Nanoscale

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Nanoscale*, 2024, **16**, 5729

Deep-learning-assisted spectroscopic single-molecule localization microscopy based on spectrum-to-spectrum denoising†

Dandan Xu,[a] Yuanjie Gu,[a] Jun Lu,[a] Lei Xu,[a] Wei Wang[b] and Biqin Dong *[a]

Received 19th November 2023,
Accepted 5th February 2024

DOI: 10.1039/d3nr05870k

rsc.li/nanoscale

Spectroscopic single-molecule localization microscopy (sSMLM) simultaneously captures spatial localizations and spectral signatures, providing the ability of multiplexed and functional subcellular imaging applications. However, extracting accurate spectral information in sSMLM remains challenging due to the poor signal-to-noise ratio (SNR) of spectral images set by a limited photon budget from single-molecule fluorescence emission and inherent electronic noise during the image acquisition using digital cameras. Here, we report a novel spectrum-to-spectrum (Spec2Spec) framework, a self-supervised deep-learning network that can significantly suppress the noise and accurately recover low SNR emission spectra from a single-molecule localization event. A training strategy of Spec2Spec was designed for sSMLM data by exploiting correlated spectral information in spatially adjacent pixels, which contain independent noise. By validating the qualitative and quantitative performance of Spec2Spec on simulated and experimental sSMLM data, we demonstrated that Spec2Spec can improve the SNR and the structure similarity index measure (SSIM) of single-molecule spectra by about 6-fold and 3-fold, respectively, further facilitating 94.6% spectral classification accuracy and nearly 100% data utilization ratio in dual-color sSMLM imaging.

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Introduction

Fluorescence microscopy is an indispensable tool for biological imaging, which can observe the structure of cells and their molecular-specific functional mechanisms. However, due to the existence of the diffraction limit, the spatial resolution of conventional fluorescence microscopy is restricted to ~250–700 nm.[1] Super-resolution microscopy has extended the spatial resolution of conventional fluorescence microscopy beyond the diffraction limit, by achieving ~20–100 nm resolution.[2,3] Among these techniques, single-molecule localization microscopy (SMLM) can overcome the diffraction barrier and provide nanometer-level (~20 nm)[4–6] spatial resolution by detecting and localizing the random subsets of fluorophores. Apart from offering a remarkable spatial resolution, multi-color imaging in SMLM is critical for investigating structural and dynamical biological processes at the nanoscale.[7,8] However, conventional multi-color SMLM, based on color filters,[9,10] requires spectrally well-separated channels (~100 nm) between dyes to minimize spectral cross-talk, which limits the number of discrete color channels. Recently, several spectroscopic SMLM (sSMLM) methods have been proposed to significantly extend the number of distinct species *via* simultaneously capturing the spatial position and full spectra of fluorescence emission from single molecules.[8,11–13] The additional spectral signatures enable sSMLM to distinguish highly overlapping dye molecules. Besides, by integrating functional information into the single-molecule spectrum using environmentally sensitive fluorophores, functional super-resolution microscopy (f-SRM)[14,15] enables probing of local physicochemical parameters (*e.g.*, chemical polarity, pH, and hydrophobicity) at nanoscale resolutions. However, the number of photons emitted by a single-molecule localization event is physically limited. Since these photons are further dispersed into spectral dimensions in sSMLM, the signal-to-noise ratio (SNR) of recorded spectra is often insufficient for multi-color and functional super-resolution imaging.[16]

To address the above concern, the most fundamental approach is to capture sufficient fluorescence photons that allow high SNR spectra to be acquired. Nevertheless, compared to SMLM, collected photons need to be further divided into spatial and spectral channels. This causes an inherent trade-off between the localization precision and the spectral fidelity. Despite these obstacles, strategies, including dual-objective sSMLM[8] and symmetrically dispersed sSMLM,[17]

[a]*Academy for Engineering and Technology, Yiwu Research Institute, Fudan University, Shanghai 200433, China. E-mail: dongbq@fudan.edu.cn*
[b]*Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3nr05870k

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

have been developed for more effective photon collection and utilization. However, applications of these schemes are often restricted in practice, as they require more sophisticated experimental setups or are limited by sample types and mounting geometries. In addition to physical strategies, deep learning methods can provide an alternative solution to solve the problems in multi-color sSMLM. Gaire *et al.* developed a machine-learning method for reducing the data acquisition time in sSMLM imaging.[18] The pre-trained network can reconstruct high-density images from low-density images, thereby reducing the number of imaging frames. Besides, instead of the widely used spectral centroid-based (SC-based) method,[8,18] Zhang *et al.* proposed a machine-learning-based method to classify single-molecule emission spectra, and achieved ten-fold reduction in misclassification and two-fold improvement in the data utilization ratio.[19] Meanwhile, Manko *et al.* presented a supervised denoising network for restoring the spatial and spectral images in sSMLM. The training network in such a supervised manner depends heavily on paired ground truth (GT) signals.[20] However, in the context of sSMLM imaging, it is challenging to obtain an unbiased spectral signal due to the fast dynamics and inherent spectral heterogeneity of stochastic single-molecule emission.[11,21]

In this paper, we propose the spectrum-to-spectrum (Spec2Spec) framework, a novel self-supervised deep-learning-based spectral denoising method to unbiasedly remove the noise from single-molecule emission spectra in sSMLM. First, as the signals of spatially adjacent pixels are closely correlated while the noise is independent, the sub-images sampled from each raw spectral image can serve as the training input and target. This scheme overcomes the difficulty of acquiring ground truth spectral images in sSMLM. Second, we quantitatively evaluated the denoising performance of Spec2Spec on simulated data, suggesting that Spec2Spec can achieve 6-fold improvement in the SNR and 3-fold enhancement in the structure similarity index measure (SSIM). Finally, we applied Spec2Spec to experimental sSMLM data of tubulin and clathrin to validate its ability in multi-color sSMLM imaging.

Methods

Experimental sSMLM system

A home-built sSMLM system was designed to simultaneously capture the spatial position and full spectra of fluorescence emission from single molecules. The schematic of the sSMLM setup is shown in Fig. 1a. Continuous-wave laser illumination (642 nm, 1100 mW, CNI) was used for excitation, and the excited fluorescence signals from the sample were collected by a silicone immersion objective (UPLSAPO60XS2, Olympus). After passing through a tube lens (TL, SWTLU-C, Olympus), the collected fluorescence was guided by a mirror (M) and an entrance slit (S), which restricted the imaging field of view. Then, the fluorescence was split into spatial and spectral channels with a standard 50 : 50 beam splitter (BS). The transmitted fluorescence was directed toward the camera by two reflecting mirrors (BB1-E02-10, Thorlabs) to record the spatial images. Besides, the reflected fluorescence passed through a 30° dispersing prism (43–649, Edmund) to record the corresponding spectrally dispersed images. We then simultaneously acquired the spatial and spectral images from the different regions in a scientific complementary metal oxide semiconductor (sCMOS) camera (Dhyana 400BSI V2, Tucsen).

We calibrated the sSMLM module using a set of lasers with emission peaks of 488, 556, and 647 nm. Besides, to characterize the transformation of spatial and spectral coordinates, we imaged fluorescent beads and performed the affine transformation. To gain the full spectrum of each single-molecule localization event, we first used Thunder-STORM[22] to obtain the coordinates of each localization event in the spatial domain, which subsequently served as the references to extract the corresponding spectral images. In addition, median filtering was employed to remove constant background patterns.

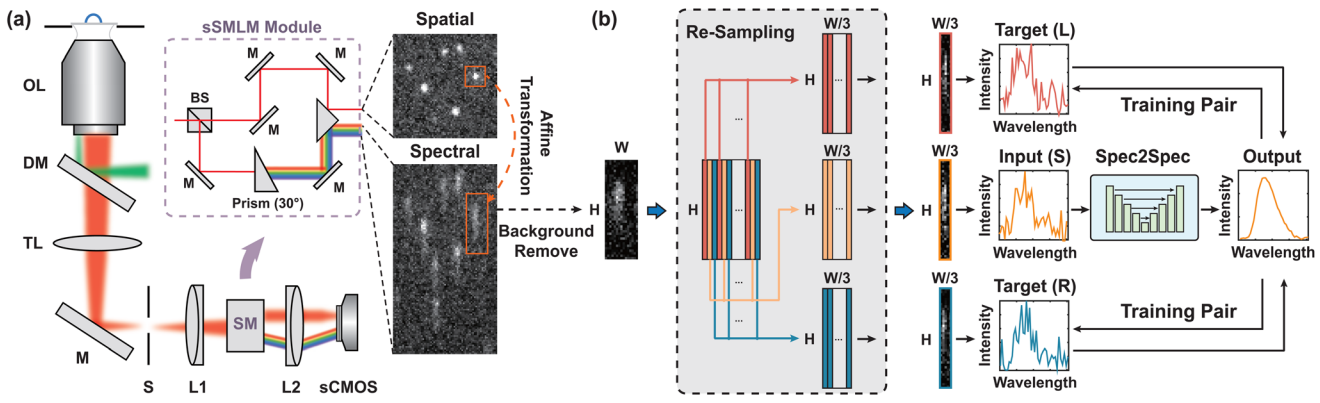

Fig. 1 Experimental sSMLM system and the proposed Spec2Spec framework. (a) Schematic of the sSMLM experimental system. The spectral module is responsible for dividing the emitted photons into two channels, which provide spatial and spectral information, respectively; (b) workflow of the self-supervised spectral denoising in sSMLM, including the re-sampling and re-integrating process from each raw spectral image to generate training pairs and the optimization process of the Spec2Spec network parameters.

This journal is © The Royal Society of Chemistry 2024

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Proposed re-sampling and re-integrating strategy

To generate spectrum-to-spectrum training pairs, we proposed a re-sampling and re-integrating scheme as depicted in Fig. 1b. More specifically, for each input spectral image with $H \times W$ pixels, we divided it into three sub-images with $H \times W/3$ pixels using an interval sampling strategy. To ensure the integrity of spectroscopic signatures, three adjacent lines with $H \times 1$ pixels were separated in order and concatenated into three sub-images with $H \times W/3$ pixels. Three sub-images were then re-integrated along the direction perpendicular to the dispersion axis to generate the corresponding re-sampled emission spectra, respectively. Since these spectra were derived from the same spectral image, they can be viewed as independent samplings containing the same underlying ground truth signal while having consistent signal intensity and SNR level. Therefore, in the training stage, one spectrum can be used as the input, while the other two serve as targets for optimizing the network parameters. Furthermore, we investigated the impact of different re-sampling strategies on the network performances. As described in the ESI,† the interval sampling strategy with three input sub-images achieves overall best performance compared to the random sampling strategy and other numbers of sub-images.

Network architecture and training strategy

We designed the network architecture of Spec2Spec based on U-Net style, as shown in Fig. 2a, which is composed of a 1D encoder module, a 1D decoder module, and four skip connections from each encoder block to the corresponding decoder block. The 1D encoder module is equipped with four encoder blocks, and each block contains two convolutional layers with 1×3 kernels, a rectified linear unit (ReLU), and a max pooling with strides of two. The 1D decoder module is composed of four decoder blocks, and each block contains a transposed

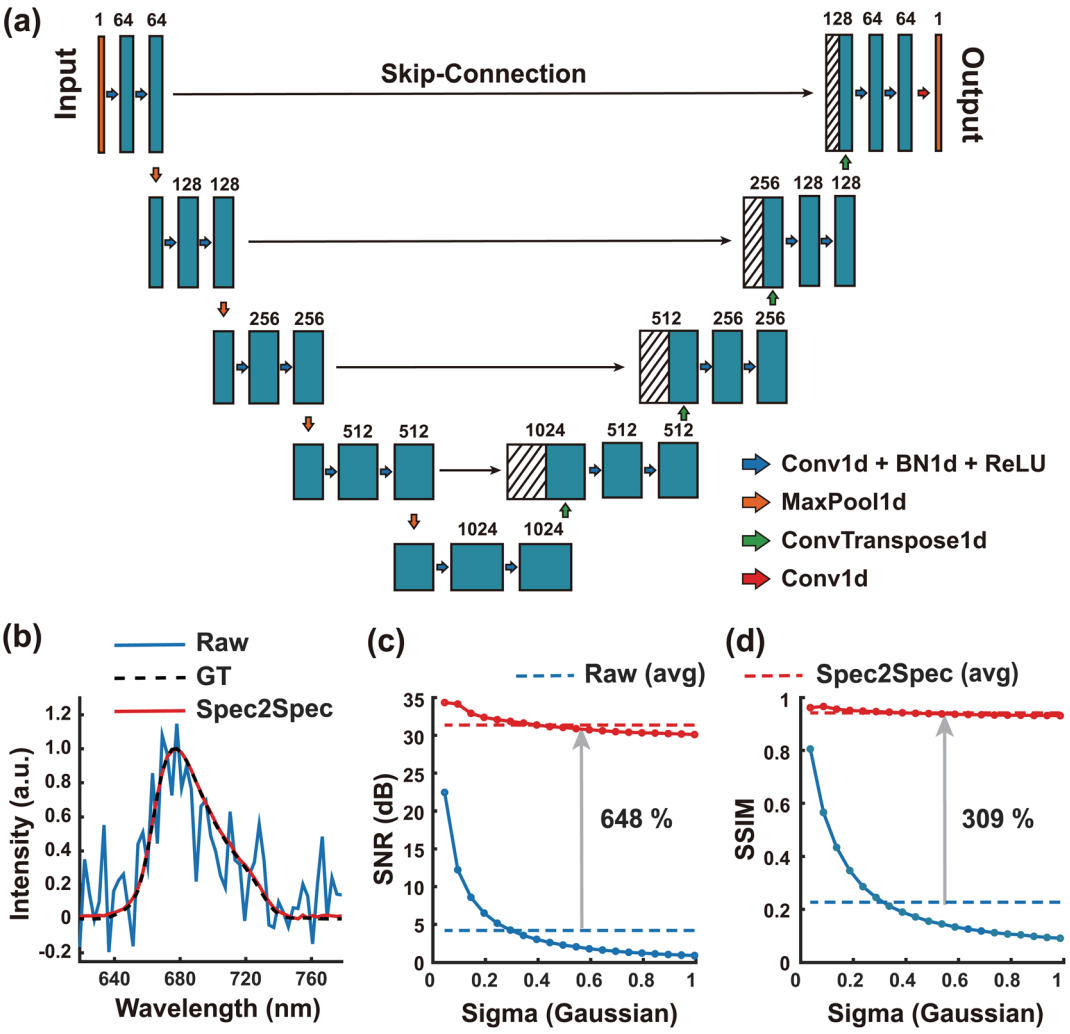

Fig. 2 Network architecture and qualitative and quantitative performance of Spec2Spec on simulated sSMLM data. (a) Network architecture of Spec2Spec. (b) Single-molecule emission spectra before (blue) and after (red) denoising; the black dashed curve is GT. Quantitative evaluation of the performance with (c) the SNR and (d) SSIM over a wide range of input SNRs before and after denoising; dashed lines show average values of the corresponding indicators, and grey arrows show the relative enhancement. Each data point is the average result of 5000 spectra.

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

convolutional layer, two convolutional layers with 1 × 3 kernels, and a ReLU. To accelerate training convergence and alleviate the gradient problem, a batch normalization layer is deployed after each convolutional layer. Skip connections enable the network to retain more high-resolution detail through the fusion of low-level and high-level features.

The network was trained using an Adam optimizer, with a learning rate of 0.0001 and a batch size of 16. We used the arithmetic mean of L1-loss and L2-loss as the loss function to optimize the parameters of the network. We defined the input spectrum as S, and the other two target spectra as L and R, respectively. The loss functions are defined as follows:

$$\mathrm{Loss_L} = |F_{\mathrm{Spec2Spec}}(S) - L|_1 + ||F_{\mathrm{Spec2Spec}}(S) - L||_2{}^2 \quad (1)$$

$$\mathrm{Loss_R} = |F_{\mathrm{Spec2Spec}}(S) - R|_1 + ||F_{\mathrm{Spec2Spec}}(S) - R||_2{}^2 \quad (2)$$

$$\mathrm{Loss_{total}} = \mathrm{Loss_L} + \mathrm{Loss_R} \quad (3)$$

We used PyTorch to construct and implement our network. To speed up the training and testing process, we used NVIDIA GTX 3080Ti graphics processing units (GPU). It took about 2 hours to train a pre-trained model on a typical dataset (50 000 spectra) on a single GPU. Once trained, the testing stage (1000 spectra) only took ~4 seconds.

Evaluation metrics

For a simulated single-molecule emission spectrum X and its ground truth Y, we use SNR and SSIM to evaluate the denoising performance of Spec2Spec.

The SNR measures the pixel-level deviation between two spectra, which can be described as follows:

$$SNR = 10\log_{10}\frac{\|Y\|_2{}^2}{\|X - Y\|_2{}^2} \quad (4)$$

The SSIM measures the perceptual-level similarity between two spectra, which can be expressed as follows:

$$\mathrm{SSIM} = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X{}^2 + \mu_Y{}^2 + C_1))(\sigma_X{}^2 + \sigma_Y{}^2 + C_2)} \quad (5)$$

where (μ_X, μ_Y) and (σ_X, σ_Y) are the means and variances of spectrum X and its ground truth Y, respectively. σ_{XY} is the covariance of X and Y. The two constants C_1 and C_2 are defined as $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$, where $K_1 = 0.01$, and $K_2 = 0.03$. L represents the dynamic range of pixel values, namely $L = 65\,535$.

Sample preparation

Cells were seeded on clean glass coverslips 2 days before fixation to reach a confluency of about 50–70% on the day of fixation. COS-7 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 100 U mL^{-1} penicillin–streptomycin solution at 37 °C with 5% CO_2. Before further processing, the growth medium was aspirated and samples were rinsed twice with PBS to remove dead cells and debris.

For two single-dye-labeled microtubule samples using either Alexa Fluor 647 (AF647) or Dyomics CF660C (CF660C), cells were first washed three times with PBS prewarmed to 37 °C and fixed with 4% paraformaldehyde, 0.1% glutaraldehyde and 0.2% Triton X-100 diluted in PBS for 15 min. After that, the cells were washed with PBS three times and permeabilized and blocked with blocking buffer (5% BSA and 0.1% Triton X-100 in PBS) for 1 h while gently rocking. Cells were incubated with mouse monoclonal anti-α-tubulin antibody in blocking buffer at room temperature for 1.5 h. After washing with PBS three times, cells were incubated with AF647 or CF660C goat anti-mouse IgG for 1 h. Samples were washed three times with PBS and stored in PBS at 4 °C until imaging.

For dual-color labelling (AF647-labeled tubulin and CF660C-labeled clathrin), the cells were first rinsed with PBS. Then, the cells were immobilized with 0.1% glutaraldehyde and 3% paraformaldehyde in PBS at room temperature for 10 min. The reduction process was conducted in PBS with 0.1% sodium borohydride for 7 min. Following washing with PBS, cell permeation was performed with 0.2% Triton X-100 in PBS for 7 min. Samples were blocked with 5% BSA and 0.05% Triton X-100 in PBS at room temperature for 90 min, and then incubated with primary antibodies (rabbit anti-clathrin heavy chain, mouse anti-α-tubulin) in blocking buffer (0.05% Triton X-100, 5% BSA in PBS) for 60 min. The cells were then washed 3 times with PBS, each time for 5 min. Secondary antibodies were labeled with AF647 and CF660C. Next, the cells were washed with PBS 5 times and fixed with 4% paraformaldehyde in PBS for 10 min.

Before sSMLM imaging, coverslips were mounted in 200 μL of blinking buffer (50 mM Tris, pH 8, 10 mM NaCl, 10% (w/v) D-glucose, 35 mM 2-mercaptoethylamine, 500 μg mL^{-1} GLOX, 40 μg mL^{-1} catalase, 2 mM COT).

Dual-color sSMLM reconstruction

First, we used Thunder-STORM to obtain the spatial coordinates of each localization event in the spatial domain, which subsequently served as the references to extract the corresponding spectral images. Simultaneously, we recorded the spatial position corresponding to each spectral image. Then, after Spec2Spec denoising and classification using principal component analysis (PCA) and K-means, we divided each spectrum (including its spatial position) into the corresponding dye categories based on the corresponding label. Finally, for better visualization, the reconstructed super-resolution images were rendered with pseudo-color and their contrast and brightness were manually adjusted to make structures in the image as clear as possible. Each localization point was convolved with a Gaussian kernel, and its kernel size was set based on localization precision provided by the Thunder-STORM software. Data pre-processing, data post-processing, and the rendering process are performed on MATLAB R2020b.

Results and discussion

Quantitative evaluations on simulated sSMLM data

To demonstrate the denoising performance of the proposed Spec2Spec framework, quantitative evaluations were first per-

 This journal is © The Royal Society of Chemistry 2024

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

formed on a simulated spectral dataset. We used the average of 20 000 experimentally acquired single-molecule spectral images of AF647 as the noise-free image and the corresponding emission spectrum was calculated to serve as the GT. Then, we applied different levels of Poisson–Gaussian noise to generate spectral images with different SNRs. Percentage normalization was only applied in the inference phase for computing performance metrics, not in the training phase. Fig. 2b shows an example of a single spectrum before and after denoising, suggesting that the result of Spec2Spec is in high agreement with GT. Quantitatively, we computed SNR and SSIM metrics to evaluate the denoising performance of the network. Moreover, we applied different levels of Gaussian noise to explore the denoising performance of the network over a wide range of input SNRs, including some extremely low SNR conditions. We first performed percentile-normalization for the GT spectrum, and then added different levels of Gaussian noise by controlling the standard deviation of the Gaussian noise, named sigma. We found that the average SNR and SSIM was improved by ~648% (4.2 dB *versus* 31.4 dB) and ~309% (0.23 *versus* 0.94), respectively, as shown in Fig. 2c and d, respectively. Furthermore, we investigated the impact of different spectral resolutions on the network denoising performance with simulated data (see details in the ESI†). Surprisingly, the advantages of Spec2Spec are even more pronounced when the sSMLM system has higher spectral resolution, which further demonstrates the generalization of the Spec2Spec framework.

Performance of Spec2Spec on the experimental sSMLM data

We applied Spec2Spec to the experimental sSMLM data of AF647-labeled tubulin and CF660C-labeled clathrin to validate its ability in restoring emission spectra. The experimentally collected single-molecule emission spectra and denoised spectra are shown in Fig. 3a. Affected by noise, the distribution of spectral centroids (SCs) of different dye molecules exists as a heavily overlapped region, as shown in Fig. 3b, causing difficulty in classifying spectra using a conventional SC-based method, whereas the Spec2Spec framework can effectively remove noise from raw spectra, thereby reducing the impact of noise on the SC distributions, as shown also in Fig. 3b. We further combined PCA and K-means clustering to classify the recovered spectral data, and the corresponding clustering result is shown in Fig. 3c. Compared to SC obtained by calculating the intensity-weighted average of wavelength, PCA can provide more distinguishable low-dimensional features for classification. Furthermore, the usage of K-means, an unsupervised clustering algorithm, can alleviate misclassifications and gain a higher UR.

Table 1 quantitatively compares the improvement of the classification accuracy and utilization ratio (UR) among different methods, where UR is defined as the number of spectra allocated into each dye category divided by the total number of spectra. To obtain the ground truth, we prepared two single-dye-labeled samples using AF647 and CF660C, respectively. We then independently acquired sSMLM data from each sample. Finally, we mixed all collected spectra and labeled each spectrum according to its origin. In the conventional SC-based method, spectral windows (SWs) are predefined based on the SC distribution for separating spectra to each dye category. If SWs were set to be 687–689 nm for AF647 and 692–694 nm for CF660C, a classification accuracy of 82.6% and a UR of 36.9% can be obtained. As expected, the SC-based method inevitably discards spectra outside SWs, resulting in a low UR, which is attributed to more discontinuous structures in reconstructed sSMLM images. Expanding SWs increases the UR to 100%; however, more spectral cross-

Table 1 Comparison of the classification accuracy and data utilization ratio

Data	Method	SWs [nm]	Accuracy	UR
Raw	SC-based	687–689, 692–694	82.6%	36.9%
		<690, >690	74.6%	100.0%
Spec2Spec		<690, >690	92.6%	100.0%
	PCA & K-means	—	94.6%	100.0%

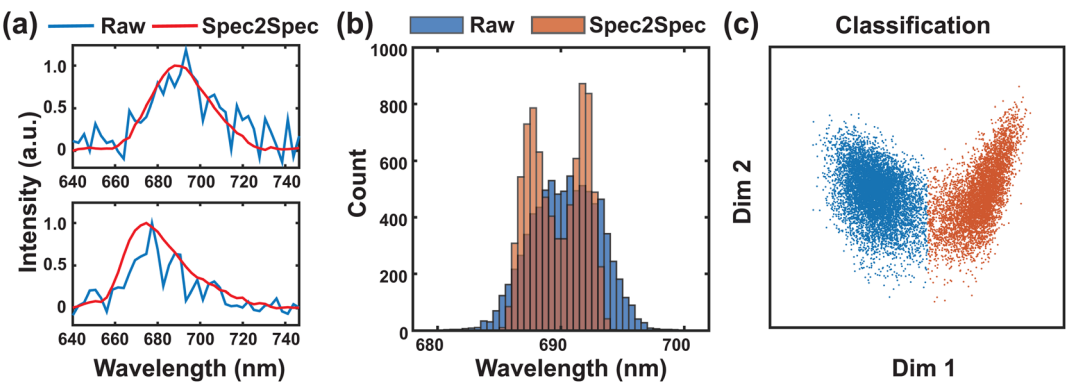

Fig. 3 Application of Spec2Spec on the experimental sSMLM data. (a) Experimentally obtained single-molecule emission spectra from AF647-labeled tubulin and CF660C-labeled clathrin before (blue) and after (red) denoising; (b) histogram of the spectral centroid distributions of mixed AF647 and CF660C spectra before and after denoising; and (c) scatter plots showing the clustering result using denoised data, respectively.

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

talk will be observed in sSMLM images due to the reduced accuracy. In contrast, after processing by Spec2Spec, the SC-based method can achieve a classification accuracy of 92.6% and a UR of 100%. Using PCA and K-means can further improve the classification accuracy, and more importantly, eliminate the difficulty in selecting appropriate SWs to balance the classification accuracy and UR. Besides, we expect that more-advanced feature extraction and clustering algorithms will further improve the classification accuracy in multi-color sSMLM imaging, especially when dealing with more colors and dyes with high spectral overlap.

Dual-color sSMLM imaging using Spec2Spec

Furthermore, we experimentally demonstrate the benefits of Spec2Spec in dual-labeled sSMLM imaging. Fig. 4a shows a conventional SMLM image reconstructed from the spatial channel. Here, we pre-defined two SWs, 683–690 nm for AF647 and 690–697 nm for CF660C, resulting in a UR of

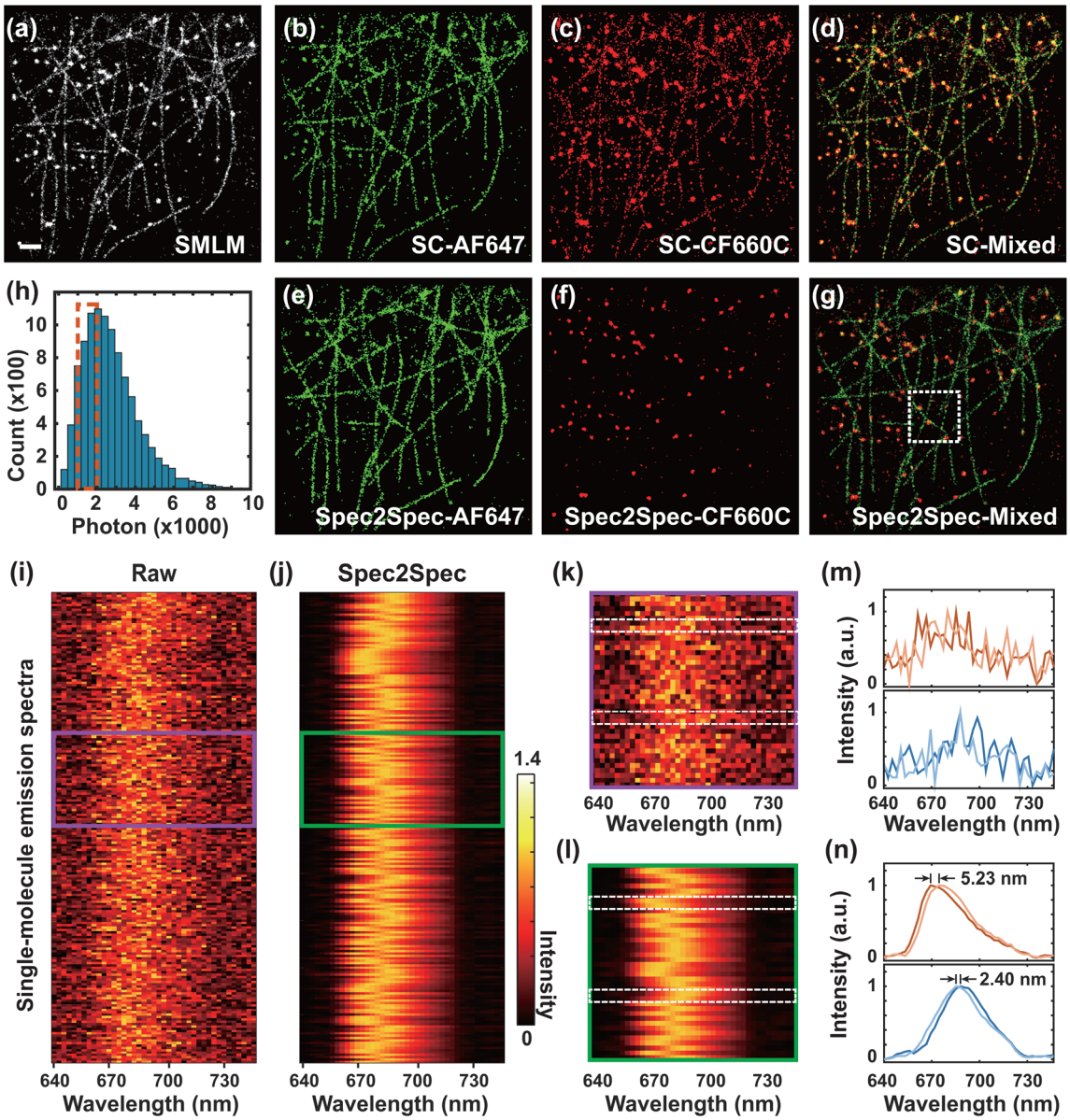

Fig. 4 The improvement of Spec2Spec in dual-color sSMLM imaging and observation of the minute spectral variations from the same type of dye molecules. (a) Reconstructed SMLM image of AF647-labeled tubulin and CF660C-labeled clathrin in a fixed COS-7 cell. Scale bar: 1 μm. (b) Separated tubulin image and (c) clathrin image processed by the SC-based method and (d) their merged dual-color sSMLM image. (e) Separated tubulin image and (f) clathrin image processed by the proposed method and (g) their merged dual-color sSMLM image. (h) Histogram of the photon number distributions of spectra from the region of interest (ROI) as highlighted in (g); (i) and (j) mixed emission spectra before and after denoising, respectively; (k) and (l) a magnified view of boxed regions in (i) and (j), respectively; (m) raw spectra from the white dashed boxes in (k) showing AF647 (upper) and CF660C (lower) spectra from different localization events. (n) Their corresponding denoised spectra show distinguishable spectral heterogeneity.

This journal is © The Royal Society of Chemistry 2024

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

98.6% in a SC-based classification. From Fig. 4b and c, reconstructed images of two dye channels show many misclassifications of tubulin and clathrin, further causing heavy cross-talk in reconstructed images (Fig. 4d). In contrast, the Spec2Spec framework can effectively suppress the noise and improve the classification accuracy. As shown in Fig. 4e and f, the distribution of tubulin and clathrin can be clearly separated into the AF647 and CF660C channels, respectively. Meanwhile, overlaying the two channels gives the reconstructed dual-color super-resolution image, as shown in Fig. 4g. This result indicates that the Spec2Spec framework can effectively reduce cross-color contamination caused by noise in multi-color sSMLM imaging.

Fig. 4h shows the histogram of the photon number of mixed AF647 and CF660C spectra (~9000) obtained from the boxed region (Fig. 4g). Fig. 4i and j displays 200 single-molecule emission spectra with 1000–2000 detected photons [highlighted in Fig. 4e]. Fig. 4k and l shows the magnified view of the boxed region in Fig. 4i and j, respectively. As raw spectra distorted by noise, we hardly determine the categories of emission spectra. Remarkably, after removing the noise, spectra of different dye molecules can be readily distinguishable on the basis of spectral shape (Fig. 4j). Additionally, Spec2Spec can restore the spectral variations caused by the underlying fluorescence heterogeneity.[11,20] As shown in Fig. 4m and n, the minute spectral variations from the same type of dye molecules, which are drowned out by noise and are difficult to distinguish from each other in the raw data, can be clearly resolved in the denoised data. After computing their spectral peaks, we observed a shift of 5.23 nm for two AF647 single-molecule spectra and 2.40 nm for two CF660C single-molecule spectra, respectively. These results prove that Spec2Spec can not only improve the classification accuracy in multi-color sSMLM, but is also able to offer more convincing results in functional sSMLM.

Conclusions

In this work, we propose Spec2Spec, a self-supervised deep-learning-based spectral denoising framework for effectively removing noise from the emission spectra of single-molecule localization events in sSMLM. The proposed re-sampling and re-integrating strategy eliminates the need for ground truth spectra in supervised learning when constructing spectrum-to-spectrum training pairs. Besides, we validate the denoising performance on simulated data, suggesting that Spec2Spec offers 6-fold improvement in SNR and 3-fold enhancement in SSIM. We further employ the Spec2Spec framework for experimental dual-color sSMLM data, proving that the combination of Spec2Spec, PCA, and K-means can achieve more than 20% improvement in classification accuracy with data UR of 100%. Therefore, we envision that the proposed deep-learning-assisted method offers a novel avenue for facilitating multiplexed and functional super-resolution imaging using sSMLM.

Data and code availability

The PyTorch codes of Spec2Spec, representative pre-trained models, as well as example data for testing, will be publicly available at **https://github.com/FDU-donglab/spec2spec**.

Conflicts of interest

Biqin Dong has a financial interest in Lishi Intelligent Science & Technology (Shanghai) Co., Ltd, which, however, did not support this work. All other authors declare no competing interests.

Acknowledgements

We thank Prof. Jiong Ma for providing the cell culture facility. This work was supported by the National Key R&D Program of China (2022YFF0708700), the Shanghai Basic Research Special Zone Program (22TQ020), the Natural Science Foundation of Shanghai (22ZR1404300), the Shanghai Science and Technology Innovation Action Plan (22S31905500), the Fudan University-CIOMP Joint Fund (FC2020-002), and the Medical Engineering Fund of Fudan University (yg2021-032, yg2022-31, and yg2022-2).

References

1 J. W. Lichtman and J. A. Conchello, *Nat. Methods*, 2005, **2**, 910–919.
2 Y. M. Sigal, R. Zhou and X. Zhuang, *Science*, 2018, **361**, 880–887.
3 B. Huang, H. Babcock and X. Zhuang, *Cell*, 2010, **143**, 1047–1058.
4 M. J. Rust, M. Bates and X. Zhuang, *Nat. Methods*, 2006, **3**, 793–795.
5 E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess, *Science*, 2006, **313**, 1642–1645.
6 B. Dong, L. M. Almassalha, Y. Stypula-Cyru, B. E. Urban, J. E. Chandler, T. Q. Nguyen, C. Sun, H. F. Zhang and V. Backman, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 9716–9721.
7 S. Moon, R. Yan, S. J. Kenny, Y. Shyu, L. Xiang, W. Li and K. Xu, *J. Am. Chem. Soc.*, 2017, **139**, 10944–10947.
8 Z. Zhang, S. J. Kenny, M. Hauser, W. Li and K. Xu, *Nat. Methods*, 2015, **12**, 935–938.
9 M. Bates, B. Huang, G. T. Dempsey and X. Zhuang, *Science*, 2007, **317**, 1749–1753.
10 H. Shroff, C. G. Galbraith, J. A. Galbraith, H. White, J. Gillette, S. Olenych, M. W. Davidson and E. Betzig, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20308–20313.
11 B. Dong, L. Almassalha, B. E. Urban, T. Q. Nguyen, S. Khuon, T. L. Chew, V. Backman, C. Sun and H. F. Zhang, *Nat. Commun.*, 2016, **7**, 1–8.

Open Access Article. Published on 07 February 2024. Downloaded on 5/22/2026 2:54:30 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

12 B. Dong, B. T. Soetikno, X. Chen, V. Backman, C. Sun and H. F. Zhang, *ACS Photonics*, 2017, **4**, 1747–1752.
13 B. Brenner, C. Sun, F. M. Raymo and H. F. Zhang, *Nano Convergence*, 2023, **10**, 14.
14 R. Yan, B. Wang and K. Xu, *Curr. Opin. Chem. Biol.*, 2019, **51**, 92–97.
15 R. Yan, S. Moon, S. J. Kenny and K. Xu, *Acc. Chem. Res.*, 2018, **51**, 697–705.
16 K. H. Song, B. Dong, C. Sun and H. F. Zhang, *Rev. Sci. Instrum.*, 2018, **89**, 123703.
17 K. H. Song, Y. Zhang, B. Brenner, C. Sun and H. F. Zhang, *Light: Sci. Appl.*, 2020, **9**, 92.
18 S. K. Gaire, Y. Zhang, H. Li, R. Yu, H. F. Zhang and L. Ying, *Biomed. Opt. Express*, 2020, **11**, 2705.
19 Z. Zhang, Y. Zhang, L. Ying, C. Sun and H. F. Zhang, *Opt. Lett.*, 2019, **44**, 5864.
20 H. Manko, Y. Mély and J. Godet, *Small*, 2023, **19**, 33.
21 Y. Zhang, K.-H. Song, B. Dong, J. L. Davis, G. Shao, C. Sun and H. F. Zhang, *Appl. Opt.*, 2019, **58**, 2248.
22 M. Ovesný, P. Křížek, J. Borkovec, Z. Švindrych and G. M. Hagen, *Bioinformatics*, 2014, **30**, 2389–2390.

This journal is © The Royal Society of Chemistry 2024