Spectral and lifetime fluorescence unmixing via deep learning

Jason T. Smith[1+], Marien Ochoa[1+], Pingkun Yan[1] and Xavier Intes[1*]

[1]Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA.

Contact: intesx@rpi.edu

+These authors contributed equally to this work

Abstract— **Hyperspectral Fluorescence Lifetime Imaging allows for the simultaneous acquisition of spectrally-resolved temporal fluorescence emission decays. In turn, the rich multidimensional data set acquired enables to image simultaneously multiple fluorescent species to facilitate high-content molecular imaging for improved diagnosis. However, to enable quantitative imaging, inherent spectral overlap between the considered fluorescent probes and potential bleed-through has to be taken into account. Such task is performed typically either via spectral or lifetime unmixing, but neither both simultaneously. Herein, we present UNMIX-ME, a deep learning fluorescence un-mixing algorithm that is tasked with performing quantitative fluorophore unmixing using both spectral and temporal signatures simultaneously. UNMIX-ME was efficiently trained and validated using an *in silico* framework replicating the characteristics of our compressive hyperspectral fluorescent lifetime imaging platform. It was benchmarked against a conventional LSQ method for tri-exponential simulated samples. Last, UNMIX-ME performances were assessed using NIR FRET *in vitro* and *in vivo* small animal experimental data.**

Index terms—Fluorescence Lifetime Imaging, Deep Learning, hyperspectral unmixing, inverse-solver optimization

Fluorescence imaging is the most employed molecular imaging technique from the wet lab to the bed side. A key strength of fluorescence imaging is its ability to image multiple fluorophores simultaneous (multiplexing) for improved understanding of the sample molecular features. Typically, multiplexing is achieved via selecting exogenous fluorophores with distinct spectral features. However, as in the case of endogenous biomarkers in which multiple species are excited simultaneously at a given excitation wavelength, spectral overlap of the fluorophores excitation and emission is unavoidable, leading to bleed-through. Hence, spectral imaging is always associated with spectral unmixing algorithms which task is to determine the contribution of the different fluorophores at each pixel by leveraging its value in each spectral channel. Such unmixing methodologies are often based on a priori fitting techniques that use publicly available or experimentally acquired "pure" spectra. [1]

Though, spectral imaging and linear unmixing are sensitive to noise, large spectral overlap and/or wrong or incomplete spectral information [2].

Besides fluorescence spectrum, it is also possible with dedicated instruments to quantify fluorescence lifetime, which is an intrinsic characteristic of the fluorophore independent of fluorescence intensity. Though, for biomedical applications, it is challenging to perform lifetime quantification beyond bi-exponential models and hence two distinct fluorophores (or fluorophore states). Recently, there has been great interest in performing multi- or hyper-spectral FLI to augment FLI imaging. Especially, coupling spectral unmixing with FLI has the potential to achieve significantly higher unmixing sensitivity than that of intensity-based methods alone. [3] Still, despite great progress in implementing new instrumental approaches to collect such multidimensional data sets, the approach to perform unmixing is still limited to one dimension. Herein, we propose, a methodology to perform fluorophore unmixing that simultaneously leverages spectral and lifetime contrast.

We report on a deep convolutional neural network (CNN), UNMIX-ME, trained entirely *in silico*, capable of retrieving the spatially resolved spectral coefficients associated with individual fluorophores. The methodology proposed is developed within the context of Hyperspectral Macroscopic Fluorescence Lifetime Imaging (HMFLI). Such approach is based on a recently proposed novel instrumental concept that leverages a single-pixel strategy to acquire 16 spectrally resolved FLI channels over large field of views (FOV) simultaneously [4]. Through the use of Deep Learning (DL), HMFLI has proven capable of probing nanoscale biomolecular interactions across large fields of view (FOV) at resolutions as high as 128x128 within minutes [5]. DL has also greatly improved the processing time for its inverse solving procedure, yielding intensity and lifetime reconstructions in a single framework, through usage of simulated training data mimicking the single-pixel data generation [6]. UNMIX-ME further enhance HMFLI hyperspectral classification by providing accurate unmixing capabilities.

First, we report on the design of UNMIX-ME architecture. Then we describe the novel data simulation routine used to efficiently generate 16-channel fluorescence temporal point-spread functions (TPSFs) used to train our CNN – which bypasses the need of collecting large quantities of experimental data and enables the enforcement of correct

parametric mapping to ground-truth in the case of mono-spectral data, instead of relying on fitting procedures. The performances of UNMIX-ME are reported in the case of a tri-component unmixing. To further validate UNMIX-ME, we report on its ability to handle experimental data sets not used during training. First, we provide evidence of its capability to unmix *in vitro* fluorophores undergoing Föster Resonance Energy Transfer (FRET) with excitation and emission spectra in the near-infrared (NIR) – i.e. fluorophores currently known to possess short lifetime values (sub-nanosecond) and correspondingly high analytic complexity. Lastly, we present lifetime unmixing results for *in vivo* acquired Trastuzumab AF700 and AF750 conjugated FRET pair, which was injected to an athymic nude mouse bearing a tumor xenograft and imaged 76 hours post-injection.
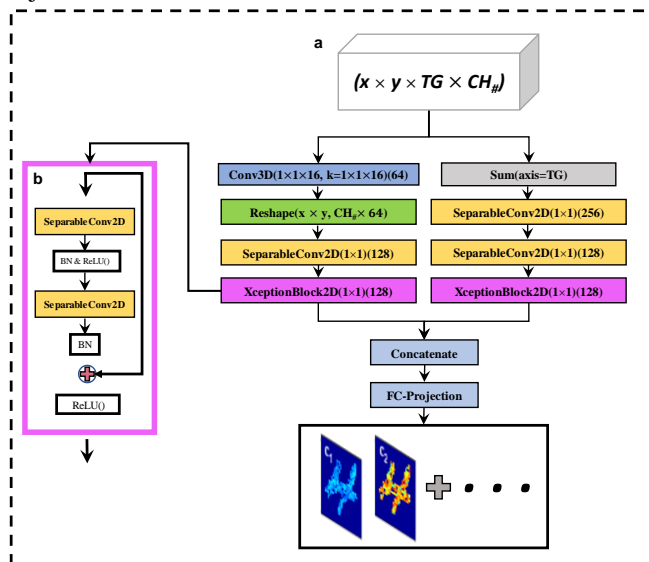

Fig 1. UNMIX-ME model architecture. All 16-channel TPSF input at every spatial location is mapped to spatially independent unmixed fluorescence coefficient values (**a**). A residual block comprised of 1×1 separable convolutions [7] (referred to as an "XceptionBlock" [8]) (**b**) is illustrated in greater detail.

UNMIX-ME's architecture (**Fig 1**) was crafted such that extraction of temporal information was made the primary focus while keeping in mind the computational burden necessary for operation on 16 channels-worth of TPSF data. Given that the use of 3D convolutional operations (Conv3D) is notoriously expensive computationally, just a single 3D convolution possessing both a kernel size and stride of 1×1×16 was included – allowing for a significant reduction in parameters within the early layers. Further, a second branch from the input immediately took the sum along the temporal axis, essentially obtaining continuous wave (CW) data for additional, computationally efficient feature extraction. Both input branches were followed by 2D separable convolutions [7] with kernel size 1×1 as a more computationally friendly alternative for spatially-independent temporal and spectral feature extraction. Moreover, "XceptionBlock" [8] operations (i.e., residual blocks with 1×1 separable convolutions) were included to ensure that our model would reap the benefits obtained through residual learning [9] while maintaining focus on the

primary objective – spatially-independent temporal and spectral feature extraction.

As **Fig 2** illustrates, the data generation workflow for efficient but comprehensive training employed in this work follows the scheme of [6], [10] where a binary handwritten number dataset MNIST was used for assignment of spatially-independent random variables during TPSF ($\Gamma(t)$, **Eq. 1**) simulation (including short-lifetime (τ_1), long-lifetime (τ_2) and intensity scalar (I)) along with output retrieval (coefficient values (c_1 and c_2 for bi-exponential case, **Eq. 2**)[1]). The data simulation can be described through:

$$\Gamma(t) = I \times IRF(t) * [a_1 e^{-t/\tau_1} + a_2 e^{-t/\tau_2}] \quad \textbf{(1)}$$

$$\begin{pmatrix} y_{1,1} \\ \vdots \\ y_{16,1} \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ \vdots & \vdots \\ a_{16,1} & a_{16,2} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \quad \textbf{(2)}$$

Where $IRF(t)$, a and y correspond to the instrument response function, spectral intensity and observed spectra (superimposed), respectively. All variables used during spatially-independent generation of TPSFs were assigned at random over wide bounds (ex., NIR case: (τ_1, τ_2, I) ∈ [0.25-0.5 ns, 0.8-1.2 ns, 50-500 p.c.] **Fig 2(c, d)**. Further, a pair of spectral intensity profiles (a_1, a_2 – obtained experimentally) were also assigned at random to allow for representative multi-component TPSF simulation across all 16 wavelength channels. It is trivial to extend these expressions to include $n_{coeff} > 2$, (illustrated in **Fig 3**).
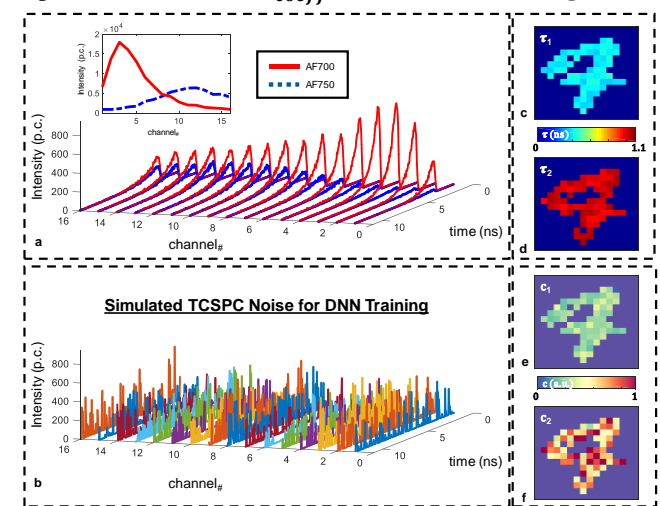

Fig 2. Simplistic illustration of simulation workflow. A binary MNIST image is assigned lifetime values within two set bounds (short-lifetime [**c**] and long-lifetime [**d**] for this bi-exponential case). Using these values, along with spatially-unique spectra for gathering intensity multipliers, 16 TPSFs are created at each non-zero spatial pixel (**a**). The coefficients are calculated shortly after (**e**, **f**). (**b**) is given to illustrate the simulated noise used during training.

Fig 2a illustrates an example of spectral profiles and corresponding simulated TPSF data of both AF700 and AF750 sans mixing (mono-exponential). Ideally, an unmixing framework would assign only a single non-zero coefficient value upon inference – but this is quite often not the case due to various sources of noise or sub-optimal

distribution of spectra across channels. Thus, each spatial location which possessed a pure mono-exponential was made to project all but a single coefficient value to zero. Furthermore, during the Ranked Hadamard inverse solving procedure, it is not uncommon to introduce problematic, high intensity noise outside of the ROIs. To forego the need of complex localized intensity thresholding, random (x, y) coordinates of each MNIST figure were also assigned noise across all 16 wavelength channels (illustrated in **Fig 2b**, details provided in *Supplementary*). Afterwards, each location was assigned corresponding ground-truth spectral coefficient values of zero (all coefficients). Thus, each 4D dataset was of size 16×16×256×16 (x, y, time-points, wavelength channels). Further research will involve the inclusion of noise models that are more pertinent to the
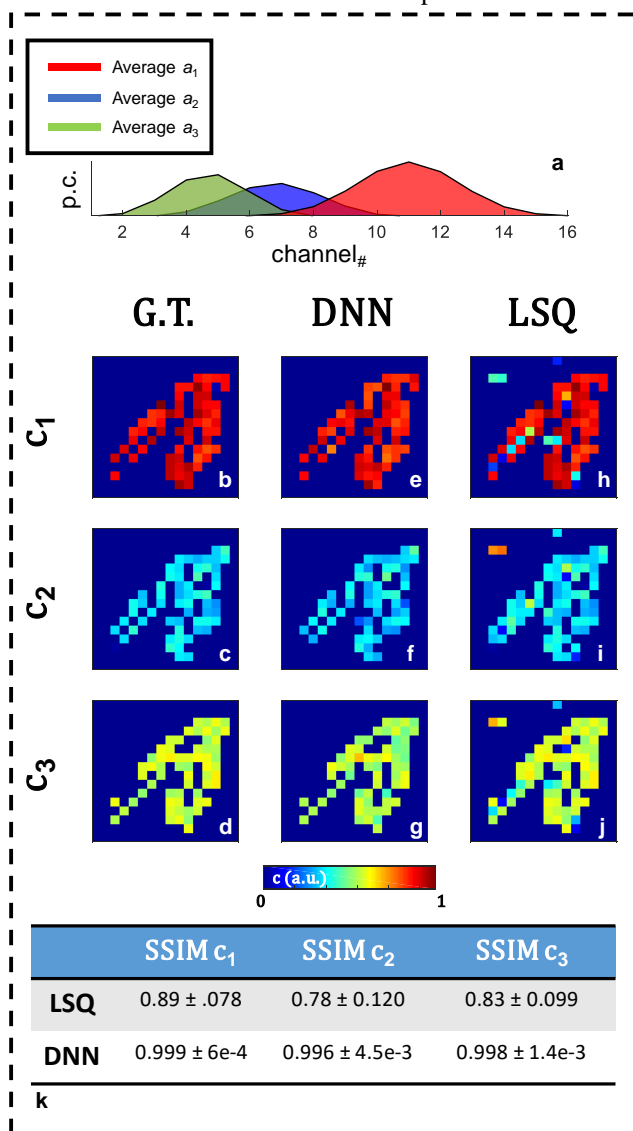

Fig 3. Three-coefficient spectral unmixing *in silico*. Averaged spectra used for simulation (**a**) are given. (**b-d**) Ground-truth values are illustrated as well as the coefficients retrieved via DNN lifetime unmixing (**e-g**) and conventional LSQ fitting (**h-j**). Table k provides average and standard deviation SSIM values calculated across 100 test samples.

system used for acquisition. To summarize, our model was trained to map three types of data to zero upon coefficient

retrieval: 1) data which possessed only zeros, 2) mono-exponential data (mapped to one scalar values with all remaining set to zero) and 3) complex noise vectors. The presented network model was built using *Tensorflow* with *Keras* backend in python. Training was performed over 50 epochs using a total number of 1,000 datasets (80%-20% training-validation) and mean-squared error (MSE) loss. The total time for data simulation and training was 15 minutes and 7.5 minutes, respectively (NVIDIA Titan GPU).

First, 100 tri-component spectral TPSF and continuous wave (CW) data were simulated to illustrate how both approaches perform during tri-spectral unmixing *in silico* (**Fig 3**). Three overlapping gaussian profiles (**Fig 3a**) were used to mimic independent 16-channel emission spectra. Further, lifetime parameters were assigned at random between three set bounds: $(\tau_1, \tau_2, \tau_3) \in [0.25\text{-}0.5, 1.5\text{-}2.0, 0.8\text{-}1.2]$ ns. Afterwards, noise was iteratively added to each spectrum to obtain thousands of spectra for use in the data simulation workflow previously described. 1,000 separate data were generated for model training, which required just one architecture alteration – assigning the number of (1×1) filters in the model's final layer to three. **Fig 3(e-g)** illustrates high spatial concordance with regards to all three coefficients, which is confirmed by the high SSIM values listed in **Fig 3k**.
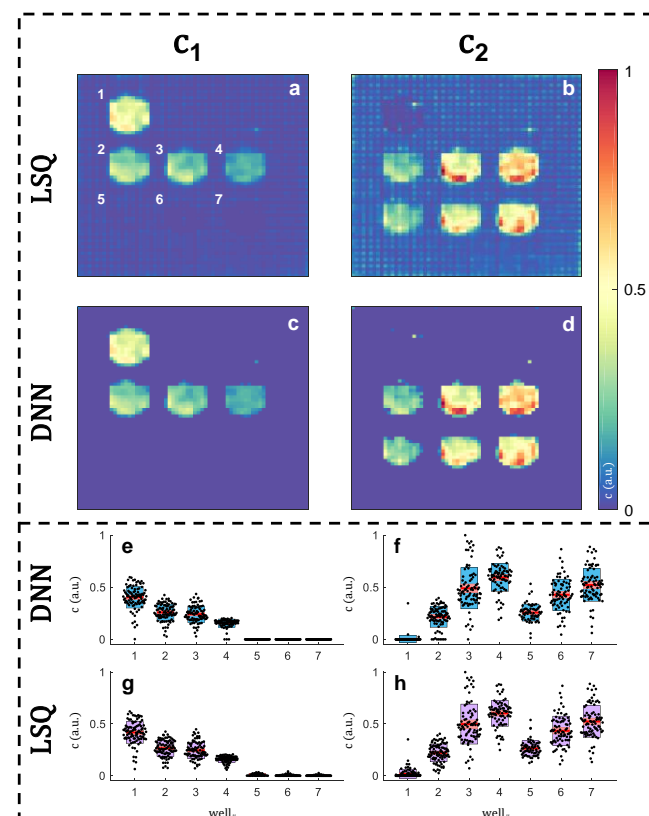

Fig 4. MFLI coefficient retrieval *in vitro*. Results obtained directly through LSQ non-linear iterative fitting (**a**, **b**) and our DNN approach (**c**, **d**) are given. Boxplots of coefficient values retrieved at each ROI (numbered, top-left) are given for both approaches (**e-h**).

Therefore, the DNN correctly assigned all noisy pixels to zero. This performance is in contrast with LSQ, which performs less accurately at noisy pixel locations and resulted

in average SSIM values < 0.9 for all three coefficients. For experimental validation, coefficient values were obtained from the HMFLI time domain reconstruction of a NIR well-plate with varying volumetric fractions of conjugated AF700 and AF750 dye, as illustrated in **Fig 4**. The time domain reconstruction process is further explained in *Supplementary*. The UNMIX-ME framework allows for retrieval of coefficient values identical to those obtained through conventional methodology at the main regions of interest (ROIs). Further, UNMIX-ME provides results essentially devoid of noise (**Fig 4b** & **Fig 4d**) in stark contrast to LSQ (**Fig 4a** & **Fig 4b**). This demonstrates UNMIX-ME's capability to map a noisy TPSF input directly to zero during inference, instead of finding the best possible point of convergence as is the case with LSQ and other conventional methods.
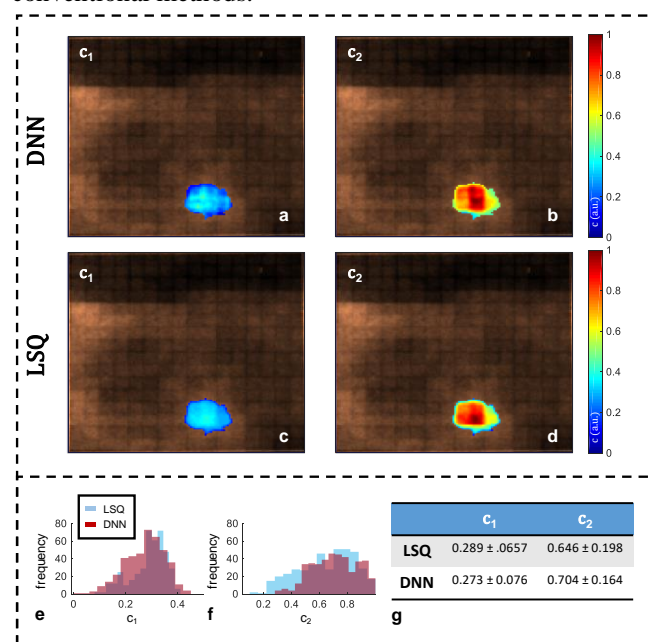

		c$_1$	c$_2$
LSQ		0.289 ± .0657	0.646 ± 0.198
DNN		0.273 ± 0.076	0.704 ± 0.164

Fig 5. MFLI bi-spectral coefficient retrieval *in vivo*. Mouse xenograft imaging took place 76-hours post-injection of NIR fluorescently labeled Trastuzumab (AF700 & AF750). Results obtained directly through LSQ (**a**, **b**, **e**) and UNMIX-ME (**c**, **d**, **f**) are shown. (**g**) Average and standard deviation of coefficient values across the xenograft ROI are given. Pixel resolution is 128×128 with compression ratio was set to 98% during TVAL reconstruction.

Finally, UNMIX-ME was used for complex case of HMFLI: FRET imaging. We have recently reported that MFLI quantitatively reports on target-receptor interaction via *in vivo* lifetime-based Förster Resonance Energy Transfer (FRET) [11], [12]. Herein, a sub-cutaneous tumor xenograft was implanted in a nude athymic mouse. A HER2-Trastuzumab AF700 and AF750 FRET pair was injected and the subject imaged 76-hours after post-injection. Of importance, Trastuzumab is currently used to treat metastatic breast cancer in the clinic. The bi-component spectral unmixing of these data was retrieved using both LSQ and UNMIX-ME (**Fig 5**). Single-pixel imaging at high resolution and reconstruction quality necessitates projection of an increasingly large number of Hadamard patterns and a corresponding increase in acquisition time.

The single-pixel imaging of non-sentient samples (e.g. **Fig 4**) can be accomplished relatively easily. Yet, *in vivo* HMFLI with these acquisition times would be demanding on the subject and thus a lower pattern count (higher compression) is necessary as highlighted in [13]. Therefore, 128×128 pixel reconstruction was performed using Ranked-Hadamard [14] basis projection with 98% compression. In contrast to the rest of this report, the *in vivo* spectral unmixing results illustrated in **Fig 5** were obtained via a global photon-count thresholding in order to focus solely on the xenograft ROI. The results obtained through LSQ and UNMIX-ME are in excellent correspondence both spatially (**Fig 5(a-d)**) and in overall coefficient distributions (**Fig 5(e, f)**). Though there is no "ground-truth" for coefficient retrieval in this case, it is worth noting that this experiment is the first in which completely identical coefficient values were not obtained through both unmixing approaches. Notably, the coefficient values obtained through LSQ exhibit higher homogeneity compared to that of the UNMIX-ME approach.

In summary, we present UNMIX-ME, a DNN-based workflow trained on simulation data for fluorescence lifetime unmixing. Furthermore, unlike intensity-based approaches, UNMIX-ME uses both intensity and lifetime features for spectral classification. The approach provided significant improvement over the conventional nonlinear intensity-based LSQ method by providing a novel means of automated de-noising and demonstrating higher concordance with ground truth during tri-coefficient retrieval. Furthermore, UNMIX-ME provided accurate unmixed profiles both for *in vitro* FRET interactions and for an *in vivo* tumor xenograft region during non-invasive HMFLI.

REFERENCES

[1]   H. Tsurui, H. Nishimura, S. Hattori, S. Hirose, K. Okumura, and T. Shirai, "Seven-color fluorescence imaging of tissue samples based on fourier spectroscopy and singular value decomposition," *J. Histochem. Cytochem.*, 2000.

[2]   T. Zimmermann, J. Marrison, K. Hogg, and P. O'Toole, "Clearing up the signal: Spectral imaging and linear unmixing in fluorescence microscopy," *Methods Mol. Biol.*, 2014.

[3]   G. J. Kremers, E. B. Van Munster, J. Goedhart, and T. W. J. Gadella, "Quantitative lifetime unmixing of multiexponentially decaying fluorophores using single-frequency fluorescence lifetime imaging microscopy," *Biophys. J.*, 2008.

[4]   Q. Pian, R. Yao, N. Sinsuebphon, and X. Intes, "Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging," *Nat. Photonics*, 2017.

[5]   M. Ochoa, R. Yao, P. Yan, and X. Intes, "Facilitating Hyperspectral Single Pixel Lifetime Imaging via deep-learning," in *Optical Molecular Probes, Imaging and Drug Delivery*, 2019, pp. OT3D-2.

[6]   R. Yao, M. Ochoa, P. Yan, and X. Intes, "Net-FLICS: fast quantitative wide-field fluorescence lifetime imaging with compressed sensing – a deep learning approach," *Light: Science and Applications*. 2019.

[7]   F. Chollet, "Deep Learning with Separable Convolutions," *arXiv Prepr. arXiv1610.02357*, 2016.

[8]   F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017.

[9]   K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016.

[10]   J. T. Smith *et al.*, "Ultra-fast fit-free analysis of complex fluorescence lifetime imaging via deep Learning," *bioRxiv*, p. http://dx.doi.org/10.1101/523928, 2019.

[11]   S. Rajoria, L. Zhao, X. Intes, and M. Barroso, "FLIM-FRET for Cancer Applications," *Curr. Mol. Imaging*, 2014.

[12]   A. Rudkouskaya, N. Sinsuebphon, X. Intes, and M. Barroso, "Role of tumor heterogeneity in imaging breast cancer targeted delivery using FLIM FRET in vivo," in *Optics InfoBase Conference Papers*, 2014.

[13]   A. Lorente Mur, M. Ochoa, J. Cohen, X. Intes, and N. Ducros, "Handling negative patterns for fast single-pixel lifetime imaging," 2019.

[14]   M. Ochoa, Q. Pian, R. Yao, N. Ducros, and X. Intes, "Assessing patterns for compressive fluorescence lifetime imaging," *Opt. Lett.*, 2018.