λSplit: Self-Supervised Content-Aware Spectral Unmixing for Fluorescence Microscopy

Federico Carrara[1,2], Mehdi Seifi[1], and Florian Jug[1]

[1] Fondazione Human Technopole, Milan, Italy
[2] Università Campus Bio-Medico, Rome, Italy

Abstract. In fluorescence microscopy, spectral unmixing aims to recover individual fluorophore concentrations from spectral images that capture mixed fluorophore emissions. Since classical methods operate pixel-wise and rely on least-squares fitting, their performance degrades with increasingly overlapping emission spectra and higher levels of noise, suggesting that a data-driven approach that can learn and utilize a structural prior might lead to improved results. Learning-based approaches for spectral imaging do exist, but they are either not optimized for microscopy data or are developed for very specific cases that are not applicable to fluorescence microscopy settings. To address this, we propose λSplit, a physics-informed deep generative model that learns a conditional distribution over concentration maps using a hierarchical Variational Autoencoder. A fully differentiable Spectral Mixer enforces consistency with the image formation process, while the learned structural priors enable state-of-the-art unmixing and implicit noise removal. We demonstrate λSplit on 3 real-world datasets that we synthetically cast into a total of 58 challenging spectral unmixing benchmarks. We compare our results against a total of 10 baseline methods, including classical methods and a range of learning-based methods. Our results consistently show competitive performance and improved robustness in high noise regimes, when spectra overlap considerably, or when the spectral dimensionality is lowered, making λSplit a new state-of-the-art for spectral unmixing of fluorescence microscopy data. Importantly, λSplit is compatible with spectral data produced by standard confocal microscopes, enabling immediate adoption without specialized hardware modifications. The code and all datasets used in the experiments are available at `https://github.com/juglab/lambdaSplit`.

Keywords: Spectral Unmixing · Hierarchical Variational Inference

1 Introduction

Fluorescence microscopy is a fundamental tool for biomedical discovery in the life sciences, enabling us to visualize selected biological structures labeled with fluorescent markers called fluorescent probes (FPs). The most common way to image biological samples is arguably multiplexed imaging, where multiple FPs labeling different structures in a sample are imaged one at a time using appropriate bandpass filtering (see Fig. 1). Due to the emission properties of FPs, the

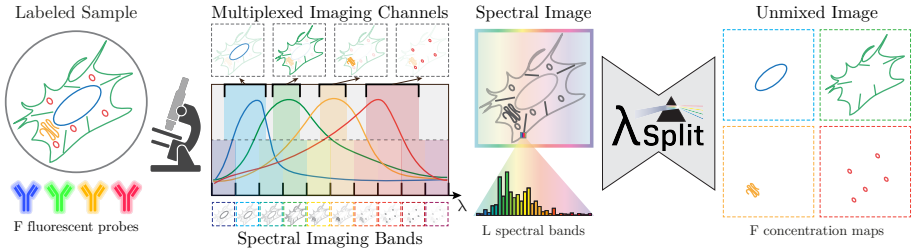

Fig. 1: Spectral Imaging and Unmixing in a nutshell. A biological sample is labeled with F fluorescent probes (FPs) whose emission spectra typically overlap. While for multiplexed imaging each FP is imaged in isolation using F appropriate band-pass filters (see Multiplexed Imaging Channels), spectral imaging, instead, acquires L spectral bands in which photons of all FPs are unfiltered and therefore mixed (see Spectral Imaging Bands). Note that multiplexed imaging can suffer from bleed-through artifacts due to overlapping emission spectra. To retrieve all F unmixed concentration maps (right), we propose λSplit, a self-supervised method that leverages learned structural priors to achieve state-of-the-art results even when spectra overlap considerably, when imaging noise is high, and when the spectral dimensionality L is lowered.

limited spectral range of visible light, and the available microscopy hardware, multiplexed imaging is, in practice, used to image 2 to 4 structures in their respective image channels. In addition, sequential acquisitions increase the required imaging time and, for live samples, result in channels that do not correspond to the same time point. Last but not least, repeated rounds of imaging can lead to increased light exposure and, therefore, a higher degree of photobleaching and phototoxic effects, which are naturally undesired [12,22,34,41]. These limitations motivate other approaches to image multiple fluorescently labeled structures in a biological sample, *i.e.*, spectral imaging approaches.

Spectral fluorescence microscopy does not image FPs one at a time but instead collects all emitted light at once [10,45]. To later recover the individual FP emissions (fluorophore concentrations), the emitted light is collected in a fixed number of spectral bands. Here, the signal at each spatial location is not a single intensity value but a spectral measurement that quantifies how much light was collected over a set of wavelength intervals (bands), as shown in Fig. 1.

Spectral imaging shifts the burden from hardware-based channel separation to computational separation of fluorophore emissions. This separation is commonly addressed through spectral unmixing, where classical approaches estimate fluorophore contributions independently at each pixel using least-squares fitting to reference spectra [44]. While conceptually simple, such pixel-wise formulations degrade when spectra overlap considerably, when noise levels are high, or when the spectral dimensionality is lowered (*i.e.*, when fewer bands are imaged). Learning-based approaches offer an alternative by incorporating spatial context and data-driven structural priors. However, existing methods either do not account for the optics and noise characteristics of fluorescence microscopy or are tailored to narrowly defined acquisition settings.

To address these limitations, we introduce λSplit, a physics-informed deep generative framework for spectral unmixing in fluorescence microscopy that leverages a fully differentiable spectral mixing module (*i.e.*, the Spectral Mixer), which enables self-supervised training. Our approach achieves robust unmixing and implicit denoising across challenging imaging conditions, as we demonstrate experimentally on a total of 58 benchmarks we defined for this work.

Importantly, λSplit operates directly on spectral acquisitions produced by standard point-scanning confocal microscopes, enabling immediate integration into existing experimental workflows without hardware modifications.

2 Problem Setting

A spectral image can be represented as $S \in \mathbb{R}^{L \times Z \times Y \times X}$, where (Z, Y, X) denotes the spatial dimensions and L denotes the spectral dimension, *i.e.*, the number of spectral bands. Depending on the microscope configuration, L can vary substantially: some systems provide relatively dense spectral sampling with a number of contiguous, equispaced, narrow bands of up to 32, while others rely on coarser spectral measurements, leading to only 4 or 5 bands that can even be flexibly tuned to cover arbitrary wavelength intervals.

The measured spectral signal at each spatial location arises from the union of all photons emitted by FPs that are present at that location. Because FPs typically exhibit relatively broad and overlapping emission spectra (see, *e.g.*, Fig. S.4), the signal in each spectral band contains contributions from multiple fluorescent sources.

The goal of spectral unmixing is therefore to recover, from an observed spectral acquisition S, a set of images representing the spatial concentration maps of the individual FPs, also referred to as the unmixed images U [44]. The spectral image formation process can be described by a linear mixing model:

$$S = MU + \varepsilon, \tag{1}$$

where $U \in \mathbb{R}^{F \times Z \times Y \times X}$ denotes the unknown fluorophore concentration maps for F FPs, $M \in \mathbb{R}^{L \times F}$ is the mixing matrix whose columns correspond to the discretized emission spectra of the fluorophores, and ε represents pixel noise. In Eq. (1), tensors are multiplied following the Einstein summation convention.

The unmixing problem at hand is inherently an ill-posed inverse problem. Overlap between emission spectra makes the unmixing harder, and together with additional noise introduces ambiguities for this inverse problem. Additionally, a limited number of spectral bands (lower spectral dimensionality) can render the system underdetermined, particularly in regimes where $L < F$. These challenges are further amplified in realistic biological samples, where fluorophores may be spatially co-localized, and experimental conditions can induce spatial and spectral variability.

3 Related Work

Classical spectral unmixing. A widely used approach is linear unmixing (LU), which estimates fluorophore concentration maps by solving a least-squares problem given reference spectra [44]. Common variants, non-negative LU (NNLU) and fully constrained LU (FCLU), incorporate additional constraints such as non-negativity and sum-to-one priors [4,9,35]. While efficient and widely adopted in fluorescence microscopy, these pixel-wise formulations remain sensitive to noise, spectral overlap, and reduced spectral dimensionality. Alternative approaches, such as Richardson–Lucy unmixing (RLU) [18] and hybrid phasor-based methods (HyU) [6] improve robustness in noisy regimes while still operating pixel by pixel.

Blind spectral unmixing. Blind unmixing methods estimate both emission spectra and fluorophore concentrations directly from spectral data. A common approach is non-negative matrix factorization (NMF) [21], including variants developed for microscopy such as NMF-RI [14,37]. These approaches still rely on pixel-wise LU once spectra are estimated. Other approaches formulate unmixing as a mutual information minimization problem between target structures, such as PICASSO and Mosaic-PICASSO [5, 33]. Still, their practical applicability remains limited to cases where the spectral dimensionality matches the number of target structures.

Learning-based approaches. Learning-based spectral unmixing methods have recently emerged in both microscopy and remote sensing. In fluorescence microscopy, clustering-based approaches such as LUMoS [23] and supervised convolutional models such as AutoUnmix [13] and UNMIX-ME [36] have been proposed. More broadly, a large body of work in remote sensing explores deep architectures for hyperspectral unmixing, including autoencoder-based models, convolutional networks exploiting spatial context, and model-inspired networks that incorporate the mixing process into the decoder or unroll iterative optimization schemes [7, 8, 11, 25, 26, 30, 31, 38]. Despite conceptual similarities, the microscopy setting differs substantially due to its high spatial resolution, limited spectral dimensionality, and distinct noise characteristics.

Content-aware models in microscopy. Recent work has shown that content-aware priors can substantially improve inverse problems in fluorescence microscopy [16, 17, 27, 28, 42]. MicroSplit [1], for example, uses hierarchical variational architectures to separate multiple structures from a single fluorescence image. While this approach does not model spectral mixing, it highlights the potential of learned structural priors for decomposing microscopy images.

A more extensive discussion of related work and methodological differences can be found in the supplementary material (Sec. S.2).

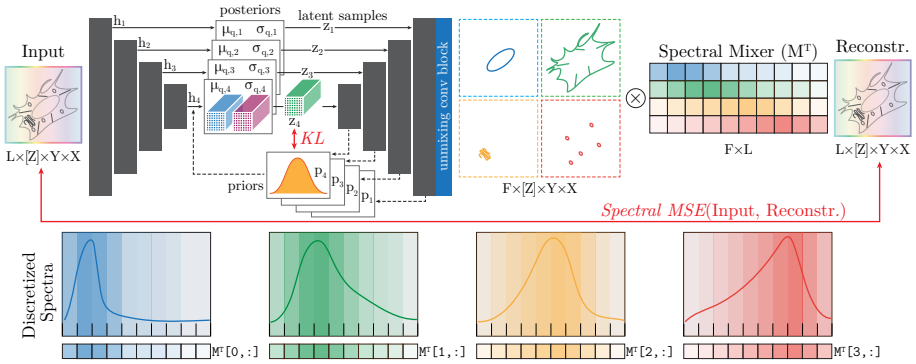

Fig. 2: Proposed architecture of λSplit. The model builds on an LVAE backbone [39], where a bottom-up encoder produces features h_i at multiple hierarchy levels. At the highest hierarchy level we employ a default multivariate Gaussian prior, followed by learnable top-down priors p_i (conditioned from above) that, combined with the respective h_i, form the final posterior distributions $(\mu_{q,i}, \sigma_{q,i})$. Latent samples are generated by drawing from the topmost Gaussian and successively propagating and resampling down the hierarchy through a deterministic decoder to, finally, generate FP concentration maps. The last decoder layer comprises an unmixing convolutional block (blue) that maps the final activation into the desired unmixed concentration map tensor. To render our method fully self-supervised, a differentiable Spectral Mixer multiplies the predicted concentration maps by the transposed mixing matrix $M^T \in \mathbb{R}^{F \times L}$, obtainable by discretizing the emission spectra of the FPs contained in the imaged sample (bottom). This models the physical image formation process and, given that the predicted concentration maps are correct, reconstructs the spectral image we received as input, rendering this setup an end-to-end trainable variational autoencoder. The model is trained by minimizing the sum of the hierarchical KL divergence (computed from latent samples) and a spectral MSE reconstruction loss.

4 Methods

Our approach builds on three key components. First, we introduce a convolutional hierarchical variational framework based on a Ladder Variational Autoencoder (LVAE) backbone [39], replacing pixel-wise estimation with a spatially aware model that learns multi-scale structural priors from data. The variational formulation implicitly regularizes the inverse problem, models a conditional distribution over fluorophore concentration maps, and improves robustness under noise, spectral overlap, and varying acquisition configurations. Second, we incorporate a fixed, fully differentiable spectral mixing module that implements the physics-based linear image formation model by mapping predicted concentration maps into spectral space using known emission profiles (see Fig. 2). This enables self-supervised training directly from mixed spectral observations while maintaining interpretability and general applicability across acquisition setups. Third, we perform extensive validation using a high-fidelity spectral simulation pipeline and three real multiplexed microscopy datasets, constructing 58

controlled unmixing benchmarks that systematically vary photon noise, spectral overlap, and spectral dimensionality, including underdetermined regimes. Across these settings, we demonstrate improved robustness and stability compared to classical and learning-based baselines.

As introduced in Sec. 2, given a spectral image $S \in \mathbb{R}^{L\times Z\times Y\times X}$, our goal is to recover FP concentration maps $U \in \mathbb{R}^{F\times Z\times Y\times X}$. Here, our objective is to learn a conditional generative model $p_\theta(U|S)$ that captures a distribution over plausible unmixed solutions while remaining consistent with the physically grounded linear mixing process introduced in Eq. (1). To enforce this consistency, we introduce a differentiable Spectral Mixer that reconstructs a spectral image $\hat{S}$ directly from the generated unmixed output U via multiplication with the spectral mixing matrix $M \in \mathbb{R}^{L\times F}$. The column $M_{:,j} = [m_{1j}, m_{2j}, \ldots, m_{Lj}]^\top$ of the mixing matrix encodes the discretized spectrum of the fluorophore j, with entries $m_{\ell j}$ obtained by averaging the continuous emission profile over the ℓ-th spectral band. In practice, these spectra can either be obtained from public databases such as FPBase [20] or measured directly from dataset-specific single-fluorophore control samples. We further ℓ_1-normalize each spectrum such that $\sum_{\ell=1}^{L} m_{\ell j} = 1$, ensuring that the Spectral Mixer preserves the total intensity between spectral and unmixed images, *i.e.*,

$$\sum_{\ell=1}^{L} \hat{S}_{\ell,p} = \sum_{\ell=1}^{L}\sum_{j=1}^{F} m_{\ell j} U_{j,p} = \sum_{j=1}^{F} U_{j,p}, \tag{2}$$

for all spatial locations p.

Modified ELBO and LVAE Loss for Spectral Unmixing. To model the generative conditional distribution $p_\theta(U|S)$, we train a hierarchical VAE, specifically an LVAE [39]. Our objective is therefore to maximize the conditional log-likelihood

$$\arg\max_\theta \sum_{i=1}^{N} \log p_\theta(U_i \mid S_i), \tag{3}$$

where θ denotes the decoder parameters defining the generative distribution. By introducing hierarchical latent variables $\mathbf{z} = \{z_1, \ldots, z_K\}$, where K is the number of hierarchy levels, and the variational posterior parameterized by the encoder $q_\phi(\mathbf{z} \mid S)$ with parameters ϕ, we obtain the standard variational evidence lower bound (ELBO), which, according to our notation, reads as

$$\mathbb{E}_{q_\phi(\mathbf{z}|S)}\big[\log p_\theta(U|\mathbf{z})\big] - \mathrm{KL}\big(q_\phi(\mathbf{z}|S) \,\|\, p_\theta(\mathbf{z})\big). \tag{4}$$

A detailed derivation of the ELBO is presented in Sec. S.3.1. As the unmixed FP concentration channels U are not directly observable, and to recover a self-supervised autoencoder setup in which the reconstructed output can be compared to the spectral measurement given as the original input, we map U back to the spectral domain as a reconstructed image $\hat{S}$, using the previously introduced Spectral Mixer. This fully differentiable spectral mixing layer also ensures consistency with the physical image formation process.

Hence, the spectral observation is now obtained through a fixed, deterministic mixing process, $S = g(U) = MU$. Given latent variables $\mathbf{z}$ sampled from the variational posterior $q_\phi(\mathbf{z} \mid S)$, the decoder predicts fluorophore concentration maps $U(\mathbf{z}) \sim p_\theta(U \mid \mathbf{z})$, which are subsequently mapped to the spectral domain to generate a reconstructed spectral image $\hat{S}(\mathbf{z}) = MU(\mathbf{z})$.

Since the spectral formation process is deterministic and we do not assume a probabilistic noise model, the conditional likelihood $p_\theta(S \mid \mathbf{z})$ cannot be directly specified in closed form. In practice, we approximate the corresponding reconstruction term using the spectral mean squared error (MSE) between the observed spectral image and its reconstruction,

$$\mathcal{L}_{sp\mathrm{MSE}}\big(S, \hat{S}\big) = \frac{1}{LP} \sum_{\ell=1}^{L} \sum_{p=1}^{P} \big(S_{\ell,p} - \hat{S}_{\ell,p}\big)^2, \tag{5}$$

where $P = Z \cdot Y \cdot X$ is the total number of voxels in the image. The resulting training objective is therefore defined as

$$\mathcal{L} = \mathbb{E}_{q_\phi(\mathbf{z}|S)} \Big[\mathcal{L}_{sp\mathrm{MSE}}\big(S, \hat{S}(\mathbf{z})\big) \Big] + \mathrm{KL}\big(q_\phi(\mathbf{z} \mid S) \,\|\, p_\theta(\mathbf{z})\big). \tag{6}$$

Implementation Details. As mentioned before, we employ an LVAE architecture [39]. Since it allows sampling from the posterior distribution over unmixed images, we infer predictions by averaging 50 posterior samples, giving us an approximate MMSE estimate [27]. We employ a latent hierarchy with 4 stochastic levels, for which the spatial resolution is progressively reduced by factors of two, while the feature channel dimensionality is kept constant at 128. The KL loss is computed per latent level as a Monte Carlo estimate over the batch by averaging the KL across all entries of the latent tensors, and is then averaged across levels. The final KL term is weighted by $\beta = 1$. The choice of the number of stochastic levels and the KL weight β is supported by ablation studies reported in the supplementary material Sec. S.5.

Our implementation can operate on 2D and 3D spectral data by employing 2D and 3D convolutional layers, respectively. While training is performed on randomly extracted patches of size 64×64 for 2D inputs and $8 \times 64 \times 64$ for 3D data, at inference time we employ inner tiling [3], with one-quarter overlap between neighboring tiles.

Models are trained using the Adamax [15] optimizer with default hyperparameters. We use a batch size of 32 and an initial learning rate of 10^{-3}. Training is performed for 150 epochs using mixed precision and early stopping with patience set to 30 epochs. The learning rate is reduced on plateau with a patience of 15 epochs. To preserve the relative intensity between spectral bands, input spectral images are normalized using global mean–standard deviation normalization. All models were trained on an NVIDIA DGX H200 system using either an 18GB or 35GB MIG partition for 2D and 3D models, respectively. Training time depends on the dataset size and dimensionality, typically ranging from 2 hours to approximately one day.

Baselines. We compare λSplit against a set of classical and learning-based spectral unmixing baselines. As direct linear methods based on least-squares, we include linear unmixing (LU), non-negative LU (NNLU) [35], and fully-constrained LU (FCLU) [9]. We further evaluate HyU [6], Richardson–Lucy Unmixing (RLU) [18], and NMF-RI [14], which are specifically developed for fluorescence microscopy and low-photon settings.

We additionally consider learning-based methods. For low-band settings (≤ 5 bands), we compare to LUMoS clustering-based unmixing [23]. Additionally, we include the Transformer Unmixing Autoencoder (TAEU) [8], developed for spectral unmixing in remote sensing. The motivation for choosing this algorithm is its popularity and success in the field, as well as the availability of an open-source implementation.

We also compare λSplit against two supervised DL-based methods: a standard supervised UNet [32] and AutoUnmix [13].

Please find additional information about baselines and their implementation in Sec. S.3.2.

Evaluation Metrics. We evaluate unmixing performance using a combination of intensity-based, structural, perceptual, and signal-quality metrics. We report range-invariant ***PSNR*** and ***MS-SSIM*** (abbreviated as MS3IM), as introduced in [42], which compensate for global intensity scaling differences between prediction and ground truth (*e.g.*, due to different output normalization). Moreover, we include the ***Pearson*** correlation coefficient, ***LPIPS*** [43], and ***SNR*** of the predicted unmixed images. Finally, we also report ***MicroMS-SSIM*** [2] (abbreviated as μMS3IM), which is a structural similarity metric tailored to microscopy data. More details about the implementation of the employed metrics are reported in Sec. S.3.3.

5 Data and Data Handling

Data Simulation Pipeline. Since ground-truth (GT) fluorophore concentration maps are typically unavailable (and even unobtainable) in microscopy settings, we rely on synthetic data to enable controlled benchmarking across acquisition conditions with variable noise, spectral overlap, and spectral dimensionality. To generate realistic spectral mixtures, we use a powerful microscopy simulation pipeline based on the open-source `microsim` Python package [19]. Multi-channel inputs, either real multiplexed fluorescence images or multiclass segmentation maps, are interpreted as spatial fluorophore concentration fields and serve as GT for evaluation. Each channel is associated with a fluorophore emission spectrum, from which a noise-free spectral volume is generated. The simulator then models optical effects (*e.g.*, point spread functions) and the image acquisition process, including photon statistics, readout noise, and configurable spectral band placements that emulate realistic spectral microscopy setups. This pipeline enables the controlled generation of physically consistent spectral mixtures across diverse experimental settings. Additional details and a schematic overview are provided in the supplementary material (Fig. S.1).

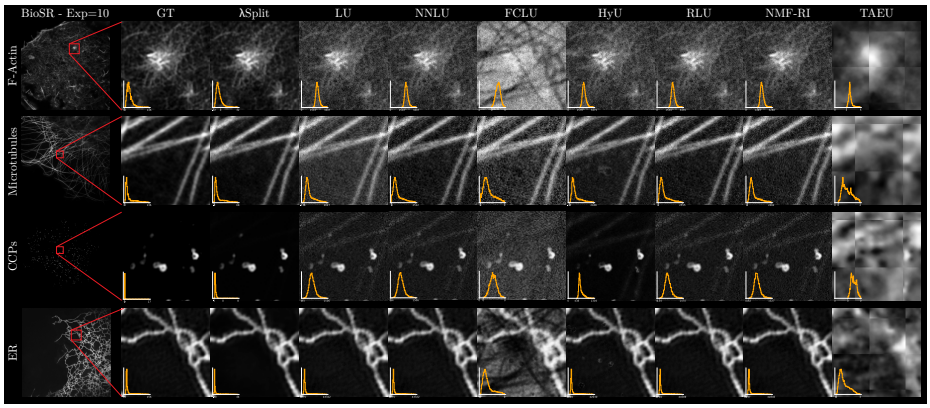

Fig. 3: Qualitative results on high-noise BioSR data (exposure 10 ms). Compared to 7 baselines, λSplit produces the cleanest results, also benefiting from its implicit noise removal capabilities. Objects and details (*e.g.*, puncta in the CCPs channel or F-Actin structures) are well recovered, enabling best possible downstream analyses.

Real-world Datasets. We use three publicly available fluorescence microscopy datasets as inputs for our simulation pipeline. The ***BioSR*** dataset [29] consists of 2D high-SNR fluorescence images (1004×1004 pixels) of four organelles (microtubules, F-actin, endoplasmic reticulum (ER), and clathrin-coated pits (CCPs)). These images provide structurally well-resolved organelle distributions suitable for controlled spectral simulations. The ***CellAtlas*** dataset contains four-channel multiplexed images from the Human Protein Atlas project [24, 40], comprising tens of thousands of 2D images across diverse cell types. In this work, we select a subset of 300, 2048×2048 images where the four channels correspond to nuclei, microtubules, endoplasmic reticulum, and mitochondria. Finally, the HHMI-D25-8bit dataset introduced in [1], here referred to as ***HHMI25***, includes 3D multiplexed fluorescence volumes (17×2048×2048) acquired from mouse liver tissue samples. Each volume contains five labeled structures: lipid droplets, nuclei, peroxisomes, cortical actin, and lysosomes, enabling us to conduct volumetric spectral simulations.

6 Experiments and Results

We evaluate λSplit across a series of controlled spectral unmixing benchmarks designed to isolate the main sources of difficulty in microscopy data. Specifically, we analyze robustness to noise, increasing spectral overlap between FP emissions, and reduced spectral dimensionality (fewer spectral bands). In all cases, we report quantitative and qualitative results against the selected baselines.

Table 1: Quantitative results on the *CellAtlas* data for various noise levels. Columns, from left to right, report: (*i*) exposure times (less causing more noise); (*ii*) the resulting SNR of the spectral input images (aggregated over all spectral bands); (*iii*) baselines considered for the set of experiments; (*iv*) quality metrics, *i.e.*, PSNR, MS-SSIM (MS3IM), MicroMS-SSIM (μMS3IM), LPIPS, Pearson correlation coefficient, and average SNR of unmixed images (see main text for details). Higher values are better, except for LPIPS. In each metric column, the best result is shown in **bold**, while the second best is underlined.

Exp	SNR_{sp}	Method	PSNR ↑	MS3IM ↑	μMS3IM ↑	LPIPS ↓	Pearson ↑	SNR_u ↑
$2ms$	5.33	LU [44]	23.96	0.799	0.647	0.485	0.779	8.44
		NNLU [35]	24.06	0.803	0.647	0.439	0.785	8.59
		FCLU [9]	20.70	0.520	0.619	0.511	0.540	5.04
		HyU [6]	24.28	0.803	0.500	0.442	0.795	11.35
		RLU [18]	23.98	0.799	0.647	0.442	0.781	8.47
		NMF-RI [14]	24.09	0.803	0.474	0.440	0.790	8.69
		TAEU [8]	24.52	0.761	**0.652**	0.520	0.800	**60.79**
		λSplit (ours)	**27.14**	**0.904**	0.625	**0.373**	**0.885**	21.35
$5ms$	11.04	LU	28.52	0.921	0.792	0.378	0.920	20.33
		NNLU	28.55	0.921	0.721	0.319	0.921	20.34
		FCLU	22.03	0.595	0.607	0.435	0.685	9.57
		HyU	27.52	0.892	0.790	0.369	0.897	29.28
		RLU	28.91	0.927	0.659	0.316	0.929	20.50
		NMF-RI	28.86	0.927	0.795	0.315	0.926	20.91
		TAEU	23.97	0.735	0.655	0.520	0.786	66.67
		λSplit (ours)	**34.03**	**0.980**	**0.979**	**0.155**	**0.975**	**2047.10**
$10ms$	22.57	LU	32.56	0.966	0.834	0.307	0.961	41.38
		NNLU	32.64	0.967	0.834	0.228	0.962	41.35
		FCLU	22.52	0.621	0.638	0.404	0.716	14.76
		HyU	29.84	0.929	0.650	0.274	0.926	62.78
		RLU	33.14	0.971	0.850	0.221	0.968	41.48
		NMF-RI	32.99	0.971	0.839	0.224	0.967	42.93
		TAEU	24.26	0.767	0.694	0.452	0.801	101.82
		λSplit (ours)	**36.66**	**0.989**	**0.988**	**0.137**	**0.986**	**2041.18**
$20ms$	43.05	LU	35.91	0.985	0.696	0.194	0.987	75.87
		NNLU	36.02	0.986	0.696	0.128	0.988	75.78
		FCLU	22.38	0.614	0.626	0.361	0.720	21.75
		HyU	30.37	0.936	0.873	0.184	0.953	109.65
		RLU	**36.68**	0.988	0.907	0.125	0.990	75.83
		NMF-RI	36.49	**0.989**	0.703	**0.124**	0.989	77.96
		TAEU	23.66	0.739	0.675	0.479	0.797	781.03
		λSplit (ours)	36.46	0.988	**0.988**	0.155	**0.988**	**2697.59**

6.1 Spectral Unmixing and Noisy Data

To assess robustness to pixel noise, we evaluate all applicable baselines on simulated 32-band spectral data derived from the BioSR, CellAtlas, and HHMI25 datasets. To simulate a realistic setting, we assigned four partially overlapping fluorophores (mTurquoise, EGFP, EYFP, and mOrange) to the four structures (fluorescent channels) in the BioSR and CellAtlas data. For the HHMI25 data, which has 5 channels, we added mScarlet. The emission spectra of all used FPs are shown in Figs. S.4 and S.5.

For each dataset, we simulate four exposure levels with increasing Poisson noise while keeping Gaussian readout noise fixed at realistic magnitudes. The highest signal-to-noise ratio (SNR) configuration is selected such that baseline

methods perform well, representing a well-conditioned regime in which classical methods perform comparably to λSplit. As shown in Tab. 1 and Tabs. S.1 and S.2, λSplit continues to perform well as SNR decreases, consistently outperforming all baselines. Note that λSplit's reported SNR_u values are much higher than those of other baseline methods. This is a result of the implicit denoising capabilities of our proposed method. An analysis supporting this observation is reported in the supplement (see Sec. S.6.2).

Qualitative comparisons also reveal that pixel-wise methods naturally propagate the noise in the input into the unmixed channels. Results obtained with λSplit, in comparison, can discriminate the imaged structures from purely stochastic pixel fluctuations, preserving boundary sharpness and leading to denoised predictions (see Fig. 3).

Finally, we observe that the data-driven TAEU [8] and classical FCLU [9] baselines, both created for spectral unmixing in remote sensing applications, seem unsuitable for application to fluorescence microscopy data, leading to subpar results in all our experiments. For this reason, and to save space, we do not report results for these two baselines in subsequent sections.

6.2 Spectral Unmixing and Overlapping Spectra

Next, we ask how unmixing performance changes as the overlap between the used FPs increases. Everything else remains unchanged, allowing us to isolate the effect of spectral overlap. To this end, we conducted a total of 40 experiments where we simulated spectral data, just as before, but moved the emission spectra closer together without changing their shape.

More concretely, we have used two pairs of the BioSR structures (Microtubules + CCPs and ER + CCPs), and assigned to both the same emission spectrum (either EGFP or mTurquoise). For the second structure (CCPs), we have then rigidly shifted its spectrum by $\Delta\lambda \in \{2\text{nm}, 5\text{nm}, 10\text{nm}, 20\text{nm}, 50\text{nm}\}$ (see Fig. 4**a,c** for an illustration).

Further, we conducted these experiments using 32- and 5-band acquisition regimes, in line with existing, commonly used commercial microscopy setups. To keep the total photon budget fixed, we left the laser power and exposure times ($20ms$) unchanged in both cases, which leads to higher per-band SNR when fewer are being imaged. This, again, leads to an initial regime (for $\Delta\lambda = 50\text{nm}$) in which classical baselines lead to results comparable to those of λSplit.

For smaller values of $\Delta\lambda$, the quality gap of predictions consistently widens, showing that λSplit can effectively utilize its learned structural prior in these more difficult settings (see Fig. 4 and Figs. S.9 to S.11 and Tabs. S.3 to S.6), where it consistently outperforms all other baselines.

6.3 Spectral Unmixing and Spectral Dimensionality

Next, we studied the effect of reducing the number of spectral bands. While classical methods desire the number of bands to be at least equal to the number

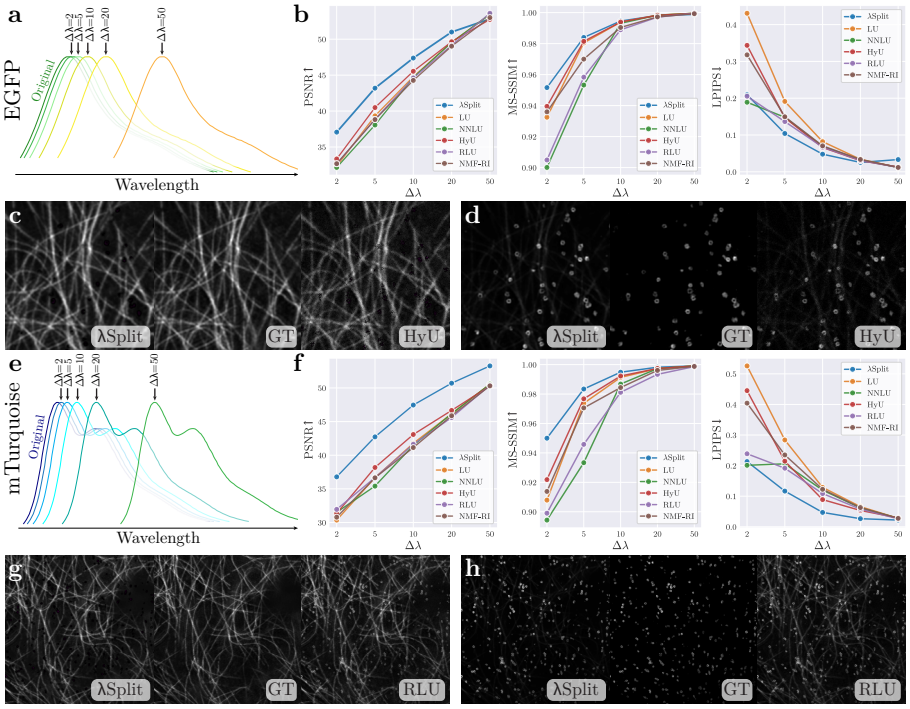

Fig. 4: Spectral unmixing with controlled spectral overlap (32 bands). For this experiment, we use the Microtubules and CCPs structures from the BioSR dataset, with 32 spectral bands imaged. **a,e** Synthetically shifted EGFP and mTurquoise emission spectra, respectively. While we fix the spectrum associated with Microtubules, the emission spectrum of CCPs is rigidly shifted with increasing $\Delta\lambda$. **b,f** Quantitative metrics over $\Delta\lambda$, showing that λSplit (blue) consistently outperforms all other baselines and that the performance gap widens when spectral overlap is larger (hence, when $\Delta\lambda$ is smaller). **c,d,g,h** Representative crops from unmixing results for $\Delta\lambda = 2\,\text{nm}$ by λSplit and the best-performing baseline (w.r.t. PSNR), plus the corresponding ground truth (GT) in the center. In c,d, we show Microtubules, while in g,h, we show CCPs. Notice that, due to the high degree of overlap, each unmixed channel could exhibit negative structures from the other channel, a common artifact in spectral unmixing. For a cleaner visual comparison, we center the colormap on the background value and we show negative pixels in magenta tones with the same scaling.

of unknowns (*i.e.*, the number of used FPs), having a higher number of bands offers numerical advantages.

We conducted two sets of experiments on the CellAtlas data: (*i*) we take the highest SNR setup from Sec. 6.1 for the 32 band setup (note that the respective column in Tab. 2 reports the same values as Tab. 1). We then reduced the number of bands to 5, 4, and 3 while keeping the per-band SNR constant. This removes SNR as a confounder, isolating the effect of spectral dimensionality, and (*ii*) we

Table 2: Quantitative results on *CellAtlas* for varying spectral dimensionality. We evaluate spectral unmixing performance when reducing the number of acquisition bands. Two experimental settings are considered: (*left*) the number of bands is reduced while keeping the per-band SNR constant, isolating the effect of spectral dimensionality; (*right*) the total photon budget is fixed, such that fewer bands lead to higher per-band SNR. Find more details in Sec. 6.3. Best results are shown in **bold**, second best underlined.

		# bands Same per-band SNR				**# bands** Same total photon budget			
Metric	**Method**	3	4	5	32	3	4	5	32
PSNR ↑	LU [44]	23.69	26.76	27.77	35.91	23.72	27.77	27.77	25.81
	NNLU [35]	26.21	27.30	28.56	36.02	26.79	28.36	28.56	25.83
	HyU [6]	23.69	26.87	27.19	30.37	23.72	27.61	27.19	25.66
	RLU [18]	26.39	27.56	28.91	**36.68**	27.13	28.63	28.91	26.15
	NMF-RI [14]	26.31	27.76	28.38	36.49	26.77	28.70	28.38	26.13
	LUMoS [23]	21.97	21.98	21.37	23.21	21.40	22.14	21.37	21.47
	λSplit (ours)	**27.25**	**29.95**	**31.12**	36.46	**27.45**	**31.54**	**31.12**	**31.32**
MS3IM ↑	LU	0.760	0.904	0.925	0.985	0.773	0.924	0.925	0.864
	NNLU	0.863	0.916	0.938	0.986	0.874	0.934	0.938	0.865
	HyU	0.759	0.895	0.900	0.936	0.773	0.906	0.900	0.855
	RLU	0.866	0.921	0.943	0.988	**0.882**	0.938	0.943	0.872
	NMF-RI	0.857	0.918	0.934	**0.989**	0.866	0.935	0.934	0.872
	LUMoS	0.723	0.764	0.743	0.796	0.704	0.766	0.743	0.748
	λSplit (ours)	**0.878**	**0.960**	**0.971**	0.988	0.868	**0.967**	**0.971**	**0.971**
LPIPS ↓	LU	0.431	0.436	0.427	0.194	0.397	0.413	0.427	0.441
	NNLU	0.239	0.267	0.232	0.128	0.192	0.238	0.232	0.392
	HyU	0.434	0.364	0.398	0.184	0.398	0.341	0.398	0.412
	RLU	0.238	0.258	**0.220**	0.125	0.187	0.229	**0.220**	**0.387**
	NMF-RI	**0.237**	**0.247**	0.227	**0.124**	**0.185**	**0.217**	0.227	0.389
	LUMoS	0.324	0.327	0.310	0.287	0.293	0.314	0.310	0.381
	λSplit (ours)	0.283	0.267	0.313	0.155	0.273	0.238	0.313	0.218

take the total photon budget of the 5-band setup from above and simulate a 3-, 4-, and 32-band setup. Note that the 5-band results in Tab. 2 are, for this reason, also identical. The question we are asking here is, therefore, how the potential downside of having fewer bands balances with the advantage of having them at a higher SNR.

Overall, we observe that, as spectral dimensionality is reduced, λSplit consistently outperforms the baselines in PSNR and MS-SSIM while remaining competitive in perceptual quality (evaluated via LPIPS). It is important to note that when the number of bands drops below the number of unknowns (*i.e.*, where the number of structures to be unmixed is larger than the spectral dimension), the inverse problem becomes underdetermined, and the performance of classical least-squares methods (*i.e.*, LU or HyU) drops significantly (see the columns showing results for 3 bands in Tab. 2). In this regime, λSplit leverages its learned structural prior and consistently leads to state-of-the-art results.

6.4 λSplit *vs.* Supervised Spectral Unmixing

Finally, we compare λSplit with a supervised UNet and AutoUnmix, a supervised convolutional neural network for spectral unmixing in fluorescence microscopy. For both baselines, we use noise-free ground-truth (GT) concentration maps as targets, which is possible because of our data simulations. Note that this enables supervised baselines to also learn to perform supervised denoising, giving them an additional advantage.

In the low spectral dimensionality experiments (Tab. S.8 and Fig. S.13), we immediately observe that the performance discrepancy between λSplit and AutoUnmix progressively decreases as the number of bands increases: richer spectral sampling provides λSplit with more information about the emission profiles, while AutoUnmix benefits less from additional bands since the supervision signal and overall signal content remain essentially unchanged. Conversely, in the increasing spectral overlap experiments (Tabs. S.9 to S.12 and Fig. S.14), we observe that, as the overlap reduces, λSplit matches or surpasses AutoUnmix.

Still, AutoUnmix is not a practical alternative for λSplit. On one hand, supervised training data is not available in practice; we could only train it due to our synthetic data generation setup. On the other hand, the model size of AutoUnmix scales combinatorially with the number of spectral bands, leading to a 450M-parameter model for 5 bands and a 10B-parameter model for 32 bands (compared to λSplit's 3M parameters for both cases).

Finally, we observe that the UNet consistently performs substantially worse than λSplit and AutoUnmix in both experiments (see Fig. S.13).

In summary, high-fidelity predictions can be obtained with a significantly more compact, physics-consistent, and self-supervised framework that can be instantly applied to real-world imaging setups.

7 Discussion

The results presented in this work indicate that spectral unmixing in fluorescence microscopy benefits substantially from incorporating learned structural priors in addition to explicit spectral modeling. Across 58 controlled experiments varying noise, spectral overlap, and spectral dimensionality, λSplit consistently remains stable in regimes where purely pixel-wise, least-squares-based baselines, as well as existing data-driven methods, lead to sub-par results. This suggests that resolving ambiguities in spectral data is not solely determined by spectral separability, but also by constraining solutions to structurally plausible patterns (by enforcing consistency with a learned structural prior).

A key consequence of our formulation is that such structural priors can be learned in a fully self-supervised manner by embedding the spectral mixing process into the model. This removes the dependence on paired unmixed ground truth while maintaining physical consistency and interpretability. In practice, this makes data-driven spectral unmixing accessible in realistic microscopy settings where labeled training data are unavailable.

More broadly, our findings support a shift away from purely analytical, pixel-wise inversion schemes toward hybrid approaches that combine explicit forward models with learned, content-aware priors. While our current formulation assumes linear mixing with known spectra, future extensions may relax these assumptions and incorporate spectrum estimation or uncertainty-aware downstream analysis.

Disclosure. A patent application related to the methods described in this work has been filed.

Acknowledgements

We sincerely thank Talley Lambert (Harvard Medical School) for developing the `microsim` software and for helping design and deploy data simulation pipelines. His contribution was key to the project's success. We also thank the entire Jug Group and the National Facility for Bioimage Analysis at Fondazione Human Technopole for their excellent feedback and insightful discussions. This work was supported by generous core funding from Fondazione Human Technopole.

References

1. Ashesh, Carrara, F., Zubarev, I., Galinova, V., Croft, M., Pezzotti, M., Gong, D., Casagrande, F., Colombo, E., Giussani, S., Restelli, E., Cammarota, E., Battagliotti, J.M., Klena, N., Di Sante, M., Adhikari, R., Feliciano, D., Pigino, G., Taverna, E., Harschnitz, O., Maghelli, N., Scherer, N., Dalle Nogare, D.E., Deschamps, J., Pasqualini, F., Jug, F.: MicroSplit: semantic unmixing of fluorescent microscopy data. Nat. Methods **23**(5), 1047–1057 (May 2026)
2. Ashesh, Deschamps, J., Jug, F.: MicroSSIM: Improved structural similarity for comparing microscopy data. arXiv [eess.IV] (Aug 2024)
3. Ashesh, Krull, A., Di Sante, M., Pasqualini, F.S., Jug, F.: μsplit: efficient image decomposition for microscopy data. arXiv [cs.CV] (Nov 2022)
4. Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., Chanussot, J.: Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **5**(2), 354–379 (Apr 2012)
5. Cang, H., Liu, Y., Xing, J.: Mosaic-PICASSO: accurate crosstalk removal for multiplex fluorescence imaging. Bioinformatics **40**(1) (Jan 2024)
6. Chiang, H.J., Koo, D.E.S., Kitano, M., Burkitt, S., Unruh, J.R., Zavaleta, C., Trinh, L.A., Fraser, S.E., Cutrale, F.: HyU: Hybrid unmixing for longitudinal in vivo imaging of low signal-to-noise fluorescence. Nat. Methods **20**(2), 248–258 (Feb 2023)
7. Gao, L., Han, Z., Hong, D., Zhang, B., Chanussot, J.: CyCU-net: Cycle-consistency unmixing network by learning cascaded autoencoders. IEEE Trans. Geosci. Remote Sens. **60**, 1–14 (2022)
8. Ghosh, P., Roy, S.K., Koirala, B., Rasti, B., Scheunders, P.: Hyperspectral unmixing using transformer network. IEEE Trans. Geosci. Remote Sens. **60**, 1–16 (2022)

9. Heinz, D.C., Chein-I-Chang: Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **29**(3), 529–545 (Mar 2001)
10. Hiraoka, Y., Shimi, T., Haraguchi, T.: Multispectral imaging fluorescence microscopy for living cells. Cell Struct. Funct. **27**(5), 367–374 (Oct 2002)
11. Hong, D., Gao, L., Yao, J., Yokoya, N., Chanussot, J., Heiden, U., Zhang, B.: Endmember-guided unmixing network (EGU-net): A general deep learning framework for self-supervised hyperspectral unmixing. IEEE Trans Neural Netw Learn Syst **33**(11), 6518–6531 (Nov 2022)
12. Icha, J., Weber, M., Waters, J.C., Norden, C.: Phototoxicity in live fluorescence microscopy, and how to avoid it. Bioessays **39**(8), 1700003 (Aug 2017)
13. Jiang, Y., Sha, H., Liu, S., Qin, P., Zhang, Y.: AutoUnmix: an autoencoder-based spectral unmixing method for multi-color fluorescence microscopy imaging. Biomed. Opt. Express **14**(9), 4814–4827 (Sep 2023)
14. Jiménez-Sánchez, D., Ariz, M., Morgado, J.M., Cortés-Domínguez, I., Ortiz-de Solórzano, C.: NMF-RI: blind spectral unmixing of highly mixed multispectral flow and image cytometry data. Bioinformatics **36**(5), 1590–1598 (Mar 2020)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv [cs.LG] (Dec 2014)
16. Krull, A., Vičar, T., Prakash, M., Lalit, M., Jug, F.: Probabilistic noise2void: Unsupervised content-aware denoising. Frontiers in Computer Science (2020)
17. Krull, A., Buchholz, T.O., Jug, F.: Noise2Void - learning denoising from single noisy images. arXiv [cs.CV] (Nov 2018)
18. Kumar, A., McNally, K.E., Zhang, Y., Haslett-Saunders, A., Wang, X., Guillem-Marti, J., Lee, D., Huang, B., Stallinga, S., Kay, R.R., Baker, D., Derivery, E., Manton, J.D.: Multispectral live-cell imaging with uncompromised spatiotemporal resolution. Nat. Photonics **19**(10), 1146–1156 (Sep 2025)
19. Lambert, T., Carrara, F., Croft, M., Ashesh, Ruzette, A.A., Galinova, V.: tlambert03/microsim: v0.0.12.post0 (2026)
20. Lambert, T.J.: FPbase: a community-editable fluorescent protein database. Nat. Methods **16**(4), 277–278 (Apr 2019)
21. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788–791 (Oct 1999)
22. Lichtman, J.W., Conchello, J.A.: Fluorescence microscopy. Nat. Methods **2**(12), 910–919 (Dec 2005)
23. McRae, T.D., Oleksyn, D., Miller, J., Gao, Y.R.: Robust blind spectral unmixing for fluorescence microscopy using unsupervised learning. PLoS One **14**(12), e0225410 (Dec 2019)
24. Ouyang, W., Winsnes, C.F., Hjelmare, M., Cesnik, A.J., Åkesson, L., Xu, H., Sullivan, D.P., Dai, S., Lan, J., Jinmo, P., Galib, S.M., Henkel, C., Hwang, K., Poplavskiy, D., Tunguz, B., Wolfinger, R.D., Gu, Y., Li, C., Xie, J., Buslov, D., Fironov, S., Kiselev, A., Panchenko, D., Cao, X., Wei, R., Wu, Y., Zhu, X., Tseng, K.L., Gao, Z., Ju, C., Yi, X., Zheng, H., Kappel, C., Lundberg, E.: Analysis of the human protein atlas image classification competition. Nat. Methods **16**(12), 1254–1261 (Dec 2019)
25. Palsson, B., Sigurdsson, J., Sveinsson, J.R., Ulfarsson, M.O.: Hyperspectral unmixing using a neural network autoencoder. IEEE Access **6**, 25646–25656 (2018)
26. Palsson, B., Ulfarsson, M.O., Sveinsson, J.R.: Convolutional autoencoder for Spectral–Spatial hyperspectral unmixing. IEEE Trans. Geosci. Remote Sens. **59**(1), 535–549 (Jan 2021)

27. Prakash, M., Krull, A., Jug, F.: Fully unsupervised diversity denoising with convolutional variational autoencoders. Int Conf Learn Represent (2021)
28. Prakash, M., Delbracio, M., Milanfar, P., Jug, F.: Interpretable unsupervised diversity denoising and artefact removal. arXiv [eess.IV] (Apr 2021)
29. Qiao, C., Li, D.: BioSR: a biological image dataset for super-resolution microscopy (Jan 2024)
30. Rasti, B., Koirala, B., Scheunders, P., Chanussot, J.: MiSiCNet: Minimum simplex convolutional network for deep hyperspectral unmixing. IEEE Trans. Geosci. Remote Sens. **60**, 1–15 (2022)
31. Rasti, B., Koirala, B., Scheunders, P., Ghamisi, P.: UnDIP: Hyperspectral unmixing using deep image prior. IEEE Trans. Geosci. Remote Sens. **60**, 1–15 (2022)
32. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. arXiv [cs.CV] (May 2015)
33. Seo, J., Sim, Y., Kim, J., Kim, H., Cho, I., Nam, H., Yoon, Y.G., Chang, J.B.: PICASSO allows ultra-multiplexed fluorescence imaging of spatially overlapping proteins without reference spectra measurements. Nat. Commun. **13**(1), 2475 (May 2022)
34. Shaner, N.C., Steinbach, P.A., Tsien, R.Y.: A guide to choosing fluorescent proteins. Nat. Methods **2**(12), 905–909 (Dec 2005)
35. Slawski, M., Hein, M.: Sparse recovery by thresholded non-negative least squares. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems. vol. 24. Curran Associates, Inc. (2011)
36. Smith, J.T., Ochoa, M., Intes, X.: UNMIX-ME: spectral and lifetime fluorescence unmixing via deep learning. Biomed. Opt. Express **11**(7), 3857–3874 (Jul 2020)
37. Ortiz-de Solórzano, C., Pengo, T., Zudaire, I.: Efficient blind spectral unmixing of fluorescently labeled samples using multi-layer non-negative matrix factorization. PLoS ONE **8**(11) (2013)
38. Su, Y., Li, J., Plaza, A., Marinoni, A., Gamba, P., Chakravortty, S.: DAEN: Deep autoencoder networks for hyperspectral unmixing. IEEE Trans. Geosci. Remote Sens. **57**(7), 4309–4321 (Jul 2019)
39. Sønderby, C.K., Raiko, T., Maaløe, L., Sønderby, S.K., Winther, O.: Ladder variational autoencoders. arXiv [stat.ML] (Feb 2016)
40. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A.K., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Pontén, F.: Proteomics. tissue-based map of the human proteome. Science **347**(6220), 1260419 (Jan 2015)
41. Waters, J.C., Wittmann, T.: Quantitative imaging in cell biology. preface. Methods Cell Biol. **123**, xix–xx (2014)
42. Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., Rocha-Martins, M., Segovia-Miranda, F., Norden, C., Henriques, R., Zerial, M., Solimena, M., Rink, J., Tomancak, P., Royer, L., Jug, F., Myers, E.W.: Content-aware image restoration: pushing the limits of fluorescence microscopy. Nat. Methods **15**(12), 1090–1097 (Dec 2018)
43. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. arXiv [cs.CV] (Jan 2018)

44. Zimmermann, T., Marrison, J., Hogg, K., O'Toole, P.: Clearing up the signal: spectral imaging and linear unmixing in fluorescence microscopy. Methods Mol. Biol. **1075**, 129–148 (2014)
45. Zimmermann, T., Rietdorf, J., Pepperkok, R.: Spectral imaging and its applications in live cell microscopy. FEBS Lett. **546**(1), 87–92 (Jul 2003)