

Single-sample image-fusion upsampling of fluorescence lifetime images

V.Kapitány^{†,1,*}, A.Fatima^{†,1}, V.Zickus^{1,2}, J.Whitelaw^{3,4}, E.McGhee^{1,3}, R.Insall^{3,5}, L.Machesky^{3,6}, D.Faccio^{1,*}

¹ School of Physics & Astronomy, University of Glasgow, Glasgow G12 8QQ, UK

² Department of Laser Technologies, Center for Physical Sciences and Technology, LT-10257, Vilnius, Lithuania

³Cancer Research UK, Beatson Institute, Glasgow, United Kingdom

⁴Present: School of Health & Life Sciences, University of the West of Scotland, Blantyre G72 0LH

⁵Present: Department of Cell and Developmental Biology,
University College London, London WC1E 6BT

⁶Present: Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA

[†]These authors contributed equally to the work

* Corresponding authors: valentin.kapitany@glasgow.ac.uk, daniele.faccio@glasgow.ac.uk

Fluorescence lifetime imaging microscopy (FLIM) provides detailed information about molecular interactions and biological processes. A major bottleneck for FLIM is image resolution at high acquisition speeds, due to the engineering and signal-processing limitations of time-resolved imaging technology. Here we present single-sample image-fusion upsampling (SiSIFUS), a data-fusion approach to computational FLIM super-resolution that combines measurements from a low-resolution time-resolved detector (that measures photon arrival time) and a high-resolution camera (that measures intensity only). To solve this otherwise ill-posed inverse retrieval problem, we introduce statistically informed priors that encode local and global dependencies between the two “single-sample” measurements. This bypasses the risk of out-of-distribution hallucination as in traditional data-driven approaches and delivers enhanced images compared for example to standard bilinear interpolation. The general approach laid out by SiSIFUS can be applied to other image super-resolution problems where two different datasets are available.

I. INTRODUCTION

Fluorescence lifetime imaging microscopy (FLIM) finds extensive applications in biological studies where the lifetimes of fluorophores can be used as indicators of the cellular metabolism [1–5], cellular environment [6–8] or changes in molecular conformation visible through Förster resonance energy transfer (FRET), enabling measurement of protein:protein interactions during processes such as cellular signalling [9–12]. In medical settings, endogenous FLIM can be used for identifying cancerous tissue [13, 14].

FLIM setups excite a sample with short-wavelength light and measure the temporal profile of long-wavelength fluorescence from the sample [15]. Excitation is achieved using a pulsed or amplitude modulated laser for timedomain and frequency domain FLIM, respectively [11], while emission is usually collected with time correlated single photon counting (TCSPC) or time-gated hardware. Fluorescence lifetime is then recovered from the temporal decay of fluorescence emission. Popular lifetime estimation schemes include least squares deconvolution [16], Laguerre expansion [17], phasor fitting [2, 3], rapid lifetime determination [18, 19], centre-of-mass estimation [20, 21] and machine learning [22–24].

Images are formed through raster-scanning or wide-field detection. Scanning systems allow confocal or 2-photon microscopy setups, giving excellent image resolution, and aligning well with TCSPC methods that give rich fluorescence information. However, scanning also presents drawbacks, such as the lack of instantaneous complete field-of-view information, and long acquisition times which are incompatible with the rapid intracellular dynamics of living cells [25, 26]. Widefield systems overcome these challenges by measuring temporal decay from the full field of view in parallel, often using time-gated cameras like intensified

charge-coupled devices (iCCDs) [27, 28], externally gated devices [29, 30], or single photon avalanche diode (SPAD) arrays [23, 31]. However, iCCD resolution is limited by the intensifier point-spread-function, whilst SPAD arrays typically have low-pixel counts and/or low fill-factors.

Computational super-resolution (SR) provides a route to overcome the trade-off between acquisition time and spatial resolution by offloading imaging from optics onto software. SR takes an undersampled image of a scene and estimates its high-resolution features. Multiple flavours of SR exist, which are generally either interpolation, reconstruction (inverse retrieval) or example- (learning) based.

Interpolation is the simplest form of upsampling, encompassing several methods for connecting data-points with some curve [32]. For images, this ranges from simple schemes like nearest, bilinear and bicubic interpolation, through frequency-based approaches sinc and Lanczos interpolation, to covariance-based algorithms like kriging (Gaussian processes) [33]. While interpolation is fast and computationally inexpensive, it does not add new information to the image.

Reconstruction-based modeling instead manipulates the detection to optically redistribute information about the high-resolution target into fewer measurements. This encoding provides a mathematical forward model that is employed to reconstruct the non-sampled points in an inverse retrieval framework, for example via point spread function (PSF) engineering [34], blurring [35], or compressed sensing [36–39].

Lastly, example-based schemes rely on computation and pattern recognition to upsample images in a data driven manner [40]. Classical approaches include neighbour embedding [41], sparse coding [42] and anchored neighbourhood regression [43]. More recently, machine learning algorithms have seen widespread adoption for super-resolution [44, 45]. These range

from super-resolution convolution neural networks [46, 47], through generative adversarial networks [48, 49], to diffusion models [50]. However, learning-based schemes traditionally need large, diverse training datasets, which can pose a bottleneck in niche fields like FLIM; further, different fluorophores behave differently, hampering generalisation in traditional machine learning methods [51]. Self-similarity-based super-resolution [52] and self-supervised clustering [53] approaches offer an alternative to external training set, deriving statistical information for super-resolution from the very image that is up-sampled.

Data from different sensing modalities can yield more information about a subject than is contained in each modality alone [45, 54]. Fusion-based inference is a growing field with applications from medical imaging using PET and MRI [55], through autonomous driving using camera and LiDAR [56], to content classification using video and text [57]. Data fusion has been applied to FLIM, by interpolating lifetime images and weighting them with intensity images for visualisation [24, 58].

Here we introduce a super-resolution method that relies on the fusion of two images: a high-resolution intensity image (no lifetime information) and a low-resolution lifetime image. Our method is called ‘single sample image fusion upsampling’ (SiSIFUS). SiSIFUS generates data-driven lifetime priors matching the resolution of the intensity image; this is relatively easy and inexpensive to acquire at high-resolution, compared to FLIM images.

Crucially, our method generates ‘single sample’ priors: all information in our scheme comes from the given field of view, not external training data. We develop two priors, which extract this information from the FLIM-intensity image pair in different ways. Local priors correlate low resolution FLIM pixels with corresponding intensity pixels in small neighbourhoods. Global priors instead exploit morphological signatures in the image, using a neural network to predict fluorescence lifetime from intensity patches.

SiSIFUS combines data fusion and self-supervised learning into a practical super-resolution framework. Like example-based self-similarity approaches, it avoids complex hardware modifications and external training data. Like reconstruction-based modelling, we optically measure high-resolution features, giving more information than is available in the low-resolution images alone.

II. RESULTS

Forward and inverse models. We apply SiSIFUS to both raster-scanning and widefield FLIM. The scanning system uses a PMT to gather both the FLIM and intensity image, while the widefield system uses a SPAD array to measure FLIM, and a complementary metal-oxide-semiconductor (CMOS) camera to measure intensity. Both of the setups are detailed in the Methods.

SiSIFUS involves two measurements. The first is the time-resolved, low-resolution datacube, $r \in \mathbb{N}^{m,n,t}$,

where m, n denote spatial position and t denotes time. The fluorescence lifetime image, $\tau_{LR} \in \mathbb{R}^{m,n}$, is estimated from r via a standard least squares deconvolution - other schemes, like phasor analysis or centre-of-mass estimation, could be used equivalently.

The second measurement is the high-spatial resolution intensity measurement, $I \in \mathbb{N}^{M,N}$, where M, N denote the pixel numbers of the high-spatial resolution sensor. SiSIFUS then super-resolves the lifetime image τ_{LR} to match the pixel count of the intensity image I .

Our setups sample fluorescence lifetime sparsely across the field of view. In the widefield setup, this arises from low fill-factor, ergo large dead spaces between the active areas of the SPAD pixels. In the scanning setup, this arises from the large sampling period relative to the spot size of the excitation beam in the object plane. We also assume that the intensity measurement has approx. 100% fill-factor. Fig. 1A and Fig. 2A depict how intensity and lifetime are sampled.

For 256×256 sized high-resolution intensity image I of the sample, the acquired dataset is integrated along the time axis. In a practical scenario for upsampling a confocal scan image, the FLIM samples would be acquired by taking a large line-average of low-resolution scans. A large line-average is needed for the fitted lifetime to have decent signal-to-noise ratio (SNR). The intensity image has decent SNR even with just a few line-averages, therefore the high-resolution intensity image could be obtained without adding an external sensor, by simply scanning a second time, with a higher pixel count but much fewer line averaging.

Consequently, τ_{LR} is decimated (sparsely sampled) from the high-resolution fluorescence lifetime target, $\tau_{HR} \in \mathbb{R}^{M,N}$, that we aim to reconstruct:

$$\tau_{LR} = \mathbf{A}\tau_{HR} \quad (1)$$

where \mathbf{A} represents sparse sampling (decimation). We feed the two images, $I(M, N)$ and $\tau_{LR}(m, n)$, to our prior-generation pipeline (explained below), which outputs a local and global prior, $\hat{\tau}_{LP}(M, N)$ and $\hat{\tau}_{GP}(M, N)$, respectively. These priors constrain an (otherwise ill-posed) inverse retrieval algorithm. We finally recover the high-resolution lifetime image τ_{HR*} by minimizing the following cost function:

$$\begin{aligned} \tau_{HR*} = & \arg \min_{\hat{\tau}_{HR}} C(\hat{\tau}_{HR}), \text{ where} \\ C(\hat{\tau}_{HR}) = & \|\mathbf{A}\hat{\tau}_{HR} - \tau_{LR}\|_2^2 + \gamma\|\hat{\tau}_{HR} - \hat{\tau}_{LP}\|_2^2 \\ & + \beta\|\hat{\tau}_{HR} - \tau_{GP}\|_2^2 + \alpha\|\mathbf{D}\hat{\tau}_{HR}\|_1 \end{aligned} \quad (2)$$

The first term in $C(\hat{\tau}_{HR})$ ensures the data fidelity between the low-resolution measured lifetime image and the downsampled optimal high resolution lifetime solution in each iteration. Prior constraints on the target high resolution lifetime image are enforced through the second and the third data fidelity term, weighed by the factor γ and β respectively, which are empirically optimised to yield best results. The fourth term is the L1-norm of the 2D total variation (TV) evaluated on the high-resolution lifetime image and weighed by α [59].

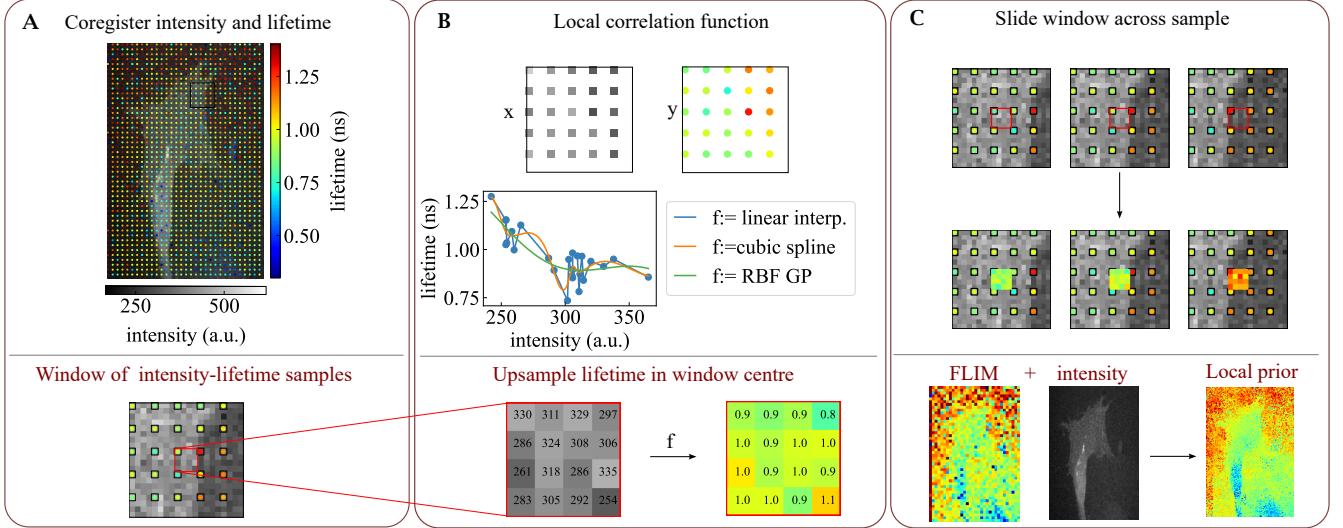


FIG. 1. Schematic of the local prior method. (A) Shown are a CMOS (fluorescence intensity) field of view, with the SPAD field of view (fluorescence lifetime), overlayed on top of it so as to match the sparse, low fill-factor pixel layout of the SPAD array. (B) We zoom in on a 5×5 window. All SPAD pixels have a corresponding CMOS measurement, but so do the areas in-between SPAD pixels. We aim to find the lifetime at points with no SPAD samples. For this, we fit a function, for instance linear interpolation, a cubic spline or a radial basis function gaussian process. Then, the high-resolution CMOS pixels x_{HR} which we wish to upsample are fitted with this function, producing a lifetime estimate $\hat{\tau}_{HR}$. (C) We slide the window across the field of view, fitting new functions for each new window, and predicting the centres, upsampling the FLIM image to the resolution of the intensity image, window-by-window.

We consider the anisotropic form of the TV [60], and so the operator \mathbf{D} represents the finite differences approximation of the horizontal and vertical image gradients.

A. Dependence between lifetime and intensity.

SiSIFUS priors exploit inter-dependence between fluorescence lifetime and intensity. Although these variables are interdependent at the single molecule level via fluorescence quantum yield, this dependence is modulated by fluorophore concentration and other complex and often unpredictable biophysical mechanisms, thus necessitating statistical methods to create our priors. Fluorescence quantum yield Q is the ratio of the number of emitted photons to the number absorbed. It depends on the radiative and non-radiative decay rates k_r and k_{nr} that depopulate excited molecules. The measured fluorescence lifetime τ also depends on these rates [16]:

$$Q = \frac{k_r}{k_r + k_{nr}} \quad \tau = \frac{1}{k_r + k_{nr}} \quad (3)$$

therefore $Q = k_r \tau$ for a single molecule. Across a given field-of-view, fluorescence intensity variations are given by the quantum yield of fluorophores (equivalently, fluorescence lifetime) multiplied by their absorbance (concentration times absorptivity times sample thickness). Absorbance is typically unknown and unpredictable, hence it acts as a confounding variable in intensity-lifetime dependencies, so fluorescence intensity alone cannot give us full lifetime information. This means that two samples might have the same lifetime, but completely different intensities, or vice versa.

However, across a single sample, fluorophore concentration typically varies slowly compared to lifetime and/or covaries with it on local scales, such that it is possible to build local priors that capture the resulting intensity-lifetime dependencies. Further, absorbance and lifetime often co-vary with cellular morphology, enabling us to create global priors. As a fail-safe, if neither local nor global dependencies exist across a specific sample or sub-region, TV-minimisation (a form of edge-preserving interpolation) in our inverse retrieval ensures that our method still performs at least as well as standard interpolation (see supplemental material for details).

Local prior. The local prior (LP) relies on direct, pixelwise dependencies between lifetime and intensity on micron scales: Fig. 1 illustrates our workflow. If the images come from different detectors, the lifetime and intensity image are first coregistered to match their fields of view. Fig. 1A shows a sparse, low-resolution lifetime image (RGB) overlayed on the corresponding intensity image (grayscale). The field of view (FOV) is divided into windows, each containing a set of corresponding intensity-lifetime samples. These samples neighbour intensity pixels in the window centre, hence this window is used to create a prior for those pixels. In each window, the intensity and lifetime pairs are vectorised and fitted with a function, f - see Fig. 1B. Thus, our lifetime estimate $\hat{\tau}$ for pixel $(\lambda i + x, \lambda j + y)$, is:

$$\hat{\tau}_{\lambda i + x, \lambda j + y} = f_{i,j}(I_{\lambda i + x, \lambda j + y}) \quad (4)$$

with samples $i \in \{0, 1, \dots, n - 1\}$ and $j \in \{0, 1, \dots, m - 1\}$, and $x \geq 0, y < \lambda$. Importantly, the functions $f_{i,j}$ are fitted *locally*, not globally. Consequently, this procedure is repeated by sliding the window across the field of view, as shown in Fig. 1C.
Global prior. Images often contain multiple examples

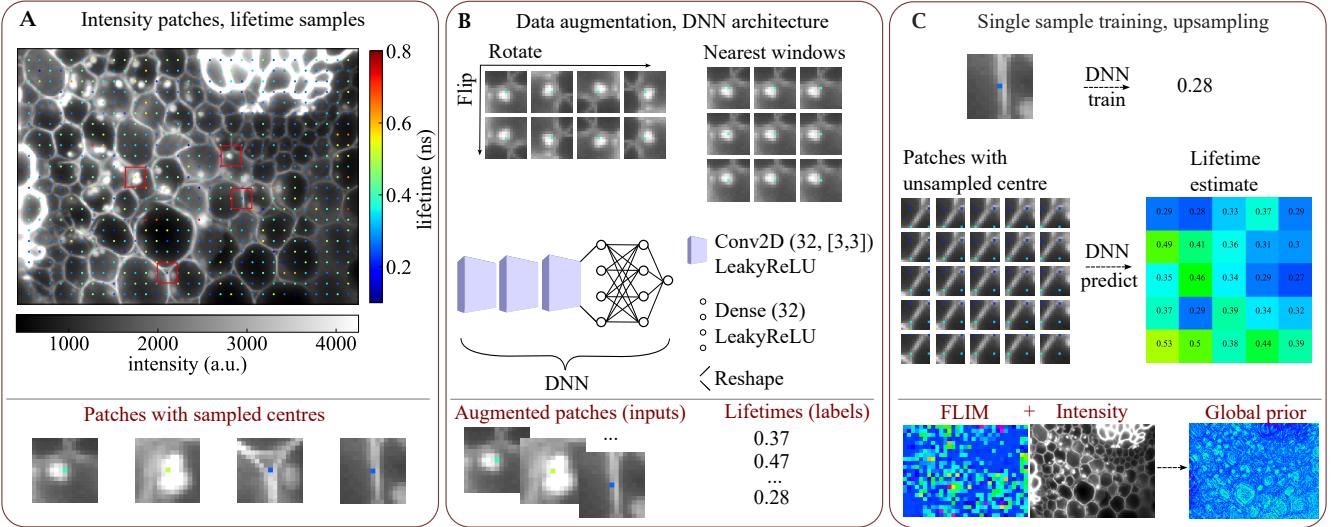


FIG. 2. Schematic of the global prior method. **(A)** Fluorescence intensity of a convallaria - acridine orange sample, with 8×8 sparse lifetime samples overlayed. We extract intensity patches from this image; a few of them correspond to a central lifetime sample. Such patches are training data, which we can use to predict the central lifetime of the rest of the patches. **(B)** Training inputs (patches) are augmented via rotation and mirroring. They can be further augmented by adding the patches which are nearest neighbours of training patches and allocating them the same label (lifetime) as the sampled patch. The deep neural network (DNN) architecture is simple, consisting of three 2D convolutional layers followed by three fully connected layers. **(C)** Finally, the trained DNN evaluates patches with unsampled centres, thus super-resolving the lifetime image.

of similar features, with similar lifetime distributions, across the field of view. This motivates our development of global priors (GPs) that exploit correlations between high-resolution morphology and lifetime.

Fig. 2 shows our pipeline. We first extract intensity patches centred on our SPAD pixels, as shown in Fig. 2A. To deal with the relatively small number of patch-lifetime pairs contained in a single sample image, we augment our training set. We use a commonly used dataset augmentation technique by reflecting and rotating the intensity windows in the training set. These operations increase our dataset 8-fold, as shown in Fig. 2B. For high upsampling factors (8×8 and 16×16), we further augment the training set by estimating the lifetimes of the patches neighbouring our sampled patches. We then label these patches with the same lifetime as their sampled neighbour. Our approach is visualised in Fig. 2B; see Methods for details.

Our global priors are designed to generalise to new samples with previously unseen morphologies, morphology-lifetime dependencies, and lifetime ranges. A deep neural network (DNN), shown in Fig. 2B, is trained from scratch for each new sample, on the intensity-patch inputs and lifetime labels obtained from the given microscope field of view. Consequently, different DNN initialisations give slightly different predictions. To estimate high-resolution lifetime, we pass each intensity patch through our trained DNN, predicting the central lifetime value, as shown in Fig. 2C.

Quality metrics. We track reconstruction quality using three metrics: learned perceptual image patch similarity (LPIPS), structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR). LPIPS measures distance between images in feature space. It has a minimum of 0 and grows with image dissimilarity

(higher values are worse). SSIM tracks the similarity in luminance, contrast and structure between two images; it is bound between -1 and 1, with larger values indicating better image similarity. Lastly, PSNR is a pixel-to-pixel comparison, where larger values are better. See the Supplement for details.

Sample 1: 16x16 (MDCK Flipper-TR). We examined a validation sample of Madin-Darby canine kidney (MDCK) cells that had been treated with Flipper-TR dye (Spirochrome Inc.), which allows quantitation of tension in living, migrating cells. Data was acquired using a commercial LaVision BioTec TriM Scope system, using two-photon excitation scanning, and detecting emission via a photo-multiplier tube (PMT). The sample was imaged at 512×512 spatial points covering a $167 \times 167 \mu\text{m}^2$ FOV, and binned into 75 time bins, giving a $512 \times 512 \times 75$ datacube. See Methods for details.

Ground truth (GT) fluorescence lifetime was estimated from this datacube using least squares deconvolution. This was decimated 16-fold to give a low-resolution FLIM image, shown in Fig. 3A. The low-resolution FLIM image is severely undersampled; a lot of the detail was lost. Fluorescence intensity (Fig. 3B) was obtained in parallel, by summing the datacube along time. In the intensity image, we see that the probe mainly localised to two types of structures: small blobs (vesicles) and edges (cell membranes). In Fig. 3(C-D) we show 3 examples of local prior windows and corresponding global prior patches, extracted as shown in Fig. 1B and Fig. 2A, respectively. Local (pixelwise) dependencies appear relatively weak, instead the global (morphological) dependencies dominate, capturing the trend that vesicles have lower lifetimes than cell membranes across the FOV.

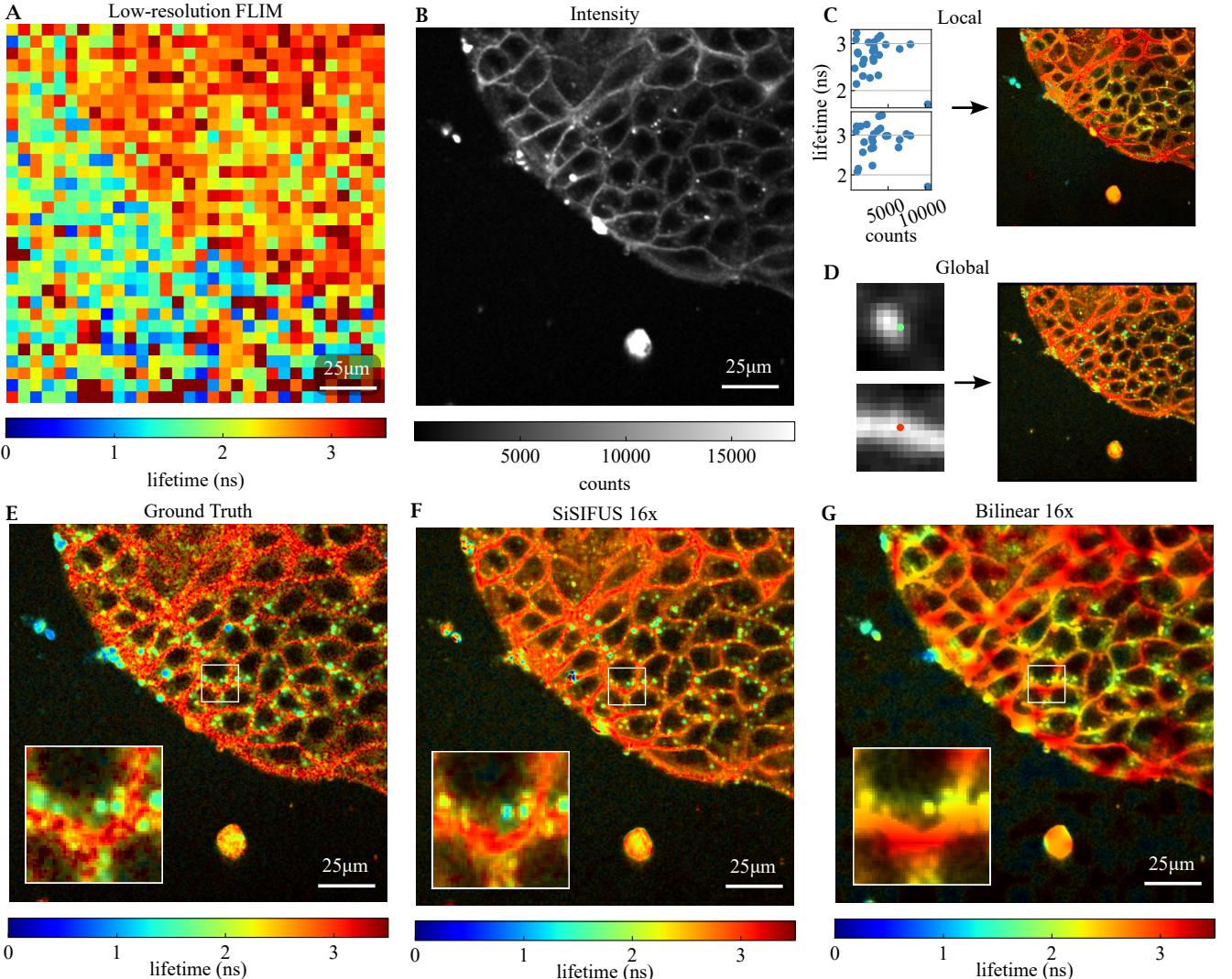


FIG. 3. 16x16 upsampling of MDCK cells. **(A)** Low resolution fluorescence lifetime image (32x32) of Madin-Darby canine kidney (MDCK) cells expressing Flipper-TR dye. **(B)** Corresponding high resolution intensity image (512x512) of the sample. **(C)** 5x5 windows of low-resolution FLIM are fitted to corresponding intensity values, to generate a local prior image (two example windows are shown). **(D)** A global prior image is generated from 13x13 intensity patches with central FLIM measurements (two examples are shown). **(E)** The ground truth high-resolution FLIM target, intensity weighted for visualisation. **(F)** The proposed method, upsampling the low-resolution measurement by a factor of 16x16. **(G)** Bilinear interpolation upsampling the FLIM measurement by 16x16.

Fig. 3E shows the ground truth lifetime, weighted with local contrast enhanced fluorescence intensity for visualisation (details in the Supplement). In Fig. 3(F-G) we show SiSIFUS and bilinear interpolation for the upsampling task. SiSIFUS automatically learns to distinguish between vesicles and cell membranes and labels them with different lifetimes, whereas interpolation fails to reconstruct fine details. SiSIFUS also maintains sharp boundaries between structures of different lifetimes, informed by the intensity image. SiSIFUS has LPIPS of 0.24, SSIM of 0.31, and PSNR of 26dB. Interpolation, instead, has LPIPS of 0.48, SSIM of 0.12, and PSNR of 24dB.

Sample 2: 8x8 (Convallaria AO). We applied SiSIFUS to a Convallaria rhizome sample dyed with Acridine Orange. The fluorescence lifetime datacube was recorded using our custom microscope setup, which uses a FLIMera 192x128-pixel SPAD array (HORIBA Sci-

entific [61]) to image an $82 \times 107 \mu\text{m}^2$ FOV, using 326 time bins. To compensate for the HORIBA camera's asymmetric pixel layout, we scanned the sample twice with a pixel shift, giving a $192 \times 256 \times 326$ datacube. Simultaneously, a high spatial resolution sCMOS camera registered a 2048×2048 -pixel image of the sample. The sCMOS camera is spatially co-registered to match the SPAD's field of view (FOV) and resolution (see Methods).

The low-resolution FLIM and high-resolution intensity guide are shown in Fig. 4(A-B). Fig. 4(C-D) shows example local and global dependencies and priors, respectively. Global dependencies appear to dominate local ones for this sample, with globules having shorter lifetimes than cell walls.

The 8×8 upsampling results in Fig. 4(E-G) illustrate that SiSIFUS recognises that globules tend to have higher lifetimes than cell walls, hence maintaining con-

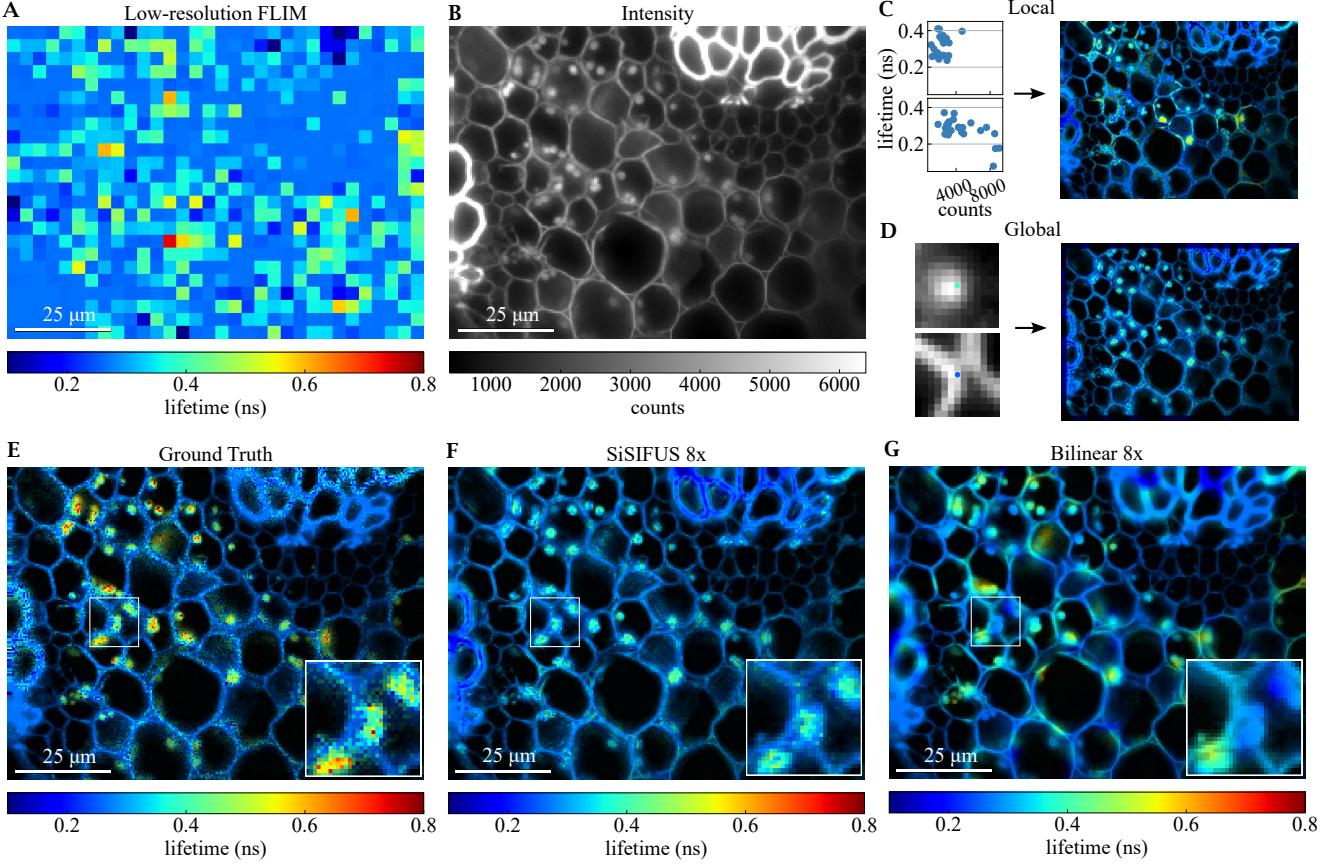


FIG. 4. 8x8 upsampling of convallaria images. (A) Low-resolution fluorescence lifetime image (24×32) of a convallaria rhizome sample stained with Acridine Orange, viewed under a widefield microscope. (B) High-resolution intensity image (192×256). (C) Example 5x5 windows of low-resolution intensity vs FLIM, used for generating the local prior shown on the right. (D) High-resolution intensity patches are labelled with lifetime, letting us create a global prior. (E-G) Ground truth, 8x8 SiSIFUS and 8x8 bilinear interpolation of the data, weighted by local contrast enhanced intensity for visualisation – see Supplementary Materials for details.

trast between these structures more consistently than bilinear interpolation. We do note though that global SiSIFUS misses certain hotspots in the GT lifetime image (high-lifetime, yellow/red coloured areas), likely because few globules in the training set have these lifetimes, hence the model treats them as outliers. SiSIFUS achieves an LPIPS of 0.11, SSIM of 0.21, PSNR of 16dB. Bilinear interpolation has an LPIPS of 0.15, SSIM of 0.29, and PSNR of 16dB.

Sample 3: 16x16 (SKOV3 - Rac1 Raichu). We further validate SiSIFUS on measurements of SKOV3 ovarian cancer cell samples expressing Rac-Raichu Clover mcherry (see Methods). The GT images are acquired using our LaVision BioTec TriM Scope two-photon scanning system. The field of view was sampled on a 256×256 square grid covering $301 \times 301 \mu\text{m}^2$ area. The temporal evolution was recorded using TCSPC with 75 timebins of 160ps duration each, giving a fluorescence datacube of size $256 \times 256 \times 75$.

The low-resolution fluorescence lifetime input is shown in Fig. 5A, alongside the high-resolution intensity guide in Fig. 5B. Fig. 5(C - D) show a set of local windows and the LP, as well as global patches and the GP, respectively.

Local dependencies exhibit plateauing positive correlations and seem to flatten fluorescence lifetimes across

the different cells, capturing inter-cellular dynamics. Global priors instead capture intra-cellular dynamics, showing that fluorescence lifetime is mostly uniform within cells, with some patterned textures. Since the upsampling factor is high (16x16), our algorithm prioritises global priors.

Fig. 5(E - G) show the ground truth compared to 16x16 SiSIFUS and bilinear interpolation; SiSIFUS gives a better estimate. SiSIFUS reconstruction has an LPIPS to the ground truth of 0.15, SSIM of 0.08, and 12dB PSNR. In contrast, interpolation has an LPIPS of 0.34, SSIM of 0.06 and PSNR of 11dB.

Sample 4: 8x8 (SKOV3 Rac1-Raichu). We applied SiSIFUS to widefield images of SKOV3 ovarian cancer cells expressing Rac-Raichu Clover mcherry, acquired with our custom FLIMera SPAD array and Zyla sCMOS setup (see Methods). The temporal evolution was recorded using TCSPC with 326 timebins, giving a fluorescence datacube of size $192 \times 128 \times 326$. This image was decimated to give the low-resolution FLIM input.

Fig. 6 shows our results. The low-resolution FLIM (Fig. 6A) and high-resolution intensity (Fig. 6B) are used to generate local and global priors, which are then fed into our inverse retrieval algorithm. Fig. 6C and Fig. 6D shows examples of local and global intensity lifetime de-

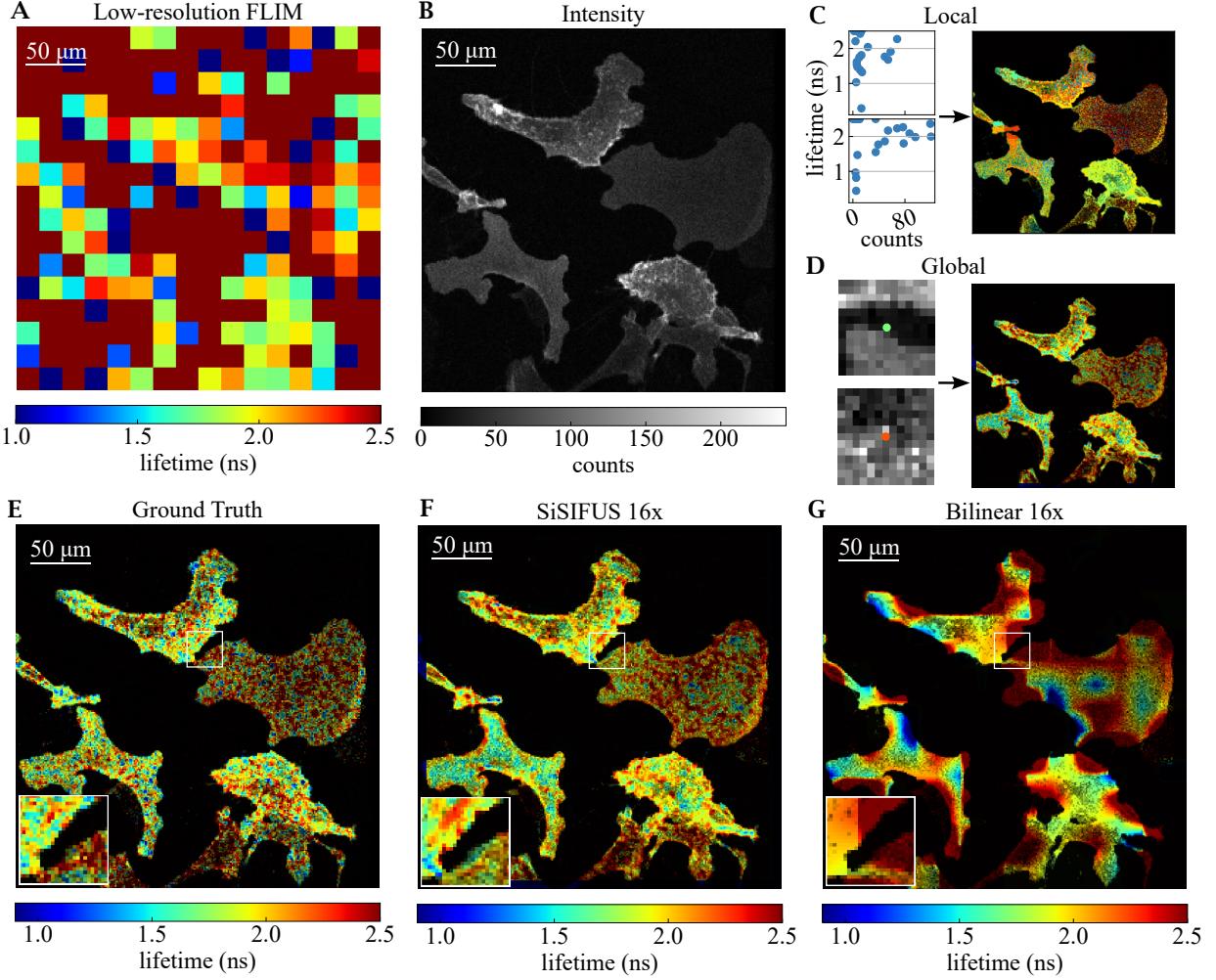


FIG. 5. 16x16 upsampling of SKOV3 cells. (A) Low-resolution fluorescence lifetime image (16×16) of a SKOV3 samples expressing Rac1-Raichu. (B) High-resolution intensity image (256×256). (C-D) Examples of local and global dependencies. (E-F) Ground truth, 16×16 super-resolved, and 16×16 interpolated images.

pendencies and priors.

This sample shows non-linear negative local interdependencies at most regions; SiSIFUS can exploit these to accurately determine the lifetime based on local intensity patterns. Conversely, global patch-lifetime dependencies are negligible. This is mainly because the field of view lacks repeating morphological features (in contrast to the MDCK and convallaria samples in Fig. 3 and Fig. 4). Our algorithm prioritises the local priors. Finally, Fig. 6(E-G) shows the ground truth compared to SiSIFUS and bilinear interpolation. SiSIFUS succeeds in reconstructing the lifetime boundaries seen at the cell edges, and also reconstructs the speckliness of the ground truth lifetime map, allowing the user to infer that there might be lifetime estimation uncertainty. Interpolation fails in these regards: edges are blurred, and lifetime estimates appear smooth give the impression of structures that are absent in the ground truth. SiSIFUS yields an LPIPS of 0.31, SSIM of 0.21, and PSNR of 16dB, whereas interpolation has an LPIPS of 0.56, SSIM of 0.22, and PSNR of 16dB.

B. Acquisition times

SiSIFUS provides an advantage in terms of acquisition times. For example, if we consider the case of measurements taken with our TriM Scope I (Figures 3 and 5), the acquisition time scales linearly with pixel number as this is a galvo-scanning system. Therefore, we have an immediate advantage given by the SiSIFUS resolution enhancement factor that is applied. Specifically, in Figure 3, where we apply 16×16 resolution enhancement, we have a $256 \times$ reduction in the number of points that need to be scanned and hence a $256 \times$ reduction in acquisition time. In a scanning system we still, however, need to perform a second scan for the high-resolution intensity measurement but this typically can be at substantially higher speed, of order $35 \times$ in our system (and this therefore remains the limiting factor). If we therefore consider the specific case of a 512×512 image (Figure 3), the total acquisition time without SiSIFUS was 73 seconds and with SiSIFUS is 2.4 seconds allowing a 0.4 fps acquisition rate. If instead we consider the case of measurements with a SPAD camera (in our case, the Horiba FLIMera system, Figures 4 and 6) this currently operates at 30 fps, i.e.

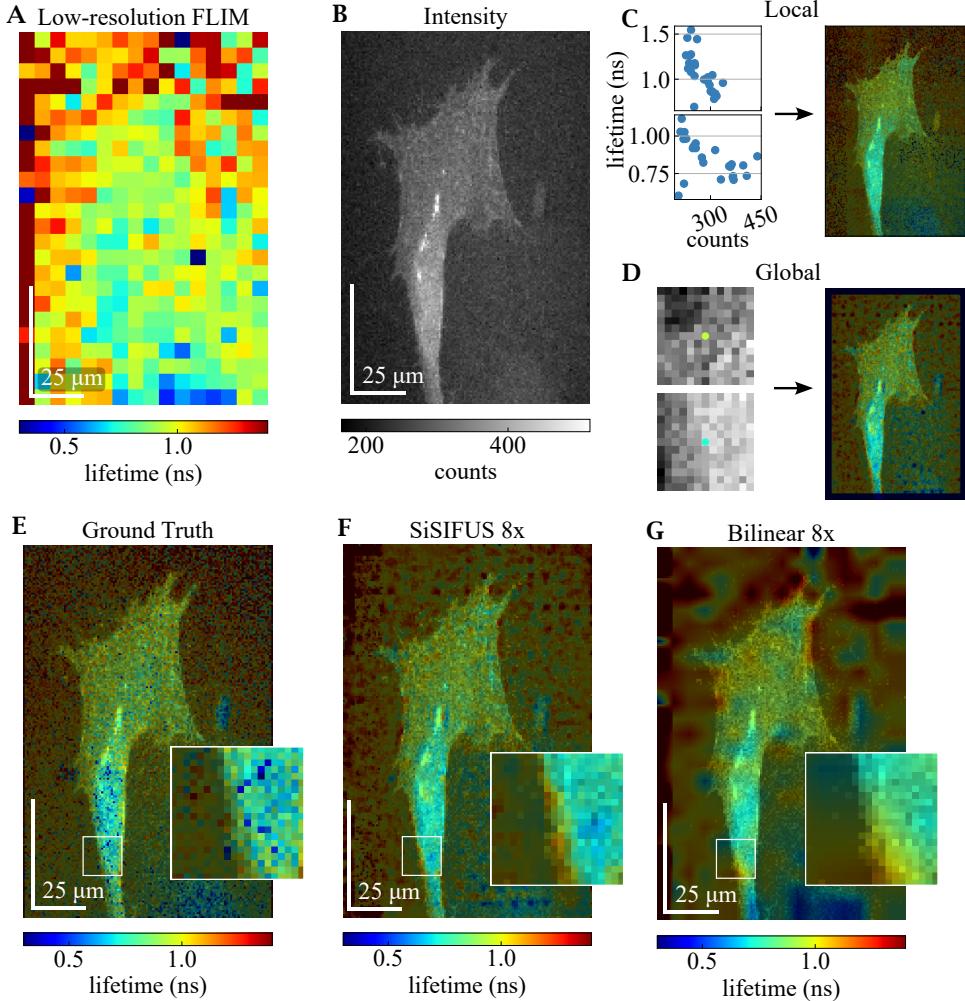


FIG. 6. 8x8 super-resolution of SKOV3 images. (A) Low-resolution FLIM image of a SKOV3 cell expressing the Rac1-Raichu probe (24×16). (B) The corresponding high-resolution fluorescence intensity image (192×128). (C-D) Comparison of local and global intensity-lifetime dependencies observed in this sample, and the corresponding local and global prior images. (E-G) High-resolution ground truth FLIM, 8x8 SiSIFUS and 8x8 bilinearly interpolated images, respectively.

33 ms acquisition time for a SiSIFUS image of any size (all pixels are acquired in parallel without any point-by-point scanning used in confocal imaging systems). We note that in this case, the intensity CMOS image is acquired in parallel and hence does not add to the acquisition time.

We may compare this also with existing commercial systems, e.g. current B&H FLIM systems can measure 512×512 pixels in 1 second [62] or previous work that operated directly with megapixel SPAD arrays in which, however, the smaller pixel size implied longer acquisition times in order to accumulate sufficient signal and was thus limited to ~ 1 second acquisition times [23].

III. DISCUSSION

We introduce SiSIFUS, a robust, data-fusion pipeline based on prior-augmented inverse retrieval for upsampling fluorescence lifetime images. We create two classes of priors that explicitly exploit a high-resolution intensity image to provide approximations for the non-sampled datapoints in a fluorescence lifetime image.

The goal of SiSIFUS is to provide a “physics inspired” approach to image resolution enhancement that performs better than standard bilinear or similar interpolation methods.

Local priors capture pixel-wise correlations between fluorescence lifetime and intensity. For this, we find a direct mapping from intensity values to lifetime in small, local neighbourhoods, and use this mapping to predict the lifetime of intensity pixels that lack corresponding lifetime pixels. This allows SiSIFUS to maintain sharp spatial boundaries, tracking the boundaries of our intensity image. The local prior is limited by measurement noise and sampling frequency. Since structures of similar intensity are assigned the same lifetime, undersampled regions may receive homogeneous lifetime estimates with sharp boundaries, as seen in the leftmost cell in Fig. 5 (c). Noisy regions can instead artificially track the intensity of an image’s noise, as in Fig. 6(c). TV-minimization and the global prior help combat these issues.

Global priors capture inter-dependence between FLIM and intensity on a morphological level. This is achieved by learning a mapping (deep neural network) from in-

tensity patches to central-pixel lifetime samples, and then using mapping to predict the lifetime of patches that have no central lifetime measurements. Thus, we capture non-linear correlations between the brightness and shape of intensity features and lifetime. In microscopic samples, there often exist strong global trends between these variables, allowing the model to predict the lifetimes of patches with unsampled centres. However, the global prior has limited ability to distinguish between similar morphologies with different lifetimes, typically assigning them with the average lifetime of such structures. This causes outliers like the low-lifetime globules on the left of Fig. 3 or the high-lifetime vesicles in Fig. 4 to be ignored in favour of global patterns – although the local prior will still retain these outliers. Consequently, the GP is most beneficial for samples that contain many examples of similar morphological structures, where these structures share similar lifetime properties. Since GPs capture sample dependent properties, they can be further exploited when the same sample is imaged across multiple regions of interest (for instance in a mosaic scan), by stacking intensity patches with corresponding lifetime labels into a common training set to improve generalisation.

The results demonstrate the fact that the introduction of the priors in the TV-based inverse retrieval algorithm, makes the latter a tractable problem. SiSIFUS gives the upsampled lifetime image sharp spatial features by extracting spatial information from an intensity image. This feature similarity is shown by perceptual metrics such as LPIPS, as features like edges, speckly textures and object shapes are captured by SiSIFUS, but not by interpolation. In contrast, pixelwise metrics penalise pixel-to-pixel estimation noise heavily, making them more lenient towards blurred, interpolated images than SiSIFUS. We therefore prefer the LPIPS metric (designed to measure image quality in a similar way to human perception) and used it to optimise our hyperparameters in validation. We note that Figures 4, reffig:5, and 6 were used as validation images that allowed us to optimise all hyperparameters (LP and GP window sizes, LP function, GP model architecture, epochs and learning rate, ADMM iterations and loss function weights). These were then fixed and used to generate the images in Fig. 3.

It is worth emphasizing that SiSIFUS is currently not implemented simultaneously with high-speed measurements like the aforementioned 30 fps SPAD video, limiting its use for direct feedback or real-time diagnostics, such as intraoperative imaging. However, it enhances measurement speed by sampling fewer lifetime points and estimating lost information afterward. Equivalently, this mitigates phototoxicity and photodamage by illuminating the sample with less light, useful for imaging live samples in research or biopsy.

Despite these limitations, SiSIFUS offers notable potential for various applications beyond its current validation scope of FLIM. It could be applied to tasks involving disparate image types but exhibiting local or global correlations underscores its versatility and applicability in diverse research settings.

A key feature that we believe will be beneficial in any such approach is that image reconstruction is never

based on statistical inference from other images - only the single image samples acquired from the two cameras are used thus strongly reducing or eliminating artefacts that may occur for example in other machine learned approaches that do indeed rely on large sets of additional data and images and thus representing a potential point of failure that is of concern in many applications.

IV. MATERIALS AND METHODS

A. Experimental design

Mammalian cell culture conditions. Both the SKOV3 ovarian cancer cells and the Madin-Darby canine kidney (MDCK) cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, 2 mM L-Glutamine and 1X PenStrep. Cell lines were maintained in 10 cm dishes at 37°C and 5% CO₂.

SKOV3 cells were transfected in the morning using Amaxa Nucleofector (Lonza) kit V, program V-001 with either 5 µg Raichu-Rac1-Clover-mCherry or pcDNA3.1-mClover DNA (adapted from [63]) following manufacturers guidelines and replated on 6 cm TC-treated dishes at 37°C and 5% CO₂. For live cell imaging, cells were collected and replated onto 35 mm glass bottom MatTek dishes that were previously coated overnight with laminin (10 µg ml⁻¹) diluted in PBS. These were left overnight at 37°C, 5% CO₂.

The next morning prior to imaging, the dishes were washed twice with pre-warmed PBS and replaced with pre-warmed FluoroBrite DMEM supplemented with 10% FBS, 2 mM L-Glutamine and 1X PenStrep. For fixed cell imaging, the cells were collected and replated onto 22 mm glass coverslips that were previously coated overnight with laminin (10 µg ml⁻¹) diluted in PBS. These were left overnight at 37°C, 5% CO₂. The next day, these cells were fixed in 4% PFA for 10 minutes and washed with PBS and mounted using Fluromount-G (Southern Biotech).

The MDCK cells were trypsinised and plated on 35 mm glass-bottom MatTek dishes and left to settle for 4 hours. Flipper-TR® probe (Cytoskeleton; CY-SC020) was resuspended in 50 µl anhydrous DMSO as per manufacturers instructions to yield a 1 mM stock. Flipper-TR was diluted in culture media to 2 µM and incubated on the cells overnight at 37°C, 5% CO₂.

The next morning prior to imaging, the dishes were washed twice with pre-warmed PBS and replaced with prewarmed FluoroBrite DMEM (ThermoFisher Scientific; A1896701) supplemented with 10% FBS, 2 mM L-Glutamine, 1X PenStrep and 2 µM Flipper-TR.

Multiphoton raster-scanning time-domain FLIM: Experimental set-up details. For the dataset shown in Fig.5, cells were left to equilibrate on a heated microscope insert at 37°C, perfused with 5% CO₂ prior to imaging. Images were acquired in the dark using a multiphoton LaVision TRIM scan head mounted on a Nikon Eclipse inverted microscope with a 20X water objective. Illumination is provided by a Ti:Sapphire femtosecond laser (Coherent Chameleon

Ultra II) used at 920 nm (12% power). The fluorescence signal was passed through band pass filters 525/50 nm emission and acquired using a FLIM X-16 Bioimaging Detector TCSPC FLIM system (LaVision BioTec). A $301 \times 301 \mu\text{m}^2$ FOV corresponding to 256×256 pixels was imaged at 600 Hz with a 10 line average in a total acquisition time of 5199 ms.

For the dataset shown in Fig. 3, cells were left to equilibrate on a heated microscope insert at 37°C, perfused with 5% CO₂ prior to imaging. Images were acquired in the dark using a multiphoton LaVision TRIM scan head mounted on a Nikon Eclipse inverted microscope with a Nikon Apo 60X oil objective, 1.4 NA. Illumination is provided by a Ti:Sapphire femtosecond laser used at 970 nm (8% power) with an acquisition delay of 5.440 ns. The fluorescence signal was passed through emission band pass filters 600/60 nm and acquired using a FLIM X-16 Bioimaging Detector TCSPC FLIM system (LaVision BioTec).

A $163 \times 163 \mu\text{m}^2$ field of view correlating to 512×512 pixels was imaged at 600 Hz with a 70-line average for a total acquisition time of 72575 ms (High-Res). A total of 100 High and Low-Res images taken from 3 independent experiments. Background images (High and Low-Res) were obtained by closing the scan-head using the above settings. Instrument response function (IRF) was obtained using carbon nanorods with the above settings and a 1% laser power.

Widefield time-domain FLIM: Experimental set-up details. For the datasets shown in Figures 4 and 6, a custom microscope system was built using high spatial resolution sCMOS sensor (Andor's Zyla) and the FLIMera SPAD array sensor. Spatial registration was achieved by identifying a set of four co-registered points on the SPAD and CMOS, and mapping the CMOS image with a perspective transformation to match the field of view of the SPAD image. See the Supplement for a schematic of the experimental set up.

B. Statistical Analysis

Inverse Retrieval Algorithm. The optimization is implemented using the alternating direction method of multipliers (ADMM) algorithm. For this the minimisation in Eq. 2 can be re-formulated as:

$$\begin{aligned} \tau_{HR*} &= \arg \min_{\hat{\tau}_{HR}} C(\hat{\tau}_{HR}), \text{ where} \\ C(\hat{\tau}_{HR}) &= \|\mathbf{A}\hat{\tau}_{HR} - \tau_{LR}\|_2^2 + \gamma\|\hat{\tau}_{HR} - \hat{\tau}_{LP}\|_2^2 \\ &\quad + \beta\|\hat{\tau}_{HR} - \hat{\tau}_{GP}\|_2^2 + \alpha\|z\|_1 \\ \text{subject to } &\mathbf{A}\hat{\tau}_{HR} - z = 0 \text{ and } \hat{\tau}_{HR} \geq 0 \end{aligned} \quad (5)$$

The Augmented Lagrangian for this problem can be written as:

$$\begin{aligned} L_\rho(\hat{\tau}_{HR}, z, y) &= \arg \min_{\hat{\tau}_{HR}} C(\hat{\tau}_{HR}), \text{ where} \\ C(\hat{\tau}_{HR}) &= \|\mathbf{A}\hat{\tau}_{HR} - \tau_{LR}\|_2^2 + \gamma\|\hat{\tau}_{HR} - \hat{\tau}_{LP}\|_2^2 \\ &\quad + \beta\|\hat{\tau}_{HR} - \hat{\tau}_{GP}\|_2^2 + \alpha\|z\|_1 \\ &\quad + y\|\mathbf{D}\hat{\tau}_{HR} - z\| + \rho/2\|\mathbf{D}\hat{\tau}_{HR} - z\|_2^2 \end{aligned} \quad (6)$$

Here y is the Lagrangian multiplier (or the dual variable) and ρ is the penalty parameter. The ADMM approach involves jointly minimizing the Lagrangian over all the primal variables followed by the updates over the dual variables. The primal updates for the variables $\hat{\tau}_{HR}$ and z are given by:

$$\begin{aligned} \hat{\tau}_{HR_{k+1}} &\leftarrow \arg \min_{\hat{\tau}_{HR}} \|\mathbf{A}\hat{\tau}_{HR} - \tau_{LR}\|_2^2 + \gamma\|\hat{\tau}_{HR} - \hat{\tau}_{LP}\|_2^2 \\ &\quad + \beta\|\hat{\tau}_{HR} - \hat{\tau}_{GP}\|_2^2 + y_k(\mathbf{D}\hat{\tau}_{HR} - z_k) \\ &\quad + \rho/2\|\mathbf{D}\hat{\tau}_{HR} - z_k\|_2^2 \end{aligned} \quad (7)$$

$$\begin{aligned} z_{k+1} &\leftarrow \arg \min_z \alpha\|z\|_1 + y(\mathbf{D}\hat{\tau}_{HR} - z) \\ &\quad + \rho/2\|\mathbf{D}\hat{\tau}_{HR_k} - z\|_2^2 \end{aligned} \quad (8)$$

The dual update is given by:

$$y_{k+1} \leftarrow y_k + \rho(\mathbf{D}\hat{\tau}_{HR} - z) \quad (9)$$

For the primal minimisation update, we use the standard optimization technique based on the fast iterative soft thresholding algorithm (FISTA). Each iteration of the ADMM hence comprises of 90 iterations of FISTA for the $\hat{\tau}_{HR}$ variable update.

The weighting factor γ for the local prior term in the cost function has been kept constant for all the cases wherein $\gamma = 1$. The factor β on the other hand is varied for different upsampling factors such that it is 0.02 for 2x and 4x upsampling factors and 0.5 for higher upsampling factor of 8x and 16x. The GP cannot predict lifetimes within 6 pixels of the edges of the sample, since one cannot extract a 13x13 window centred on these pixels. Consequently, the GP's contributions from these regions are removed from the IR reconstruction. A total of 20 iteration steps are used for minimizing the cost function - further iteration typically gives insignificant change in the solution. A Python implementation of the IR reconstruction for upsampling a 256×256 image to 512×512 (2x upsampling) takes approximately 80s.

Local prior window size and function selection. In lieu of an analytical formula linking lifetime and intensity, the extent of the windows and the form of these local functions must be found empirically from the data. A set of window sizes, spanning from 2×2 to 8×8 were tested. Likewise, a set of different function forms were tried, including Gaussian Processes with RBF kernels, B-spline fitting, and Interpolation (Nearest, Linear, Cubic). The best form was found by comparing mean upsampling metrics over a validation set of four biological samples (shown in Fig. 3, 4, 5, 6) for four upsampling factors each (2x, 4x, 8x and 16x) - see Supplement for details. 5×5 windows and linear interpolation yielded the best results, hence these are used in all results shown in this work.

Global prior data augmentation and training. Our global priors are generated from a neural network trained on a training set of intensity patches. The training set includes all patches with a central lifetime estimate, and their rotated and mirrored copies. Further, for high upsampling factors, their nearest neighbours are added to the training set (these neighbours are also in the test set), as well as the rotated and mirrored

copies of these neighbours. Once trained, the network is used to evaluate all the original intensity patches, which includes all the training patches and all patches with unknown lifetimes. A 3×3 neighbourhood gives 9x more windows to train on; combined with augmentation for rotation and reflection invariance, this yields 72x more data. This dataset augmentation method assumes that within the pixel-pitch of the intensity image, lifetime varies slowly. This does not hold for every pixel, therefore, estimated labels have inherent uncertainty. Nonetheless, this form of blind labelling increases the diversity of the training input set. This was empirically found to be necessary for 8×8 and 16×16 upsampling not because of the upsampling factor, but rather because the decimated low-resolution input was so small, that the training set size was a severe limitation.

The 13×13 training and testing patches are copied into two channels, producing $13 \times 13 \times 2$ input instances for the neural network. One channel is normalised on a per-patch basis, drawing focus to the shape and texture of the patch's content. The other channel is divided by the maximum of the original intensity image, maintaining absolute intensity variations.

For the results shown in this paper, the network was trained 3 times with different random initialisations, and the prior of median quality was selected. We train on intensity patches $13 \times 13 \times 2$ large, with ADAM [64], using a batch size of 100 over 150 epochs with mean absolute error (MAE) as the training loss, on an NVIDIA GeForce RTX 2080 Ti. Training on a 125×125 FLIM

image (i.e. $125 \times 125 \times 8 = 125000$ intensity patches due to $8 \times$ data augmentation for rotation and reflection invariance), takes ~ 25 minutes of training on this hardware using TensorFlow, irrespective of the target image size. Trained DNNs tend to have negligible bias compared to the standard deviation of prediction. Since they are trained from scratch on each new sample, their training and validation losses vary from sample to sample, up-sampling factor to up-sampling factor, and initialisation to initialisation. Example test set performances can be seen in inset D of Fig. 3, 4, 5 and 6.

Acknowledgements.

The authors acknowledge funding from EPSRC (UK, grant no. EP/T002123/1), HORIBA Ltd., the Research Council of Lithuania (project No S-PD-24-35), the Royal Academy of Engineering (Chairs in Emerging Technology programme) and QuantIC (UK, grant no. EP/T00097X/1).

Data and materials availability.

All data needed to evaluate the conclusions in the paper are present in the paper. Raw data, SISI-FUS code, and analysis code can be found at <https://doi.org/10.5281/zenodo.10955467>.

-
- [1] I. Georgakoudi and K. P. Quinn, Optical imaging using endogenous contrast to assess metabolic state, *Annual review of biomedical engineering* **14**, 351 (2012).
 - [2] C. Stringari, H. Wang, M. Geyfman, V. Crosignani, V. Kumar, J. S. Takahashi, B. Andersen, and E. Gratton, In vivo single-cell detection of metabolic oscillations in stem cells, *Cell reports* **10**, 1 (2015).
 - [3] C. Stringari, L. Abdeladim, G. Malkinson, P. Mahou, X. Solinas, I. Lamarre, S. Brizion, J.-B. Galey, W. Sappato, R. Legouis, *et al.*, Multicolor two-photon imaging of endogenous fluorophores in living tissues by wavelength mixing, *Scientific reports* **7**, 3792 (2017).
 - [4] M. A. Yaseen, J. Sutin, W. Wu, B. Fu, H. Uhlirova, A. Devor, D. A. Boas, and S. Sakadžić, Fluorescence lifetime microscopy of nadh distinguishes alterations in cerebral metabolism in vivo, *Biomedical optics express* **8**, 2368 (2017).
 - [5] P. M. Schaefer, S. Kalinina, A. Rueck, C. A. von Arnim, and B. von Einem, Nadh autofluorescence—a marker on its way to boost bioenergetic research, *Cytometry Part A* **95**, 34 (2019).
 - [6] K. Okabe, N. Inada, C. Gota, Y. Harada, T. Funatsu, and S. Uchiyama, Intracellular temperature mapping with a fluorescent polymeric thermometer and fluorescence lifetime imaging microscopy, *Nature communications* **3**, 705 (2012).
 - [7] S. Ogikubo, T. Nakabayashi, T. Adachi, M. S. Islam, T. Yoshizawa, M. Kinjo, and N. Ohta, Intracellular ph sensing using autofluorescence lifetime microscopy, *The Journal of Physical Chemistry B* **115**, 10385 (2011).
 - [8] K. Suhling, Y. Teijeiro-Gonzalez, I. E. Steinmark, A. L. James, A. M. Economou, G. Yahiglu, A. Le Marois, L. M. Hirvonen, J. Nedbal, J. A. Levitt, *et al.*, Fluorescence lifetime imaging for viscosity and diffusion measurements, in *Multiphoton Microscopy in the Biomedical Sciences XIX*, Vol. 10882 (SPIE, 2019) pp. 115–123.
 - [9] E. B. van Munster and T. W. Gadella, Fluorescence lifetime imaging microscopy (flim), *Microscopy Techniques*, 143 (2005).
 - [10] B. T. Bajar, E. S. Wang, S. Zhang, M. Z. Lin, and J. Chu, A guide to fluorescent protein fret pairs, *Sensors* **16**, 1488 (2016).
 - [11] R. Datta, T. M. Heaster, J. T. Sharick, A. A. Gillette, and M. C. Skala, Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications, *Journal of biomedical optics* **25**, 071203 (2020).
 - [12] M. Y. Berezin and S. Achilefu, Fluorescence lifetime measurements and biological imaging, *Chemical reviews* **110**, 2641 (2010).
 - [13] Y. Sun, N. Hatami, M. Yee, J. Phipps, D. S. Elson, F. Gorin, R. J. Schrot, and L. Marcu, Fluorescence lifetime imaging microscopy for brain tumor image-guided surgery, *Journal of biomedical optics* **15**, 056022 (2010).
 - [14] D. Gorpas, J. Phipps, J. Bec, D. Ma, S. Dochow, D. Yankelevich, J. Sorger, J. Popp, A. Bewley, R. Gandour-Edwards, *et al.*, Autofluorescence lifetime augmented reality as a means for real-time robotic surgery guidance in human patients, *Scientific reports* **9**, 1187 (2019).

- [15] J. R. Lakowicz, Time-Domain Lifetime Measurements, in *Principles of Fluorescence Spectroscopy*, edited by J. R. Lakowicz (Springer US, Boston, MA, 2006) pp. 97–155.
- [16] J. R. Lakowicz, *Principles of fluorescence spectroscopy* (Springer, 2006).
- [17] J. A. Jo, Q. Fang, and L. Marcu, Ultrafast method for the analysis of fluorescence lifetime imaging microscopy data based on the laguerre expansion technique, *IEEE Journal of Selected Topics in Quantum Electronics* **11**, 835 (2005).
- [18] R. M. Ballew and J. Demas, An error analysis of the rapid lifetime determination method for the evaluation of single exponential decays, *Analytical Chemistry* **61**, 30 (1989).
- [19] D. Elson, I. Munro, J. Requejo-Isidro, J. McGinty, C. Dunsby, N. Galletly, G. Stamp, M. Neil, M. Lever, P. Kellett, *et al.*, Real-time time-domain fluorescence lifetime imaging including single-shot acquisition with a segmented optical image intensifier, *New Journal of Physics* **6**, 180 (2004).
- [20] S. P. Poland, A. T. Erdogan, N. Krstajić, J. Levitt, V. Devauges, R. J. Walker, D. D.-U. Li, S. M. Ameen-Beg, and R. K. Henderson, New high-speed centre of mass method incorporating background subtraction for accurate determination of fluorescence lifetime, *Optics express* **24**, 6899 (2016).
- [21] R. F. Laine, C. Poudel, and C. F. Kaminski, A method for the fast and photon-efficient analysis of time-domain fluorescence lifetime image data over large dynamic ranges, *Journal of Microscopy* **287**, 138 (2022).
- [22] J. T. Smith, R. Yao, N. Sinsuebphon, A. Rudkouskaya, N. Un, J. Mazurkiewicz, M. Barroso, P. Yan, and X. Intes, Fast fit-free analysis of fluorescence lifetime imaging via deep learning, *Proceedings of the national academy of sciences* **116**, 24019 (2019).
- [23] V. Zickus, M.-L. Wu, K. Morimoto, V. Kapitany, A. Fatima, A. Turpin, R. Insall, J. Whitelaw, L. Machesky, C. Bruschini, *et al.*, Fluorescence lifetime imaging with a megapixel spad camera and neural network lifetime estimation, *Scientific Reports* **10**, 20986 (2020).
- [24] V. Kapitany, V. Zickus, A. Fatima, G. Carles, and D. Faccio, Single-shot time-folded fluorescence lifetime imaging, *Proceedings of the National Academy of Sciences* **120**, e2214617120 (2023).
- [25] T. Knöpfel and C. Song, Optical voltage imaging in neurons: moving from technology development to practical tool, *Nature Reviews Neuroscience* **20**, 719 (2019).
- [26] L. Dvinskikh, H. Sparks, K. T. MacLeod, and C. Dunsby, High-speed 2d light-sheet fluorescence microscopy enables quantification of spatially varying calcium dynamics in ventricular cardiomyocytes, *Frontiers in Physiology* **14**, 1079727 (2023).
- [27] J. McGinty, N. P. Galletly, C. Dunsby, I. Munro, D. S. Elson, J. Requejo-Isidro, P. Cohen, R. Ahmad, A. Forsyth, A. V. Thillainayagam, *et al.*, Wide-field fluorescence lifetime imaging of cancer, *Biomedical optics express* **1**, 627 (2010).
- [28] D. J. Kelly, S. C. Warren, S. Kumar, J. L. Lagarto, B. T. Dyer, A. Margineanu, E. W.-F. Lam, C. Dunsby, and P. M. French, An automated multiwell plate reading flim microscope for live cell autofluorescence lifetime assays, *Journal of Innovative Optical Health Sciences* **7**, 1450025 (2014).
- [29] A. J. Bowman, B. B. Klopfer, T. Juffmann, and M. A. Kasevich, Electro-optic imaging enables efficient wide-field fluorescence lifetime microscopy, *Nature communications* **10**, 4561 (2019).
- [30] A. J. Bowman, C. Huang, M. J. Schnitzer, and M. A. Kasevich, Wide-field fluorescence lifetime imaging of neuron spiking and subthreshold activity in vivo, *Science* **380**, 1270 (2023).
- [31] C. Bruschini, H. Homulle, I. M. Antolovic, S. Burri, and E. Charbon, Single-photon avalanche diode imagers in biophotonics: review and outlook, *Light: Science & Applications* **8**, 87 (2019).
- [32] Matplotlib, Interpolations for imshow.
- [33] S. Fadnavis, Image interpolation techniques in digital image processing: an overview, *International Journal of Engineering Research and Applications* **4**, 70 (2014).
- [34] Q. Sun, J. Zhang, X. Dun, B. Ghanem, Y. Peng, and W. Heidrich, End-to-end learned, optically coded super-resolution spad camera, *AcM Transactions on Graphics (TOG)* **39**, 1 (2020).
- [35] C. Callenberg, A. Lyons, D. d. Brok, A. Fatima, A. Turpin, V. Zickus, L. Machesky, J. Whitelaw, D. Faccio, and M. Hullin, Super-resolution time-resolved imaging using computational sensor fusion, *Scientific reports* **11**, 1689 (2021).
- [36] E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on information theory* **52**, 489 (2006).
- [37] Q. Sun, X. Dun, Y. Peng, and W. Heidrich, Depth and transient imaging with compressive spad array cameras, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018) pp. 273–282.
- [38] F. Soldevila, A. Lenz, A. Ghezzi, A. Farina, C. D’Andrea, and E. Tajahuerce, Giga-voxel multidimensional fluorescence imaging combining single-pixel detection and data fusion, *Optics Letters* **46**, 4312 (2021).
- [39] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, Diffusercam: lensless single-exposure 3d imaging, *Optica* **5**, 1 (2018).
- [40] W. T. Freeman, T. R. Jones, and E. C. Pasztor, Example-based super-resolution, *IEEE Computer graphics and Applications* **22**, 56 (2002).
- [41] H. Chang, D.-Y. Yeung, and Y. Xiong, Super-resolution through neighbor embedding, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, Vol. 1 (IEEE, 2004) pp. I–I.
- [42] J. Yang, J. Wright, T. S. Huang, and Y. Ma, Image super-resolution via sparse representation, *IEEE transactions on image processing* **19**, 2861 (2010).
- [43] R. Timofte, V. De Smet, and L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, in *Proceedings of the IEEE international conference on computer vision* (2013) pp. 1920–1927.
- [44] Z. Wang, J. Chen, and S. C. Hoi, Deep learning for image super-resolution: A survey, *IEEE transactions on pattern analysis and machine intelligence* **43**, 3365 (2020).
- [45] V. Kapitány, *AI for time-resolved imaging: from fluorescence lifetime to single-pixel time of flight*, Ph.D. thesis, University of Glasgow (2023).
- [46] C. Dong, C. C. Loy, K. He, and X. Tang, Learning a deep convolutional network for image super-resolution, in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part IV 13* (Springer, 2014) pp. 184–199.
- [47] C. Dong, C. C. Loy, and X. Tang, Accelerating the super-resolution convolutional neural network, in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14* (Springer, 2016) pp. 391–407.

- [48] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, Photo-realistic single image super-resolution using a generative adversarial network, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017) pp. 4681–4690.
- [49] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, Esgan: Enhanced super-resolution generative adversarial networks, in *Proceedings of the European conference on computer vision (ECCV) workshops* (2018) pp. 0–0.
- [50] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, Image super-resolution via iterative refinement, *IEEE transactions on pattern analysis and machine intelligence* **45**, 4713 (2022).
- [51] V. Mannam, Y. Zhang, X. Yuan, C. Ravasio, and S. S. Howard, Machine learning for faster and smarter fluorescence lifetime imaging microscopy, *Journal of Physics: Photonics* **2**, 042005 (2020).
- [52] D. Glasner, S. Bagon, and M. Irani, Super-resolution from a single image, in *2009 IEEE 12th international conference on computer vision* (IEEE, 2009) pp. 349–356.
- [53] A. C. Quiros, N. Coudray, A. Yeaton, X. Yang, L. Chiriboga, A. Karimkhan, N. Narula, H. Pass, A. L. Moreira, J. L. Quesne, *et al.*, Self-supervised learning in non-small cell lung cancer discovers novel morphological clusters linked to patient outcome and molecular phenotypes, arXiv preprint arXiv:2205.01931 (2022).
- [54] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines, *NPJ digital medicine* **3**, 136 (2020).
- [55] K.-H. Thung, P.-T. Yap, and D. Shen, Multi-stage diagnosis of alzheimer’s disease with incomplete multi-modal data via multi-task deep learning, in *International Workshop on Deep Learning in Medical Image Analysis* (Springer, 2017) pp. 160–168.
- [56] M. Person, M. Jensen, A. O. Smith, and H. Gutierrez, Multimodal fusion object detection system for autonomous vehicles, *Journal of Dynamic Systems, Measurement, and Control* **141**, 071017 (2019).
- [57] T. Trzcinski, Multimodal social media video classification with deep neural networks, in *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018*, Vol. 10808 (SPIE, 2018) pp. 879–886.
- [58] K. Samimi, D. E. Desa, W. Lin, K. Weiss, J. Li, J. Huisken, V. Miskolci, A. Huttenlocher, J. V. Chacko, A. Velten, *et al.*, Light-sheet autofluorescence lifetime imaging with a single-photon avalanche diode array, *Journal of Biomedical Optics* **28**, 066502 (2023).
- [59] C. Ji, X. Wang, K. He, Y. Xue, Y. Li, L. Xin, W. Zhao, J. Tian, and L. Sheng, Compressed fluorescence lifetime imaging via combined tv-based and deep priors, *Plos one* **17**, e0271441 (2022).
- [60] A. Chambolle, An algorithm for total variation minimization and applications, *Journal of Mathematical imaging and vision* **20**, 89 (2004).
- [61] R. K. Henderson, N. Johnston, F. M. Della Rocca, H. Chen, D. D.-U. Li, G. Hungerford, R. Hirsch, D. Mcloskey, P. Yip, and D. J. Birch, A 192x128 time correlated spad image sensor in 40-nm cmos technology, *IEEE Journal of Solid-State Circuits* **54**, 1907 (2019).
- [62] W. Becker, Spc-qc-104: Precision flim and fast flim in one (2023).
- [63] R. E. Itoh, K. Kurokawa, Y. Ohba, H. Yoshizaki, N. Mochizuki, and M. Matsuda, Activation of rac and cdc42 video imaged by fluorescent resonance energy transfer-based single-molecule probes in the membrane of living cells, *Molecular and cellular biology* **22**, 6582 (2002).
- [64] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
- [65] H. Szmacinski and J. R. Lakowicz, Fluorescence lifetime-based sensing and imaging, *Sensors and Actuators B: Chemical* **29**, 16 (1995).
- [66] K. Suhling, L. M. Hirvonen, J. A. Levitt, P.-H. Chung, C. Tregidgo, A. Le Marois, D. A. Rusakov, K. Zheng, S. Ameer-Beg, S. Poland, *et al.*, Fluorescence lifetime imaging (flim): Basic concepts and some recent developments, *Medical photonics* **27**, 3 (2015).
- [67] J. R. Lakowicz, Radiative decay engineering: biophysical and biomedical applications, *Analytical biochemistry* **298**, 1 (2001).
- [68] A. Schwartz, L. Wang, E. Early, A. Gaigalas, Y.-z. Zhang, G. E. Marti, and R. F. Vogt, Quantitating fluorescence intensity from fluorophore: The definition of mesf assignment, *Journal of research of the National Institute of Standards and Technology* **107**, 83 (2002).
- [69] M. Stöckl and A. Herrmann, Detection of lipid domains in model and cell membranes by fluorescence lifetime imaging microscopy, *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1798**, 1444 (2010).
- [70] A. Pierzyńska-Mach, P. A. Janowski, and J. W. Dobrucki, Evaluation of acridine orange, lysotracker red, and quinacrine as fluorescent probes for long-term tracking of acidic vesicles, *Cytometry Part A* **85**, 729 (2014).
- [71] A. K. Estandarte, S. Botchway, C. Lynch, M. Yusuf, and I. Robinson, The use of dapi fluorescence lifetime imaging for investigating chromatin condensation in human chromosomes, *Scientific reports* **6**, 31417 (2016).
- [72] F. H. van der Linden, E. K. Mahlandt, J. J. Arts, J. Beumer, J. Puschhof, S. M. de Man, A. O. Chertkova, B. Ponsioen, H. Clevers, J. D. van Buul, *et al.*, A turquoise fluorescence lifetime-based biosensor for quantitative imaging of intracellular calcium, *Nature communications* **12**, 7159 (2021).
- [73] D. R. J. Verboogen, N. González Mancha, M. Ter Beest, and G. van den Bogaart, Fluorescence lifetime imaging microscopy reveals rerouting of snare trafficking driving dendritic cell activation, *Elife* **6**, e23525 (2017).
- [74] R. Dosselmann and X. D. Yang, A comprehensive assessment of the structural similarity index, *Signal, Image and Video Processing* **5**, 81 (2011).
- [75] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on image processing* **15**, 3440 (2006).
- [76] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018) pp. 586–595.
- [77] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, High-resolution image synthesis with latent diffusion models, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022) pp. 10684–10695.
- [78] R. Zhang, Lpips v0.1.4 (2021).

SUPPLEMENTARY INFORMATION

I. FLUORESCENCE INTENSITY AND LIFETIME

Lifetime and quantum yield. Fluorescence lifetime is described in literature as being independent of fluorescent intensity [12, 65], and of fluorophore concentration [66] and excitation intensity. Here, we examine the context of these claims, and demonstrate the limitations of these generalisations.

In fluorescence, a photon excites a ground-state electron into an excited state, which then decays back to the ground state radiatively at a rate known as the decay rate. Other decay pathways compete with fluorescence, such as non-radiative decay and inter-system energy transfer between the fluorescent molecule and its environment. The probability of emitting a fluorescent photon per excitation event is called the quantum yield of fluorescence. Fluorescent intensity is the product of excitation intensity, the absorbance of the fluorophores (which depends strongly on their concentration) and fluorescence quantum yield.

The decay rate is the inverse of fluorescence lifetime, which is the expected time that an electron spends in the excited state before decaying via fluorescence. This is an intrinsic property of the molecule, and thus, is assumed to be independent of factors like fluorophore concentration. Consequently, fluorescence lifetime can be used to distinguish between different molecule populations.

However, fluorophores interact with their environment. The environment, in turn, can modulate both the excitation and emission pathways, changing both intensity and lifetime. Excitation can be enhanced or quenched by metallic surfaces or particles within the sample such as silver [67] via plasmonic resonance. Emission is modulated via nonradiative (or alternative) decay pathways, quenching the molecule's radiative fluorescence as well as its lifetime, as derived in the main section of the paper.

Fluorescence intensity. An imaging system generates a fluorescence intensity signal that depends on the spectral radiance $L_f(\lambda_o)$ of the sample and the net photon detection efficiency $PDE(\lambda_o)$ of the imaging system.

Let us consider a thin sample within the focal length of the optical system, using an epifluorescence setup. Using nomenclature from [68], the spectral radiance $L_f(\lambda_o)[Wsr^{-1}m^{-2}nm^{-1}]$ emitted by the sample at wavelength λ_o from excitation light at λ_x is given by:

$$L_f(\lambda_o) = I_x N \Omega \epsilon(\lambda_x) Q(\tau, \lambda_o, \lambda_x) \quad (10)$$

where I_x is the incident excitation power [W], $\tilde{N}(x, y)$ is the 2D concentration of fluorophores [m^2] (the integral of the 3D concentration N of fluorophores along the length of the sample along the optical axis z , $\tilde{N}(x, y) = \int_z N(x, y, z) dz$), Ω is the solid angle through which emitted light is collected from the sample [sr], $\epsilon(\lambda_x)$ is the absorptivity [m^2] of the fluorophore as per the Beer-Lambert Law, and $Q(\tau, \lambda_o, \lambda_x)$ is the lifetime-dependent, spectral quantum yield of fluorescence.

This spectral radiance is imaged onto a detector that has a response $R[AW^{-1}]$ using a system with some étendue $\Gamma[m^2sr]$. To obtain the signal generated by the emission spectrum, we must integrate over the emission spectrum, giving:

$$s_o = \int_{\lambda_o} L_f(\lambda_o) \Gamma R(\lambda_o) d\lambda_o \quad (11)$$

Substituting Eq. 10 into Eq. 12, and integrating over the acquisition time t_a gives us the measurement $M[C]$:

$$M = \int_{\lambda_o} \int_{t=0}^{t_a} I_x \tilde{N} \Omega \epsilon(\lambda_x) Q(\tau, \lambda_o, \lambda_x) \Gamma R(\lambda_o) d\lambda_o dt \quad (12)$$

A fixed excitation and detection system allows us to calibrate the intensity I_x , the collection solid angle Ω , the étendue Γ , the response $R(\lambda_o)$, and the acquisition time t_a . Therefore, variations of intensity across the field of view will depend on molecular concentration \tilde{N} and absorptivity $\epsilon(\lambda_x)$ (whose product is the absorbance of the fluorophores), as well as the spectral quantum yield of fluorescence Q , which depends on fluorescence lifetime.

Dependence of intensity on lifetime. Absorbance and fluorescence lifetime appear to be unrelated, hence absorbance (ergo, fluorophore concentration) is an unpredictable confounding variable in intensity-lifetime dependencies.

Consequently, a fluorescence intensity measurement alone cannot give us full lifetime information. We therefore must use statistical priors to extract intensity-lifetime dependencies in the presence of biological confounding variables. A local prior is developed to extract dependencies when lifetime varies more rapidly in space than these confounding variables, or when they correlate with lifetime on local scales (either positively or inversely).

Further, many biological samples absorb fluorophores into particular subcellular compartments such as the cell membrane [69], vesicles [70] or the nucleus [71]. This results in lifetime patterns that often track cellular morphology. A global prior is developed to extract such dependencies. If absorbance were completely randomly distributed (which tends not to be the case in real samples), our method would not offer improvement over interpolation, instead our methods might overfit on noise patterns. To prevent this, our algorithm uses TV-filtering to prevent

very unrealistically noisy lifetime estimates.

The question is whether recognisable intensity-lifetime dependencies actually exist in biological samples, or if absorbance renders them unusable. Below, we consider a series of case studies of fluorophore-environment interactions reported in literature, focusing on how these interactions modulate intensity and lifetime.

Case studies. Okabe et. al. [6] used a complex fluorescent molecule made of a thermosensitive unit, a hydrophilic unit and a fluorescent unit to monitor temperature. In response to higher temperature, the molecule becomes hydrophobic, curling up and increasing both fluorescence quantum yield (thereby, intensity) and fluorescence lifetime. Fluorophore concentration still affects fluorescence intensity; however, locally (in regions of uniform concentration or at organelle edges), intensity and lifetime covary. Indeed, the authors use this probe to demonstrate temperature differences between the nucleus and cytoplasm of cells, which are visibly differentiable on both the lifetime and intensity maps.

Ogikubo et. al. [7] used cellular auto-fluorescence of NADH to monitor intracellular pH. Their results show evident covariance of fluorescent intensity with fluorescence lifetime within cells; even though intensity is not a marker of pH, both intensity and fluorescence lifetime depend on the location of NADH within the cell. The reason for this is not explicitly explored, but different works have shown that the ratio of bound to free NADH depends on the local metabolism of the cell, which influences both the fluorescence lifetime and concentration of NADH autofluorescence [11]. Correlations are similarly visible between NADH fluorescence intensity and lifetime in works by Stringari et. al. [12], as both of these parameters are covariate with cellular redox ratio.

Van der Linden et. al. [72] use FLIM as a tool for a quantitative measurement of calcium levels, independent of hardware. However, for a given hardware, fluorescent intensity spikes clearly show calcium spikes, even if they do not give absolute calcium concentrations on their own. Indeed, the authors demonstrate that their FLIM probe works by showing Supplementary videos of fluorescent intensity and lifetime side-by-side, which both show synchronised flickering. Lifetime and intensity are strongly temporally correlated and are also locally correlated: cellular organoids have quasi uniform intensity and lifetime, both of which experience sudden gradients at organoid boundaries.

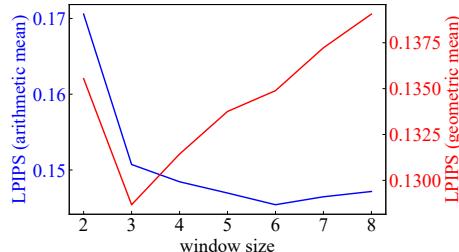
Verboogen et. al. [73] demonstrate a FLIM-FRET probe for the imaging of SNARE trafficking in dendrites. For example, Förster resonance energy transfer (FRET) relies on this phenomenon. In FRET, the fluorophore, known as the donor, is linked to another fluorophore known as the acceptor, such that their relative conformation can change. The donor molecule is excited and its fluorescence measured. If the donor and acceptor are far, the donor will decay as if it were alone. If the donor and acceptor are in close vicinity, excited electrons can transfer energy from the donor onto the acceptor molecule, providing an alternative decay path for electrons, decreasing both fluorescence quantum yield (thus, intensity) and lifetime.

Gorpas et. al. [14] use skin autofluorescence to determine qualitative boundaries between cancerous and healthy skin tissue. They demonstrate that FLIM shows skin cancer; they do so by overlaying an augmented reality image of FLIM onto a visibly melanated patch of skin, whose colour correlates strongly with its lifetime.

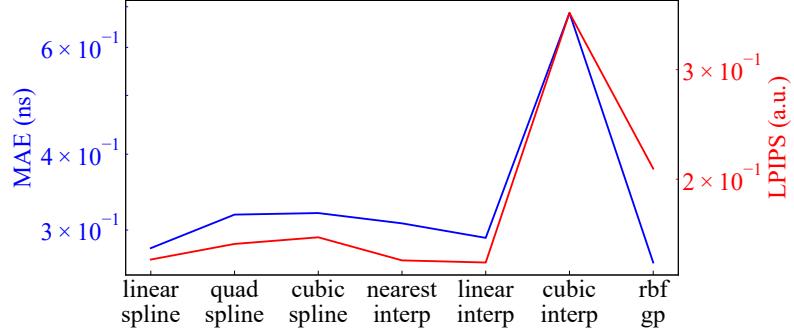
II. LOCAL PRIOR

We performed a study to find the best window size and best function to map fluorescent intensity onto fluorescence lifetime with local priors. The window sizes were in the range 2 to 8, while the functions were a set of common schemes, ranging from B-splines (linear, quadratic and cubic), through regular interpolation (nearest, linear and cubic); and kriging (radial basis function Gaussian process fitting).

We applied these window sizes and functions to 4 samples (including the three shown in Figures 4, 5, and 6) and 4 upsampling factors (2,4,8, and 16x). We evaluated the methods based on mean-absolute-error and LPIPS between the reconstruction and ground truth, averaged over these 4×4 scenarios. Our results are shown in Supplementary Fig. 1 and Supplementary Fig. 2. Based on these results, we decided to use a window size of 5 and linear interpolation for generating LPs.



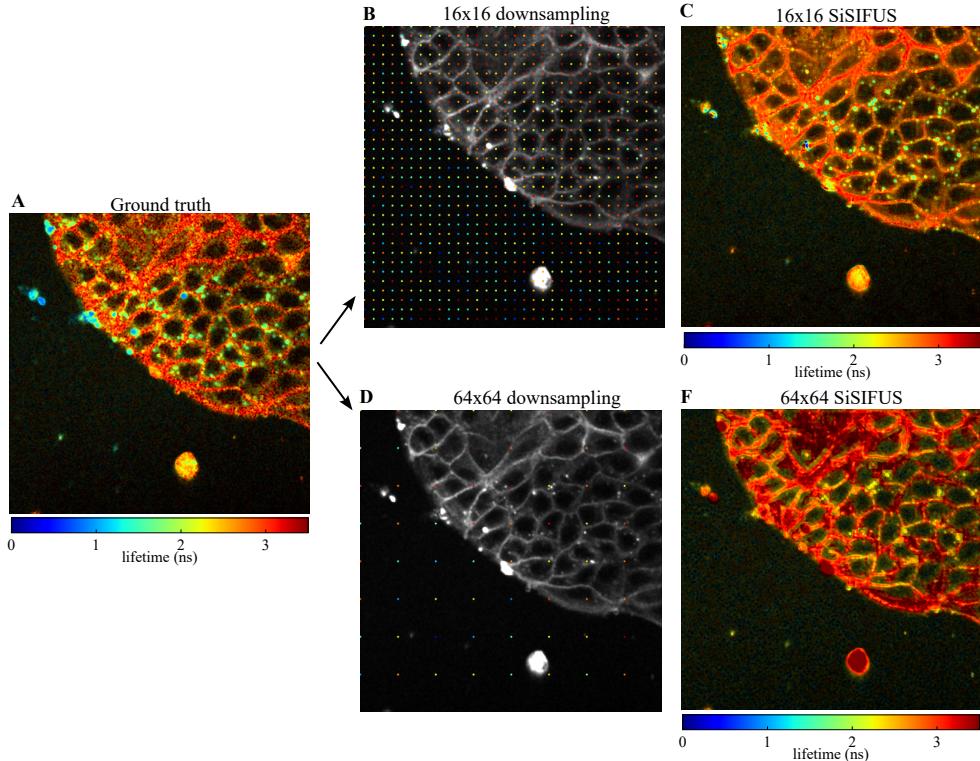
Supplementary Fig. 1. We found the mean LPIPS for priors generated using various window sizes, averaged across our 4 samples and 4 upsampling factors (2,4,8,16). We plot both the geometric and arithmetic mean.



Supplementary Fig. 2. We found the mean MAE and LPIPS for priors generated using various LCP types, averaged across our 4 samples and 4 upsampling factors (2,4,8,16).

III. SAMPLING CONSIDERATIONS

SiSIFUS relies on either local dependencies, where there are gradual changes in confounding variables such as fluorophore concentration, or global dependencies between structure and lifetime. Local dependencies involve scenarios like free fluorophores in the cytoplasm, which diffuse to achieve a locally uniform density, or beads coated in fluorescent dyes. Another example is continuous cell membranes treated with diffusive dyes like Flipper TR. Global dependencies assume that objects with similar shapes in the FOV have similar lifetimes, as illustrated in Fig. 3 and Fig. 5. For instance, in a scenario with a mixture of small and large fluorescent beads coated with different fluorophores, by sparsely sampling their lifetimes, we can infer the lifetimes of other beads in the image. Employing machine learning algorithms, SiSIFUS automates and enhances this pattern-matching process, foregoing manual matching.



Supplementary Fig. 3. **Illustration of the impact of under-sampling on image resolution.** (A) The original image is 512x512 with a pixel pitch of $0.33\mu\text{m}$. (B) Data is decimated by 16x16, with lifetime samples overlaid. (C) SiSIFUS can reasonably reconstruct the sample from intensity-lifetime pairs. (D) The ground truth lifetime is instead decimated by 64x64, resulting in only 8x8 lifetime measurements. (E) This extreme under-sampling causes SiSIFUS to fail in accurately recovering the lifetime distribution.

Capturing these dependencies is crucial for SiSIFUS; thus, sampling density must be adequate to measure them. Whether sampling locally or across the FOV, denser sampling is required for small or rare biological or mechanical structures. In general, the sparser the structures in the FOV and the greater the variety of distinct lifetimes expected to be resolved, the denser the sampling required. In the bead example, if a medium-sized bead is present but never sampled for lifetime, accurate estimation requires external knowledge. We illustrate the effect of undersampling with a 64x64 super-resolved example of the MDCK-Flipper TR sample in Fig. 3 of the main text (from 8x8 to 512x512). Results are shown in Supplementary Fig. 3.

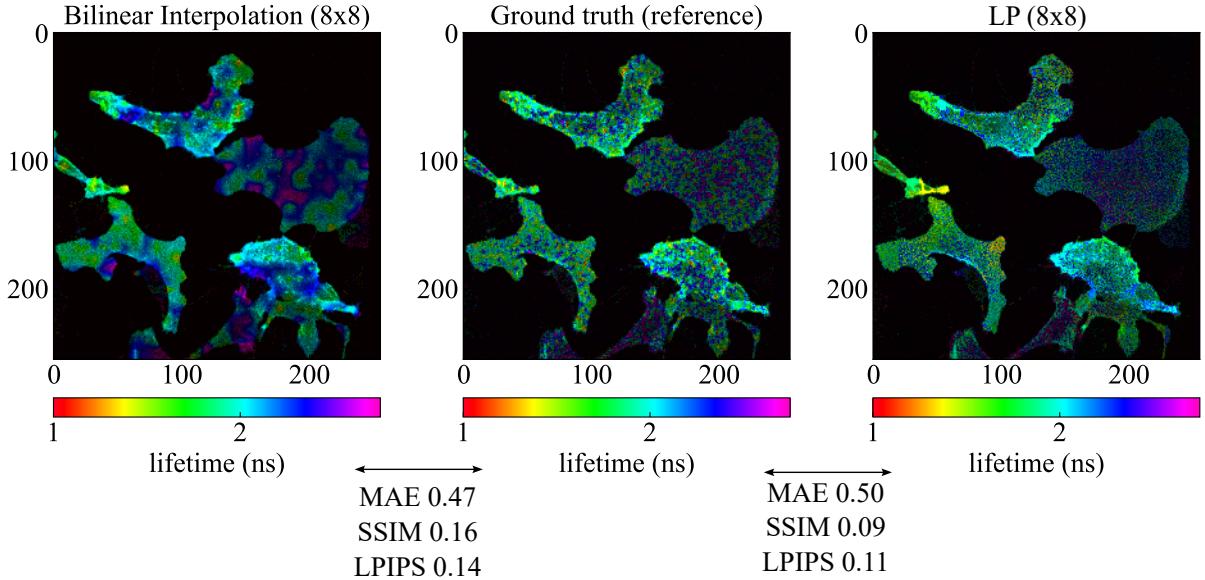
IV. VISUALISATION

As stated in the main paper, we visualise our FLIM data by overlaying (weighting) it with local contrast enhanced intensity. This allows us to see lifetime patterns more clearly than using the raw lifetime image. For this, we first choose a colormap to plot the lifetime data in and use this colormap to convert it into an RGB image. Separately, we apply Contrast Limited Adaptive Histogram Equalization (CLAHE) to the intensity image. This makes it so that the image has good contrast across the field of view, without saturating bright spots or rendering dark regions imperceptible. We then scale each channel of the RGB image with this contrast enhanced intensity, preserving the RGB image's color, but manipulating its brightness.

V. METRIC DETAILS

Pixelwise metrics such as mean absolute error (MAE), mean squared error (MSE) and peak signal-to-noise ratio (PSNR) are commonly used in image processing. Compared to MAE and MSE, PSNR is adjusted for image scale, letting it generalise image similarity across lifetime maps of different ranges. PSNR is guaranteed to favour the same method as MSE for a given sample, whilst allowing us to compare different samples as well. A quantitative and straight-forward pixel-to-pixel comparison, PSNR is valuable for automated tasks, however, a widely known issue of metrics is how poorly they reflect the human perspective of image similarity. A particularly infamous example is blurring, which produces relatively low pixelwise errors even when deteriorating image quality severely.

To address this, simple perceptual metrics such as the structural similarity index measure (SSIM) and multiscale SSIM (MS-SSIM) have been proposed. Indeed, SSIM is more sensitive to blurring than MSE and less sensitive to noise [74], akin to human perception. That said, SSIM and MS-SSIM are strongly correlated with MSE; in fact, their ‘luminance’ term is equivalent to MSE. Further, a statistical evaluation [75] found that, while SSIM correlates with human perceptual score to a factor of 0.9393, the PSNR (equivalently, MSE) baseline yields a score of 0.8709. This suggests that structural metrics, while an improvement on pixelwise metrics, still can be improved upon. In our study, we evaluate SSIM on windows of size 25. One of the early realisations regarding



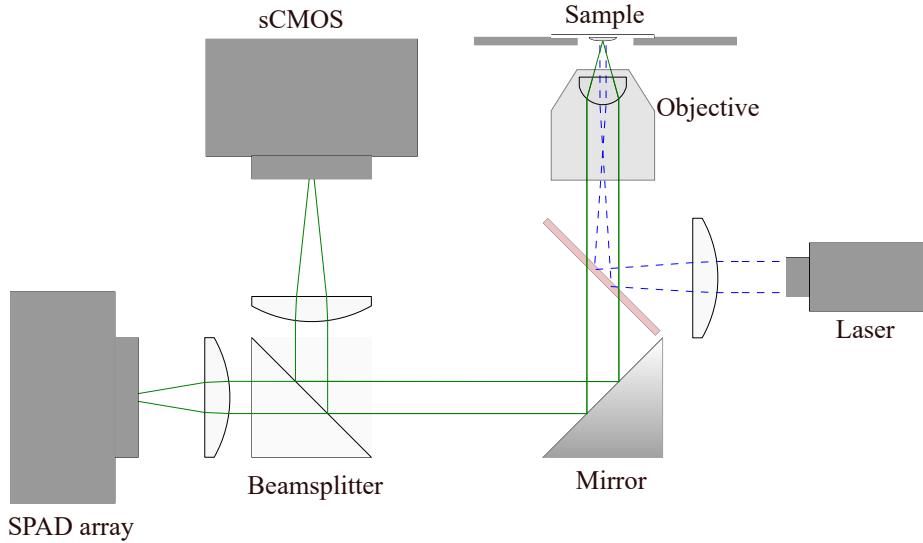
Supplementary Fig. 4. We show a comparison of interpolation vs an LP for 8×8 upsampling of a Rac-Raichu sample. Left to right: bilinear upsampling; the ground truth lifetime; LP. Both MAE and SSIM favour interpolation over the LP. However, the interpolated image is blurred and has artefacts matching the low-resolution sample grid, so, a user would likely say the LP is closer to the ground truth. LPIPS captures this perceptual similarity, favouring the LP.

convolutional neural networks was that the first few convolutional layers in a machine learning architecture tend to perform feature extraction, of growing abstractive complexity with increasing layer depth. This understanding led to the development of learned perceptual image patch similarity (LPIPS) by Zhang et. al. [76]. The authors discovered that distance in the feature space of some image-trained neural network correlates well with human perception. Since then, LPIPS has gained widespread traction, for instance the recently published image-to-image and text-to-image benchmark, Stable Diffusion, uses LPIPS in training its autoencoder structure [77]. The authors have added a Python implementation via the lpips package [78] including 3 pre-trained neural networks (In our works, we use the AlexNet network). LPIPS is open-source and provides fixed-weight neural networks with a fixed protocol for calculating image similarity, which makes our LPIPS values exactly replicable.

We show an example from experimental data to illustrate the disparity between human vision and metrics with pixelwise components in Supplementary Fig. 4.

VI. EXPERIMENTAL SETUP DETAILS

The SPAD array datasets shown in Fig. 2 and Fig. 5 of the main paper were obtained by our bespoke microscope system comprising of the 192×128 pixels SPAD array sensor (FLIMera) and the sCMOS sensor (Andor Zyla), shown in Supplementary Fig. 5. The system is a widefield epifluorescence setup, observing the sample using a 60x 1.4NA Nikon oil objective, alongside a 250mm focal length tube lens for the Zyla sensor and a 89.9mm focal length tube lens for the SPAD sensor. In this arrangement, we obtain $192 \times 128 \times 326$ datacube for the dataset shown in Fig. 2. However, the detector's pixel layout is such that two-columns of SPAD active areas are followed by a 'dead space' which is two pixel-pitches wide. For the dataset shown in Fig. 5 of the main paper, we correct for this irregular sampling. The full FOV is sampled in two shots, such that the image plane is translated by two pixel-pitches between the shots, and the outputs are then fused computationally. This yields a fluorescence datacube of size $192 \times 256 \times 326$.



Supplementary Fig. 5. Schematic representation of the experimental set up: the sample is illuminated by a pulsed laser. The fluorescence signal from the sample is split by a dichroic beamsplitter, then collected and imaged on to the high spatial resolution sCMOS sensor and the low spatial resolution SPAD array sensor.