 mm. **(E,G)**: 0.67 mm. **(I)**: 2.3 mm. **(J)**: 3.2 mm. **(M,O)**: 0.68 mm.



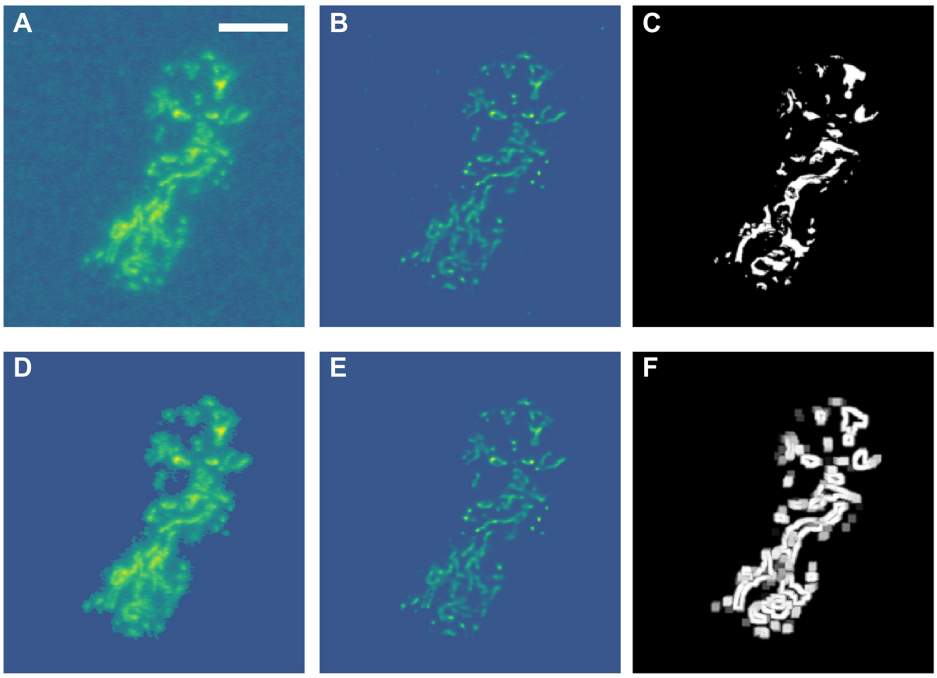
Supplementary Figure 17. PG-PEM improves DAR images of alpha particle radiotherapy. **(A,E)** H&E stained histopathological images of two human bone biopsy sections. **(B,F)** Corresponding Raw DAR images of the same section in **(A,E)**. **(C,G)** PG-PEM restored images of **(B,F)**. In **(B,C,F,G)**, the lower two images are the zoomed-in regions of the two dashed boxes in the upper images. **(D,H)** Line profiles in **(B,C,F,G)** along the corresponding dashed lines. Scale bar: **(A)**: 1 mm. **(E)**: 1.1 mm. **(B,F)**: first row: 2.3 mm, second row: 0.44 mm.



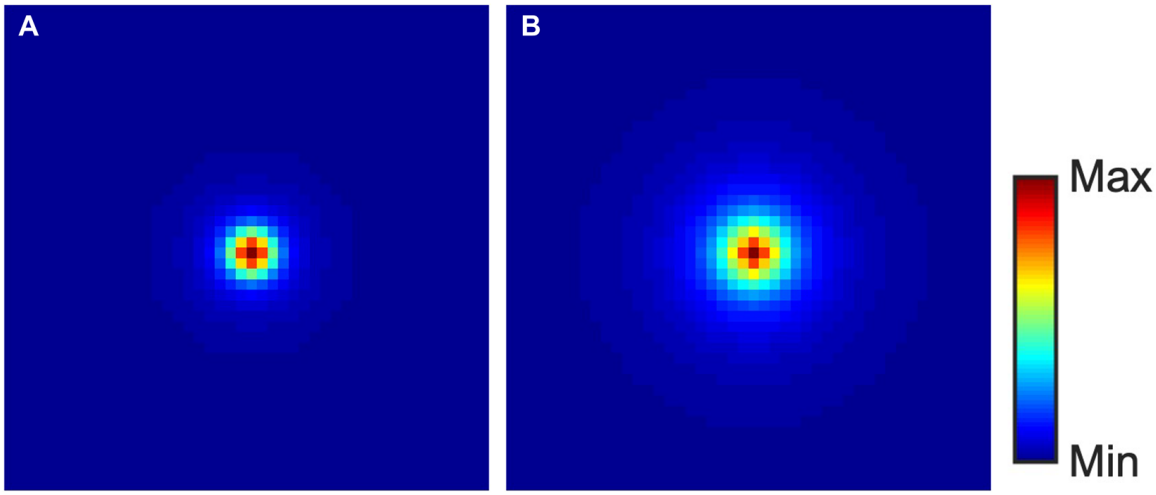
Supplementary Figure 18. Image resgistration process. **(A)** The segmented bone mask from a H&E stained histopathological image. **(B)** The corresponding PG-PEM restored DAR image. **(C)** Fused image before registration. **(D)** Initial registration: the bounding boxes of the two images are first found. Then the centers, areas and angles of the two boxes can be estimated. Next, the initial registration can be achieved by registering these two boxes including scaling, rotation and translation. **(E)** Final registration: maximize the mutual information between the two modalities by scaling, rotation and translation. Because the final registration algorithm is very easy to converge to a local minimum, the initial registration is essential. In our registration, we made the DAR image the fixed image and the bone mask the moving one. This is reasonable because 1) the pixel number of the bone mask is much bigger than that of the DAR image, and 2) the activity of the DAR image will be re-sampled and not the same as the original distribution if it is the moving image. Scale bar: **(A)**: 1 mm. **(B)**: 2.3 mm.



Supplementary Figure 19. Manually segmented bone mask, co-registered images of bone mask and DAR before and after restored by PG-PEM. (A) Figure 6. (B) Supplementary Fig 17(A)–(C). (C) Supplementary Fig 17(E)–(G).



Supplementary Figure 20. SSIM and Fusion index estimation. **(A)** A raw DAR image of the human bone biopsy section. **(B)** The corresponding PG-PEM restored image. **(C)** The segmented bone mask from the corresponding H&E stained histopathological image. **(D)** High activity region of **(A)**. **(E)** High activity region of **(B)**. The estimation of the high activity region is achieved by a simple threshold method followed by removing small region. **(F)** The estimated bone surface from **(C)** using morphological operations: The bone mask is dilated and eroded by a small disk separately. Then the bone surface is estimated by subtracting the two results. We compare **(A)**, **(B)** with **(F)** to compute Fusion index and **(D)**, **(E)** with **(C)** to compute SSIM. Scale bar: **(A)**: 2.3 mm.



Supplementary Figure 21. PSF comparison. **(A)** The estimated PSF from a $^{223}RaCl_2$ treated human bone biopsy image. **(B)** The estimated PSF from a ^{18}F-NaF treated mouse hindlimb image. To increase the image contrast, we have applied a false coloured "jet" colormap. The estimated full-width half-maximum and full-width tenth-maximum of **(A)** are approximately 0.159 mm and 0.344 mm, while those of **(B)** are 0.216 mm and 0.66 mm, respectively. These results are consistent with the physics of alpha/positron transport: alpha particles have a significantly shorter path length than positrons. This is also why the raw autoradiographic images from ^{223}Ra treated human bone biopsy have higher resolution than those from ^{18}F treated sections. It should be noted that while several beta particles are produced by its daughters, the vast majority of the ^{223}Ra decay energy is in the form of alpha particles.

References

1. Janesick JR. Photon transfer noise sources. In: *Photon transfer DN→ λ.* Bellingham, Wash: SPIE Press; 2007: 21–25.

2. Mandracchia B, Hua X, Guo C, Son J, Urner T, Jia S. Fast and accurate sCMOS noise correction for fluorescence microscopy. *Nat Commun.* 2020;11:94.

3. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *KDD.* AAAI; 1996:226–231.

4. Crain BR. Estimating the parameters of a truncated normal distribution. *Appl Math Comput.* 1979;5:149–156.

5. Chouzenoux E, Jezierska A, Pesquet JC, Talbot H. A convex approach for image restoration with exact Poisson-Gaussian likelihood. *SIAM J Imaging Sci.* 2015;8:2662–2682.

6. Fish D, Brinicombe A, Pike E, Walker J. Blind deconvolution by means of the Richardson-Lucy algorithm. *J Opt Soc Am A Opt Image Sci Vis.* 1995;12:58–65.

7. Kenig T, Kam Z, Feuer A. Blind image deconvolution using machine learning for three-dimensional microscopy. *IEEE Trans Pattern Anal Mach Intell.* 2010;32:2191–2204.

8. Keuper M, Schmidt T, Temerinac-Ott M, *et al.* Blind deconvolution of widefield fluorescence microscopic data by regularization of the optical transfer function (OTF). In: *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* IEEE; 2013:2179–2186.

9. Koho S, Tortarolo G, Castello M, Deguchi T, Diaspro A, Vicidomini G. Fourier ring correlation simplifies image restoration in fluorescence microscopy. *Nat Commun.* 2019;10:3103.

10. Lefkimmiatis S, Bourquard A, Unser M. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Trans Image Process.* 2011;21:983–995.

11. Huang X, Fan J, Li L, *et al.* Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. *Nat Biotechnol.* 2018;36:451–459.

12. Ikoma H, Broxton M, Kudo T, Wetzstein G. A convex 3D deconvolution algorithm for low photon count fluorescence imaging. *Sci Rep.* 2018;8:11489.

13. Brune C, Sawatzky A, Burger M. Bregman-EM-TV methods with application to optical nanoscopy. In: *SSVM.* Elsevier; 2009:235–246.

14. Resmerita E, Anderssen RS. Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems. *Math Models Methods Appl Sci.* 2007;30:1527–1544.

15. Brownson JR. Chapter 03-Laws of light. In: *Solar Energy Conversion Systems.* Boston: Academic Press; 2014: 41–66.

16. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* 2004;13:600–612.

17. Descloux AC, Grussmayer KS, Radenovic A. Parameter-free image resolution estimation based on decorrelation analysis. *Nat Methods.* 2019;16:918–924.

18. Boussion N, Le Rest CC, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. *Eur J Nucl Med Mol Imaging.* 2009;36:1064–1075.