Opto-Electronic Advances

CN 51-1781/TN ISSN (Print) 2096-4579 ISSN (Online) 2097-3993

Statistical SERS spectroscopy based on concentration-encoded intensity distributions for deep-learning-assisted quantitation

Meng Zhang, Hui Ding, Luyuan Xie, Chaowei Sun, Shiyu Zhang, Yue Wei, Di Zheng, Ferruccio Pisanello, Longkun Yang*, Hongxing Xu* and Zhipeng Li*

Citation: Zhang M, Ding H, Xie LY et al. Statistical SERS spectroscopy based on concentration-encoded intensity distributions for deep-learning-assisted quantitation. *Opto-Electron Adv* **9**, 260163 (2026).

https://doi.org/10.29026/oea.2026.260163

Received: 19 May 2026; Accepted: 21 July 2026; Published online: 28 August 2026

Related articles

Single-beam optical trap-based surface-enhanced Raman scattering optofluidic molecular fingerprint spectroscopy detection system

Ning Sun, Yuan Gan, Yujie Wu, et al

Opto-Electronic Advances 2025, **8**(2): 240182 doi: 10.29026/oea.2025.240182

Quantitative detection of trace nanoplastics (down to 50 nm) via surface-enhanced Raman scattering based on the multiplex-feature coffee ring

Xinao Lin, Fengcai Lei, Xiu Liang, et al

Opto-Electronic Advances 2025, **8**(6): 240260 doi: 10.29026/oea.2025.240260

Timeshare surface-enhanced Raman scattering platform with sensitive and quantitative mode

Qianqian Ding, Xueyan Chen, Yunlu Jia, et al

Opto-Electronic Advances 2026, **9**(1): 250269 doi: 10.29026/oea.2026.250269

Enrichment strategies in surface-enhanced Raman scattering: theoretical insights and optical design for enhanced light-matter interaction

Zhiyang Pei, Chang Ji, Mingrui Shao, et al

Opto-Electronic Science 2025, **4**(12): 250015 doi: 10.29026/oes.2025.250015

More related articles in Opto-Electronic Journals Group website

Opto-Electronic Journals Group

https://www.oejournal.org

OE_Journal

@OptoElectronAdv

OEA | Opto-Electronic Advances

Article

November 2026, Vol. 9, No. 11

Open Access

Statistical SERS spectroscopy based on concentration-encoded intensity distributions for deep-learning-assisted quantitation

Meng Zhang[1†], Hui Ding[2†], Luyuan Xie[3], Chaowei Sun[4], Shiyu Zhang[2], Yue Wei[1], Di Zheng[1], Ferruccio Pisanello[5], Longkun Yang[1*], Hongxing Xu[4*] and Zhipeng Li[1*]

Abstract: Reliable quantitative SERS analysis is limited by strong intensity fluctuations arising from stochastic molecule-substrate-light interactions. Here, we reveal that SERS intensity distributions carry concentration-dependent statistical information, exhibiting a characteristic evolution from an asymmetric long-tailed profile at low concentration to a nearly symmetric Gaussian-like pattern at high concentration. This concentration-dependent evolution originates from stochastic molecular adsorption at plasmonic "hot spots" and can be reproduced by Poisson-statistics-based Monte Carlo simulations. On this basis, we propose a statistical SERS strategy for concentration-level quantitation, in which concentration-encoded intensity distribution patterns are identified using a ResNet4 deep learning model. Accurate concentration identification of R6G is achieved, and the generality of this method is further demonstrated by the analysis of methamphetamine in methanol, thiram in bean sprout extract, and rosiglitazone maleate in serum. These results transform intrinsic SERS intensity fluctuations from a limitation of conventional intensity-based quantitation into useful statistical fingerprints, which certainly deepens our understanding about the stochastic nature of SERS and provides a new route toward reliable SERS quantitation for chemical and biological analysis.

Keywords: statistical spectroscopy; surface-enhanced Raman scattering; poisson distribution; quantitative detection

DOI: 10.29026/oea.2026.260163 | CSTR: 32247.14.oea.2026.260163

Citation: Zhang M, Ding H, Xie LY et al. Statistical SERS spectroscopy based on concentration-encoded intensity distributions for deep-learning-assisted quantitation. *Opto-Electron Adv* **9**, 260163 (2026).

1 Introduction

Statistical spectroscopy has emerged as an interdisciplinary field that integrates statistical methodologies with spectroscopic techniques, offering powerful solutions for analyzing complex spectral data across physics, chemistry, and astronomy[1–5]. A particularly groundbreaking application lies in quantitative analysis at the single-molecule level[6–8], where Nilsson et al. pioneered digital counting quantitation by converting analog fluorescence signals into binary digital outputs (0/1)[9]. This approach leverages Poisson statistics, where quantitation accuracy scales with the square root of event counts[7,9], establishing a robust framework for ultrasensitive quantitation.

Compared with the fluorescence, Raman spectroscopy offers rich vibrational fingerprints for molecular identification, which is particularly advantageous for analyzing in complex matrices[10–16]. With plasmonic enhancement, surface-enhanced Raman scattering (SERS) can further improve detection sensitivity down to the single-molecule level[17–19]. While, SERS signals are highly sensitive to stochastic molecule-substrate-light interactions[12,20–22]. The random adsorption of molecules at plasmonic "hot spots" and the heterogeneous enhancement distribution of SERS substrates lead to pronounced intensity fluctuations[23,24], making conventional intensity-based quantitation challenging. Recently, a breakthrough on quantitative SERS has been

Received: 19 May 2026

Accepted: 21 July 2026

Published online: 28 August 2026

[1]Beijing Key Laboratory for Nano-Photonics and Nano-Structure (NPNS), Department of Physics, Capital Normal University, Beijing 100048, China; [2]Beijing Key Laboratory of Electronic System Reliability Technology, College of Information Engineering, Capital Normal University, Beijing 100048, China; [3]School of Software and Microelectronics, Peking University, Beijing 100091, China; [4]Henan Academy of Sciences, Zhengzhou 450046, China; [5]Istituto Italiano di Tecnologia, Center for Biomolecular Nanotechnologies, Arnesano, Lecce 73010, Italy.

[†]These authors contributed equally to this work.

*Correspondence: ZP Li, E-mail: zpli@cnu.edu.cn; LK Yang, E-mail: lkyang@cnu.edu.cn; HX Xu, E-mail: hxxu@hnas.ac.cn

made by introducing statistical digital counting method[25,26]. By gauging SERS probability rather than absolute intensity, quantitation at single-molecule concentration can even be achieved with high precision. It has demonstrated practical applicability in fields such as environmental protection and food safety, which indicates the rising of statistical SERS spectroscopy[27–31]. Beyond binary event counting, SERS intensity distributions contain richer statistical information associated with both molecular-number fluctuations and "hot-spot" dependent enhancement variations[32,33]. The former is related to the number of molecules contributing to each SERS measurement[34], whereas the latter reflects the enhancement characteristics of the SERS substrate[35]. Therefore, deciphering how SERS intensity distributions evolve with analyte concentration may provide a new paradigm for quantitative SERS analysis.

In this work, we propose a statistical SERS quantitation strategy based on deep-learning-assisted identification of concentration-encoded SERS intensity distributions (Fig. 1). We first reveal that SERS intensity distributions exhibit a clear concentration-dependent evolution. As the R6G concentration increases from 1×10^{-11} M to 1×10^{-7} M, the intensity distribution gradually evolves from an asymmetric long-tailed profile to a nearly symmetric Gaussian-like pattern. This evolution can be explained by stochastic molecular adsorption at plasmonic "hot spots" and is reproduced by Poisson-statistics-based Monte Carlo simulations. The subtle but concentration-specific distribution features can be effectively identified by the ResNet4 deep learning model, enabling accurate concentration-level quantitation of R6G with an accuracy up to 100%. The generality and practical applicability of this strategy are further demonstrated by the analysis of methamphetamine in methanol (1 ppb–10 ppm), thiram in bean sprout extract (1×10^{-8} M–1×10^{-6} M), and rosiglitazone maleate in serum (1×10^{-7} M–1×10^{-5} M). This statistical strategy transforms intrinsic SERS intensity fluctuations from an obstacle in conventional intensity-based quantitation into useful concentration-encoded fingerprints, providing a new route toward reliable SERS quantitation in chemical and biological analysis.

2 Results and discussion

2.1 Statistical distribution characteristics of SERS intensity

Statistical SERS quantitation requires a large number of spectra acquired from substrates with good spatial uniformity. To meet this requirement, well-ordered silver nanoparticle arrays were fabricated using the Langmuir-Blodgett (LB) technique, as described in the Experimental section[36,37]. Scanning electron microscopy (SEM) images (Fig. S3 in the Supplementary information) show that the silver nanoparticles (~60 nm in diameter) are closely packed and uniformly distributed over a micrometer-scale area, providing a suitable platform for SERS mapping measurements. R6G was selected as a model analyte. The SERS samples were prepared by immersing the substrates in R6G solutions with different concentrations for 2 h, followed by drying under ambient conditions. Unless otherwise stated, SERS measurements were performed under 532 nm laser excitation with a power of 7.5 μW and an integration time of 1 s. The mapping area was 30 × 30 μm² with a step size of 1 μm. Figure 2(a) shows the SERS mapping images of R6G at 612 cm^{-1} over the concentration range from 1×10^{-11} M to 1×10^{-7} M. At each concentration, 100 pixels were randomly selected from the mapping image, and the corresponding SERS intensities at 612 cm^{-1} are plotted in Fig. 2(b). The average SERS intensity generally increased with increasing R6G concentration, and the fitted calibration curves gave R^2 values of approximately 0.90 (Fig. S4(a)). However, large intensity fluctuations were observed at each concentration. For example, the relative standard deviation (RSD) of the SERS intensity at 1×10^{-9} M reached 37.9%. Since

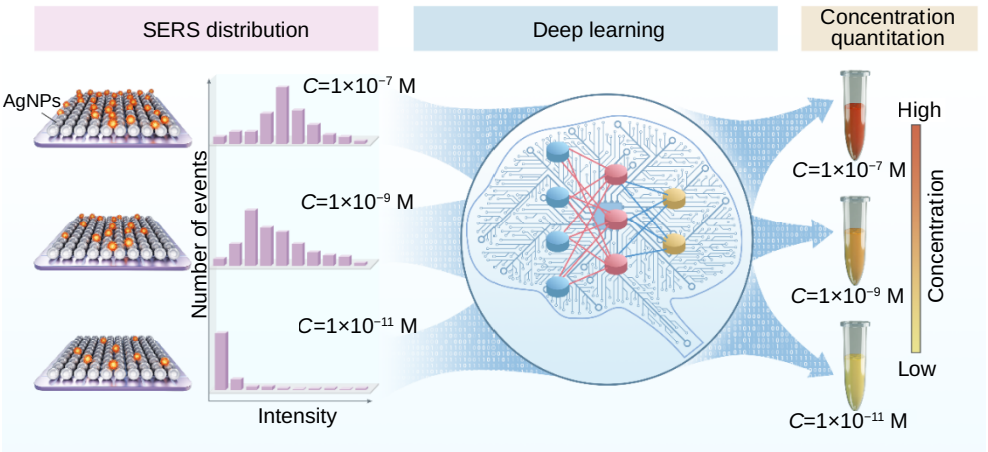


Fig. 1 | The scheme of SERS quantitation via deep learning identified statistical distributions of intensity.

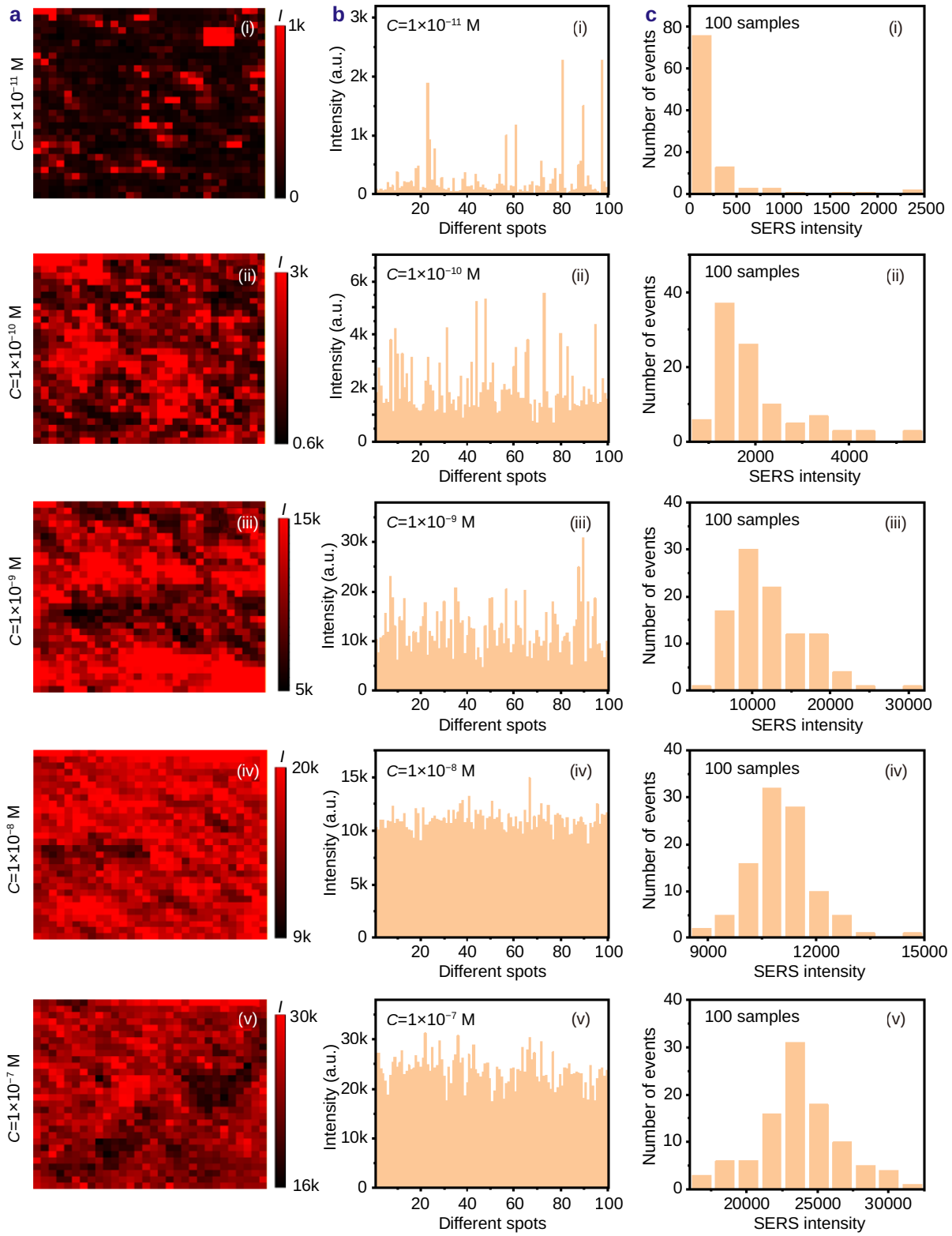


Fig. 2 | Statistical distributions of R6G SERS intensity. (a) SERS mapping images of R6G acquired at 612 cm^{-1} at different concentrations. Panels (i)–(v) correspond to 1×10^{-11} M, 1×10^{-10} M, 1×10^{-9} M, 1×10^{-8} M, and 1×10^{-7} M, respectively. (b) SERS intensities at 612 cm^{-1} extracted from 100 randomly selected pixels in the corresponding mapping images (a). (c) Statistical histograms of SERS intensity in (b).

conventional intensity-based quantitation uses the SERS intensity as the sole quantitative metric, such large signal fluctuations inevitably reduce the precision of concentration determination. Moreover, the calibration curves obtained from different batches exhibited different slopes and intercepts (Fig. S4(b)), further indicating that robust quantitation based solely on absolute SERS intensity remains challenging.

To overcome this limitation, we analyzed the statistical distribution of SERS intensities rather than individual intensity values. At each R6G concentration, the SERS intensity values at 612 cm^{-1} were uniformly divided into 10 bins from I_{min} to I_{max}, where I_{min} and I_{max} were determined from the corresponding full mapping dataset. The number of SERS events falling into each bin was then counted to construct a statistical intensity histogram. The obtained intensity distributions are shown in Fig. 2(c). Interestingly, the distribution patterns exhibited a clear dependence on analyte concentration. At 1×10^{-11} M, most SERS events accumulated in the first bin, forming an asymmetric long-tailed, Pareto-like distribution (panel i)[38,39]. As the concentration increased, the population of SERS events gradually shifted toward higher-intensity bins (panels ii–iv), and finally evolved into a nearly Gaussian-like distribution at 1×10^{-7} M (panel v). To quantitatively describe this concentration-dependent evolution, the skewness of the raw SERS intensity data was calculated at each concentration. As shown in Table S1, the skewness decreased from 3.515 at 1×10^{-11} M to 0.086 at 1×10^{-7} M, confirming that the intensity distribution gradually evolved from a highly asymmetric right-skewed profile to a nearly symmetric one with increasing analyte concentration.

The concentration-dependent evolution of SERS intensity distributions can be understood from the stochastic combination between molecules and plasmonic "hot spot". The number of molecules contributing to each SERS spectrum can be described by Poisson statistics. The probability of measuring (n) molecules in one spectrum follows the relation:

$$p(n,\alpha) = \frac{\alpha^n}{n!}e^{-\alpha}\,, \tag{1}$$

where α represents the average number of molecules per spectrum at a given analyte concentration. The electromagnetic "hot spot" was modeled using a coupled Ag nanoparticle dimer with a diameter of 60 nm and a gap size of 2 nm, as shown in Fig. S5(a). Assuming a total number of spectral samplings N_s, the number of spectra containing n molecules is given by $N_s{\cdot}p(n)$. In the Monte Carlo simulations, these n molecules were randomly distributed on the surface of the nanoparticle dimer structure. For each sampling event containing n molecules, the SERS intensity will be proportional to

$$I_j \sim \sum\nolimits_{i=1}^{n} |E_i|^4\,, \tag{2}$$

where E_i is the local electric field at the location of the ith molecule. For each α value, corresponding to a given analyte concentration in the experiment, a set of simulated SERS intensities $\{I_1, I_2, \ldots, I_{Ns}\}$ can be obtained by performing N_s times Monte Carlo simulations. Histograms were then constructed to analyze the simulated intensity distributions. As shown in Fig. S5(b), when α increased from 40 to 2000, the simulated intensity distribution gradually evolved from a Pareto-like long-tailed pattern to a near-Gaussian profile, well reproducing the experimental observations. To further examine the influence of near-field enhancement, additional simulations were performed using Ag nanoparticle dimers with different gap sizes. As shown in Fig. S5(c), reducing the gap size from 2 nm to 1 nm generated a stronger and more localized electromagnetic field in the junction region. Accordingly, the simulated SERS intensities increased, and the intensity distributions become broader and more skewed, especially at low and intermediate α values. These results indicate that the gap size or near-field enhancement strength can affect the absolute intensity scale, distribution width and detailed distribution profile. Nevertheless, the overall α-dependent evolution remains unchanged: the intensity distribution evolves from a skewed long-tailed pattern at low α to a more symmetric profile at high α. Therefore, the near-field enhancement modulates the detailed intensity distributions, but does not alter the fundamental statistical mechanism responsible for the concentration-dependent evolution of SERS intensity distributions.

2.2 SERS intensity distributions identified by deep learning model

To establish statistical SERS intensity distributions as a reliable quantitation metric, their reproducibility should first be verified. We calculated the Pearson correlation coefficients between SERS intensity distributions obtained from independent measurements at the same concentration. As shown in Fig. S6, strong positive correlations ($\rho \approx 1$) were observed across all tested concentrations, indicating the good reproducibility of the concentration-encoded intensity distributions. This result suggests that statistical intensity distributions can serve as robust fingerprints for SERS quantitation. However, direct concentration identification based solely on Pearson correlation coefficients between an unknown distribution and standard distributions at different concentrations was found to be unreliable, especially for adjacent concentrations with highly similar distribution profiles (Fig. S7). This is because Pearson correlation mainly evaluates global linear similarity, whereas the concentration information in SERS intensity distributions is encoded in more subtle and nonlinear features, including the relative population of different intensity bins, the peak position, the distribution width, and the local variation among neighboring bins. These features are difficult to fully capture using simple

intensity calibration or correlation-based comparison.

To better extract concentration-specific distribution features, we employed the ResNet4 deep learning model to identify the statistical SERS intensity distributions. The one-dimensional convolutional layers in ResNet4 can efficiently extract local bin-level correlations among neighboring histogram bins, while the stacked residual blocks enable hierarchical feature learning from local intensity-bin variations to higher-level distribution representations. In addition, the residual connections help preserve feature information and improve optimization stability during training. Therefore, ResNet4 was selected as an appropriate classifier for distinguishing one-dimensional concentration-encoded SERS intensity distribution patterns derived from SERS mapping data[40–45]. The architecture of the ResNet4 model is illustrated in Fig. 3. Briefly, the input histogram data are first downsampled by a one-dimensional convolution and a max-pooling layer, then passed through four residual blocks, each with one-dimensional convolutions, batch normalization, ReLU activation, and shortcut connections. After feature extraction, adaptive average pooling generates a fixed-size feature representation, which is then fed into a fully connected layer to output the concentration class. More details of the ResNet4 model are provided in the *Experimental* section.

A sufficient number of distribution datasets is essential for training a stable deep learning model[46,47]. As described in the Experimental section, we developed a random-sampling extraction strategy to generate multiple statistical distribution patterns from the SERS mapping data. For each R6G concentration, 200 distribution datasets were generated, with each histogram constructed from 100 randomly selected spectra using 10 bins. These datasets were divided into training and testing sets at a 1:1 ratio, yielding 100 training datasets and 100 testing datasets for each concentration. To evaluate the robustness of the quantitation strategy, the training and testing datasets were sourced from independent batches of SERS mapping measurements. The ResNet4 was trained using the AdamW optimizer with an initial learning rate of 0.01 and a batch size of 64. Cross-entropy loss with label smoothing was used as the loss function. The training set was further divided into training and validation subsets for model selection and early stopping, while the independent testing set was used only for final performance evaluation. As shown in Table 1, the ResNet4 model achieved 100% recognition accuracy for all five R6G concentrations, demonstrating its strong capability for concentration-level identification. For comparison, classical machine-learning methods, including KNN and SVM, were also used to classify the statistical distributions. As summarized in Table 1, ResNet4 achieved the best overall performance among the tested models. This advantage may arise from its

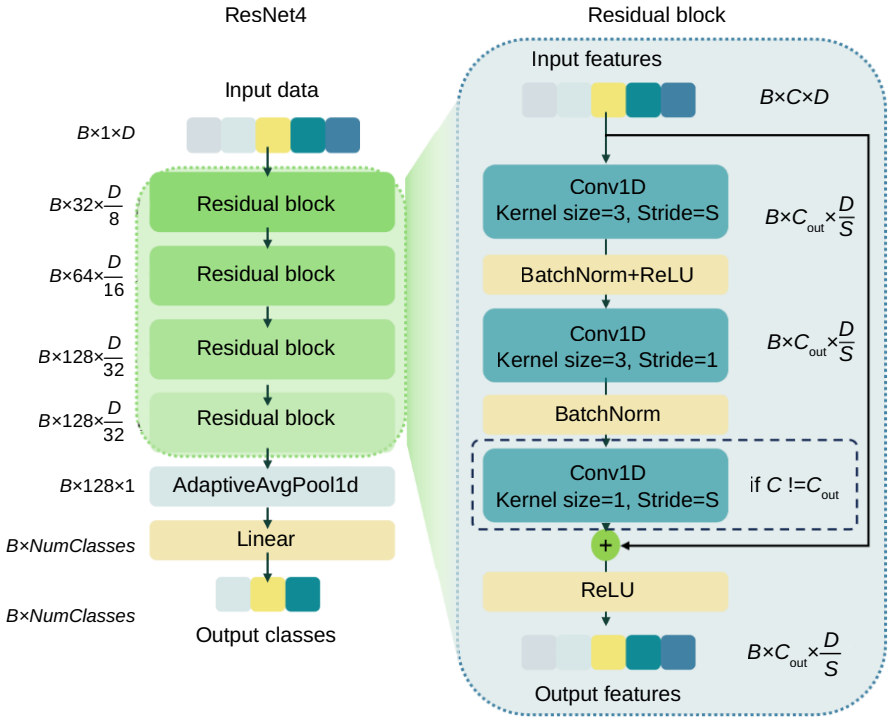


Fig. 3 | The architecture of the ResNet4 model. Left panel is the overall architecture of ResNet4, illustrating how the input data passes through multiple residual blocks, an adaptive average pooling layer, and a linear layer to ultimately output class predictions. On the right, the internal structure of a residual block is detailed, including one-dimensional convolutional layers, batch normalization layers, activation functions, and the implementation of residual connections.

ability to learn both local bin-level correlations and global distribution features from histogram inputs, whereas conventional methods based on predefined distance metrics or kernel functions are less effective in capturing subtle nonlinear differences among concentration-encoded SERS intensity distributions.

To further evaluate the concentration resolution of the proposed statistical quantitation method, we expanded the classification system by introducing four intermediate concentrations, namely 0.5×10^{-10} M, 0.5×10^{-9} M, 0.5×10^{-8} M, and 0.5×10^{-7} M, in addition to the original five concentration levels. Consistent with Table 1, the statistical intensity histograms at each concentration were constructed from 100 spectral samples using 10 bins. Each concentration contained 100 training datasets and 100 testing datasets, giving a total of 900 training datasets and 900 testing datasets. As shown in Table 2, when nine closely spaced concentration levels were included, the average recognition accuracy decreased to 77%. This decrease is expected because the statistical distribution differences between adjacent concentrations become more subtle under the current sampling density. These results indicate that finer concentration resolution requires more spectral samplings to better characterize the distribution features.

To obtain more detailed statistical information, we further increased both the number of spectral samplings and the number of histogram bins. At each concentration, 200 intensity values were randomly selected from the SERS mapping images shown in Fig. 2(a) and Fig. S8(a). These intensity values were then divided into 20 bins to generate statistical distribution patterns, as shown in Fig. 4. It was noted that increasing the bin number did not change the fundamental concentration-dependent evolution of the distributions, but the larger number of SERS events provided richer statistical features and improved the distinguishability among adjacent concentrations. As summarized in Table 2, using 200 samples and 20 bins, the ResNet4 model achieved an average concentration identification accuracy of 100%. Further increasing the sampling number and bin number to 400 samples and 40 bins yielded a comparable average accuracy of 99% (Fig. S10). These results demonstrate that statistical distributions generated from sufficient SERS samplings can significantly improve concentration-level recognition performance. The high accuracy and robustness of this strategy highlight the potential of ResNet4-identified statistical intensity distributions for reliable SERS quantitation.

2.3 Applications of AI-driven statistical SERS quantitation methodology

To evaluate the applicability and generality of the proposed statistical SERS quantitation strategy, we first applied it to the analysis of methamphetamine (MA) in methanol as a proof-of-concept demonstration. MA is a representative amphetamine-type stimulant, and its rapid and sensitive detection is of practical importance[48,49]. The SERS spectrum, mapping images, and corresponding statistical intensity distributions of MA over the concentration range from 1 ppb to 10 ppm are shown in Fig. S11. The characteristic

Table 1 | Accuracy of concentrations identification by deep learning model (ResNet4) and machine learning models (KNN and SVM). Input data contains 5 concentrations, 1×10^{-11} M, 1×10^{-10} M, 1×10^{-9} M, 1×10^{-8} M, 1×10^{-7} M. At each concentration, statistical histograms are obtained from 100 spectral samples with bins number 10.

	Models	1×10^{-11} M	1×10^{-10} M	1×10^{-9} M	1×10^{-8} M	1×10^{-7} M	Average accuracy
Deep learning	ResNet4	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Machine learning	KNN (*K*=1)	92%	91%	90%	91%	91%	91%
	KNN (*K*=3)	91%	90%	89%	90%	90%	90%
	KNN (*K*=5)	88%	88%	87%	88%	89%	88%
	SVM (linear)	98%	97%	97%	99%	99%	98%
	SVM (rbf)	44%	42%	43%	46%	45%	44%

Table 2 | Accuracy of concentrations identification by ResNet4 model, where the input data contains 9 concentration levels, 1×10^{-11} M, 0.5×10^{-10} M, 1×10^{-10} M, 0.5×10^{-9} M, 1×10^{-9} M, 0.5×10^{-8} M, 1×10^{-8} M, 0.5×10^{-7} M, 1×10^{-7} M. At each concentration, statistical histograms are obtained from 100/200/400 spectral samples with bins number 10/20/40.

	Bins number	1×10^{-11} M	0.5×10^{-10} M	1×10^{-10} M	0.5×10^{-9} M	1×10^{-9} M	0.5×10^{-8} M	1×10^{-8} M	0.5×10^{-7} M	1×10^{-7} M	Average accuracy
Res Net4	10 bins (100 samples)	74%	64%	78%	68%	75%	92%	96%	44%	100%	77%
	20 bins (200 samples)	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	40 bins (400 samples)	95%	100%	100%	100%	100%	100%	100%	100%	100%	99%

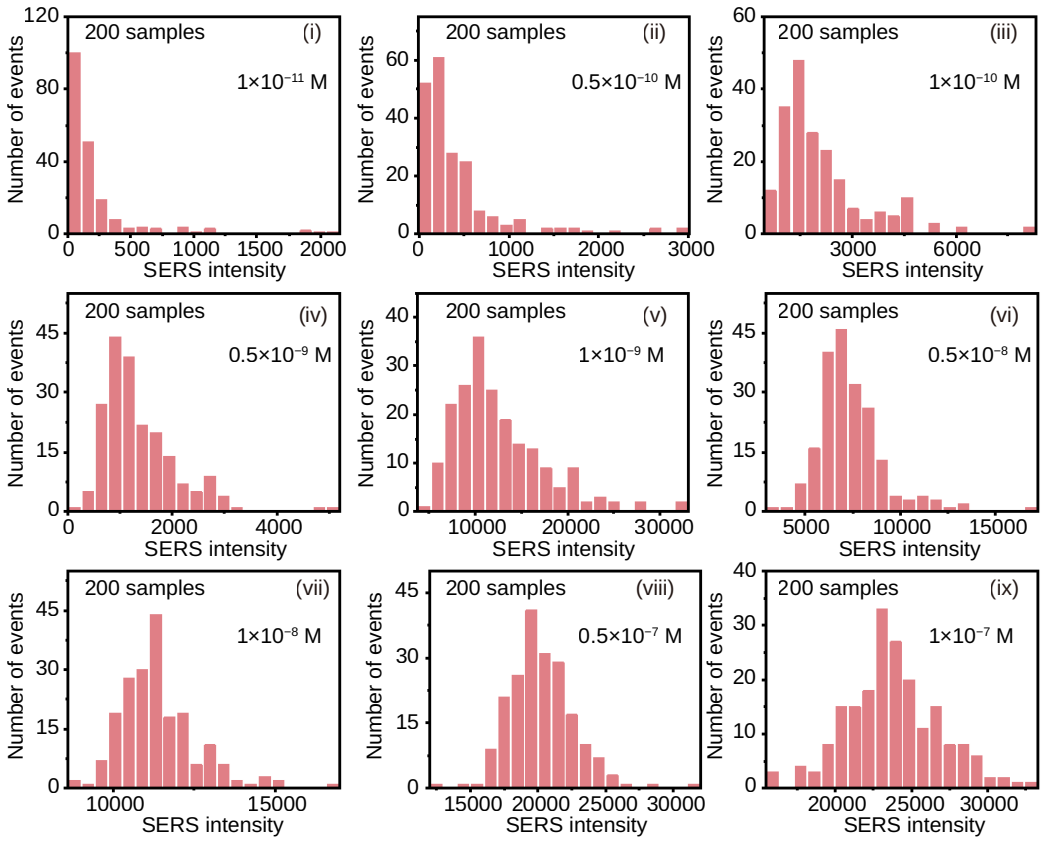


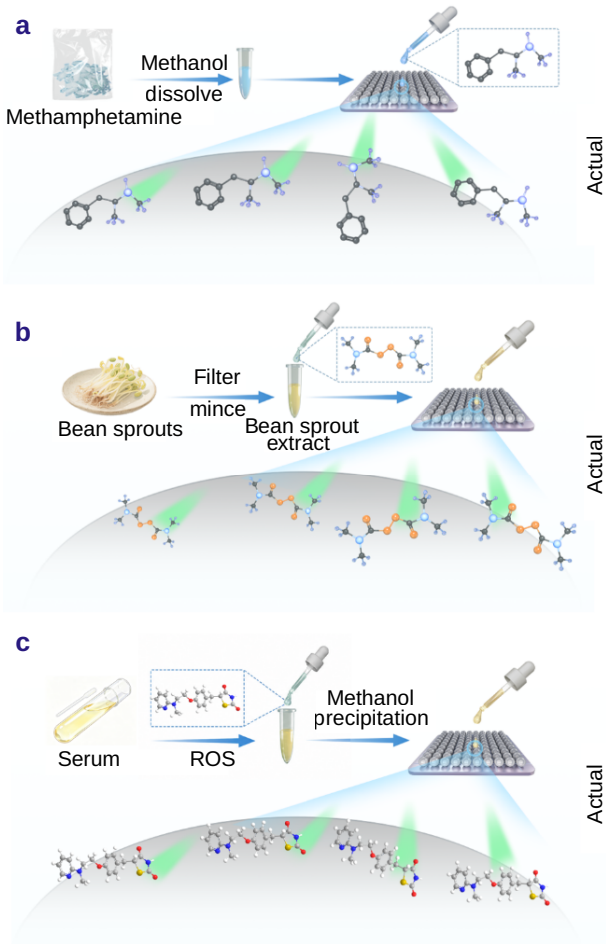
Fig. 4 | Statistical distributions of 200 SERS intensity samplings within 20 bins, where the concentrations are panel (i) 1×10^{-11} M, panel (ii) 0.5×10^{-10} M, panel (iii) 1×10^{-10} M, panel (iv) 0.5×10^{-9} M, panel (v) 1×10^{-9} M, panel (vi) 0.5×10^{-8} M, panel (vii) 1×10^{-8} M, panel (viii) 0.5×10^{-7} M, panel (ix) 1×10^{-7} M, respectively.

peak at 1002 cm^{-1}, assigned to the phenyl ring breathing vibration of MA, was selected to construct the intensity distributions. Using the concentration-encoded intensity distributions as inputs, the trained ResNet4 model achieved recognition accuracies higher than 98% for all tested concentrations, with an average accuracy of 99.6%, as summarized in Fig. 5(a). These results demonstrate that the proposed statistical SERS strategy is effective for concentration-level identification of small molecules in a simple solvent system.

We next evaluated the method in a more complex food matrix by detecting thiram in bean sprout extract[50,51]. Fresh bean sprouts were physically crushed, filtered, and centrifuged to obtain a clear extract, which was then spiked with different concentrations of thiram to simulate pesticide residues. The SERS spectrum, mapping images, and corresponding statistical intensity distributions are provided in Fig. S12. The bean sprout extract showed no obvious SERS interference at the characteristic thiram bands. The peak at 1387 cm^{-1}, assigned to the C-N stretching vibration of thiram, was selected to construct the intensity distributions. For thiram concentrations from 1×10^{-8} M to 1×10^{-6} M, the statistical SERS intensity distributions exhibited clear concentration-dependent features. As shown in Fig. 5(b), the trained ResNet4 model achieved recognition accuracies higher than 95% for all tested concentrations, with an average accuracy of 99%.

To further examine the applicability of this method in a more complex biological matrix, we carried out concentration-level analysis of rosiglitazone maleate (ROS) in artificial serum. Before SERS measurements, serum samples were pretreated with methanol to precipitate proteins and reduce matrix interference. The corresponding SERS spectrum and peak assignments are shown in Fig. S14. Both characteristic Raman fingerprints of ROS and background signals from serum components were observed, reflecting the complexity of the biological matrix. The peak at 739 cm^{-1}, assigned to the benzene ring C-H deformation vibration of ROS, was selected to construct the intensity distributions. Despite the presence of serum-derived background signals, the statistical intensity distributions still showed concentration-dependent evolution as the ROS concentration increased from 1×10^{-7} M to 1×10^{-5} M. Based on these distribution patterns, the ResNet4 model achieved an average recognition accuracy of ~96%, as shown in Fig. 5(c). This result suggests that the proposed statistical SERS methodology can retain quantitative capability even in a complex biological matrix.

It should be noted that the present ResNet4 model



Predicted

Actual

MA (20 bin)	1 ppb	5 ppb	100 ppb	1 ppm	10 ppm
1 ppb	100%	0	0	0	0
5 ppb	0	100%	0	0	0
100 ppb	0	0	98%	2%	0
1 ppm	0	0	0	100%	0
10 ppm	0	0	0	0	100%

Predicted

Actual

Thiram in bean sprout extract (20 bin)	1×10^{-8} M	0.5×10^{-7} M	1×10^{-7} M	0.5×10^{-6} M	1×10^{-6} M
1×10^{-8} M	100%	0	0	0	0
0.5×10^{-7} M	0	100%	0	0	0
0.6×10^{-7} M	0	100%	0	0	0
1×10^{-7} M	0	0	100%	0	0
0.5×10^{-6} M	0	0	0	100%	0
1×10^{-6} M	0	0	0	5%	95%

Predicted

Actual

ROS in serum (20 bin)	1×10^{-7} M	0.5×10^{-6} M	1×10^{-6} M	0.5×10^{-5} M	1×10^{-5} M
1×10^{-7} M	92%	8%	0	0	0
0.5×10^{-6} M	0	100%	0	0	0
0.6×10^{-6} M	4%	96%	0	0	0
1×10^{-6} M	0	0	100%	0	0
0.5×10^{-5} M	0	0	0	93%	7%
1×10^{-5} M	0	0	0	4%	96%

Fig. 5 | Applications of AI-driven statistical SERS quantitation. (a) Concentration-level identification accuracy of MA in methanol. (b) Concentration-level identification accuracy of thiram in bean sprout extract. (c) Concentration-level identification accuracy of ROS in methanol-treated serum.

performs concentration-level quantitation based on predefined concentration classes, rather than continuous regression prediction. To evaluate the response of the model to concentrations not included in the training set, several untrained intermediate concentrations were further tested. As shown in Fig. 5(b) and Fig. 5(c), the unknown samples at 0.6×10^{-7} M for thiram and 0.6×10^{-6} M for ROS were assigned mainly to their nearest trained concentration levels. This result indicates that the proposed statistical SERS strategy can estimate the concentration range of unknown samples within a predefined concentration grid. In principle, finer quantitative resolution can be achieved by introducing more concentration levels and sufficient statistical sampling data, as discussed in Table 2.

In addition, for all quantitative analysis, the training and testing datasets were derived from independent batches of SERS mapping measurements, further supporting the robustness of this statistical quantitation strategy against measurement-to-measurement and batch-to-batch variations. Taken together, the successful analysis of MA in methanol, thiram in bean sprout extract, and ROS in methanol-treated serum demonstrate the generality of concentration-encoded SERS intensity distributions for reliable concentration-level quantitative analysis in different application scenarios.

3 Conclusions

In summary, we developed a statistical SERS quantitation strategy by combining concentration-encoded intensity distributions with deep-learning-assisted pattern identification. We demonstrated that SERS intensity distributions exhibit a clear concentration-dependent evolution, changing from asymmetric long-tailed profiles at low analyte concentrations to nearly symmetric Gaussian-like profiles at high concentrations. This evolution originates from stochastic molecular adsorption at plasmonic “hot spots” and can be reproduced by Poisson-statistics-based Monte Carlo simulations. By using the full intensity distributions as inputs, the ResNet4 model effectively identified

concentration-specific distribution patterns, enabling accurate concentration-level quantitation of R6G with an accuracy up to 100%. The applicability of this strategy was further verified by the analysis of methamphetamine in methanol, thiram in bean sprout extract, and rosiglitazone maleate in serum, demonstrating its potential for concentration-level quantitation in different analytical scenarios.

More importantly, this work extends statistical SERS quantitation beyond single-molecule digital counting. In digital SERS counting, analog SERS signals are typically converted into binary events, and the concentration is quantified from event probability based on Poisson statistics. In contrast, the present method preserves the full analog SERS intensity distribution as a concentration-encoded statistical fingerprint. This intensity-distribution metric captures richer statistical information, including molecular-number fluctuations, "hot spot" dependent enhancement variations, distribution skewness, and the overall evolution of the distribution profile. In this way, intrinsic SERS intensity fluctuations are transformed from a limitation of conventional intensity-based quantitation into useful statistical information. This work therefore proposes intensity distribution as a new statistical metric for SERS quantitation, deepens the understanding of stochastic nature of SERS spectroscopy, and provides a new route toward reliable SERS quantitative analysis beyond strict single-molecule conditions[52−56].

References

1. Mark H, Workman Jr J. Statistics in Spectroscopy. 2nd ed. Amsterdam: Elsevier, 2003.
2. Maher AD, Fonville JM, Coen M et al. Statistical total correlation spectroscopy scaling for enhancement of metabolic information recovery in biological NMR spectra. *Anal Chem* **84**, 1083–1091 (2012).
3. Morampudi SC, Turner AM, Pollmann F et al. Statistics of fractionalized excitations through threshold spectroscopy. *Phys Rev Lett* **118**, 227201 (2017).
4. Mazuritskiy MI, Kozakov AT, Kumar N et al. Interface heterogeneity of periodic multilayer mirrors investigated by X-ray fluorescence, scattering spectroscopy, and mathematical methods of statistics. *Surf Interfaces* **33**, 102258 (2022).
5. Lin XN, Lei FC, Liang X et al. Quantitative detection of trace nanoplastics (down to 50 nm) via surface-enhanced Raman scattering based on the multiplex-feature coffee ring. *Opto-Electron Adv* **8**, 240260 (2025).
6. Hindson CM, Chevillet JR, Briggs HA et al. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* **10**, 1003–1005 (2013).
7. Rissin DM, Walt DR. Digital concentration readout of single enzyme molecules using femtoliter arrays and Poisson statistics. *Nano Lett* **6**, 520–523 (2006).
8. Rissin DM, Kan CW, Campbell TG et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat Biotechnol* **28**, 595–599 (2010).
9. Jarvius J, Melin J, Göransson J et al. Digital quantification using amplified single-molecule detection. *Nat Methods* **3**, 725–727 (2006).
10. Li N, Shen FZ, Cai ZH et al. Target-induced core-satellite nanostructure assembly strategy for dual-signal-on fluorescence imaging and Raman quantification of intracellular MicroRNA guided photothermal therapy. *Small* **16**, e2005511 (2020).
11. Lim DK, Jeon KS, Hwang JH et al. Highly uniform and reproducible surface-enhanced Raman scattering from DNA-tailorable nanoparticles with 1-nm interior gap. *Nat Nanotechnol* **6**, 452–460 (2011).
12. Chen XY, Cui AR, He MY et al. Slippery Au nanosphere monolayers with analyte enrichment and SERS enhancement functions. *Nano Lett* **23**, 6736–6743 (2023).
13. Koike K, Bando K, Ando J et al. Quantitative drug dynamics visualized by alkyne-tagged plasmonic-enhanced Raman microscopy. *ACS Nano* **14**, 15032–15041 (2020).
14. Shao MR, Ji C, Tan JB et al. Ferroelectrically modulate the Fermi level of graphene oxide to enhance SERS response. *Opto-Electron Adv* **6**, 230094 (2023).
15. Pei ZY, Ji C, Shao MR et al. Enrichment strategies in surface-enhanced Raman scattering: theoretical insights and optical design for enhanced light-matter interaction. *Opto-Electron Sci* **4**, 250015 (2025).
16. Ji C, Yue Y X. Principle and research progress of surface-enhanced Raman scattering (SERS) laser spectroscopy technology combination. *Opto-Electron Eng* **52**, 250225 (2025).
17. Lee Y, Choi K, Kim JE et al. Integrating, validating, and expanding information space in single-molecule surface-enhanced Raman spectroscopy for biomolecules. *ACS Nano* **18**, 25359–25371 (2024).
18. Benz F, Schmidt MK, Dreismann A et al. Single-molecule optomechanics in "picocavities". *Science* **354**, 726–729 (2016).
19. Yang SK, Dai XM, Stogin BB et al. Ultrasensitive surface-enhanced Raman scattering detection in common fluids. *Proc Natl Acad Sci USA* **113**, 268–273 (2016).
20. Shegai T, Vaskevich A, Rubinstein I et al. Raman spectroelectrochemistry of molecules within individual electromagnetic hot spots. *J Am Chem Soc* **131**, 14390–14398 (2009).
21. Wang X, Huang SC, Hu S et al. Fundamental understanding and applications of plasmon-enhanced Raman spectroscopy. *Nat Rev Phys* **2**, 253–271 (2020).
22. Ding QQ, Chen XY, Jia YL et al. Timeshare surface-enhanced Raman scattering platform with sensitive and quantitative mode. *Opto-Electron Adv* **9**, 250269 (2026).
23. Ding QQ, Wang J, Chen XY et al. Quantitative and sensitive SERS platform with analyte enrichment and filtration function. *Nano Lett* **20**, 7304–7312 (2020).
24. Schmidt MM, Farley EA, Engevik MA et al. High-speed spectral characterization of single-molecule SERS fluctuations. *ACS Nano* **17**, 6675–6686 (2023).
25. Bi XY, Czajkowsky DM, Shao ZF et al. Digital colloid-enhanced Raman spectroscopy by single-molecule counting. *Nature* **628**, 771–775 (2024).
26. Zhang H, Yang LK, Zhang M et al. A statistical route to robust SERS quantitation beyond the single-molecule level. *Nano Lett* **24**, 11116–11123 (2024).
27. Bi XY, He ZC, Luo ZW et al. Digital colloid-enhanced Raman spectroscopy for the pharmacokinetic detection of bioorthogonal drugs. *Chem Sci* **15**, 13998–14008 (2024).
28. de Albuquerque CDL, Sobral-Filho RG, Poppi RJ et al. Digital protocol for chemical analysis at ultralow concentrations by surface-enhanced Raman scattering. *Anal Chem* **90**, 1248–1254 (2018).
29. Thrift WJ, Ragan R. Quantification of analyte concentration in the single molecule regime using convolutional neural networks. *Anal Chem* **91**, 13337–13342 (2019).
30. Liu HM, Lei Y. A critical review: recent advances in "digital" biomolecule detection with single copy sensitivity. *Biosens Bioelectron* **177**, 112901 (2021).
31. Xie ZL, Meng C, Yue DH et al. Tip-enhanced Raman scattering of glucose molecules. *Opto-Electron Sci* **4**, 240027 (2025).
32. Etchegoin PG, Meyer M, Le Ru EC. Statistics of single molecule SERS signals: is there a Poisson distribution of intensities? *Phys Chem Chem Phys* **9**, 3006–3010 (2007).
33. Le Ru EC, Etchegoin PG, Meyer M. Enhancement factor distribution around a single surface-enhanced Raman scattering hot spot

and its relation to single molecule detection. *J Chem Phys* **125**, 204701 (2006).
34. Neumann O, Zhou JY, Ju YL et al. Surface-enhanced Raman spectroscopy: from the few-analyte limit to hot-spot saturation. *J Phys Chem C* **128**, 8649–8659 (2024).
35. Fang Y, Seong NH, Dlott DD. Measurement of the distribution of site enhancements in surface-enhanced Raman scattering. *Science* **321**, 388–392 (2008).
36. Zhou ZR, Liu WX, Yao ZY et al. Uniform and dense hotspots in Au rough-nanocube monolayer for sensitive and reproducible SERS detection. *ACS Appl Nano Mater* **7**, 17009–17016 (2024).
37. Wang KQ, Sun DW, Pu HB et al. Stable, flexible, and high-performance SERS chip enabled by a ternary film-packaged plasmonic nanoparticle array. *ACS Appl Mater Interfaces* **11**, 29177–29186 (2019).
38. dos Santos DVN, dos Santos DP. Correlating single-molecule SERS statistics to plasmonic nanocluster structure and near-field properties. *Quím Nova* **42**, 1029–1037 (2019).
39. Le Ru EC, Etchegoin PG. Phenomenological local field enhancement factor distributions around electromagnetic hot spots. *J Chem Phys* **130**, 181101 (2009).
40. Shin H, Oh S, Hong S et al. Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS Nano* **14**, 5435–5444 (2020).
41. Thrift WJ, Ronaghi S, Samad M et al. Deep learning analysis of vibrational spectra of bacterial lysate for rapid antimicrobial susceptibility testing. *ACS Nano* **14**, 15336–15348 (2020).
42. Cui FY, Yue Y, Zhang Y et al. Advancing biosensors with machine learning. *ACS Sens* **5**, 3346–3364 (2020).
43. Chen XJ, Xie LY, He YH et al. Fast and accurate decoding of Raman spectra-encoded suspension arrays using deep learning. *Analyst* **144**, 4312–4319 (2019).
44. Bi XY, Lin L, Chen Z et al. Artificial intelligence for surface-enhanced Raman spectroscopy. *Small Methods* **8**, 2301243 (2024).
45. Rho E, Kim M, Cho SH et al. Separation-free bacterial identification in arbitrary media via deep neural network-based SERS analysis. *Biosens Bioelectron* **202**, 113991 (2022).
46. Leong YX, Lee YH, Koh CSL et al. Surface-enhanced Raman scattering (SERS) taster: a machine-learning-driven multireceptor platform for multiplex profiling of wine flavors. *Nano Lett* **21**, 2642–2649 (2021).
47. Lin RH, Zhai YF, Xiong CX et al. Inverse design of plasmonic metasurfaces by convolutional neural network. *Opt Lett* **45**, 1362–1365 (2020).
48. Hong Y, Zhu L, Zhang B, et al. Fabrication of an optoplasmonic raft with improved SERS performance detecting methamphetamine through bubble enrichment. *ACS Appl Mater Interfaces* **16**, 5245–5254 (2024).
49. Yu BR, Ge MH, Li P et al. Development of surface-enhanced Raman spectroscopy application for determination of illicit drugs: towards a practical sensor. *Talanta* **191**, 1–10 (2019).
50. Liu XF, Ma JM, Jiang PF et al. Large-scale flexible surface-enhanced Raman scattering (SERS) sensors with high stability and signal homogeneity. *ACS Appl Mater Interfaces* **12**, 45332–45341 (2020).
51. Chen M, Luo W, Liu Q et al. Simultaneous in situ extraction and fabrication of surface-enhanced Raman scattering substrate for reliable detection of Thiram residue. *Anal Chem* **90**, 13647–13654 (2018).
52. Hong LH, Yang HY, Zhang JZ et al. Coulomb attraction driven spontaneous molecule-hotspot pairing enables universal, fast, and large-scale uniform single-molecule Raman spectroscopy. *Opto-Electron Adv* **8**, 240309 (2025).
53. Bell SEJ, Charron G, Cortés E et al. Towards reliable and quantitative surface-enhanced Raman scattering (SERS): from key parameters to good analytical practice. *Angew Chem Int Ed* **59**, 5454–5462 (2020).
54. Yi J, You EM, Liu GK et al. AI-nano-driven surface-enhanced Raman spectroscopy for marketable technologies. *Nat Nanotechnol* **19**, 1758–1762 (2024).
55. Xu YK, Aljuhani W, Zhang YR et al. A practical approach to quantitative analytical surface-enhanced Raman spectroscopy. *Chem Soc Rev* **54**, 62–84 (2025).
56. Yi J, You EM, Hu R et al. Surface-enhanced Raman spectroscopy: a half-century historical perspective. *Chem Soc Rev* **54**, 1453–1551 (2025).

Acknowledgements

The authors gratefully acknowledge financial support from the Beijing Natural Science Foundation (No. Z240005), National Key Research and Development Program of China (No. 2021YFA1400800), National Natural Science Foundation of China (No. 12374028 and 12474033), Extreme Light Field Manufacturing Science and Engineering (No. 52488301), Research on Water-guided Laser Processing System and Key Technologies for Superhard Materials (No. 231723002), Natural Science Foundation for Youths of Inner Mongolia Autonomous Region (No. 2026QC0418), Elite Revitalizing Inner Mongolia Program (No. 2025TGL05), Interdisciplinary Project of Capital Normal University (No. 2026JCYY04).

Author contributions

This endeavor was accomplished through the collective efforts of all authors: Zhipeng Li and Hongxing Xu designed and supervised the research. Meng Zhang, Yue Wei and Longkun Yang performed the experiments. Hui Ding, Luyuan Xie and Shiyu Zhang carried out the deep learning model training/testing. Zhipeng Li, Chaowei Sun, Di Zheng, Ferruccio Pisanello and Longkun Yang analyzed the data. Longkun Yang, Zhipeng Li and Meng Zhang wrote the paper. All authors reviewed the paper.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at
https://doi.org/10.29026/oea.2026.260163

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. To view a copy of this license, visit
http://creativecommons.org/licenses/by/4.0/
©The Author(s) 2026.
Published by Editorial Office of *Opto-Electronic Advance*, Institute of Optics and Electronics, Chinese Academy of Sciences.

Opto-Electronic Journals Group
www.oejournal.org
Scan for Article PDF