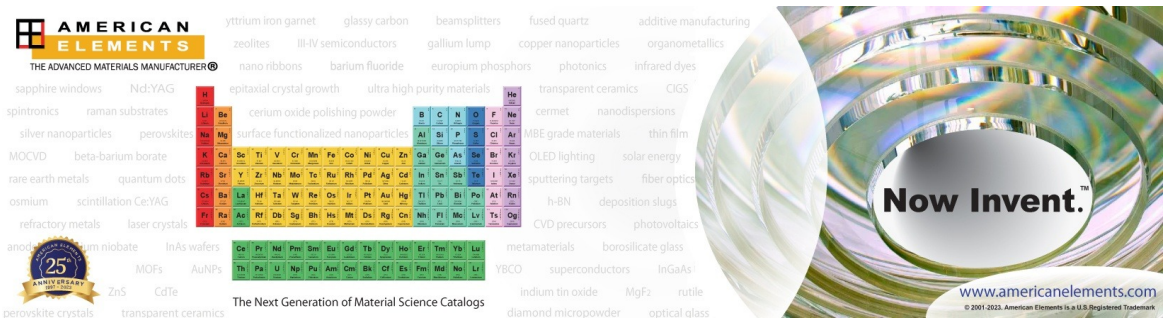
APL Photonics

RESEARCH ARTICLE | SEPTEMBER 05 2023

Deep learning-enhanced single-molecule spectrum imaging

Hao Sha ; Haoyang Li ; Yongbing Zhang ; Shangguo Hou

APL Photonics 8, 096102 (2023)
https://doi.org/10.1063/5.0156793



Deep learning-enhanced single-molecule spectrum imaging

Cite as: APL Photon. 8, 096102 (2023); doi: 10.1063/5.0156793
Submitted: 3 May 2023 • Accepted: 15 August 2023 •
Published Online: 5 September 2023

Hao Sha,[1,2] Haoyang Li,[2,3] Yongbing Zhang,[1,a)] and Shangguo Hou[2,a)]

AFFILIATIONS

[1] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Guangdong 518006, China
[2] Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518055, China
[3] Center for Ultrafast Science and Technology, School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

[a)] **Authors to whom correspondence should be addressed:** ybzhang08@hit.edu.cn and shangguo.hou@szbl.ac.cn

ABSTRACT

Fluorescence is widely used in biological imaging and biosensing. Rich information can be revealed from the fluorescence spectrum of fluorescent molecules, such as pH, viscosity and polarity of the molecule's environment, and distance between two FRET molecules. However, constructing the fluorescence spectrum of a single fluorescent molecule typically requires a significant number of photons, which can suffer from photobleaching and, therefore, limit its potential applications. Here, we propose a deep learning-enhanced single-molecule spectrum imaging method (SpecGAN) for improving the single-molecule spectrum imaging efficiency. In SpecGAN, the photon flux required to extract a single-molecule fluorescence spectrum can be reduced by 100 times, which enables two orders of magnitude higher temporal resolution compared to the conventional single-molecule spectrometer. The concept of SpecGAN was validated through numerical simulation and single Nile Red molecule spectrum imaging on support lipid bilayers (SLBs). With SpecGAN, the super-resolution spectrum image of the COS-7 membrane can be reconstructed with merely 12 000 frames of single-molecule localization images, which is almost half of the previously reported frame count for spectrally resolved super-resolution imaging. The low photon flux requirement and high temporal resolution of SpecGAN make it a promising tool for investigating the molecular spectrum dynamics related to biological functions or biomolecule interactions.

I. INTRODUCTION

Since the invention of optical microscopy, it has become an essential tool in life science research due to its non-invasive nature and specificity. Over the past several decades, the emerging super-resolution fluorescence microscopy has overcome limits of optical diffraction and pushed the spatial resolution up to several nanometers,[1–5] which can reveal unprecedented fine structure details of organelles in cells. In fluorescence microscopy, the fluorescence intensity, i.e., fluorescence photon emission rate, is usually used to build a morphology image of the sample. Aside from the intensity information, the fluorescence signal also has many other characteristics, such as fluorescence spectrum,[6] fluorescence lifetime,[7] and fluorescence polarization.[8] These characteristics indicate the microenvironment biophysical properties of the fluorescent molecules, such as pH, viscosity, and polarity.

In recent years, various spectrum imaging methods have been developed. The fluorescence spectrum can provide additional information that cannot be acquired by intensity imaging, such as molecular structure,[9,10] membrane potential,[11] spatial arrangement of biomolecules,[12,13] and the microenvironment surrounding molecules.[14–16] Spectrum imaging has been applied to various research fields. For instance, with fluorescence spectrum imaging, the stereoisomers of borondipyrromethene (BODIPY) can be well resolved and their photophysical properties and excitation dynamics can be further investigated.[10] Combining the pH-sensitive photochromic dye and spectrum imaging is able to disclose the alternative route of nanoparticles reaching lysosomes and to observe the pH changes in real-time.[16] The surface hydrophobicity of α-synuclein (αS), a Parkinson's disease-related protein, can also be characterized with the fluorescence spectrum of solvatochromic dyes.[17] Usually, the fluorescence spectrum is measured using a fluorescence

spectrophotometer. However, the fluorescence spectrophotometer is not applicable to cellular experiments, which precludes its further cellular applications.

To acquire the morphological and spectral information of cells simultaneously, spectrally resolved super-resolution microscopy (SR-SRM) has been proposed in recent years.[15,18,19] This method has been applied to reveal the nanoscale spectral features of different dyes on cell membranes.[15,18] In addition, the polarity heterogeneity of the plasma membrane and organelle membranes with different cholesterol levels was revealed with Nile Red labeling.[20] On the other hand, owing to the limited fluorescence emission rate and total emission photons of a single molecule, obtaining a highly precise fluorescence spectrum of a single molecule in a short collection time period is challenging. In particular, in intracellular imaging, the presence of autofluorescence further deteriorates the spectrum imaging. To obtain an accurate fluorescence spectrum, multiple spectra are usually collected and averaged. For SR-SRM imaging of mammalian cells, reconstructing a spectral mapping image typically requires 15 000–30 000 frames of localization image, resulting in a poor temporal resolution and precluding the spectral dynamics analysis.[21]

Deep learning technology has become increasingly popular in the realm of microscopic imaging, with a broad range of applications, including denoising,[7,22] modal conversion,[23,24] resolution enhancement,[25] downstream task analysis,[26,27] and high-speed imaging.[28–30] In single-molecule spectral imaging, deep learning has also been employed to address the problem of spectral noise. Convolution neural networks (CNNs) have been found to be highly effective in feature extraction, enabling the accurate classification of three-color labeled cell samples with low rates of molecular misidentification.[31] However, this approach was specifically designed for classifying different dyes and cannot be applied to identify the continuous spectral changes in different environments. In image generation tasks, the generative adversarial network (GAN)[32] has emerged as a popular generative model. While originally proposed for unsupervised learning, GAN has been shown to be effective in supervised tasks, such as super-resolution and semantic segmentation.[33,34]

In this paper, we propose a novel deep learning-based single-molecule spectrum imaging method called SpecGAN, which is tailored to denoise single-molecule spectra and accurately identify the spectral characteristics of environment-sensitive fluorescent dyes. The SpecGAN is validated with simulation, support lipid bilayers (SLBs) imaging, and cell membrane super-resolution spectrum imaging. Notably, implementing SpecGAN does not require modifications to the current spectrally resolved super-resolution microscopy (SR-SRM) system, and it holds promising potential for real-time investigations of single-molecule spectral dynamics owing to its low photon flux requirement to construct the spectrum. The single molecule spectral denoising processes is shown in Fig. S14.

II. RESULTS

A. Workflow of SpecGAN

Generative Adversarial Networks (GANs) consist of two main components: a generator and a discriminator. The generator is responsible for creating synthetic data instances by learning the underlying distribution of the real dataset, while the discriminator, on the other hand, is a neural network designed to distinguish between real and synthetic data instances. Similar to the GAN model, SpecGAN contains a generator and discriminator. The generator of SpecGAN takes the 1D spectral data with noise into a UNet-based network and outputs clean spectrum data. The UNet is a convolutional neural network (CNN) architecture designed for semantic segmentation tasks. The network architecture consists of a contracting path that captures context and a symmetric expanding path that enables pixel-to-pixel mapping.[35] In UNet, the encoder uses convolution layers and maximum pooling layers for feature extraction, while the decoder uses binary linear interpolation and convolution layers for upsampling and information recovery, respectively. For the discriminator of SpecGAN, auxiliary classification tasks[36] are adopted to enhance the model's focus on environmental information. In addition to determining whether the spectral curve is generated from the generator, the discriminator is also responsible for classifying the environments represented by the spectrum. More details about SpecGAN can be found in Sec. IV.

To validate the performance of SpecGAN, we built a single-molecule spectrum imaging microscopy [Figs. 1(a) and S1]. The fluorescence is directed to two paths using a beam splitter. The fluorescence in path 1 is used for recording the location information of molecules, while path 2 is used for recording the fluorescence spectrum. An example of fluorescent bead spectrum imaging is shown in Fig. 1(a). As previously mentioned, to obtain the fluorescence spectrum of single molecules, it is necessary to average multiple spectral data from the same environment to counteract the influence of noise. As a demonstration, we simulated the average fluorescence spectrum with a different amount of fluorescent spectrum data [Fig. 1(b)]. It can be seen that several hundred single-molecule spectra data are required to produce a well-resolved averaged spectrum. In this study, we choose Nile Red as a fluorescence probe to evaluate the performance of SpecGAN. Nile Red is a widely used fluorescent marker to monitor the changes of polarity in spectrally resolved super-resolution imaging because it exhibits a significant blue shift in the lipid order (Lo) phase compared to the lipid disorder (Ld) phase,[37] as shown in Fig. 1(c).

The flowchart of spectrum imaging with SpecGAN is illustrated in Fig. 1(d). Due to the low signal-to-noise ratio (SNR) of the raw spectrum of a single molecule, it is challenging for the SpecGAN model to reach stabilization gradually during training. Thus, before feeding the raw spectral data into SpecGAN, the variational modal decomposition (VMD) algorithm is used to adaptively decompose a spectral curve into several intrinsic mode functions (IMFs) and a residual item.[38] The residual item of VMD, which contains the inherent feature of the raw spectral curve and is always smooth due to the removal of high-frequency noise components, is used as the input for the SpecGAN. For the construction of training datasets, the average of all spectral curves is taken as the ground truth. The workflow of data preprocess is illustrated in Fig. S4. To further validate whether the main information of the ground truth is contained in the residual item after VMD decomposition, we reconstructed the averaged spectral curve with raw data and residual item in different solvents. It can be seen that the average curve calculated using the residual data is similar that calculated using the original data, as shown in Fig. S13. This workflow aims to generate a fluorescence

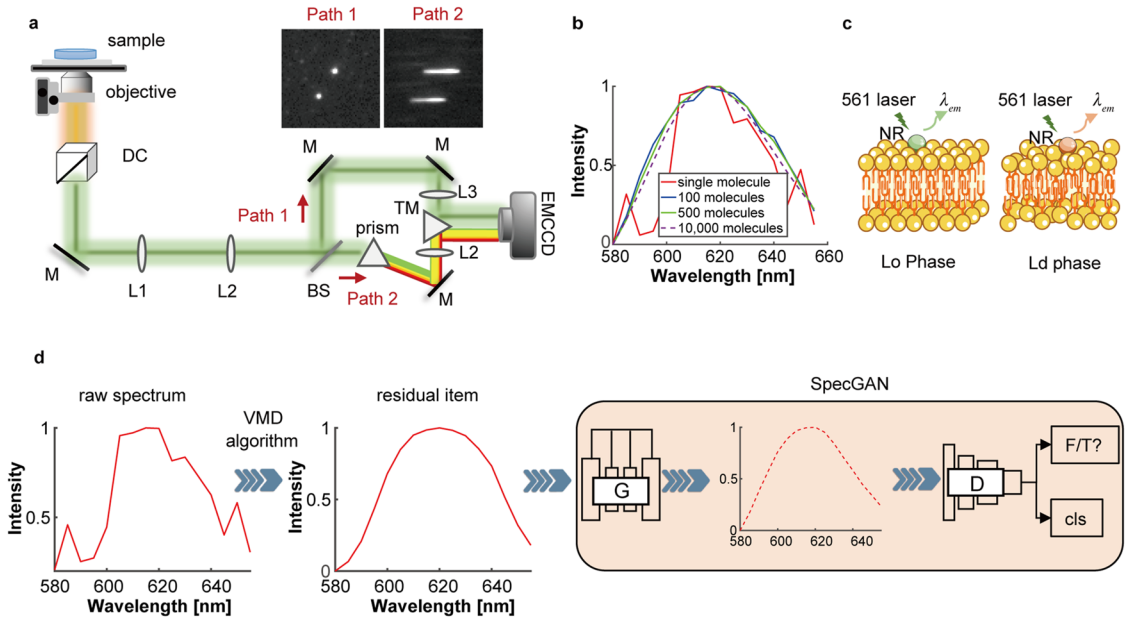


FIG. 1. Experimental setup and data process overview. (a) Schematic diagram of the single-molecule spectrum imaging microscopy setup. The inset shows the spectra and position images of the fluorescent bead. (b) Averaged spectra with different amounts of recorded molecules. (c) Schematic of Nile Red in different lipid environments. The emission spectra peak of Nile Red (NR) change with the lipid orders. Lo: lipid order; Ld: lipid disorder. (d) Overview of the SpecGAN structure, where the residual item is extracted from the raw spectrum by the VMD algorithm, and the SpecGAN outputs the denoised spectral curve of the fed residual one.

spectrum similar to the ground truth spectrum using the fewest single-molecule spectral data.

B. SpecGAN performance evaluation with simulated data

To assess the effectiveness of SpecGAN, we conducted experiments on three different types of datasets, namely simulation data, SLBs data, and cellular data. The VMD decomposition-based preprocess is applied to all datasets before inputting to the SpecGAN model. We trained our proposed SpecGAN on the simulation and SLBs datasets. The performance of SpecGAN is first tested with the simulation data based on the spectral shifts of Nile Red in different solvents.[39] The Nile Red spectra in different solvents (acetone, CHCl3, C3H8O, MeOH, and EtOH) are measured with a fluorescence spectrophotometer (Lumina, Thermo Scientific) and used as the ground truths during the training stage [Fig. 2(a)]. The fluorescence centroid in different solvents show different spectral characters, with the blue shift being most pronounced in chloroform and the red shift in MeOH [Fig. 2(b)]. In addition, these solvent categories are further labeled as the environmental classes for the supervision of the auxiliary discriminator. During the imaging process, noise is inevitable. To reduce the differences between simulation and real-world acquisition conditions, these ground truths are added with Gaussian, Poisson, and salt and pepper noises (Fig. S15). The simulation dataset comprises ~50 000 spectral curves with varying noise levels and spectral centroids. In deep learning, the division of the dataset is important because it determines the effectiveness of model learning and generalization capabilities. The simulation and SLBs datasets are randomly divided into three subsets according to the convention of deep learning method. In this paper, 80% of the data are used for training, 10% of the data are used for validation, and 10% of the data are used for testing. As for the generalization validation, all results shown in this paper are based on the test dataset rather the training dataset.

As shown in Fig. 2(c), the averaged spectrum of ten raw fluorescence spectra significantly deviated from the ground truth and the centroid histogram is dispersed. This result indicates that the random noise in the simulation data may interfere with the analysis of the molecular fluorescence spectrum, resulting in an incorrect spectral centroid measurement. In contrast, the average spectrum of SpecGAN outputs is closer to the ground truth, and the centroid of the spectrum can be easily identified. Correspondingly, the outputs of SpecGAN show smaller variances, indicating that SpecGAN can give a centroid prediction with high confidence [Fig. 2(d)]. We then compared the fluorescence spectrum maxima identification accuracy of SpecGAN output data and raw data. We found that the identified centroids with SpecGAN nearly 100% deviate less than 5 nm from the ground truth. Compared with raw data, SpecGAN output data show a significant accuracy improvement, specifically, with a maximum of 20% accuracy improvement for the spectra in acetone solvent measurement [Fig. 2(e)].

C. Evaluation of specGAN's performance in spectral imaging of SLBs

To demonstrate the biological compatibility of SpecGAN, the SLBs spectrum imaging was conducted. SLBs are artificially

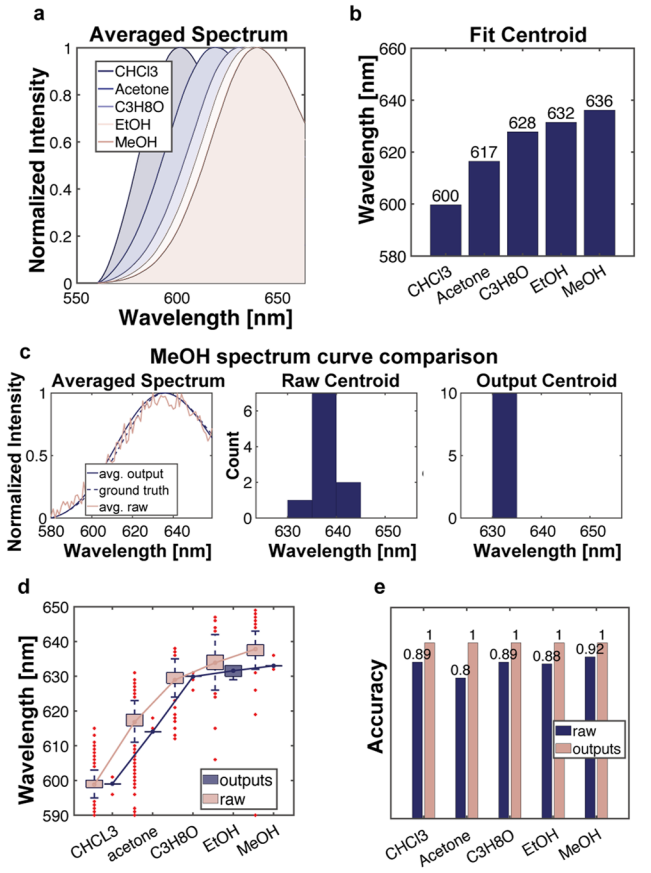


FIG. 2. Performance evaluation of SpecGAN with simulated data. (a) Spectra of Nile Red in different solvents measured using a fluorescence spectrophotometer. (b) The spectral centroids of Nile Red spectra in different solvents. (c) Comparison of raw spectrum with noise and the spectrum output from SpecGAN. Solvent: MeOH. (d) Fluorescence spectrum centroid of Nile Red in different solvents calculated with raw spectra and SpecGAN outputs. (e) Fluorescence spectrum maxima identification accuracy comparison.

constructed, two-dimensional lipid structures that mimic the properties of natural cell membranes. SLBs serve as valuable tools for studying the biophysical and biochemical properties of cell membranes, as well as for investigating processes such as membrane–protein interactions, lipid organization, and the dynamics of membrane components.[8] We made the SLBs by mixing the lipids 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), sphingomyelin (SM), and cholesterol (Chol) in chloroform in different proportions. These proportions were selected to induce the emission spectrum shift of Nile Red. We utilized five distinct mixing ratios of DOPC, SM, and Chol in this study, namely 1:0:1, 1:1:0, 1:1:1, 1:3:1, and 0:1:1, representing the respective proportions of the DOPC, SM, and Chol components. These mixtures are abbreviated as DC, DS, DSC, DSC311, and SC, respectively. To prepare the SLBs, we employed the bicelle adsorption and fusion method (Fig. S2), which facilitates the formation of stable bilayer structures on glass surfaces.[40]

To keep the imaging system consistent, the spectra of Nile Red in five different SLBs compositions were collected using single-molecule spectrum imaging microscopy. Approximately 60 000 single-molecule spectral data points for each category are averaged to create the ground truths [Figs. 3(a) and S6]. For the SLBs Nile Red spectra dataset recorded by single-molecule spectrum imaging microscopy, we found that the noise has a non-negligible effect on signal processing, which may make the deep learning model hard to converge during training. Therefore, it is necessary to exclude the spectral data with significant noise. In the training of SpecGAN, the SLBs dataset is manually pre-screened to exclude poor-quality spectra data. The parameters were empirically selected to produce a stably converged training loss. Specifically, a signal is considered usable if the difference between the centroid of the residual item and the average spectrum is smaller than ±15 nm or the root mean square error (RMSE) is smaller than 0.3. Compared with raw spectra data, the fluorescence spectral centroid localization accuracy of SpecGAN outputs is improved by up to 62% and the variance of the SpecGAN outputs is reduced by about four times (Figs. S7 and S8). Moreover, using SpecGAN, the Nile Red spectrum can be reconstructed with just ten molecular spectral data points, whereas ~500 to 1000 molecules were necessary for reconstructing a similar Nile Red spectrum in raw data. This indicates that SpecGAN can potentially reduce the photon flux requirements by up to 100 times.

SpecGAN has demonstrated strong denoising capabilities when applied to manually screened data. To automate the data

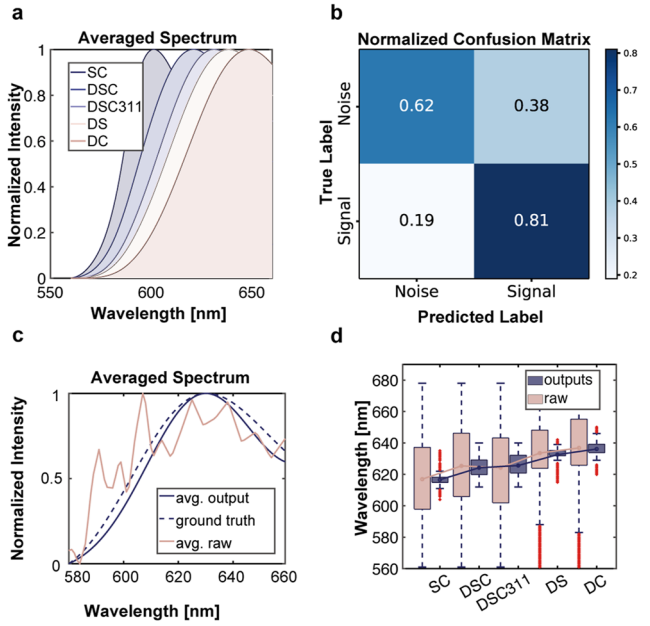


FIG. 3. Performance evaluation of SpecGAN with SLBs data. (a) Averaged spectra of Nile Red in SLBs with different components acquired by single-molecule spectrum imaging microscopy. (b) Normalized confusion matrix of ResNet-based classification model. (c) The comparison between raw spectrum and the output of SpecGAN in DOPC:SM:Chol (3:1:1). (d) Boxplot of centroid calculated by raw spectra and SpecGAN outputs.

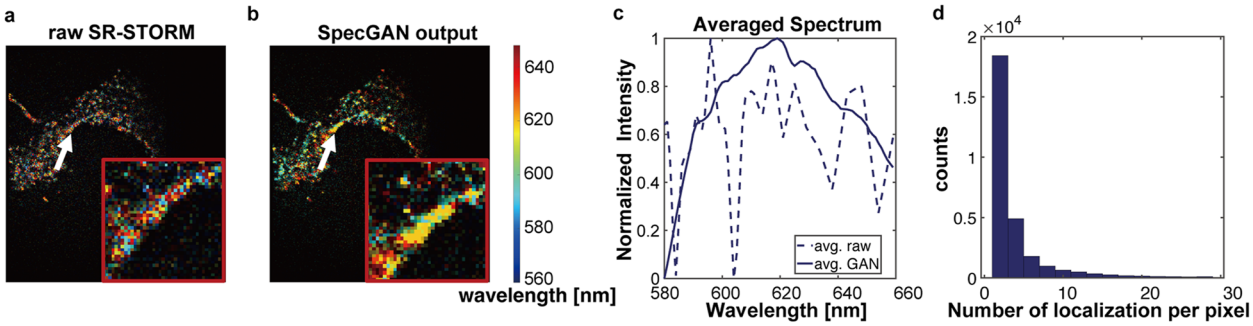


FIG. 4. SpecGAN enhanced super-resolution spectrum imaging. (a) Super-resolution spectrum imaging with raw-STORM. (b) Super-resolution spectrum imaging with SpecGAN. Each detected single molecule is color-coded according to its spectral centroid. The inset is a magnified view of the region indicated by the white arrow. (c) Comparison of the averaged single-molecule spectral curve at the cell membrane pointed by the arrow in (a) and (b). (d) Histogram of number of localization per pixel. Total frames: 12 000.

screening process, we trained a ResNet-based classifier as a replacement. This classifier can automatically determine the signal quality, and the screened data are then fed to the SpecGAN, which has already been trained on manually screened datasets. The normalized confusion matrix of the ResNet-based classification model is displayed in Fig. 3(b). After 150 training iterations, the model achieves a precision and recall rate of 68% and 81%, respectively. Then, we test its denoising capabilities with ten single-molecule spectra data points [Figs. 3(c) and 3(d)]. In the raw SLBs dataset, the centroid cannot be identified from the limited numbers of single molecule spectrum, and the variance is much higher than that of the SpecGAN outputs. Moreover, the wavelength of DSC centroid shows a slight red shift compared to DSC311, which is not consistent with the ground truth. As a comparison, SpecGAN achieves 81% identification accuracy, showing an improvement of 56% over the raw data analysis (Figs. S9–S11). It should be noted that the centroid determined by SpecGAN for the SC category deviates slightly from the average fluorescence spectrum but still has a lower variance compared with raw data. The deviation may be caused by the poor performance of the classifier in the SC dataset.

D. SpecGAN enhanced super-resolution spectrum imaging

The super-resolution spectrum imaging technique or spectrally resolved stochastic optical reconstruction microscopy (SR-STORM) has been developed in recent years.[15,18,19] In SR-STORM, the wavelength corresponding to the extreme point of the spectral curve is color-coded to each pixel and then mapped to the image reconstructed by the single molecule localization algorithm. Incorporating spectral information into super-resolution imaging not only adds an additional dimension for analysis but also enhances localization precision. The SpecGAN has the capability to further improve super-resolution spectrum imaging. Here, we performed a super-resolution spectrum imaging of Nile Red labeled COS-7 cells to demonstrate the performance of SpecGAN. Compared to super-resolution spectral imaging reconstructed using raw data (raw SR-STORM), SpecGAN significantly improves spectral precision [Figs. 4(a) and 4(b) and S12]. The spectral distribution in COS-7 cells appears discontinuous and noisy for raw SR-STORM [Fig. 4(a)], while the SpecGAN output image displays a more continuous and concentrated spectral distribution [Fig. 4(b)]. In addition, we compared the spectrum of a pixel for raw SR-STORM and SpecGAN output and observed that the spectrum in raw SR-STORM is heavily affected by noise, ultimately impacting the accuracy of spectrum maxima localization [Fig. 4(c)]. In contrast, the SpecGAN output yields a smoother spectrum curve, facilitating more precise spectrum maxima localization. In a set of 12 000 localization images, the majority of pixels have fewer than ten localizations [Fig. 4(d)]. As a result, addressing noise interference is crucial for fast spectrally resolved super-resolution imaging. Herein, SpecGAN provides a novel approach to analyze spectral features at the single-molecule level.

III. DISCUSSION AND CONCLUSION

In this study, we presented a novel deep learning-enhanced single-molecule fluorescence spectrum imaging technique called SpecGAN, which allows for the effective extraction of single-molecule spectra with significantly improved signal-to-noise ratio and accuracy. Compared to conventional spectral super-resolution imaging methods, SpecGAN has a high temporal resolution of spectrum imaging. The spectrum imaging speed improvement can be reflected on two aspects. First, for reconstructing a fluorescence spectrum, SpecGAN can considerably decrease the photon flux requirements—by as much as 100 times when compared to conditions that do not utilize SpecGAN. This significant reduction translates to less time required to gather photons for creating a fluorescence spectrum, thereby yielding a higher temporal resolution. Second, in a super-resolution spectrum imaging experiment, SpecGAN only takes 12 000 frames of single-molecule localization images to construct a super-resolution spectrum image, which is almost half of the previously reported frame count. Therefore, SpecGAN has twofold temporal resolution improvement and enables fast spectrum imaging. Due to the improvement of the temporal resolution, we believe that SpecGAN has the potential to monitor real-time biochemical interaction processes when combined with the single molecule tracking technology, such as changes in the tumor microenvironment during drug delivery.[41,42] On the other

hand, the spectral information of single photon localization has been proven to be useful for spatial resolution enhancement.[19] In future work, it is possible to further improve the temporal and spatial resolution of single-molecule localization microscopy when considering the correlations of the associated emission spectrum and the location of centroids. Overall, this work introduces a promising approach for fast spectrum imaging for weak fluorescence signals, which holds significant potential for a wide range of applications in the study of biological interaction dynamics.

IV. METHODS

A. Pixel shifts and wavelength mapping

A bandpass filter (FF01-591/6-25, Semrock) is used to collect at least six paired points for calculating the mapping matrix. The code implementation is based on the work of Xu *et al.*[43] The mapping matrix can transform the pixel coordinates in the position channel to the position of 591 nm in the spectral channel. Furthermore, to establish the mapping relationship between the pixel shifts and wavelength, fluorescent beads (BangsLab FSDG003 and FSSY002) are imaged and the pixel coordinates of three specific wavelengths in the spectral channel are recorded using bandpass filters (FF01-532/3-25, LL02-561-12.5, and LL01-638-12.5, Semrock). The pixel–wavelength calibrated curve is fitted by a third-order polynomial finally, as shown in Fig. S3 and Table S1. The center coordinates of single molecules in each frame are detected using the ThunderSTORM plugin.[44] Therefore, the spectral curve of each molecule and the average value that is considered the ground truth can be obtained.

B. The VMD-based pre-processing

In order to enhance the performance of our model by reducing the impact of noise, we propose a data pre-processing method based on VMD. As a signal process method without prior knowledge, VMD has been applied to various signal tasks.[45] It assumes that a signal is composed of a series of sub-signals with specific center frequencies and limited bandwidths. The solution to this variational problem involves using the Weiner filter and the Hilbert transform.[38] Utilizing the VMD algorithm, the spectral information corrupted by noise can be separated into different IMFs and a residual item. The residual signal is relatively smoother and partially removes the interference of high-frequency noise. Thus, we use the residual component derived from VMD as the input of the SpecGAN model, which can lead to superior performance. More details about the VMD algorithm can be found in Fig. S5 and supplementary material, Note 4.

C. ResNet-based data screening

In this work, a two-stage network is proposed to identify the accurate centroid of the spectrum collected by the SR-SRM system at the single-molecule level. The structure of the SpecGAN is shown in Fig. 5. The presence of environmental noise results in a

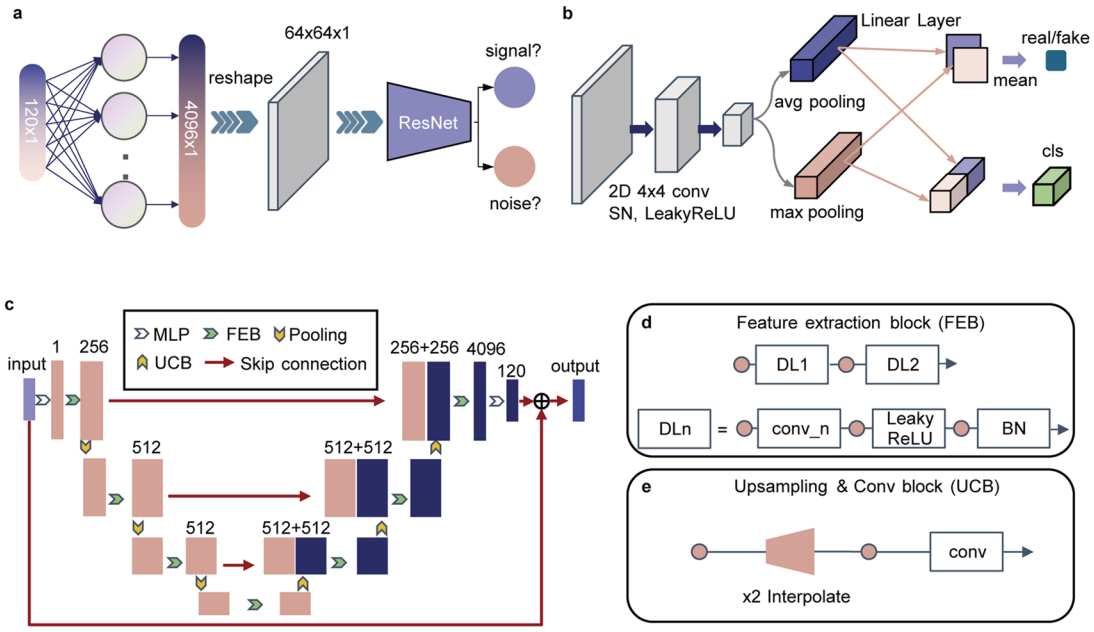


FIG. 5. Workflow of SpecGAN. (a) A two-stage network model for filtering low-SNR spectral signals. The ResNet-based classification model is employed to judge whether the inputs are valid signals or noise and only the valid signals are fed to the subsequent SpecGAN training and testing in SLBs and cell data. (b) Discriminator of SpecGAN with the task of determining whether the data are real or fake and the corresponding environmental category of the current Nile Red molecular spectrum. (c) Generator of SpecGAN based on a simplified UNet architecture, with details of the feature extraction block (FEB) and the upsampling and convolution block (UCB) shown in (d) and (e). The FEB extracts abstract features and maps them to a lower-dimensional representation, while the UCB uses a bilinear interpolation layer to upsample the lower-dimensional representation.

low SNR in single-molecule spectrum images, which precludes the precise single-molecule spectral features analysis. To address this issue, a classifier, as depicted in Fig. 5(a), is proposed to exclude signals with a too low SNR from the SpecGAN training and testing processes. The traditional 2-D convolution layer is adapted to the 1D spectral data by converting the number of parameters from 120 to 4096 using a fully connected network and reshaping it to 64 × 64 × 1. Subsequently, feature extraction is performed through a simplified ResNet.[46] The output of the classification model is a binary result indicating whether the current signal is useful or not. This model is trained using a binary cross-entropy loss function.

D. Details of the SpecGAN

SpecGAN aims to produce a pure spectral curve that can identify environmental features from the VMD residual component of the raw signal. It contains a discriminator and a generator. The generator in Fig. 5(c) is based on the UNet architecture and is composed of eight sub-blocks. The encoder uses convolution layers and maximum pooling layers for feature extraction, while the decoder uses binary linear interpolation and convolution layers for upsampling and information recovery. The corresponding sub-blocks between the encoder and decoder are concatenated for information interaction. The output of the UNet is then passed through a linear layer to ensure that it has the same dimensions as the input spectral data and is then added to the original input to produce the final denoised spectral curve. In particular, the LeakyReLu activation layer and batch normalization (BN) layer are utilized in the feature extraction block (FEB).

In the discriminator depicted in Fig. 5(b), auxiliary classification tasks are designed to further improve the performance. The discriminator needs to determine the confidence of the pure spectral curve and classify the environments represented by it. The labels have been previously described in detail. For other environment-sensitive dye molecules, changes in environmental conditions, such as pH, viscosity, and microenvironment can also be encoded as labels. The discriminator initially extracts features using three convolution layers with the spectral normalization (SN) method.[47] Subsequently, a global average pooling and a global maximum pooling layer reduce the feature map to a size of 1 × 1 × 128. Finally, the downscaled pooling features are processed using linear layers based on the requirements, and the corresponding averages are used for auxiliary classification or GAN tasks.

E. Loss function

In terms of the loss function of SpecGAN, we use the Wasserstein GAN loss function[48] to improve the stability during training, while the $L1$ loss is introduced to determine the distance between the input spectrum and the average spectrum. The auxiliary classification loss is employed to measure the accuracy of category classification. In addition, the 1D total variation (TV) regular term loss function is included to promote output spectral curves smoother. During training, the generator and discriminator parameters are updated in an alternating manner. The overall loss function of the network can be expressed as follows:

$$\mathcal{L} = \arg\min_{G}\min_{D}\left[\alpha_0 \mathcal{L}_{WGAN}(G,D) + \alpha_1 \mathcal{L}_{L1}(G) + \alpha_2 \mathcal{L}_{Aux}(G,D) + \alpha_3 \mathcal{L}_{TV}(G)\right], \quad (1)$$

where G and D represent the generator and discriminator, respectively. The WGAN loss $\mathcal{L}_{WGAN}(G,D)$ is divided into two parts according to Ref. 48. The $\mathcal{L}_{L1}(G)$ is the mean absolute error (MAE) loss between the generator and the ground truth. The auxiliary loss $\mathcal{L}_{Aux}(G,D)$ is a multi-classification cross-entropy loss that measures the accuracy of category classification, as described in the following equation:

$$\mathcal{L}_{Aux} = -\sum_{\{i=1\}}^{\{N\}} y_i \log \hat{y}_i, \quad (2)$$

where y_i is the one-hot coding label, N is the number of categories, and $\hat{y}_i$ represents the probability predicted by the network. The 1D TV loss is defined as follows:

$$\mathcal{L}_{TV}(\mathrm{G}) = \sum_{\mathrm{j}} \left(G(x)_{j+1} - G(x)_j\right)^2, \quad (3)$$

where $G(x)_j$ is the jth intensity value of the spectral curve. During training, the generator and discriminator parameters are updated in an alternating manner with the coefficients of the four loss components set to 1.0, 1.0, 1.0, and 2.0, respectively.

SUPPLEMENTARY MATERIAL

See the supplementary material for more results and details of experiments.

ACKNOWLEDGMENTS

S. Hou acknowledges the support from the National Natural Science Foundation of China (Grant No. 22204106), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2021A1515110710), and the Guangdong Pearl River Talents Program (Grant No. 2021QN02Z631). Y. Zhang acknowledges the support from the National Natural Science Foundation of China (Grant No. 62031023), Shenzhen Science and Technology Project (Grant No. GXWD20220818170353009). S. Hou and S. Hao wish to thank Dr. Yan Zhao in Shenzhen Bay Laboratory for her help on fluorescence spectrophotometer and thank Professor Chenguang Wang in Jilin University for a helpful discussion. S. Hou and S. Hao would also like to express appreciation to Professor Xu at the University of California, Berkeley, for the discussion on the alignment of the spectral and position channels.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Hao Sha: Conceptualization (equal); Data curation (equal); Investigation (lead); Methodology (equal); Software (equal); Validation

(equal); Visualization (equal); Writing – original draft (lead); Writing – review & editing (equal). **Haoyang Li**: Data curation (equal); Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Yongbing Zhang**: Conceptualization (equal); Investigation (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Shangguo Hou**: Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The code implementation and the data used for training and testing in this work are available at our GitHub repository, https://github.com/hitsh95/SpecGAN. You can get the download link for the dataset in the readme.md file. The raw images acquired by the SR-SRM system are available from the corresponding authors upon reasonable request.

REFERENCES

[1] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," Nat. Methods **3**, 793–796 (2006).
[2] A. Sharonov and R. M. Hochstrasser, "Wide-field subdiffraction imaging by accumulated binding of diffusing probes," Proc. Natl. Acad. Sci. U. S. A. **103**, 18911–18916 (2006).
[3] M. D. Lew *et al.*, "Three-dimensional superresolution colocalization of intracellular protein superstructures and the cell surface in live *Caulobacter crescentus*," Proc. Natl. Acad. Sci. U. S. A. **108**, E1102–E1110 (2011).
[4] F. Balzarotti *et al.*, "Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes," Science **355**, 606–612 (2017).
[5] T. Deguchi *et al.*, "Direct observation of motor protein stepping in living cells using MINFLUX," Science **379**, 1010–1015 (2023).
[6] K. H. Song, Y. Zhang, G. Wang, C. Sun, and H. F. Zhang, "Three-dimensional biplane spectroscopic single-molecule localization microscopy," Optica **6**, 709–715 (2019).
[7] Y. I. Chen *et al.*, "Generative adversarial network enables rapid and robust fluorescence lifetime image analysis in live cells," Commun. Biol. **5**, 18 (2022).
[8] M. Mazaheri, J. Ehrig, A. Shkarin, V. Zaburdaev, and V. Sandoghdar, "Ultrahigh-speed imaging of rotational diffusion on a lipid bilayer," Nano Lett. **20**, 7213–7219 (2020).
[9] D. Kim, Z. Zhang, and K. Xu, "Spectrally resolved super-resolution microscopy unveils multipath reaction pathways of single spiropyran molecules," J. Am. Chem. Soc. **139**, 9447–9450 (2017).
[10] L. Sansalone *et al.*, "High-throughput single-molecule spectroscopy resolves the conformational isomers of BODIPY chromophores," J. Phys. Chem. Lett. **10**, 6807–6812 (2019).
[11] W. Y. Kao, C. E. Davis, Y. I. Kim, and J. M. Beach, "Fluorescence emission spectral shift measurements of membrane potential in single cells," Biophys. J. **81**, 1163–1170 (2001).
[12] A. S. Klymchenko, "Solvatochromic and fluorogenic dyes as environment-sensitive probes: Design and biological applications," Acc. Chem. Res. **50**, 366–375 (2017).
[13] C. Phelps, T. Huang, J. Wang, and X. Nan, "Multipair Förster resonance energy transfer via spectrally resolved single-molecule detection," J. Phys. Chem. B **126**, 5765–5771 (2022).
[14] K. Zhanghao *et al.*, "High-dimensional super-resolution imaging reveals heterogeneity and dynamics of subcellular lipid membranes," Nat. Commun. **11**, 5890 (2020).
[15] M. N. Bongiovanni *et al.*, "Multi-dimensional super-resolution imaging enables surface hydrophobicity mapping," Nat. Commun. **7**, 13544 (2016).
[16] R. L. Zhang *et al.*, "Simultaneous single-particle tracking and dynamic pH sensing reveal lysosome-targetable mesoporous silica nanoparticle pathways," ACS Appl. Mater. Interfaces **12**, 42472–42484 (2020).
[17] J. E. Lee *et al.*, "Mapping surface hydrophobicity of α-synuclein oligomers at the nanoscale," Nano Lett. **18**, 7494–7501 (2018).
[18] Z. Zhang, S. J. Kenny, M. Hauser, W. Li, and K. Xu, "Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved super-resolution microscopy," Nat. Methods **12**, 935–938 (2015).
[19] B. Dong *et al.*, "Super-resolution spectroscopic microscopy via photon localization," Nat. Commun. **7**, 12290 (2016).
[20] S. Moon *et al.*, "Spectrally resolved, functional super-resolution microscopy reveals nanoscale compositional heterogeneity in live-cell membranes," J. Am. Chem. Soc. **139**, 10944–10947 (2017).
[21] D. I. Danylchuk, S. Moon, K. Xu, and A. S. Klymchenko, "Switchable solvatochromic probes for live-cell super-resolution imaging of plasma membrane organization," Angew. Chem., Int. Ed. Engl. **58**, 14920–14924 (2019).
[22] H. Guan *et al.*, "Deep-learning two-photon fiberscopy for video-rate brain imaging in freely-behaving mice," Nat. Commun. **13**, 1534 (2022).
[23] N. Wagner *et al.*, "Deep learning-enhanced light-field imaging with continuous validation," Nat. Methods **18**, 557–563 (2021).
[24] Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," Light: Sci. Appl. **8**, 85 (2019).
[25] X. Huang *et al.*, "Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy," Nat. Biotechnol. **36**, 451–459 (2018).
[26] S. Puntener and P. Rivera-Fuentes, "Single-molecule peptide identification using fluorescence blinking fingerprints," J. Am. Chem. Soc. **145**, 1441–1447 (2023).
[27] Q. Wang *et al.*, "Deep-learning-assisted single-molecule tracking on a live cell membrane," Anal. Chem. **93**, 8810 (2021).
[28] Z. Wang *et al.*, "Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning," Nat. Methods **18**, 551–556 (2021).
[29] X. Chen *et al.*, "Accelerated phase shifting for structured illumination microscopy based on deep learning," IEEE Trans. Comput. Imaging **7**, 700–712 (2021).
[30] A. Speiser *et al.*, "Deep learning enables fast and dense single-molecule localization with high accuracy," Nat. Methods **18**, 1082–1090 (2021).
[31] Y. Zhang *et al.*, "Minimizing molecular misidentification in imaging low-abundance protein interactions using spectroscopic single-molecule localization microscopy," Anal. Chem. **94**, 13834–13841 (2022).
[32] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 27 (MIT Press, 2014), pp. 2672–2680.
[33] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *30th IEEE Conference on Computer Vision And Pattern Recognition (CVPR 2017)* (IEEE, 2017), pp. 105–114.
[34] T. C. Wang *et al.*, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *2018 IEEE/CVF Conference on Computer Vision And Pattern Recognition (CVPR)* (IEEE, 2018), pp. 8798–8807.
[35] O. Ronneberger, P. Fischer, and T. U.-N. Brox, "Convolutional networks for biomedical image segmentation," in *Medical Image Computing And Computer-Assisted Intervention, Pt III*, 9351 (Springer, 2015), pp. 234–241.
[36] Y. Y. Lin *et al.*, "Unpaired multi-domain stain transfer for kidney histopathological images," in *AAAI Conference on Artificial Intelligence (AAAI)* (AAAI, 2022), pp. 1630–1637.
[37] R. Kreder *et al.*, "Solvatochromic Nile Red probes with FRET quencher reveal lipid order heterogeneity in living and apoptotic cells," ACS Chem. Biol. **10**, 1435–1442 (2015).
[38] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," IEEE Trans. Signal Process. **62**, 531–544 (2014).
[39] W. Teo *et al.*, "Nile Red fluorescence spectroscopy reports early physicochemical changes in myelin with high sensitivity," Proc. Natl. Acad. Sci. U. S. A. **118**, e2016897118 (2021).

[40]J. A. Jackman and N. J. Cho, "Supported lipid bilayer formation: Beyond vesicle fusion," Langmuir **36**, 1387–1400 (2020).
[41]S. Hou, J. Exell, and K. Welsher, "Real-time 3D single molecule tracking," Nat. Commun. **11**, 3607 (2020).
[42]S. Hou, X. Lang, and K. Welsher, "Robust real-time 3D single-particle tracking using a dynamically moving laser spot," Opt. Lett. **42**, 2390–2393 (2017).
[43]K. Xu, H. P. Babcock, and X. Zhuang, "Dual-objective STORM reveals three-dimensional filament organization in the actin cytoskeleton," Nat. Methods **9**, 185–188 (2012).
[44]M. Ovesny, P. Krizek, J. Borkovec, Z. Svindrych, and G. M. Hagen, "ThunderSTORM: A comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging," Bioinformatics **30**, 2389–2390 (2014).
[45]J. Cheng, X. Chen, L. X. Xu, and Z. J. Wang, "Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," IEEE J. Biomed. Health Inf. **21**, 1422–1433 (2017).
[46]K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision And Pattern Recognition (CVPR)* (IEEE, 2016) , pp. 770–778.
[47]T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in International Conference on Learning Representations (ICLR), 2018.
[48]I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Advances in Neural Information Processing Systems (NIPS)*, 30 (MIT Press, 2017).