Research Article Vol. 14, No. 9 / 1 Sep 2023 / Biomedical Optics Express 4814
Biomedical Optics EXPRESS

AutoUnmix: an autoencoder-based spectral unmixing method for multi-color fluorescence microscopy imaging

YUAN JIANG,[1,†] HAO SHA,[1,†] SHUAI LIU,[2] PEIWU QIN,[3,4] AND YONGBING ZHANG[1,*]

[1] *School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, Guangdong 518055, China*
[2] *Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong Province 518055, China*
[3] *Center of Precision Medicine and Healthcare, Tsinghua-Berkeley Shenzhen Institute, Guangdong Province, 518055, China*
[4] *Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong Province, 518055, China*
[†] The authors contributed equally to this work.
[*] *ybzhang08@hit.edu.cn*

Abstract: Multiplexed fluorescence microscopy imaging is widely used in biomedical applications. However, simultaneous imaging of multiple fluorophores can result in spectral leaks and overlapping, which greatly degrades image quality and subsequent analysis. Existing popular spectral unmixing methods are mainly based on computational intensive linear models, and the performance is heavily dependent on the reference spectra, which may greatly preclude its further applications. In this paper, we propose a deep learning-based blindly spectral unmixing method, termed AutoUnmix, to imitate the physical spectral mixing process. A transfer learning framework is further devised to allow our AutoUnmix to adapt to a variety of imaging systems without retraining the network. Our proposed method has demonstrated real-time unmixing capabilities, surpassing existing methods by up to 100-fold in terms of unmixing speed. We further validate the reconstruction performance on both synthetic datasets and biological samples. The unmixing results of AutoUnmix achieve the highest SSIM of 0.99 in both three- and four-color imaging, with nearly up to 20% higher than other popular unmixing methods. For experiments where spectral profiles and morphology are akin to simulated data, our method realizes the quantitative performance demonstrated above. Due to the desirable property of data independency and superior blind unmixing performance, we believe AutoUnmix is a powerful tool for studying the interaction process of different organelles labeled by multiple fluorophores.

© 2023 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

1. Introduction

Fluorescence microscopy imaging is an essential tool in biomedical research to obtain cellular tissue characteristics of samples. Labeling imaging samples with multiple fluorescent dyes may result in spectral overlap between emission signals detected by distinct channels, leading to cross-talk and blurring of signals, and thereby limiting the number of fluorophores that can be employed simultaneously [1]. To address this problem, various spectral unmixing algorithms have been developed for multispectral imaging to obtain the spectral features, and separate them into corresponding image sets of individual dyes, thus outputting pure sample images.

To identify the individual fluorescent dyes of biomarkers and the corresponding ratios of each pixel in each channel, researchers have proposed several spectral unmixing methods, which are mainly divided into linear unmixing and non-linear unmixing. Based on the assumption of linear mixing of emission fluorescence, various linear spectral unmixing methods usually require known

#498421 https://doi.org/10.1364/BOE.498421
Journal © 2023 Received 21 Jun 2023; revised 12 Aug 2023; accepted 14 Aug 2023; published 22 Aug 2023

Biomedical Optics EXPRESS

emission spectral profiles of fluorophores. Linear unmixing [2] can be formulated as the inverse problem of $Y = AX$, where Y, X, and A represent the mixed image, unmixed image, and mixing matrix respectively. Although linear unmixing is employed widely in multi-color microscopy imaging [3,4], the unmixing performance mainly relies on the quality of the calibrated emission spectra and the mixing matrices. When the calibrated environment is not consistent with the actual experimental environment, the measured emission spectra may be shifted significantly, affecting the unmixing results greatly [5]. Additionally, traditional iterative optimization methods to solve the inverse problem of linear unmixing suffer from heavy computational costs. To improve computational efficiency, some researchers have proposed unmixing methods using vertex component analysis (VCA) [6] and independent component analysis (ICA) [7]. Meanwhile, based on the property that both the fluorescent intensity and the ratios of different channels are non-negative, the non-negative matrix factorization (NMF) method [8,9] has been proposed to solve the inverse problem.

In order to solve the problem of requiring prior spectral information and low computational efficiency, many non-linear spectral unmixing methods have been proposed in recent years. Seo et al. [10] design an information theory-based iterative optimization method by minimizing mutual information between channels of mixed images to blindly unmix 15 colors without reference spectra. With the wide spread of machine learning, related methods have been applied in the field of fluorescent microscopy imaging. Some studies utilize support vector machine (SVM) [11] to learn the pixel feature of fluorescent images and cluster them to realize spectral unmixing and separation of cellular locations [12,13]. In addition, McRae et al. [14] propose Learning Unsupervised Means of Spectra (LUMoS) without the need for prior measurements of their emission spectra and constraints on the number of fluorophores that can be utilized. This work uses K-Means to learn the spectral characteristics of individual fluorophores from mixed images to achieve blind spectral unmixing. However, the latter two machine learning methods require hand-craft iterating parameters and the settings of hyperparameters may impact the final effects of unmixing. Besides, methods based on phasor analysis [15] show some advantages in this regard. Cutrale et al. [16] develop a spectral phasor unmixing algorithm that employed Fourier transform to map spectral images to a two-dimension plane and perform spectral unmixing with clustering methods. Furthermore, Gratton et al. [17] realize good performance in fluorescent lifetime imaging with spectral phasor analysis. However, these phasor-based nonlinear unmixing methods are still dependent on the performance of clustering methods such as K-Means, though they are analytically easy and computationally fast in phase space.

In recent years, some studies have turned to the use of deep learning for fluorescence microscopy to achieve spectral unmixing. These works use data-driven methods to extract the features of emission spectra instead of utilizing prior emission spectra directly and have advantages of noise insensitivity and good generalization performance. Smith et al. [18] combine 3-D convolution and Xception [19] architecture to propose a multi-layer network for the unmixing of fluorescence lifetime imaging. Their network is trained on synthetic datasets and achieves better unmixing results in vivo imaging than that of iterative fitting methods. Li et al. [20] incorporate the physical process of multi-color fluorescence imaging into the spectral unmixing framework which couples several generator networks to recover the pure images. Manifold et al. [21] propose an UwU-Net model based on classical U-Net and demonstrate impressive unmixing performance for hyperspectral imaging, mass spectrometry imaging, and Raman scattering imaging. Overall, the deep learning methods have the potential to solve the problems of the prior emission spectra requirement and high computational load in spectral unmixing.

Here, we propose an autoencoder-based fast asymmetric spectral unmixing method (AutoUnmix), which can blindly separate mixed images without reference spectra. Compared with the previous methods that focus on only the spectral unmixing process, our proposed method leverages the physical process of both spectral unmixing and mixing into two networks respectively.

Biomedical Optics EXPRESS

Specifically, for the unmixing process in actual microscopy imaging, we design a spectral learning module to learn the mutual information features between channels and a spatial learning module to extract morphological features. In addition, we employ a U-Net [22] to extract the features of spectral mixing process to further help reconstruct unmixed images accurately. In experiments, we validate that AutoUnmix can separate highly mixed images precisely in around 100ms after training the model on simulated datasets. In addition, a new transfer learning framework is also proposed to further improve the image quality of spectral unmixing by fine-tuning network parameters for the case of real images. On the real-acquired images of mouse cell samples and mouse brain slice samples, we demonstrate that our method has great unmixing reconstruction performance and good generalization. More importantly, AutoUnmix solves the problem that ground truth is not available for the real multi-color fluorescence imaging system and can be applied to different microscopic imaging systems without retraining, thus reducing training cost and improving unmixing efficiency.

2. Materials and methods

2.1. Overview of the workflow

We propose a supervised asymmetric autoencoder unmixing network (AutoUnmix) to learn the spectral characteristics of each fluorophore, correct the fluorescent crosstalk from other channels, and output pure images that only contain the corresponding labeled samples for each channel. The AutoUnmix contains the unmixing stage and the mixing stage to imitate the physical process of spectral unmixing and mixing during imaging, as shown in Fig. 1(a). In the first stage, the unmixing network learns the features of spectral unmixing by taking mixed images y as input and outputting unmixed image $\hat{x}$. A spectral learning module is employed with channel attention mechanism to learn spectral features between channels and a U-Net is used to extract spatial features. We concatenate and convolve all the generated feature maps in the reconstruction module to obtain the final unmixed output. Then, in the mixing stage, the mixing network extracts the spectral mixing features from the generated unmixed images. Since the physical principle of spectral mixing is not complex, the mixing network adopts a vanilla U-Net to reduce computational complexity and realize fast spectral mixing. Compared to the mixing stage, the unmixing network is more sophisticated. This asymmetric model design helps our method maintain the unmixing performance while reducing the computational cost of the entire model.

The unmixing network, as shown in the left part of Fig. 1(a), takes the mixed image y with dimension (H, W, C) as input, where H and W denote the spatial size of y, and C denotes the number of channels of y. We design a Spectral Learning Module (SLM) containing a convolution and a Channel Attention Module (CAM) [23]. The operation of CAM can be formulated as:

$$CAM(feat) = MLP_2(MLP_1(AvgPooling(feat))) + feat, \tag{1}$$

where $feat$ denotes the high-dimension feature map, MLP_1 containing a fully connected layer maps the input of size $1 \times 1 \times C_{feat}$ to size of $1 \times 1 \times C_{feat}/16$ and MLP_2 is also a fully connected layer performing the inverse operation of MLP_1. The SLM sequentially selects each channel pair i and j and generates the corresponding high-dimension feature maps as follows:

$$SLM(y(:,:,i), y(:,:,j)) = CAM(LeakyReLU(Conv(Concat(y(:,:,i), y(:,:,j))))), \tag{2}$$

where $y(:,:,i)$ and $y(:,:,j)$ denote the corresponding channel i and j of mixed image y. Each feature map extracted by SLM is then fed into a U-Net that shares parameters to reduce computational cost to learn the spatial features. Finally, in the spectral reconstruction module, all of the U-Net outputs are concatenated and convolved to obtain the unmixed image $\hat{x}$, which can be expressed

Biomedical Optics EXPRESS

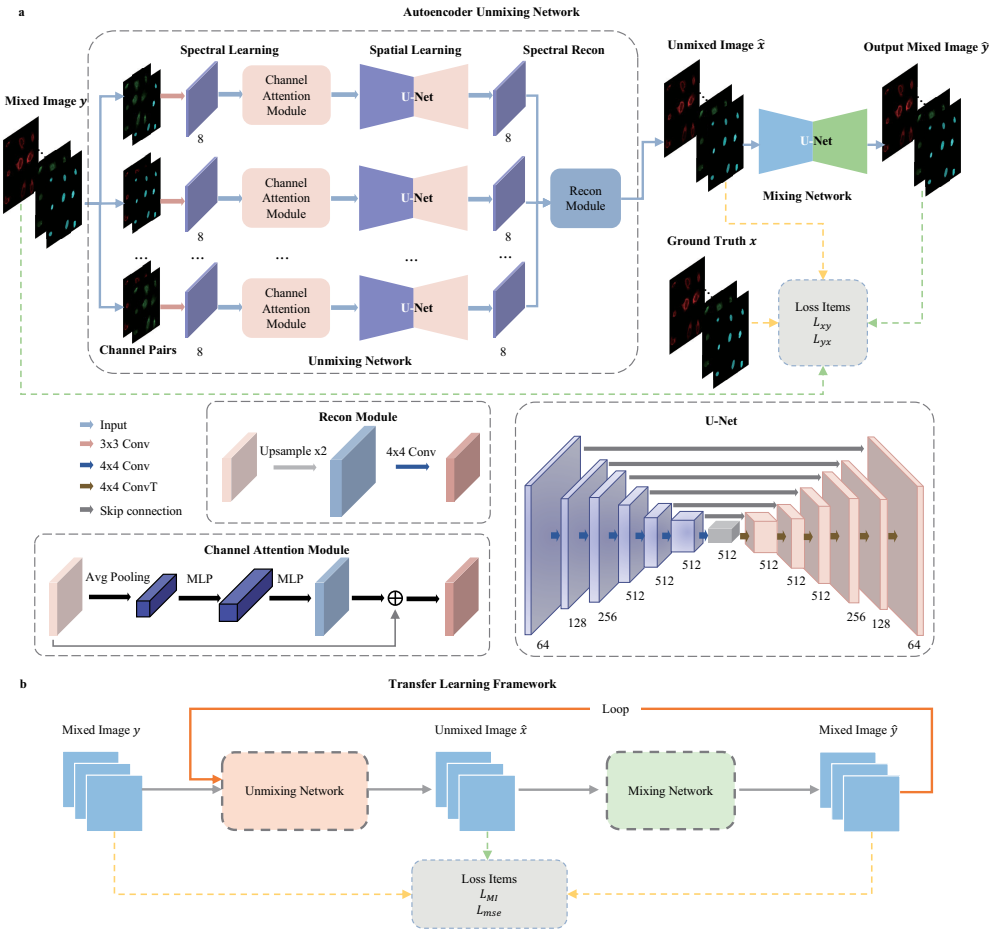

Fig. 1. Overview of training procedure and transfer learning procedure of AutoUnmix. **a.** The training procedure of spectral unmixing model. In Unmixing Network, every two channels of input mixed image are selected to extract spectral features by Convolution and Channel Attention Module. A U-Net with shared parameters learns the spatial features in the feature map and then the unmixed image is reconstructed by Recon Module. The backbone of Mixing Network is a U-Net. **b.** The procedure of transfer learning. Mixed image is the initial input. The loop consists of Unmixing Network and Mixing Network. The output from the final loop is the unmixed image for real-acquired mixed images.

as:

$$\hat{x} = Conv(Concat(UNet(SLM(y(:,:i), y(:,:,j))), forall\ i \neq j)), \tag{3}$$

The mixing network, as shown in the right part of Fig. 1(a), takes the unmixed image $\hat{x}$ generated in the previous stage as input and employs a lightweight U-Net to recover the mixed image.

To ensure better reconstruction image quality, we reduce the difference between the output unmixed image and ground truth while preserving the morphological structure of tissue cells. The loss function during training includes Mean Squared Error (MSE) loss and Structural Similarity Index Measure (SSIM) [24] loss, shown as follows,

$$Loss = \alpha L_{xy} + (1 - \alpha) L_{yx}, \tag{4}$$

$$L_{xy} = m.s.e(y, \hat{y}) + \gamma(1 - SSIM(y, \hat{y})), \tag{5}$$

Biomedical Optics EXPRESS

$$L_{yx} = m.s.e(x, \hat{x}) + \gamma(1 - SSIM(x, \hat{x})), \tag{6}$$

where $m.s.e$ represents the MSE loss, L_{xy} is the loss term guiding the mixing network training and L_{yx} is the loss term guiding the unmixing network training. α controls the relative importance between L_{xy} and L_{yx} and it can be selected empirically. Here, we set it as 0.5.

Transfer learning framework. In Fig. 1(b), a transfer learning framework is designed for the situation that it is impossible to obtain the spectral unmixing ground truth in real experiments. Therefore, the input is the acquired mixed image y and the output is the generated unmixed image $\hat{x}$. The intermediate result is the mixed image $\hat{y}$ reconstructed by $\hat{x}$. The mixing stage and unmixing stage are iteratively repeated to improve spectral unmixing quality. We design our custom loss function by minimizing MSE and mutual information loss between every two channels. This conforms to the independent and exclusive characteristics of the cell tissues labeled with fluorescence dyes. The design of the loss function also enables us to adjust and optimize the model for different real experimental environments based on the model trained on simulated data, improving the generalization ability and ensuring high-quality spectral unmixing in various situations.

The loss function in transfer learning is different from the one in the training with simulated data. In addition to computing the MSE loss between the input and output mixed images, mutual information loss [25] also needs to be computed.

$$Loss = \beta m.s.e(y, \hat{y}) + (1 - \beta) \sum_{i,j} MI(\hat{x}_i, \hat{x}_j), \tag{7}$$

where β represents the coefficient, and the latter term represents the loss calculating the mutual information between every two channels and summing them up.

Network training and testing. The model used in this paper is trained on the simulated datasets using randomly initialized parameters and optimized using our custom loss function with stochastic gradient descent. Our proposed spectral unmixing framework utilizes the Adam optimizer with the learning rate of 0.0002, the beta of 0.5 and 0.999, and is trained for 200 epochs. The simulated datasets are split into training and testing sets in a 4:1 ratio. All the images are cropped into a patch size of 256×256 and augmented using rotations and flips to enlarge training sets. All the training and prediction are performed on a server running Ubuntu 18.04, equipped with an Intel Xeon 4216 processor and Nvidia RTX 2080Ti GPU. On our machine, training the 3-channel and 4-channel unmixing model takes approximately 3 and 5 hours respectively, while the prediction time for a single test image sized of 256×256 is around 100 milliseconds.

2.2. Metrics

Following the numerical evaluation methods proposed in the previous papers, the SSIM, PSNR and MSE are used to evaluate the results of the spectral unmixing methods. The values of the SSIM metrics are between 0 and 1. The higher SSIM and lower MSE values represent a high-quality output result. MSE is a general objective evaluation index in measuring image quality, while SSIM assures a high degree of structural consistency between output and ground truth. The SSIM is defined as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_1)}, \tag{8}$$

where μ_x, μ_y and σ_x^2, σ_y^2 are the averages and variances of images x, y; σ_{xy} is the covariance of x and y; c_1, c_2 are the variables used to stabilize the division with a small denominator. We also use PSNR to measure the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. The higher PSNR represents

Biomedical Optics EXPRESS

better image reconstruction quality. PNSR is defined as follows based on MSE:

$$PSNR(x, y) = 10 \log_{10} \frac{MAX_x^2}{MSE}, \tag{9}$$

where MSE is the mean square error between image x and image y, MAX_x is the maximum possible gray-scale value of the image, which is 255 in our experiments.

2.3. Data acquisition

Simulated datasets. Although some existing methods can acquire images with as few spectral crosstalk as possible, such methods involve tedious biological sample preparation process. Here, we generate datasets with various fluorophores and spectra profiles to simulate different imaging environments, which are then used for training and testing. After that, we apply the trained model to unmixing tasks for real sample images. The generation procedure is shown in Fig. S8. Firstly, we select 345 4-channel images of U-2 OS cells from the DULoc [26] dataset as ground truth, which were obtained from the HPA [27] database and processed by the cell masks tool. Each image has 9 cells and 4 channels, where each channel corresponds to protein, nucleus, ER, and microtubule. Second, we select multiple fluorophore groups and download the reference spectra curve of them from FPBase [28] database, including DAPI, AF488, CF488, AF532, and so on. We also set the center wavelength and width of bandpass filters according to the selected fluorophores. The intensity values of each channel for each pixel of the ground truth image are mixed using linear mixing, and then additional Poisson noise is added to each channel to simulate the shot noise.

$$Y = filter(AX + n_{possion}), \tag{10}$$

where Y represents generated mixed image, X is the original unmixed image, A is the mixing matrix, $n_{poisson}$ is Poisson noise, and filter represents Gaussian filter and median filter. Images are convolved with a Gaussian filter with a standard deviation of 0.5 and a 3×3 median filter to represent the real-world diffusion effect to obtain mixed images. The purpose of this workflow is to closely simulate the real conditions that biological samples often exhibit overlapping tissue structures. The spatial dimensions of each synthesized image are 512×512, and 3- and 4-channel images are generated, respectively.

Real biological sample acquisition. Real biological experiments were performed on mouse cells stained with DAPI for nuclei, FITC for actin fibers, and MitoTracker Red for mitochondria. Multicolor images (Fig. S8a) were obtained using a TIRF microscope, Nikon ECLIPSE Ti2-E. The fluorophores were excited with three lasers at 405, 488, and 561nm, respectively. The size of the acquired mixed images is 512×512×3. Then, we normalize each channel of the images to facilitate the subsequent unmixing tasks. We also used 3-color (CF488, ATTO514, ATTO531) and 4-color (CF488, ATTO488, ATTO514, ATTO532, Fig. S8a) fluorescence images of mouse brain slices [10] to test the unmixing performance of our method on real samples that are not consistent with the distribution of the training datasets.

3. Results

3.1. Unmixing performance on simulated datasets

We first compare AutoUnmix with linear unmixing [2] and LUMoS [14] in terms of unmixing accuracy on the simulated datasets. To generate simulated data, we select three fluorophores (DAPI, Alexa Fluor 488, MitoTracker Red) with emission spectra highly overlapped and choose three bandpass filters (470/40, 520/40, 600/50nm) corresponding to the spectra peaks of selected fluorophores (Fig. 2(a)). Then, 345 three-color labeled U-2 OS cell images are artificially synthesized based on the above spectral curves. Each channel of the cell images corresponds to protein, nucleus, and endoplasmic reticulum (ER), respectively. We randomly select 35 images

Biomedical Optics EXPRESS

to compare the unmixing performance of linear unmixing, LUMoS, and AutoUnmix (Ours). Linear unmixing is performed according to the theoretical reference spectra (Fig. 2(a)), while LUMoS does not need emission spectra to unmix images. We quantify the unmixing accuracy by computing the SSIM, PSNR, and MSE between the output unmixed images and the ground truth. Table 1 shows the quantitative evaluation of the unmixing results of the three methods on the synthetic datasets. The performance of linear unmixing and LUMoS methods are closely comparable, with average SSIM values of 0.92 and 0.91 respectively. However, linear unmixing demonstrates significant variance, suggesting a lack of robustness in this approach. Our method achieves an average SSIM of 0.99 on 35 mouse cell images of different shapes, with a small variance, indicating our stable performance and the ability to learn the spectral features of mixed images well. In addition, AutoUnmix completes unmixing within 60ms, which is up to 100 times faster than the other two methods. Figure 2(c) shows the unmixing results of the three methods and the ground truth. It is obvious that the unmixing performance of linear unmixing is poor, as it cannot correctly unmix the images in channel 2 and channel 3, and even fails to unmix them completely. In contrast, our method is almost identical to the ground truth. As shown in Fig. 2(d), we measured the normalized intensity values along the white arrows in each channel, and our method shows the highest consistency with the ground truth, while linear unmixing and LUMoS do not.

Table 1. Quantitative comparisons of unmixing performance between the three methods

	Methods	MSE	SSIM	PSNR	Time
Simulated Dataset with 3 Channels	Linear unmixing	0.0046±0.0014	0.92±0.015	23.6±1.3	815s
	LUMoS	0.0028±0.00059	0.91±0.01	25.7±.9	8s
	Ours	0.00039±0.00017	0.99±0.0052	34.5±1.6	60ms
Simulated Dataset with 4 Channels	Linear unmixing	0.0068±0.0015	0.78±0.033	22±0.94	817s
	LUMoS	0.0067±0.002	0.82±0.025	22±1.3	9s
	Ours	0.00077±0.00013	0.98±0.0035	31.2±0.7	90ms

Then, we validate the unmixing performance over four dyes highly spectral overlapped: CF488, ATTO488, ATTO514, and ATTO532 (Fig. 2(b), bandpass filter of 515/10, 525/10, 535/10, and 555/10nm). A simulated dataset is generated using four-color labeled U-2 OS cell images with each channel corresponding to protein, nucleus, ER, and microtubules. We still use part of the simulated dataset for training and randomly select 35 mixed images for testing. Although the fluorescence dyes have highly overlapping emission spectra, our method (Fig. 2(e)) still shows good consistency with the ground truth, with an average SSIM of 0.987. In Fig. S1 and S2, we also present the comparison of each channel image between these three unmixing methods as well as the normalized intensity values along the white arrows. The quantified results in Table 1 demonstrate our AutoUnmix can successfully handle various fluorescence mixing situations with fast computational speed.

3.2. Validation of blind spectral unmixing

We next validate the ability of AutoUnmix for blind spectral unmixing by choosing fluorophore groups different from those used for training. Our experiments show that our AutoUnmix can successfully separate mixed images with absolutely different emission spectra curves. We select three groups of fluorophores. The emission spectra of the first group (CF488, ATTO514, ATTO532) are similar to the previous four-channel simulated dataset, but it only has three highly overlapped fluorophores. The second group (tagBFP, Cerulean, Citrine) is commonly used for colorful cell imaging, where tagBFP is for nucleus, Cerulean is for cell membrane, and Citrine is for mitochondria. The third group (CF633, Alexa647, CF660R, CF680R) is overlapped relatively

Biomedical Optics EXPRESS

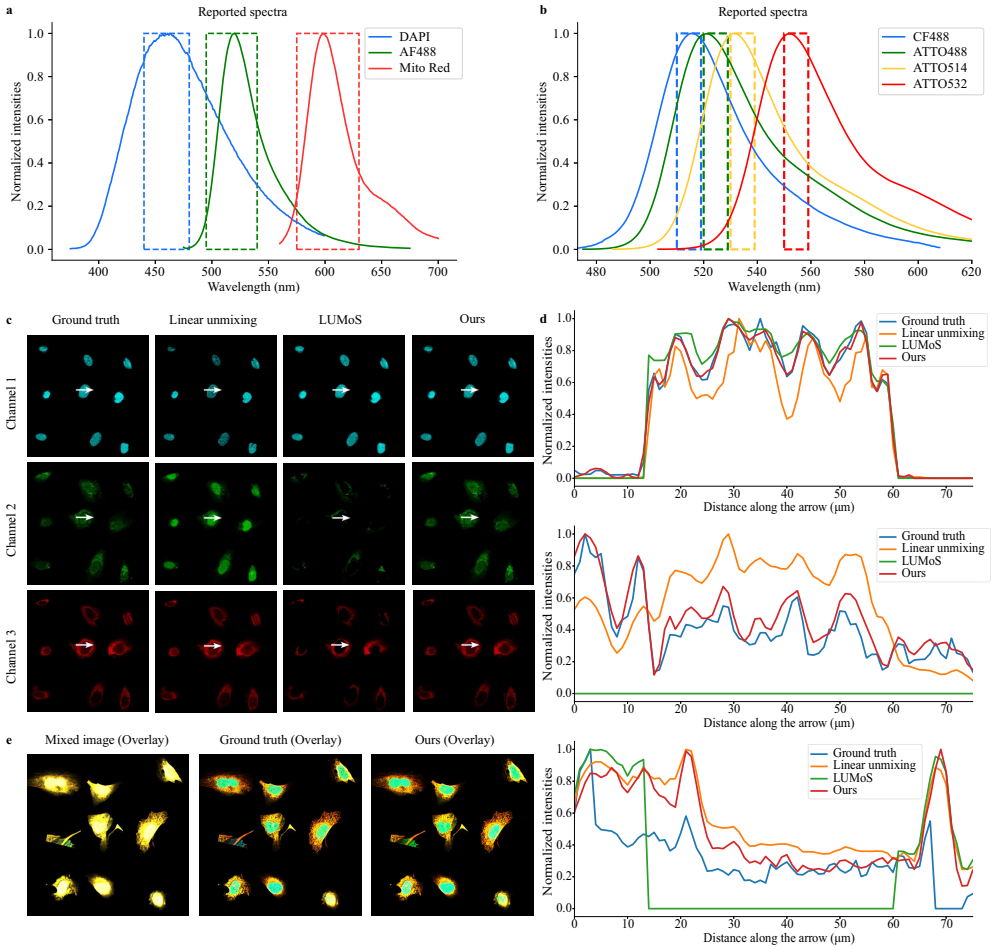

Fig. 2. Comparison between linear unmixing, LUMoS and our methods. a. The theoretical emission spectral curve of 3 fluorophores (CF488, ATTO514, ATTO532), and the selection of bandpass filter. **b.** The emission spectra of the 4 fluorophores used in the simulated dataset and the detection channels. **c.** The unmixed images of ground truth, linear unmixing, LUMoS and ours. Each row represents channel 1, 2, and 3 respectively. **d.** From top to bottom, normalized intensity values along the white arrows for channel 1, 2 and 3 in **c.** **e.** Overlay of the results of 4-color unmixing, including the mixed image, ground truth, and the output unmixed image.

balanced. The test datasets are generated based on the three groups of spectral curves (Fig. 3). Following the testing routine in section 3.1, we directly use the model for 3-channel and 4-channel unmixing trained in section 3.1 to test on the data of the corresponding number of channels. The quantitative results on different combinations of fluorophores are shown in Table S1 and the unmixed images are displayed in Fig. 3.

The statistics align with the experiment results in section 3.1. For all three groups, linear unmixing and LUMoS both exhibit worse performance with SSIM below 0.95 and PSNR below 28. On the four-color unmixing task (Fig. 3(c)), linear unmixing and LUMoS achieve SSIM values below 0.9 and PSNR values approximately at 22, a result that concurs with Table 1. Conversely, our proposed AutoUnmix attains higher performance, demonstrating SSIM exceeding 0.97 and PSNR exceeding 30 on the generated three-channel mixed images, and SSIM 0.95 and

Biomedical Optics EXPRESS

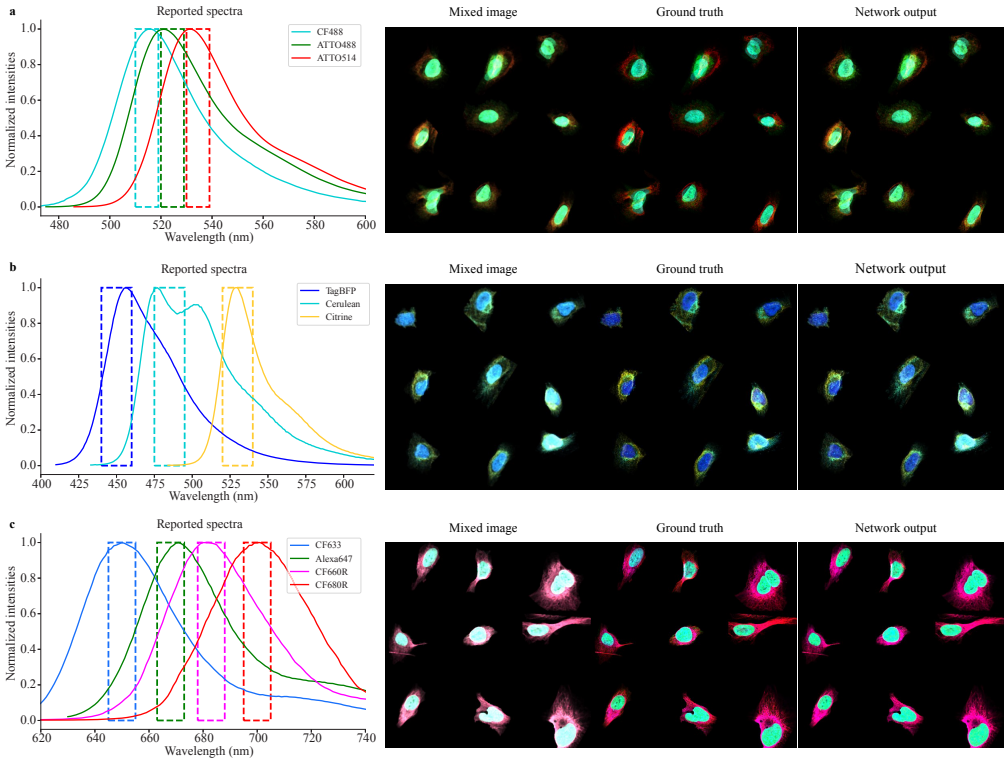

Fig. 3. Reference spectra and blind unmixing results of different groups. Different groups of theorectical spectral curve and corresponding bandpass filter. The overlayed images of mixed, ground truth and our output in **a** (CF488, ATTO514, ATTO532), **b** (tagBFP, Cerulean, Citrine) and **c** (CF633, Alexa647, CF660R, CF680R).

PSNR 28 for the mixed data of four fluorophores with a diminishment of SSIM by 0.02 and PSNR by 3 compared to Table 1. Our method still shows superiority over the other two methods for all three groups. We can see from the overlay outputs that AutoUnmix can separate the mixed images and have good consistency with the ground truth (Fig. S4, S5, and S6 show each channel of the comparison results). We also evaluate the unmixing results for the same sample image as Fig. 3(a) generated with largely different ratios of fluorophore combinations in Fig. S3. Whatever the degree of spectral overlapping, our method can still perform well in such tests. The results above show that our AutoUnmix has impressive blind unmixing performance, indicating our unmixing method enables correctly separating fluorescent dyes with distinct emission spectra from training sets.

Comparing the quantitative and visual results of mixing network output (Fig. S9), we also find that our network can actually learn the process of spectral mixing. AutoUnmix is capable of acquiring knowledge regarding the mutual information of adjacent channels through the Spectral Learning module, thus learning invariant spectral characteristics and establishing independence from the training dataset (Table S2). We assume that our sequential combination of the mixing network and unmixing network might possess the capability to discern the fundamental physical mechanisms governing spectral mixing processes. Consequently, AutoUnmix may not require retraining neural networks for different fluorescent dyes, reducing the heavy cost of re-training while maintaining good unmixing results.

Biomedical Optics EXPRESS

3.3. Generalization performance on real biological samples

In real biological experiments, it is impossible to obtain the ground truth of spectral unmixing, which imposes a great challenge to supervised deep learning methods. Here, we propose a transfer learning framework (Fig. 1(b)) to remedy the absence of ground truth by applying finetuning. First, we train a base model using the simulated datasets generated from reference spectra of the same number of fluorophores. Then we initialize the parameters of AutoUnmix with the base model and iteratively finetune the parameters on the real-acquired images to learn the distribution gap between synthetic data and real data. The finetuning is stopped when the MSE loss between input mixed image and output mixed image is smaller than a threshold. We demonstrate the unmixing performance of our proposed method on three real biological samples.

We tested the mouse cell samples with DAPI, FITC and MitoTracker Red. The 3-channel mixed images were obtained using a TIRF microscope. We employ the base model in section 3.1 where the spectral curve for training is quite similar to that of mouse cell samples and then finetune the model on the real acquired mixed images. Figure 4(a) displays the original mixed image, linear unmixing results, LUMoS results, and Ours. In channel 1, the nucleus structure cannot be completely reconstructed by linear unmixing method, and LUMoS can maintain the cell structure to some extent while losing some clear details. In channel 2, crosstalk from channel 1 is visible in the mixed image. Only our method can remove all nuclei while reducing the noise signal in the image simultaneously. In channel 3, it can be seen that LUMoS is indeed able to remove the crosstalk from channel 1 and 2, but destroys the mitochondrial structure. However, our method can both maintain the mitochondrial structure and remove the crosstalk from the nuclei, indicating that AutoUnmix can indeed separate real mixed images.

Next, we verify the unmixing quality of AutoUnmix on samples exhibiting different morphology. Figure 4(b) displays the mixed images of mouse brain slices with highly overlapped spectra (CF488, ATTO488, ATTO514, ATTO532), which shows quite different morphological structures from training datasets. Compared with linear unmixing and LUMoS, AutoUnmix with transfer learning can achieve better unmixing performance as shown in Fig. S7. Our method without transfer learning cannot completely remove some of the cell structures that crosstalk into channel 1. In contrast, by employing transfer learning framework, more cell details can be recovered in the white rectangle in Fig. 4(b), indicating that our transfer learning framework can significantly improve the output quality.

We further test the proposed method for samples with different morphology and emission spectra as shown in Fig. 4(c). The three fluorophores are CF488, ATTO514, and ATTO532. We directly use the previously trained network to finetune on the real-acquired images, and the impressive unmixing results can be obtained. Although the three methods can remove the crosstalk from channel 2 into other channels, linear unmixing loses a lot of tissue structure in channel 1, and LUMoS lacks some details in channel 3, while our method maintains the structure of the cellular tissue well. These results show that our method can still maintain high unmixing performance for samples with different morphologies by performing transfer learning on real sample images. This demonstrates AutoUnmix does not need to retrain networks for real experimental conditions or rely on reference spectra.

In order to establish the validity of our spectral unmixing approach, we acquire the typical single-stain images corresponding to distinct channels for comparison purposes (Fig. S11). These images effectively showcase the distinct morphological attributes of each structure and function as reference to evaluate the accuracy of our unmixing method's outputs. While our method may not perfectly reconstruct all correct structures, it is noteworthy that AutoUnmix outperforms the other two methods when compared against the reference images.

Biomedical Optics EXPRESS

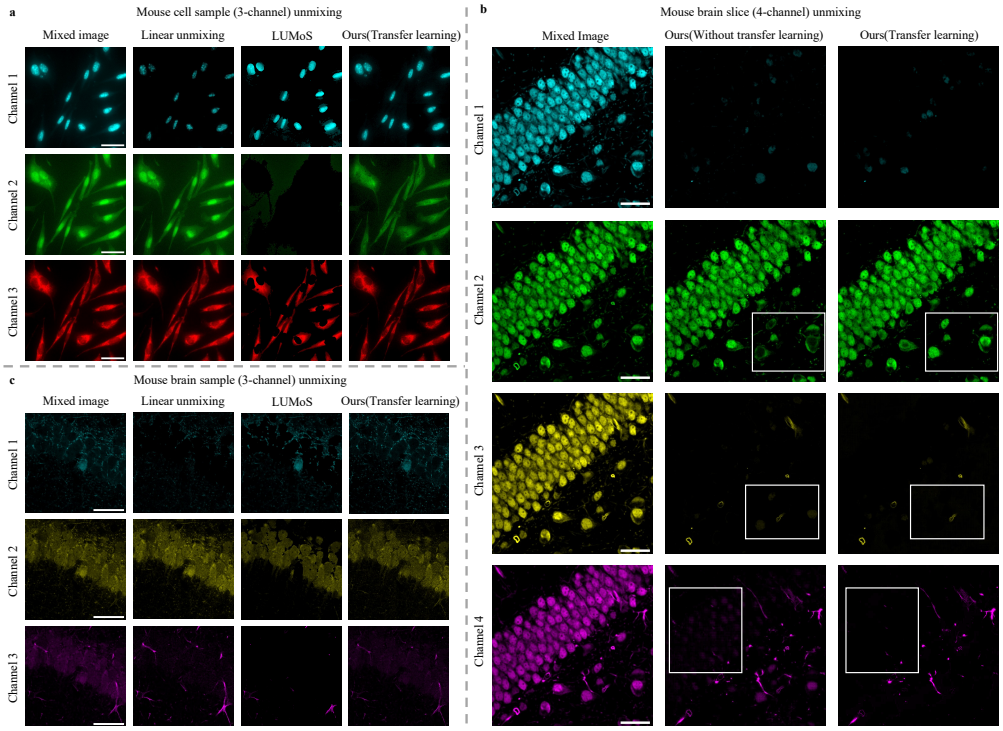

Fig. 4. Unmixing results on real biological samples. a. Unmixing results on 3-channel mouse cell samples exhibiting similar spectra and morphology to training data. Each column is results of the three methods and each row represents different spectral channel. Scale bar: 50μm **b.** Unmixing results of the AutoUnmix with and without transfer learning on 4-channel mouse brain slice samples with different morphology. Scale bar: 30μm **c.** Unmixing results on 3-channel mouse brain slice samples with different morphology and emission spectra. Scale bar: 40μm.

4. Conclusion

In this work, we propose a new deep learning method for fluorescence spectral unmixing, termed AutoUnmix, as an alternative flexible choice for separating mixed images. Our proposed method designs an asymmetric autoencoder model that incorporates the process of spectral mixing and unmixing as two stages in deep neural networks to simultaneously learn spectral characteristics during this procedure. On simulated datasets with 3- and 4-color mixed images, we can achieve SSIM of 0.99 and PSNR of 30+, which show great superiority over traditional linear unmixing methods and machine learning-based LUMoS method.

Further, we demonstrate the unmixing performance of our AutoUnmix with the proposed transfer learning framework. By fine-tuning the network on single real-acquired image, our method successfully unmixes images on both real mouse cell samples and mouse brain slice samples that have inconsistent morphology with training sets and largely different emission spectra of labeling fluorophores from the theoretical spectra. Compared to the other two unmixing methods, our method shows impressive unmixing performance and generalization ability. This result highlights that we present a novel approach to address the issue that the ground truth of real samples is not available and training images differ from real biological images. This also can be attributed to the autoencoder architecture in AutoUnmix that is naturally applicable to network training without ground truth and improves unmixing performance by iterative looping between

Biomedical Optics EXPRESS

the encoder and decoder. To some extent, our approach solves the problem that the independent identically distribution assumption between synthetic data and real data is not satisfied. However, it is noted that since our method does not learn morphological features during training, some channels of real samples cannot be absolutely separated, which deserves further improvement in subsequent research.

Finally, we aim to extend the application of our AutoUnmix to various unmixing imaging modalities. Our transfer learning method can further adapt to different real imaging systems and bio-samples labeled with different dyes, such as mRNA imaging [29–31], cell tracing [32], fluorescent barcoding techniques [13,33], multidimensional cytometry [34] and sub-cellular structures locations [35]. Moreover, our unmixing network architecture can be potentially used in label-free detection of intracellular structures [21,36]. Our method may also perform well in biomedical tasks such as virtual staining [37–39], as they also involve multi-spectral imaging tasks. Additionally, our method can be combined with spectral phasor methods [17,40,41] to achieve spectral unmixing while utilizing the computational efficiency in the Fourier space. We can also explore replacing modules in our mixing network and unmixing network with more efficient modules like vision transformer blocks [42,43] to better extract spatial features and address deficiencies on unseen morphologies. Overall, we anticipate that our AutoUnmix will be useful in medical imaging, microscopy imaging, remote sensing [44,45], and an even broader range of computer vision tasks.

Funding. National Natural Science Foundation of China (62031023); Shenzhen Science and Technology Project (GXWD20220818170353009, JCYJ20200109142808034, JSGG20191129110812708).

Disclosures. The authors declare no conflicts of interest.

Data Availability. The data that support the findings of this study are available from our GitHub repository [46] along with code.

Supplemental document. See Supplement 1 for supporting content.

References

1. 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 **1075**, 129–148 (2014).
2. T. Zimmermann, "Spectral imaging and linear unmixing in light microscopy," Adv Biochem Eng Biotechnol **95**, 245–265 (2005).
3. A. M. Valm, S. Cohen, W. R. Legant, J. Melunis, U. Hershberg, E. Wait, A. R. Cohen, M. W. Davidson, E. Betzig, and J. Lippincott-Schwartz, "Applying systems-level spectral imaging and analysis to reveal the organelle interactome," Nature **546**(7656), 162–167 (2017).
4. K. Chen, R. Yan, L. Xiang, and K. Xu, "Excitation spectral microscopy for highly multiplexed fluorescence imaging and quantitative biosensing," Light: Sci. Appl. **10**(1), 97 (2021).
5. S. L. Jacques, "Optical properties of biological tissues: a review," Phys. Med. Biol. **58**(11), R37–R61 (2013).
6. J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," IEEE Trans. Geosci. Remote Sensing **43**(4), 898–910 (2005).
7. J.-F. Cardoso, "Blind signal separation: statistical principles," Proc. IEEE **86**(10), 2009–2025 (1998).
8. Y. Ma, X. Li, P. Huang, D. Hou, Q. Wang, and G. Zhang, "The spectral data analysis and components unmixing based on non-negative matrix factorization methods," Spectrochim. Acta, Part A **177**, 49–57 (2017).
9. D. Jimenez-Sanchez, M. Ariz, J. M. Morgado, I. Cortes-Dominguez, and C. Ortiz-de Solorzano, "Nmf-ri: blind spectral unmixing of highly mixed multispectral flow and image cytometry data," Bioinformatics **36**(5), 1590–1598 (2020).
10. J. Seo, Y. Sim, J. Kim, H. Kim, I. Cho, H. Nam, Y.-G. Yoon, and J.-B. Chang, "Picasso allows ultra-multiplexed fluorescence imaging of spatially overlapping proteins without reference spectra measurements," Nat. Commun. **13**(1), 2475 (2022).
11. C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn **20**(3), 273–297 (1995).
12. Q. Yang, H.-Y. Zou, Y. Zhang, L.-J. Tang, G.-L. Shen, J.-H. Jiang, and R.-Q. Yu, "Multiplex protein pattern unmixing using a non-linear variable-weighted support vector machine as optimized by a particle swarm optimization algorithm," Talanta **147**, 609–614 (2016).
13. H. Shi, Q. Shi, B. Grodner, J. S. Lenz, W. R. Zipfel, I. L. Brito, and I. De Vlaminck, "Highly multiplexed spatial mapping of microbial communities," Nature **588**(7839), 676–681 (2020).
14. T. D. McRae, D. Oleksyn, J. Miller, and Y. R. Gao, "Robust blind spectral unmixing for fluorescence microscopy using unsupervised learning," PLoS One **14**(12), e0225410 (2019).

Biomedical Optics EXPRESS

15. F. Fereidouni, A. N. Bader, and H. C. Gerritsen, "Spectral phasor analysis allows rapid and reliable unmixing of fluorescence microscopy spectral images," Opt. Express **20**(12), 12729 (2012).
16. F. Cutrale, V. Trivedi, L. A. Trinh, C. L. Chiu, J. M. Choi, M. S. Artiga, and S. E. Fraser, "Hyperspectral phasor analysis enables multiplexed 5d in vivo imaging," Nat. Methods **14**(2), 149–152 (2017).
17. L. Scipioni, A. Rossetta, G. Tedeschi, and E. Gratton, "Phasor s-flim: a new paradigm for fast and robust spectral fluorescence lifetime imaging," Nat. Methods **18**(5), 542–550 (2021).
18. J. T. Smith, M. Ochoa, and X. Intes, "Unmix-me: spectral and lifetime fluorescence unmixing via deep learning," Biomed. Opt. Express **11**(7), 3857–3874 (2020).
19. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258.
20. X. Li, *Blind Stain Separation Using Model-Aware Generative Learning and Its Applications on Fluorescence Microscopy Images* (Springer International Publishing, 2021), pp. 98–107.
21. B. Manifold, S. Men, R. Hu, and D. Fu, "A versatile deep learning architecture for classification and label-free prediction of hyperspectral images," Nat. Mach. Intell. **3**(4), 306–315 (2021).
22. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer), pp. 234–241.
23. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19.
24. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Process. **13**(4), 600–612 (2004).
25. M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, "Mutual information neural estimation," in *International conference on machine learning, (PMLR)*, pp. 531–540.
26. M. Q. Xue, X. L. Zhu, G. Wang, and Y. Y. Xu, "Duloc: quantitatively unmixing protein subcellular location patterns in immunofluorescence images based on deep learning features," Bioinformatics **38**(3), 827–833 (2022).
27. J. Li, J. Y. Newberg, M. Uhlén, E. Lundberg, and R. F. Murphy, "Automated analysis and reannotation of subcellular locations in confocal images from the human protein atlas," PLoS One **7**(11), e50514 (2012).
28. T. J. Lambert, "Fpbase: a community-editable fluorescent protein database," Nat. Methods **16**(4), 277–278 (2019).
29. A. Raj, P. Van Den Bogaard, S. A. Rifkin, A. Van Oudenaarden, and S. Tyagi, "Imaging individual mrna molecules using multiple singly labeled probes," Nat. Methods **5**(10), 877–879 (2008).
30. J. R. Moffitt, J. Hao, D. Bambah-Mukku, T. Lu, C. Dulac, and X. Zhuang, "High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing," Proc. Natl. Acad. Sci. **113**(50), 14456–14461 (2016).
31. E. L. Sylwestrak, P. Rajasethupathy, M. A. Wright, A. Jaffe, and K. Deisseroth, "Multiplexed intact-tissue transcriptional analysis at cellular resolution," Cell **164**(4), 792–804 (2016).
32. B. Andreiuk, A. Reisch, M. Lindecker, G. Follain, N. Peyriéras, J. G. Goetz, and A. S. Klymchenko, "Fluorescent polymer nanoparticles for cell barcoding in vitro and in vivo," Small **13**(38), 1701582 (2017).
33. A. M. Valm, R. Oldenbourg, and G. G. Borisy, "Multiplexed spectral imaging of 120 different fluorescent labels," PLoS One **11**(7), e0158495 (2016).
34. D. Schraivogel, T. M. Kuhn, B. Rauscher, M. Rodrìguez-Martìnez, M. Paulsen, K. Owsley, A. Middlebrook, C. Tischer, B. Ramasz, and D. Ordo nez-Rueda, "High-speed fluorescence image-enabled cell sorting," Science **375**(6578), 315–320 (2022).
35. H. Yuan, L. Cai, Z. Wang, X. Hu, S. Zhang, and S. Ji, "Computational modeling of cellular structures using conditional deep generative networks," Bioinformatics **35**(12), 2141–2149 (2019).
36. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," Nat. Methods **15**(11), 917–920 (2018).
37. N. Bayramoglu, M. Kaakinen, L. Eklund, and J. Heikkila, "Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision workshops*, pp. 64–71.
38. Y. Rivenson, H. Wang, Z. Wei, K. De Haan, Y. Zhang, Y. Wu, H. Günaydin, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nat. Biomed. Eng. **3**(6), 466–477 (2019).
39. Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. De Haan, and A. Ozcan, "Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning," Light: Sci. Appl. **8**(1), 23 (2019).
40. P. N. Hedde, R. Cinco, L. Malacrida, A. Kamaid, and E. Gratton, "Phasor-based hyperspectral snapshot microscopy allows fast imaging of live, three-dimensional tissues for biomedical applications," Commun. Biol. **4**(1), 721 (2021).
41. T. Vu, A. Vallmitjana, J. Gu, K. La, Q. Xu, J. Flores, J. Zimak, J. Shiu, L. Hosohama, and J. Wu, "Spatial transcriptomics using combinatorial fluorescence spectral and lifetime encoding, imaging and analysis," Nat. Commun. **13**(1), 169 (2022).
42. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022.

Biomedical Optics EXPRESS

43. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv, arXiv:2010.11929 (2020).
44. J. S. Bhatt and M. V. Joshi, "Deep learning in hyperspectral unmixing: A review," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, (IEEE), pp. 2189–2192.
45. S. Ozkan, B. Kaya, and G. B. Akar, "Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," IEEE Trans. Geosci. Remote Sensing **57**(1), 482–496 (2019).
46. Y. Jiang, H. Sha, S. Liu, P. Qin, and Y. Zhang, "AutoUnmix: an autoencoder-based spectral unmixing method for multi-color fluorescence microscopy imaging: code," Github, (2019), https://github.com/AlphaYuan/AutoUnmix