A deep-learning–based method to spectrally separate overlapping fluorophores based on their fluorescence lifetime

L. Cuneo(1)(2)(*), M. Castello(3)(2), S. Piazza(3)(4), I. Nepita(2)(3),
I. Cainero(2), G. Tortarolo(4), L. Lanzanò(5)(2), P. Bianchini(2),
G. Vicidomini(4)(3) and A. Diaspro(1)(2)(3)

(1) *DIFILAB, Department of Physics, University of Genoa - Genoa, Italy*
(2) *Nanoscopy, Istituto Italiano di Tecnologia - Genoa, Italy*
(3) *Genoa Instruments Srl - Genoa, Italy*
(4) *Molecular Microscopy and Spectroscopy, Istituto Italiano di Tecnologia - Genoa, Italy*
(5) *Department of Physics and Astronomy "Ettore Majorana", University of Catania*
    Catania, Italy

Summary. — The simultaneous labelling and imaging of different bio-molecules are required to understand the relationships between the various sub-cellular components and macro-molecular complexes constituting a cell. In fluorescence microscopy, a careful selection of fluorophores is required to prevent spectral overlap, which limits the number and types of fluorophores that can be simultaneously used. This limitation can be overcome with the fluorescence lifetime, able to separate the fluorescence signal. In this study, the authors used deep learning to separate two fluorophores based on their fluorescence lifetime, taking advantage of non-linear spatial-temporal information. The training was carried out on synthetic images, and the results were evaluated on test synthetic images.

1. – Introduction

Modern optical microscopy takes advantage of the developments of experimental approaches based on the phenomena of fluorescence as a mechanism of contrast, reaching excellent performances in spatial and temporal resolution [1]. Due to the relevance of the fluorescence process, here we focus on separating the contribution of two overlapping fluorescence signals collected from biological samples using optical microscopy. The commonly used strategy of using spectrally separable fluorophores can have limitations due to spectral overlap between them. Fluorescence lifetime, as a spectroscopic property, is useful to separate the emission signals. Traditional methods based on frequency domain, such as the phasor approach [2] and SPLIT [3], are limited by linear separation and sensitivity to noise. Therefore, they are limited in terms of the information contained in the non-linear components and can fail in case of large spectral overlap and incorrect or incomplete prior information. The authors propose using a deep neural network (NN) to overcome these limitations and separate the signals based on fluorescence lifetime. The

(*) Corresponding author. E-mail: lisa.cuneo@iit.it

(a) Full model.

(b) Temporal convolution.

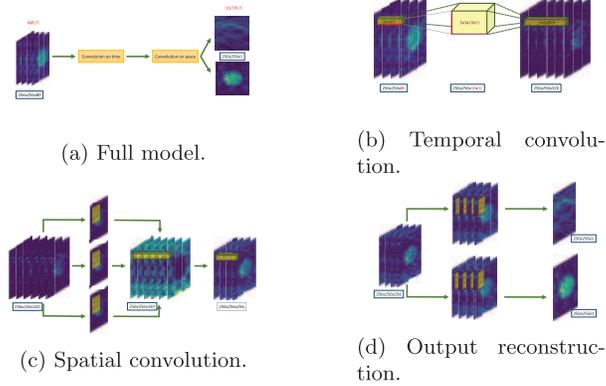

(c) Spatial convolution.

(d) Output reconstruction.

Fig. 1. – Neural Network model architecture. Panel (a) shows the full and schematic representation of the model; the temporal and spatial convolution are explained in panels (b) and (c), respectively, while panel (d) is the last part of the network.

NN takes into account non-linear temporal components and no *a priori* information on the temporal decay model is required. The network was trained with synthetic images to avoid a long acquisition process. In this paper, the design of the neural network architecture, the simulation used to generate the data for training our NN, and the training parameters are described. The results of several tests are shown.

2. – Methods

In this study, a neural network architecture was designed to emphasise the extraction of temporal information. The network was constructed to perform a convolution along the temporal axis first (fig. 1(b)). This was followed by the use of a 2D separable convolution with a kernel size of 3×3 to infer spatially-independent features (fig. 1(c)). Finally, the network was split into two separate paths (fig. 1(d)) in order to reconstruct the images by separating the contributions, in terms of fluorescent photons, associated with the two different fluorophores.

2˙1. *Data simulation*. – The decision to train the neural network with synthetic images was made for two reasons. Firstly, it is easier and faster to generate large quantities of data instead of collecting them experimentally. Secondly, this strategy allows for the reliance on a ground truth rather than fitting procedures. In detail, $15'000$ synthetic images of tubulin and nuclei were created by varying the number and size of filaments, the number and size of molecules within the nucleus, the dimensions of the nucleus, and the intensities. To further enhance the realism of the synthetic images, they were blurred through a convolution with a realistic Gaussian Point Spread Function (PSF) defined as

$$(1) \qquad PSF(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right)},$$

where the standard deviation $\sigma_x = \sigma_y = \frac{\text{FWHM}}{2\sqrt{2\ln 2}}$ assuming the Full Width at Half Maximum $\text{FWHM} = \frac{\lambda_{\text{exc}}}{2NA}$, the wavelength excitation $\lambda_{\text{exc}} = 635\,[\text{nm}]$ and the numerical aperture $NA = 1.49$. Both tubulin and nuclei images, referred to as $Phantom_1$ and $Phantom_2$, respectively, are convoluted with the PSF to produce blurred and realistic
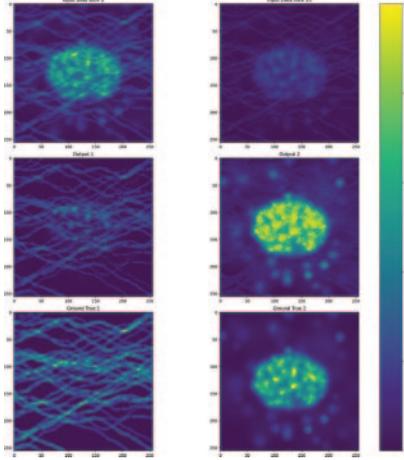
Fig. 2. – The first row includes the input image sampled at two different instant $t_1 = 5$ and $t_2 = 15$; the second row includes the predicted pairs of images; the last row includes the ground true images.

images as the ground truth. The temporal decay of both images is obtained through a convolution between the function $e^{-\frac{t}{\tau}}$ and the Instrument Response Function (IRF).

$$(2) \qquad IRF(t) = \frac{\sigma_t}{2\pi} e^{\left(-\frac{t^2}{2\sigma_t^2}\right)},$$

where the standard deviation $\sigma_t = \frac{\text{FWHM}}{2\sqrt{2\ln 2}}$ with FWHM $= 0.5$ [ns]. We assumed the temporal resolution of Time to Digital Converter (TDC) as 400 [ps] and the laser repetition frequency equal to 20 [MHz]. The two lifetimes were set as $\tau_1 = 2.8$ [ns] and $\tau_2 = 3.4$ [ns]. Lastly, a Poisson noise was added to the temporal decay. Finally, the input data was generated as the sum of the two normalised components.

$$Data(x,y,t) = \frac{Image_1(x,y,t)}{\max\limits_{x,y} Image_1(x,y,t)} + \frac{Image_2(x,y,t)}{\max\limits_{x,y} Image_2(x,y,t)},$$

$$(3)$$
$$Image_i(x,y,t) = [Phantom_i(x,y) * PSF(x,y)] \times \left[\mathcal{P}\left(e^{-\frac{t}{\tau_i}} * IRF(t)\right)\right] \quad i \in \{1,2\}.$$

2˙2. Training. – To further improve the robustness of the neural network, data augmentation was performed on the synthetic images. This was done to ensure that the network was robust with respect to symmetries. The data was augmented by flipping the images vertically and horizontally, as well as rotating them by 90, 180, and 270 degrees. The training was conducted over 200 epochs using the Adam optimiser [4]. A loss function was also chosen to balance the numeric values and the visual perception of the reconstructed image. This function was a weighted sum of the Mean Absolute Error (MAE) and the Structural Similarity Index Measure (SSIM). The goal was to ensure that the network could reconstruct the image in a manner that was both similar to the ground truth in terms of numeric values and visually appealing to the human eye.

(a) Same τ.   (b) Inverted τ.

(c) Only cells.   (d) Only tubulin.
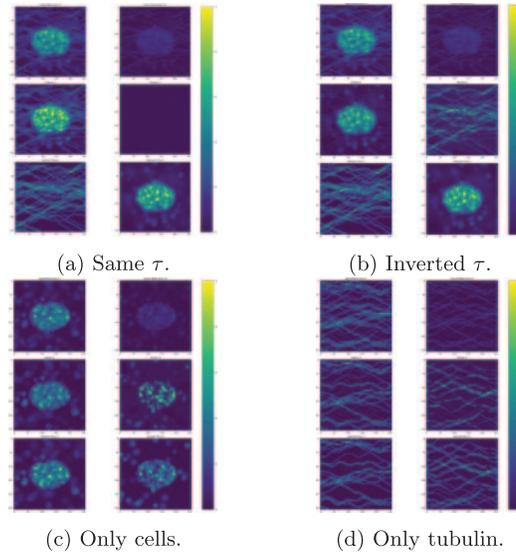
Fig. 3. – Panel (a) shows the test with only one lifetime; in panel (b), the test with inverted lifetime is presented; panels (c) and (d) are, respectively, test on images of only cell or tubulin.

3. – Results and discussion

To validate the effectiveness of our method, we first tested it on data generated as the training set and described in eq. (3). Specifically, we used the same lifetime values $\tau_1 = 2.8$ [ns] and $\tau_2 = 3.4$ [ns] and the same shape. The results are shown in fig. 2. All subsequent results will be presented in a similar manner. To further validate the effectiveness of the model, we generated four different sets of data. The first two sets were related to the lifetime values. In fig. 3(a), we used the same value as $\tau_2 = \tau_1 = 2.8$ [ns], while in fig. 3(b) we inverted the lifetime values with $\tau_1 = 3.4$ [ns] and $\tau_2 = 2.8$ [ns]. The last two sets were designed to ensure that the network was not learning morphological information. In fig. 3 panels (c) and (d), both phantoms were nuclei-like and tubulin-like, respectively. In conclusion, the article presents a deep neural network (NN) method able to precisely separate the contribution of two overlapping fluorophores in fluorescence lifetime images. Moreover, the NN architecture, which includes spatial convolution, is not dependent on the user input and performs well on low signal-to-noise pixels. However, the method only works for data with the same lifetime values used in training, therefore, the lifetime must be known *a priori*. Future works will include testing on real data, extending the method to more than two overlapping species and multiple wavelengths, and accounting for fluorophore crosstalk.

REFERENCES

[1] Diaspro A. and Bianchini P., *Riv. Nuovo Cimento*, **43** (2020) 385.
[2] Lanzanò L. *et al.*, *Nat. Commun.*, **6** (2015) 6701.
[3] Digman M. A. *et al.*, *Biophys. J.*, **94** (2008) L14.
[4] Kingma D. P. and Ba J., *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980v9 (2015).