ACS NANO

www.acsnano.org

ARTICLE

Identifying Surface-Enhanced Raman Spectra with a Raman Library Using Machine Learning

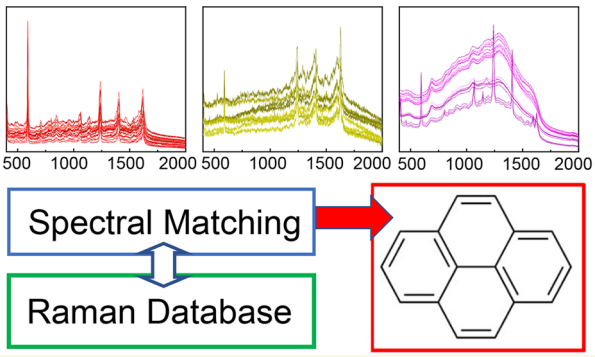
Yilong Ju,[□] Oara Neumann,[□] Mary Bajomo, Yiping Zhao, Peter Nordlander, Naomi J. Halas,* and Ankit Patel*

Cite This: *ACS Nano* 2023, 17, 21251−21261

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Since its discovery, surface-enhanced Raman spectroscopy (SERS) has shown outstanding promise of identifying trace amounts of unknown molecules in rapid, portable formats. However, the many different types of nanoparticles or nanostructured metallic SERS substrates created over the past few decades show substantial variability in the SERS spectra they provide. These inconsistencies have even raised speculation that substrate-specific SERS spectral libraries must be compiled for practical use of this type of spectroscopy. Here, we report a machine learning (ML) algorithm that can identify chemicals by matching their SERS spectra to those of a standard Raman spectral library. We use an approach analogous to facial recognition that utilizes feature extraction in the presence of multiple nuisance variables for spectral recognition. The key element is a metric we call "Characteristic Peak Similarity" (CaPSim) that focuses on the characteristic peaks in the SERS spectra. It has the flexibility to accommodate substrate-specific variability when quantifying the degree of similarity to a Raman spectrum. Analysis shows that CaPSim substantially outperforms existing spectral matching algorithms in terms of accuracy. This ML-based approach could greatly facilitate the spectroscopic identification of molecules in fieldable SERS applications.



KEYWORDS: *surface-enhanced Raman scattering, machine learning, nanoparticles, polycyclic aromatic hydrocarbons, characteristic peak similarity*

INTRODUCTION

Raman spectroscopy is a vibrational spectroscopy that provides molecular structure information based on the specific spectral features of inelastic light scattered from molecules.[1−3] Surface-enhanced Raman spectroscopy (SERS) is performed when molecules are positioned on a metallic nanoparticle- or nanostructure-based (plasmonic) substrate that significantly enhances the spectral signal from those molecules, typically by several orders of magnitude. Due to this greatly enhanced sensitivity and straightforward sample preparation, SERS has continued to be developed as a powerful technique for applications such as live cell and drug analysis, planetary rock examination, and pollutant detection.[4−9] However, a major drawback of SERS is that it displays large spectral variabilities due to the substrate's specific local electromagnetic properties (i.e., nanoparticle type, size, shape, and refractive index of the surrounding medium) and the analyte molecules' orientation, concentration, and other related factors (e.g., frequency of incident light, laser power, and exposure time).[10−14] With many laboratories worldwide developing and using many different types of SERS substrates,[2] it could be challenging to identify analyte molecules using a single standard spectral library due to substrate-specific nuisance variation. Such variation is task irrelevant and is caused by nuisance variables such as the intensity of background fluorescence and noise and the location, intensity, and even existence of a spurious peak or peaks. It has been proposed that for SERS to become a mainstream tool for rapid chemical identification it is necessary to create SERS libraries specific to each SERS substrate.

With vast Raman spectral libraries already available, it is worthwhile to examine whether one can circumvent the issues of SERS substrate specificity that arise when using a variety of different SERS-based substrates for chemical identification (i.e., NP aggregates (hot spots) and complex 3D nanostruc-

Received: June 18, 2023
Revised: October 19, 2023
Accepted: October 23, 2023
Published: November 1, 2023

Downloaded via NATL RESEARCH COUNCIL CANADA on May 12, 2025 at 18:24:02 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications © 2023 American Chemical Society

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

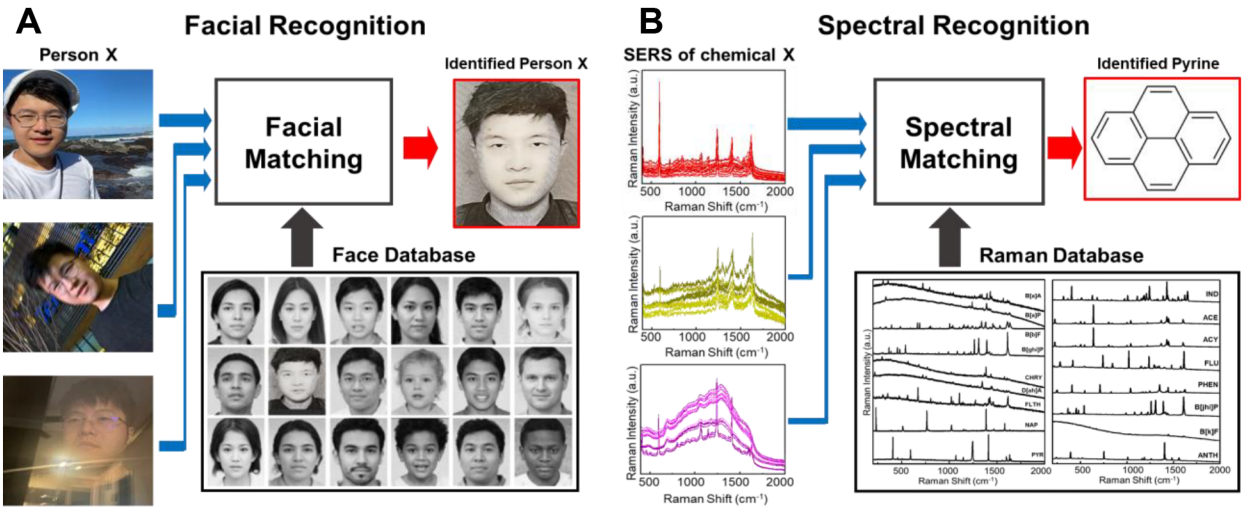


Figure 1. Schematic illustration of facial and spectral recognition. (A) Identifying a person X involves a facial matching algorithm to extract facial features and a face database for face matching. The algorithm should be robust to variations in the orientation, brightness, etc., in the query of face images and be able to match them to a portrait of person X in the database. (Face Database composed of AI-generated images: https://generated.photos/faces. Figure 1A contains 4 photos of Yilong Ju, taken with permission. Copyright belongs to Yilong Ju.) (B) Similarly, identifying an unknown molecule X using its SERS spectrum involves a spectral matching algorithm to extract spectral features and a Raman database. The spectral matching algorithms should be robust to variations in SERS substrates and able to match any query spectrum to the Raman spectrum of the corresponding chemical. (For this illustration, the faces in the face database, except Yilong Ju, the first author of this paper, have been artificially generated).

tures with sharp corners or junctions).[12,13,15−21] These nanostructured geometries each offer their own specific near-field properties. Here, we report a machine learning (ML) approach for identifying unknown chemicals from their SERS spectrum using a standard Raman library. A similarity score between the unknown SERS "query" spectrum and each spectrum in the Raman library is computed, and the unknown chemical is identified based on the relative magnitude of the scores. We call this process Spectral Recognition, in analogy with Facial Recognition (Figure 1). Past efforts have largely focused on applying spectral recognition algorithms to match unknown Raman spectra with Raman libraries (i.e., matching Raman to Raman or SERS to SERS)[22−29] and have frequently employed various similarity metrics, such as Pearson's correlation, Euclidean distance, cosine similarity, and other methods.[30−39] These approaches are poorly suited for dealing with different spectroscopic variants such as Raman and SERS. These metrics tend to be sensitive to nuisance variations such as noise and uncharacteristic background peaks, which can be induced by substrate type and molecular adsorbate conformation. When matching SERS to Raman spectra, a robust method must prioritize the characteristic peaks (CPs) corresponding to the underlying vibrational modes for accurate spectral recognition while suppressing spurious information obtained from nuisance variations.

Recently, an unsupervised ML strategy for demixing a complex mixture of SERS spectra of various molecules was proposed,[40] which relies on Characteristic Peak Extraction (CaPE). CaPE is a spectral feature extraction algorithm that yields characteristic SERS peaks based on counts of detected peaks from spectra of a mixture, taking relatively small frequency shifts into consideration. In this work, we leverage CaPE to accommodate the peak shifts that we observe in our data (typically 1−8 cm^{-1}) and propose a similarity metric, namely, Characteristic Peak Similarity (CaPSim). It focuses only on the locations of CPs when computing the similarity score and, hence, is robust to most nuisance variations in SERS spectra. Using CaPSim, we develop a spectral recognition algorithm that can match the query (unknown analyte) SERS spectrum to a library of standard (known) Raman spectra with high accuracy.

RESULTS AND DISCUSSION

Design of the Spectral Recognition Algorithm. A facial recognition algorithm (Figure 1A) identifies an unknown person based on facial similarity and characteristics among an extensive collection of well-known faces in a face database. The facial matching algorithm should be able to extract robust characteristic facial features from images, ignoring or filtering out nuisance variations in the input photos. This includes but is not limited to changes in brightness, orientation, ethnicity/gender, facial expression, etc. The algorithm then identifies a person by matching the input images to a standardized image (e.g., a forward-facing portrait).[41] Analogously, we developed a SERS spectral recognition algorithm where a query SERS spectrum (of an unknown analyte) will be matched to one of many known Raman spectra in a library (database) (Figure 1B). This involves the analysis of multiple Raman spectra and determining their CP locations.

SERS Substrates. Plasmonic nanostructure-based SERS substrates play an important role as a Raman signal amplifier in SERS. In this study, three different kinds of plasmonic substrate geometries were prepared: gold colloid (Au NPs), silica−gold core−shell "nanoshell" nanoparticles (Au NSs), and oriented silver nanorods (Ag NRs). Each of these NP geometries offers its own characteristic near-field optical properties, based on its specific geometry. The Au NSs (SiO_2 radii 60 nm and Au shell thickness 14 nm) in aqueous solutions have a plasmon resonance at 785 nm, overlapping with the excitation laser wavelength used in the SERS measurements. Monomer Au NPs have a maximum plasmon resonance at 550 nm, and we tune the plasmon resonance

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

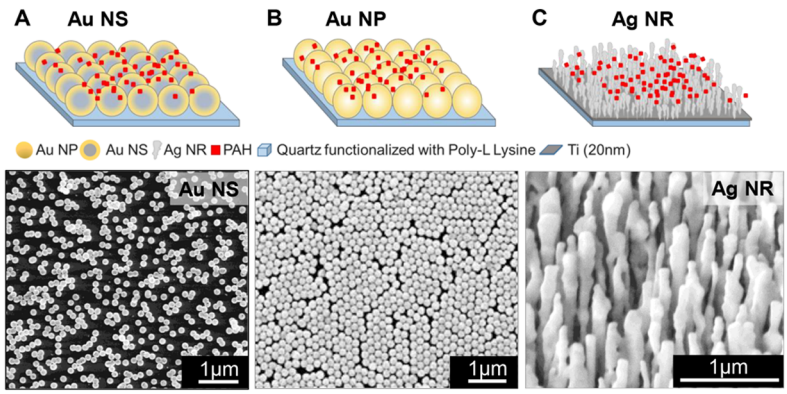


Figure 2. SEM Image of the SERS substrates. The three SERS substrate geometries are (A) Au NSs, (B) Au NPs, and (C) Ag NRs. Top: Schematic representations depicting substrate plus analyte. Bottom: Scanning electron microscopy (SEM) images showing the morphology of the nanostructures.

toward the near IR (NIR) region by creating Au NP aggregates. Therefore, the SERS signal is enhanced by the field enhancement in the gap "hot spot" regions between closely packed Au NP aggregates. Extinction spectra of aqueous solutions of Au NSs and Au NPs and Au NP aggregates on a quartz substrate are presented in Figure S1, Supporting Information (SI). The aligned Ag NR array has an average diameter of approximately 80 nm at the tips, an average rod length of 1000 nm, and a broad absorption peak in the NIR region.[42] All of these three SERS-based substrate geometries provide similar electric field enhancement factors (EFs), essential for enhanced SERS sensitivity.[10,12,21,42−44]

Scanning electron microscopy (SEM) was performed to image the substrate morphologies (Figure 2). The chemically synthesized Au NSs have a plasmon resonance at 785 nm, overlapping with the excitation laser wavelength used in the SERS measurements (Figure 2A). One significant advantage of Au NSs is the tunability of plasmon resonance, attained by fine-tuning the geometrical parameters of both the shell and the core. Chemically synthesized Au NPs (Figure 2B) were adsorbed onto a quartz substrate in an aggregated state using poly-L-lysine (PLL) to generate a junction that provides enhanced local optical fields that will enhance the SERS signal. The aligned Ag NR array (Figure 2C) was prepared by an oblique angle-deposition technique on a quartz surface at a fixed angle in the range of 78°−85°.[42,45−49] The average diameter of the Ag NRs measured at the tips is approximately 80 nm, and the average rod length is nominally 1000 nm.

Peak Extraction from SERS and Raman Spectra. Just as facial recognition relies on feature extraction from images, the key to spectral recognition lies in the extraction of characteristic spectral features. For SERS/Raman spectra, such features are essentially CPs that correspond to the underlying Raman active vibrational modes of the analytes. Therefore, we focus on extracting peaks for the query (SERS spectrum of the unknown chemical) and the references (Raman spectra in the library) (Figure 3). For preprocessing, we first apply a baseline removal algorithm to remove slow-moving trends in the data because they are typically substrate specific.[50] Then, we apply a l_2-normalization to each spectrum by dividing the data by its l_2-norm, making the data scale consistent. We show the preprocessed reference Raman spectrum (Figure 3(i)) and SERS spectra on the Au NS, Ag NR, and Au NP substrates, respectively, in Figure 3(iii, v, vii), for the four molecules

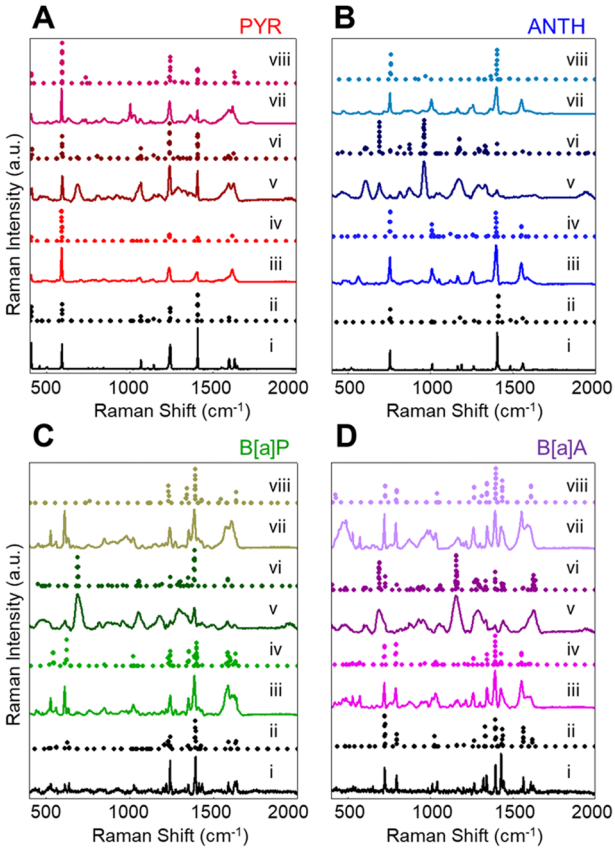


Figure 3. Examples of raw PAH spectra and extracted peaks. (A) Pyrene, (B) anthracene, (C) benzo[a]pyrene, (D) benzo[a]-anthracene. Each panel shows (i) the Raman spectrum, the SERS spectrum of the analyte molecule on (iii) the Au NS substrate, (v) the Ag NR substrate, (vii) the Au NP substrate, and (ii, iv, vi, and viii) the corresponding extracted peaks. Spectra after removing baselines are shown as solid lines, while the peaks extracted using CaPE are shown as dotted lines.

pyrene (PYR) (Figure 3A), anthracene (ANTH) (Figure 3B), benzo[a]pyrene (B[a]P) (Figure 3C), and benzo[a]anthracene (B[a]A)) (Figure 3D). For each PAH molecule, 25−30 SERS spectra were collected from different areas of the substrates to account for the spectral and substrate variability needed for the spectral recognition algorithms. Following data acquisition and

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

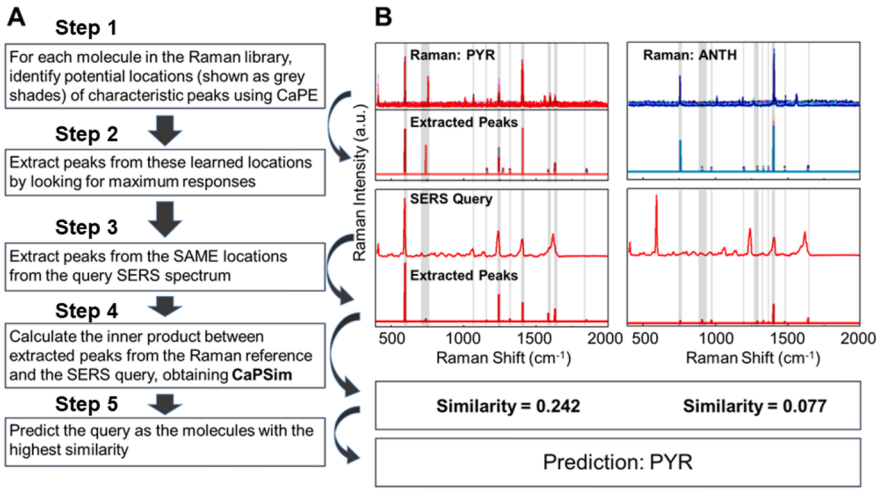


Figure 4. Flowchart of the proposed spectral recognition algorithm. (A) Step-by-step procedure for obtaining the proposed similarity metric and characteristic peak similarity (CaPSim) and using it for spectral matching. (B) Visualization of the corresponding steps including two matchings: Raman reference PYR (Pyrene) vs SERS query PYR (Pyrene) and Raman reference ANTH (Anthracene) vs SERS query PYR (Pyrene). The corresponding CaPE-extracted peaks are shown below each set of Raman/SERS spectra.

preprocessing, we used CaPE to extract the CPs.[40] The extracted CPs of pyrene, anthracene, benzo[a]pyrene, and benzo[a]anthracene on all three SERS substrates are presented in Figure 3 (ii, iv, vi, and viii) (dotted lines). By comparing the wavenumber of known characteristic peaks[51−54] and CaPE-extracted peaks, all peaks present in the Raman spectra match the peak locations of the corresponding PAH SERS spectra very well in the case of pyrene (Figure 3A) and anthracene (Figure 3B) on all three substrates. In the case of the Ag NR SERS-based substrates, several spectral variations from their Raman spectra were observed. For example, the broad pyrene peak at ∼700 cm^{-1} (Figure 3A (i,v)) corresponding to the C−H out-of-plane and C−C out-of-plane bending modes is presented in the SERS spectrum but not the Raman spectrum.[52] The anthracene peaks at ∼1557 cm^{-1} (C−C stretching and C−H bending) and 1621 cm^{-1} (C−C asymmetric stretching and C−H bending) modes (Figure 3B (i,v)) were not present in the SERS spectrum, while the peak at ∼690 cm^{-1} was present in the SERS spectrum but not in the Raman spectrum. Several factors can induce SERS spectral deviation from the Raman spectra, such as weak vibrational modes undetectable with Raman but enhanced in SERS; interaction between the analyte molecules and the Ag NR substrate can introduce, suppress, or broaden vibrational modes; or molecular configuration on the metal surface. For the benzo[a]anthracene (Figure 3C) and benzo[a]pyrene (Figure 3D) spectra, the SERS spectra obtained using the Ag NR array possessed several minor features not present in the CaPE-extracted CP data: this may be due to the highly fluorescent background and air or solvent contaminants.[55−58] For example, the SERS peaks of benzo[a]anthracene at 1299 cm^{-1} assigned to C−C stretching and C−H in-plane bending modes are absent in the Raman spectrum, while the 793 cm^{-1} peak appears only in the Raman data. Nevertheless, despite some discrepancies between the SERS and Raman spectra of pyrene, anthracene, benzo[a]pyrene, and benzo[a]anthracene, most of the intense SERS peaks of PAHs on all three substrates match the CP data very well. Therefore, CaPE shows high accuracy in extracting the CP locations and peak intensities. We subsequently performed spectral recognition by matching the query to each reference chemical in the library, focusing only on the characteristic features and eliminating the influence of the remaining baselines and noisy peaks.

Spectral Recognition with Characteristic Peak Similarity (CaPSim). Although there are many similarity metrics in the literature, SERS−Raman spectral recognition remains a challenge because the spectra may contain irrelevant spectral features.[30] For example, slow-changing trends may correspond to the underlying substrate and persist even after removing the baseline since no algorithm guarantees complete removal, not to mention low-intensity noisy peaks that are not suppressed by the baseline removal algorithm and may vary randomly in every measurement. Therefore, we designed a CaPSim algorithm that tackles these issues (Figure 4). In Figure 4A, we show a flowchart explaining step-by-step the CaPSim algorithm mechanism, and in Figure 4B, we show the corresponding visualizations and a comparison between a good match with a high similarity metric (SERS PYR vs Raman PYR) and a poor match with a substantially lower similarity metric (SERS PYR vs Raman ANTH). Our guiding strategy is that if the potential locations of the characteristic features are known the matching algorithm only needs to inspect and compare "close-ups" of the query and reference spectra at these locations. In practice, we first used CaPE to learn these potential CP locations from the reference Raman spectra, as shown in Figure 4A, Step 1, and Figure 4B. Gray shades represent the learned potential CP locations. Note that some shades have significant width due to peak shifting.[30,31] CaPSim deals with this issue by extracting the peak intensities (Figure 4A Step 2 and extracted peaks on the right) using max pooling, a technique widely used in convolutional neural networks, which makes the models invariant to translations in input data when the amount of translation is not greater than the size of the pooling window.[59] This window size is a parameter of the CaPE algorithm and can be tuned using a training data set in order to properly deal with most of the shifts in the data. The key step (Figure 4A Step 3 and SERS extracted peaks on the right (Figure 4B)) is where we use the potential CP locations learned from reference Raman spectra to extract peaks from the query SERS spectrum. When the query chemical matches

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

the reference chemical, the response, indicated by the intensity of the extracted peaks, is likely to be significantly higher (Figure 4A, Steps 4 and 5) than the case where the query does not match the reference spectrum. The response approaches 0 as peaks are sought in the wrong locations. We can then obtain a similarity metric naturally by taking the inner product between the extracted peaks from the query and the reference, as shown in the bottom part of Figure 4. Indeed, in the case of a poor match (Figure 4B, bottom row), the extracted peaks are all very small except for one that happens to be shared between PYR and ANTH. This leads to a clear difference (>3×) in the similarity score.

Comparing CaPSim to Existing Similarity Metrics. Next, we quantitatively evaluated CaPSim and compared it to other existing similarity metrics. We performed our quantitative evaluation using a Raman library of 16 chemicals (see Materials for the complete list) (Figure 5). For the queries, nine chemicals are in the SERS data using a Au NS substrate, while SERS data using Au NP and Ag NR substrates contain four chemicals. For each chemical, we have 25 to 30 spectra recordings (both Raman and SERS). Thus, we performed multiple tests for each query chemical using these recordings. For each query spectrum, when matching it to a specific chemical in the Raman library, we also averaged its similarity to every recording of the reference Raman spectra. We observe that the standard deviation of the similarity values over all recordings of one compound is typically 10% of the mean value. Hence, averaged values led to a more stable and reliable estimation of the similarity. Our evaluation metric is top-k accuracy. For each query, if the correct answer shows up in the top k of the returned list of matching chemicals, it counts as a success; otherwise, it is a failure (see Methods for details). Figure 5 illustrates the average of the top-k accuracies of all SERS PAHs on each substrate (Au NS, Au NP, and Ag NR). Each panel shows the accuracy vs k, where $k = 1,\cdots, 5$, and each color represents a different similarity metric. Our proposed CaPSim algorithm is shown in red. CaPSim outperforms every other existing similarity metric we studied, especially in top-1 accuracy. For example, the top-1 matching accuracy using the best existing similarity metric is improved from 69.8% to 82.4% on the Au NP substrate, from 61.9% to 76.8% on the Au NS substrate, and from 50.0% to 75.0% on the Ag NR substrate. Top-1 accuracy is essential because it can serve as a proxy for matching performance in settings where we have a far larger Raman reference library. Since we have access to only a Raman library of 16 chemicals here, the top-k saturates quickly as k grows. In addition to CaPSim's high accuracy, it is also efficient. CaPSim processes each query in 1.7 ms—only 17% more time than a plain dot product—and it scales only linearly with the number of compounds in the library. For very large libraries, we can also borrow techniques from mature and successful subfields of computer science/machine learning such as GPU acceleration, parallelization, and approximate nearest neighbors to accelerate this process.[60−62] Note that for all methods considered in this paper we do not deal with

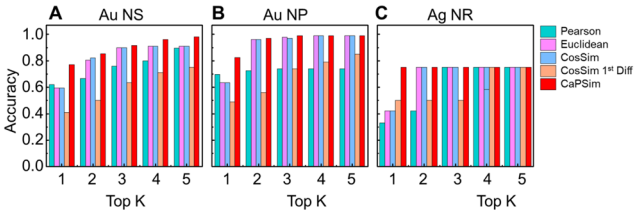


Figure 5. Comparison of the CaPSim performance to the existing metrics. (A) Au NS, (B) Au NP, and (C) Ag NR substrates. Top-k accuracy is a way to measure the performance of a similarity metric's performance. We look at the top-k matching and check whether the correct PAH is among them. For each substrate, the accuracy is averaged over all available PAHs on that substrate. The following PAHs are included: (A) pyrene, anthracene, benzo[a]-anthracene, benzo[a]pyrene, chrysene, fluoranthene, Indeno-[123cd]perylene, naphthalene, and phenanthrene on a Au NS SERS substrate; (B) pyrene, anthracene on a Au NP SERS substrate; (C) pyrene, anthracene, benzo[a]anthracene, and benzo[a]pyrene on a Ag NR SERS substrate.

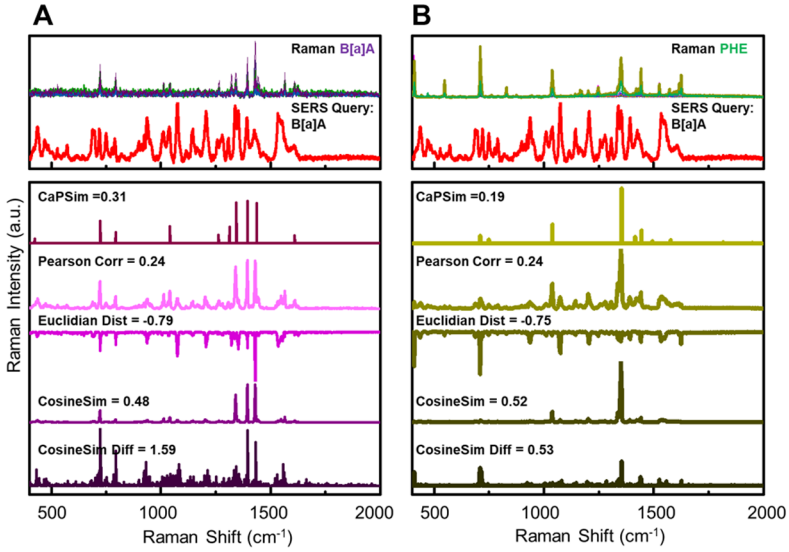


Figure 6. Comparison of similarity attribution between CaPSim and existing similarity metrics. Left: Matching between Raman benzo[a]anthracene and a SERS benzo[a]anthracene. Right: Matching between Raman benzo[a]anthracene and a SERS pyrene. Note that we consider negative Euclidean distance as a similarity score (higher is a better match).

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

queried compounds that are not present in the library. By setting a proper minimum threshold for the similarity score, we can identify those queries that do not match any compound in the library. We leave this as a topic for future work.

Close Inspection of the Calculation of CaPSim and Existing Similarity Metrics. In Figure 6, we show the performance of five similarity metrics (e.g., CosineSim Diff, CosineSim, Pearson Corr, Euclidean Dist, and CaPSim). For each similarity metric, we show how much each specific part of the spectrum contributes to the final similarity score (see Methods for details). Two matches are shown in the top panels, indicating a good match between Raman and SERS benzo[a]anthracene (CaPSim = 0.311, Pearson Corr = 0.240, Euclidean Dist = −0.787, CosineSim = 0.476, Cosine Sim Diff = 1.588) and an imperfect match between Raman phenanthrene (PHE) and SERS benzo[a]anthracene (CaPSim = 0.188, Pearson Corr = 0.238, Euclidean Dist = −0.747, CosineSim = 0.520, CosineSim Diff = 0.530). In the bottom panels, we show the contribution at each wavenumber, whose sum across all wavenumbers equals the similarity score. In this example, we see that Euclidean distance and cosine similarity produce poor matching, while CaPSim, Pearson's correlation coefficient, and first-order different cosine similarity lead to correct matching (see Methods for a detailed definition and attribution approaches of these similarity metrics).

An error analysis of the similarity attribution reveals that existing metrics focus on irrelevant locations, mainly noise, leading to mismatches for some similarity metrics (e.g., CosineSim and EuclideanDist). Specifically, between Raman and SERS benzo[a]anthracene, we find that Euclidean distance penalizes the peak at ∼1,400 cm^{-1} too much since, at major peaks, a slight horizontal shift of the peak location will lead to a significant change in the vertical signal intensity. As for the cosine similarity, we can see that it fails to assign sufficiently large similarity to the modest peaks at ∼800 cm^{-1}, 950 cm^{-1}, 1,200 cm^{-1}, and 1,600 cm^{-1}, which are all common features between the Raman and SERS spectra, as we can see from the top panel. It is worth noting that although Pearson's correlation coefficient produces the correct result the similarity scores are very close for good and bad matching. A closer inspection shows that it is affected by the interaction between the low-intensity peaks in the query and the background noise in the reference spectra. Of all similarity metrics investigated in this study, CaPSim produces the cleanest attribution that is not affected by unimportant peaks and background noise, making it much more interpretable. Due to its high performance and interpretability, CaPSim allows for matching spectra of different types with high confidence, irrespective of the measurement settings/context, enabling accurate spectral recognition. We also perform an ablation study in which we remove important steps of CaPSim one at a time to estimate their contribution to the final performance. The result is shown in Figure S2, SI. We found that baseline removal improves the matching accuracy a lot, even for the plain dot product. On top of it, adding the ability to extract the characteristic peaks already boosts the spectral recognition performance to a level above CosSim. Although smoothing or max pooling does not seem to improve the performance on its own, their collective effect notably enhances the matching accuracy.

CONCLUSIONS

In analogy with facial recognition, we developed a spectral recognition algorithm that can identify an unknown chemical from a Raman library based on its SERS spectrum. Its key strategy, CaPSim, focuses only on the locations of CPs when computing the similarity score. Hence, it is robust to most variations that appear frequently in SERS, such as noise and uncharacteristic background peaks. We compared CaPSim to several existing similarity metrics that we applied to the same spectra, and CaPSim outperforms all others regarding the top-1 to top-5 accuracy. Given the expansive Raman databases currently available, we believe that this approach will more rigorously connect the various SERS substrate-specific spectra obtained by multiple research groups with these currently available libraries. This demonstration focuses on small molecules, such as those frequently of interest in environmental sensing. There will, of course, be limitations in the identification and differentiation of molecules with extremely similar molecular structure, as with many analytical approaches. The challenges of biomolecule detection and characterization and the identification of even larger structures, such as viruses and bacteria, are not addressed in this proof-of-principle study. The direction reported here is essential for advancing the utility of SERS as a research tool and as a sensing modality that yields chemically specific information. We foresee substantial opportunities and benefits of further testing, improving and refining this algorithm and applying it to other surface-enhanced spectroscopies, such as surface-enhanced IR absorption spectroscopy,[63−65] and other procedures that require spectral recognition, such as identifying chemicals from mixtures across different types of spectroscopies.[40]

MATERIALS AND METHODS

Materials. (3-Aminopropyl) triethoxysilane (ES, 99%), tetrachloroauric acid ($HAuCl_4{\cdot}3H_2O$), tetrakis hydroxymethyl phosphonium chloride, poly-L-lysine hydrobromide (MW 150,000−300,000), poly-L-lysine (PLL) pyrene, anthracene, benzo[a]anthracene, benzo[a]pyrene, chrysene, fluoranthene, indeno[123cd]perylene, naphthalene, and phenanthrene were purchased from Sigma-Aldrich. Formaldehyde (37%), sulfuric acid (H_2SO_4, 100%), hydrogen peroxide (H_2O_2, 30%), potassium dihydrogen phosphate (KH_2PO_4), and 200 proof ethanol were obtained from Fisher Scientific. All chemicals were used as received without further purification. Water was deionized and filtered with a Milli-Q water system (18.2 MΩ·cm at 25 °C, Millipore). Quartz slides were purchased from Fisher Scientific. The 120 nm diameter aminated SiO_2 cores in ethanol (10 mg/mL) were purchased from nanoComposix, Inc., San Diego, CA.

Synthesis of Nanoparticles. *Au Nanoshells (Au NSs).* The Au NSs were synthesized using a bottom-up wet chemistry synthesis.[66] At first, small Au colloids (1−3 nm) in size were synthesized by reducing chloroauric acid using tetrakis (hydroxymethyl) phosphonium chloride (THPC) as a reducing agent. Then, 300 μL aminated SiO_2 cores were incubated with 20 mL of THPC-Au solution for 24 h to enable the attachment of the Au colloids onto the SiO_2 surface, followed by multiple cleaning cycles to remove the excess reaction chemicals. Following this step, an electroless plating process reduced Au onto the SiO_2 surface, forming a continuous Au shell. The electroless deposition was achieved by reducing Au from a 3 mL solution of 1.8 mM potassium carbonate and 0.4 μM chloroauric acids with 15 μL of formaldehyde. The Au NSs were fabricated with the core and shell radii $[r_1, r_2] = [63, 75]$ nm, corresponding to a Au NS plasmon resonance of 785 nm (Raman laser wavelength) in H_2O.

Au Nanoparticles (Au NPs). The Au NPs were fabricated by reducing 30 mL of the aged chloroauric acid solution (3 mL of 25 mM $HAuCl_4$ with 50 mg of K_2CO_3 dissolved in 200 mL of water and aged in the dark for 12 h before being used) with 200 μL of formaldehyde (37%) under rapid stirring for 30 min. Au NPs were

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

redispersed in 20 mL of water. The concentration of the Au NPs was estimated using absorbance measurements and Mie scattering theory.

Ag NRs. The Ag NRs were fabricated through the oblique angle-deposition (OAD) process in a custom-built electron beam evaporation system.[48] The entire evaporation process was conducted under high vacuum conditions (~3×10^{-6} Torr), and the deposition rate and thickness of the deposited films were monitored by a quartz crystal microbalance. A 20 nm layer of Ti was first deposited on a (9 mm × 9 mm) clean glass substrate at a rate of 0.2 nm/s, followed by a 200 nm Ag film deposited at 0.3 nm/s. Then, the substrate was rotated by 86°, and another 2000 nm layer of Ag was deposited at the same rate. The density of the Ag NRs was approximately 15−25 μm^{-2}. Before use, the Ag NR substrates were cleaned for 2 min in a custom-built inductively coupled radio frequency plasma chamber, operated at 30 W under a constant flow of ultrapure argon with a chamber pressure of ~600 mTorr. Argon plasma cleaning allowed the removal of carbonaceous and organic contaminants during fabrication and storage of the substrates.[55]

SERS Substrate Preparation. Cleaned quartz slides (from Thermo Scientific Quartz microscope slide, fused, 76.2 × 25.4 mm) were immersed in 0.01% w/v aqueous solution of PLL (MW 150,000 to 300,000) for 3 h to facilitate subsequent attachment of nanoparticles on the quartz surface. The surfaces were cleaned with ethanol and dried in a flow of N_2. Cut silicone isolators (Grace Bio-Laboratories Press-To-Seal silicone isolator, No PSA, 9 mm diameter) were placed on the quartz-PLL substrate, followed by a dry drop deposition of 100 μL of 10^{10} part/mL Au NS or NP aqueous solution into the isolator well. Then 20 μL of 100 μM PAHs in acetone solutions were directly dried on the Au NP/NS SERS substrates. Before acquiring the SERS spectra, the substrates were fully immersed in Milli-Q water.

Measurements. SERS measurements were acquired with a Renishaw inVia Raman microscope (Renishaw) with a 785 nm excitation wavelength and 55 μW laser power at the samples. Backscattered light was collected using a 63× water immersion objective lens (Leica) with a 20 s exposure time. The extinction measurements were performed on a Cary 5000 UV/vis/NIR Varian spectrophotometer. SEM measurements were performed by using an FEI Quanta 400 field emission SEM at a 20 kV scanning electron microscope acceleration voltage.

Computational Methods. In this section, we describe the proposed spectral matching strategy, similarity metric, and evaluation procedure in detail.

Notation. Let $q \in \mathbb{R}^D$ be the *query spectra* where D is the dimensionality of spectra; $M(q; \mathcal{S}, \mathcal{L})$ is the molecule matched by a similarity metric $\mathcal{S}: \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ given a Raman library $\mathcal{L}$; and $M(q)$ be the ground truth identity of q. The similarity metric $\mathcal{S}(q, r)$ takes a pair of spectra $q, r \in \mathbb{R}^D$ as an input and returns a scalar representing the similarity between the two. In the case where the library contains multiple spectra for the same molecule, we define $\bar{\mathcal{S}}$ as the mean similarity between q and multiple spectra $R \in \mathbb{R}^{D\times n}$, where n is the number of spectra in R. We have

$$\bar{\mathcal{S}}(q, R) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{S}(q, r_i)$$

where r_i represents the ith column of R, supposing the library $\mathcal{L}$ contains N different molecules denoted by $\mathcal{L} = \{R_1, \cdots, R_N\}$, where $R_j \in \mathbb{R}^{D\times n_j}$, $\forall j \in [N]$ and $[N] = \{1, \cdots, N\}$. We call $R_1, \cdots, R_N$ the reference spectra.

Spectral Matching. Given a query q, a similarity metric $\mathcal{S}$, and a Raman library $\mathcal{L}$, the full spectral matching procedure between q and each $R \in \mathcal{L}$ is:

1. Preprocess q and R. Let $\mathcal{P}$ denote the preprocessing function. We have

$$q \leftarrow \mathcal{P}(q),\ R \leftarrow \mathcal{P}(R)$$

Here if $\mathcal{P}$ only works with a single spectrum, $\mathcal{P}(R)$ is equivalent to first applying $\mathcal{P}$ to each individual spectrum in R and then concatenating the results.

2. Before calculating the similarity, some metrics may require extracting peaks from the spectra. Let $\mathcal{A}$ denote the peak extraction algorithm. We have

$$q \leftarrow \mathcal{A}(q),\ R \leftarrow \mathcal{A}(R)$$

3. Compute the mean similarity between q and each $R \in \mathcal{L}$. We have

$$\bar{\mathcal{S}}(q, R) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{S}(q, r_i)$$

After we perform the above steps for each $R \in \mathcal{L}$, we select the matching molecule based on the similarity values. Here, we use C to denote the choice function, which takes the similarity values between q and every molecule in the library and spits out the prediction. We can write the predicted identity of q as

$$\hat{M}(q; \mathcal{S}, \mathcal{L}) = C(\{\bar{\mathcal{S}}(q, R) | R \in \mathcal{L}\})$$

Next, we introduce our choices for each of the four steps: (1) the preprocessing function $\mathcal{P}$, (2) the peak extraction algorithm $\mathcal{A}$, (3) the similarity function $\mathcal{S}$, and (4) the choice function C.

Preprocessing. We follow a standard preprocessing in related work, where we first use a baseline removal algorithm $\mathcal{B}$ to remove the baselines (i.e., slow-changing trend) from all of our SERS and Raman spectra because these baselines are shared across different molecules and may confuse the spectral matching algorithm.[50,67] We compare the raw spectra and their baseline-removed version in Figure S3 and find that our algorithm is able to remove the trend without introducing spurious peaks.[50] This step can be denoted by $q \leftarrow \mathcal{B}(q)$ and $R \leftarrow \mathcal{B}(R)$. Next, we normalize each spectrum by its l_2-norm. We chose l_2-normalization based on our early exploratory data analysis. Some spectra in the Raman library do not contain any significant characteristic peaks. They look like pure noise and would still have a large intensity if we were using min/max. For peak area-based normalization, we need to choose the peak detection criteria, and we think it would eventually have a similar effect as the l_2-normalization—scaling down the intensity of spectra that contain a lot of noise and hence reduce their impact on learning potential locations for characteristic peaks. Therefore, using l_2-normalization makes our algorithm simple and robust to noisy data. Using $\mathcal{N}$ to denote the normalizing function, we have $\mathcal{N}(x) = x/\|x\|_2$, where x can be q or an individual spectrum $r \in R$. Therefore, the preprocessing function $\mathcal{P}(x) = \mathcal{N}(\mathcal{B}(x))$.

Existing Similarity Metrics $\mathcal{S}$. We consider four similarity metrics that are widely used in the literature: Pearson's correlation, Euclidean distance, cosine similarity, and first-difference cosine similarity.[30] We show their definitions here with the following helper variables: $\bar{q} = \sum_{k=1}^{D} q_k/D$, $\bar{r} = \sum_{k=1}^{D} r_k/D$, $\overline{q^2} = \sum_{k=1}^{D} q_k^2/D$, $\overline{r^2} = \sum_{k=1}^{D} r_k^2/D$, $\overline{qr} = \sum_{k=1}^{D} q_k r_k/D$, $\Delta q_k = q_{k+1} - q_k$ and $\Delta r_k = r_{k+1} - r_k$.

Pearson's correlation:

$$\mathcal{S}_{\text{Pearson}}(q, r) = \frac{\overline{qr} + \bar{q}\bar{r}}{\sqrt{(\overline{q^2} - \bar{q}^2)(\overline{r^2} - r^2)}}$$

Euclidean distance:

$$\mathcal{S}_{\text{Euclidean}}(q, r) = -\|q - r\|_2$$

Cosine similarity:

$$\mathcal{S}_{\text{cossim}}(q, r) = \frac{\overline{qr}^2}{\overline{q^2}\,\overline{r^2}}$$

First-difference cosine similarity:

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

$$\mathcal{S}_{\text{cossim 1st diff}}(q, r) = \frac{\overline{\Delta q \Delta r}^2}{\overline{\Delta q}^2 \overline{\Delta r}^{2'}}$$

We follow Samuel et al. (2021) and use the squared form of cosine similarity and first-difference cosine similarity.[30]

Proposed Similarity Metric: Characteristic Peak Similarity (CaPSim). Due to the nature of (surface-enhanced) Raman spectroscopy, most high-intensity peaks in the spectra of a molecule correspond to specific vibrational modes that can serve as a fingerprint of the molecule. We call these characteristic peaks (CPs). CaPSim is calculated based on CPs, supposing we know the potential location of each individual CP. The high-level idea is that we first use an algorithm $\mathcal{A}$ to find the potential CP locations, then extract the CPs for both query and reference spectra, and finally compute the similarity value based only on these extracted peak intensities.

Formally, for a molecule j, let $\mathcal{K}_j = \{K_{1,j}, \cdots, K_{m_j,j}\}$ denote the set of potential CP locations, where m_j is the number of CPs in molecule j, and let $K_{l,j} = \{k_1, \cdots, k_{s_l}\}$, and $\forall l \in [m_j]$ denotes the set of potential CP indices. These indices correspond to the dimensionality of the molecules' spectra, hence $1 \le k_{l'} \le D$, $\forall l' \in [s_l]$, and $\forall l \in [m_j]$.

Then, CaPSim is then calculated in the following procedure:

1. Given a query q and a library $\mathcal{L} = \{R_1, \cdots, R_N\}$, we use an algorithm $\mathcal{A}$ to extract the potential CP locations from each molecule's spectra, i.e., $\mathcal{K}_j = \mathcal{A}(R_j)$, $\forall j \in [N]$.
2. After obtaining $\mathcal{K}_j$,[40] we perform a spectral max-pooling[59] over each spectrum $r_{i,j}$ in R_j, $\forall i \in [n_j]$, and obtain the compressed spectrum $\tilde{r}_{i,j} = [\max_{k \in K_{1,j}} r_{i,j,k}, \max_{k \in K_{2,j}} r_{i,j,k}, \cdots, \max_{k \in K_{m_j,j}} r_{i,j,k}]^{\mathrm{T}}$. In other words, for each set of potential CP indices $K_{l,j}$, $\forall l \in [m_j]$, we take the maximum intensity in this specific area of the spectrum $r_{i,j}$. The result $\tilde{r}_{i,j}$ is a vector of intensity of CPs of the ith recording of molecule j. We also normalize $\tilde{r}_{i,j}$ to range [0,1] for a consistent scale across different examples, i.e. $\tilde{r}_{i,j} \leftarrow (\tilde{r}_{i,j} - \min_k \tilde{r}_{i,j,k})/(\max_k \tilde{r}_{i,j,k} - \min_k \tilde{r}_{i,j,k})$, and we use $\tilde{R}_j$ to simply denote all compressed reference spectra, i.e., $\tilde{R}_j = [\tilde{r}_{1,j}, \cdots, \tilde{r}_{n_j,j}]$.
3. Let $\tilde{q}_j$ denote the compressed query using potential CP locations of molecule j, i.e., $\tilde{q}_j = [\max_{k \in K_{1,j}} q_k, \max_{k \in K_{2,j}} q_k, \cdots, \max_{k \in K_{m_j,j}} q_k]^{\mathrm{T}}$. We also normalize $\tilde{q}_j$ for the same reason, i.e., $\tilde{q}_j \leftarrow (\tilde{q}_j - \min_k \tilde{q}_{i,j,k})/(\max_k \tilde{q}_{j,k} - \min_k \tilde{q}_{j,k})$.
4. Compute CaPSim as $\mathcal{S}_{\text{CaPSim}}(q, R_j) = \frac{1}{n_j} \sum_{i=1}^{n_j} \tilde{q}_j^{\mathrm{T}} \tilde{r}_{i,j}$, which is simply the mean of the inner product between the compressed query and reference spectra.

In this work, the algorithm $\mathcal{A}$ we use for extracting CPs is called Characteristic Peak Extraction (CaPE).[40] We chose CaPE among other alternatives for its interpretability and generalizability.[68]

Characteristic Peak Extraction (CaPE). CaPE is a spectral compression algorithm that is able to extract important peaks for spectra identification while removing noisy and unimportant peaks. In the original work, it used peak counts as the peak selection criterion to account for CPs with relatively low intensity in the mixture. However, in our case, since we are performing spectral matching and assuming that each query spectrum has just one component, it is straightforward to use peak intensity as the criterion of selecting CPs. We replace the counting of peaks, which is either 0 or 1 for each wavenumber, with peak intensity values, which are nonnegative real numbers. CaPE has three hyperparameters: (1) the smoothing kernel size K_{smooth}, (2) the number of top peaks to keep N_{peak}, and (3) the maximum width of potential CP locations $w_{\max}$. We choose $K_{\text{smooth}} = 5$, $N_{\text{peak}} = 10$, and $w_{\max} = 36$, which are determined by our preliminary experiments on a small subset of the data and kept the same across all experiments.

Choice Function and Evaluation Metric: Top-k Accuracy. Our choice function C is argtop_k, which selects the top-k molecules based on their similarity to the query. In the experiment, we use top-k accuracy as the evaluation metric. Here we formally define this metric. Recall that $M(q)$ gives the molecule identity $M(q) = \underset{j \in [N]}{\text{argmax}} \overline{\mathcal{S}}(q, R_j)$, which is the best matching molecule in the library. However, it is usually worth looking at the second, third, ..., best matching molecule since these are also possible placements of the correct matching in practice. We define the set of top-k matching molecules as

$$\hat{\mathcal{M}}(q; \mathcal{S}, \mathcal{L}, k) = \underset{j \in [N]}{\text{argtop}_k} \mathcal{S}(q, R_j)$$

which is a set of size k and contains the indices of the first k molecules if we sort $\overline{\mathcal{S}}(q, R_j)$ in descending order, and we define a correct matching as $\mathbb{I}[M(q) \in \hat{\mathcal{M}}(\boldsymbol{q}; \mathcal{S}, \boldsymbol{\mathcal{L}}, \boldsymbol{k})]$, where $\mathbb{I}[\cdot]$ is the indicator function and has a value of 1 if the argument is true and 0 otherwise. Then, given a set of query Q, we perform spectral matching for each $q \in Q$ and calculate the top-k accuracy as

$$\frac{1}{|Q|} \sum_{q \in Q} \mathbb{I}[M(q) \in \hat{\mathcal{M}}(q; \mathcal{S}, \mathcal{L}, k)]$$

In this work, we use top-k as the choice function for simplicity. In practice, we can associate each matched compound with a confidence level, which can be calculated with the mean and standard deviation of the similarity values over multiple recordings of the same compound.

Summary. Recall that there are four choices we have to make for spectral matching: (1) the preprocessing function $\mathcal{P}$, (2) the peak extraction algorithm $\mathcal{A}$, (3) the similarity function $\mathcal{S}$, and (4) the choice function C. We use the same $\mathcal{P}$, baseline removal and l_2-normalization, and C, the argtop_k function, for all the experiments. We use various existing $\mathcal{S}$ and compare their performance to the proposed $\mathcal{S}_{\text{CaPSim}}$. Among all the similarity functions, only $\mathcal{S}_{\text{CaPSim}}$ requires a nontrivial $\mathcal{A}$, CaPE, and others only need $\mathcal{A}$ to be the identity function.

Similarity Attribution Methods. In Figure 6, we show the similarity attribution of all 5 similarity metrics used in this work. Here, we explain how we calculated the attributions. Since the attribution is meant to show how a similarity metric responds to a pair of spectra at a specific wavenumber, we take the vector before the aggregation step of each similarity metric, which is typically a sum or average over all wavenumbers. We ignore the scaling factors in all metrics. For Pearson's correlation, we also ignore the term $\overline{q}\,\overline{r}$, which is calculated from two scalars. To summarize, we have the attribution formulas for

CaPSim:

$$\text{Attr}_{\text{CaPSim}}(\tilde{q}, \tilde{r}) = \tilde{q} \circ \tilde{r}$$

Pearson's correlation:

$$\text{Attr}_{\text{Pearson}}(q, r) = q \circ r$$

Euclidean distance:

$$\text{Attr}_{\text{Euclidean}}(q, r) = -(q - r) \circ (q - r)$$

Cosine similarity:

$$\text{Attr}_{\text{cossim}}(q, r) = q \circ q \circ r \circ r$$

First-difference cosine similarity:

$$\text{Attr}_{\text{cossim 1st diff}}(q, r) = \Delta q \circ \Delta q \circ \Delta r \circ \Delta r$$

where $\circ$ is the Hadmard product.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsnano.3c05510.

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

Spectra of the optical properties of the gold plasmonic nanoparticles; ablation study results for CaPSim; and processed Raman and SERS spectra for the four PAH molecules used in the experiments (PDF)

AUTHOR INFORMATION

Corresponding Authors

Ankit Patel − *Department of Electrical and Computer Engineering, Rice University, Houston, Texas 77005, United States; Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, United States;* orcid.org/0000-0001-9678-496X; Email: abp4@rice.edu

Naomi J. Halas − *Department of Electrical and Computer Engineering, Department of Chemistry, Department of Physics and Astronomy, and Laboratory for Nanophotonics, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-8461-8494; Email: halas@rice.edu

Authors

Yilong Ju − *Department of Computer Science, Rice University, Houston, Texas 77005, United States*

Oara Neumann − *Department of Electrical and Computer Engineering and Laboratory for Nanophotonics, Rice University, Houston, Texas 77005, United States*

Mary Bajomo − *Department of Chemistry and Laboratory for Nanophotonics, Rice University, Houston, Texas 77005, United States*

Yiping Zhao − *Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602, United States;* orcid.org/0000-0002-3710-4159

Peter Nordlander − *Department of Electrical and Computer Engineering, Department of Physics and Astronomy, and Laboratory for Nanophotonics, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-1633-2937

Complete contact information is available at: https://pubs.acs.org/10.1021/acsnano.3c05510

Author Contributions

□These authors contributed equally. The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

Funding

This work was financially supported by the National Institute of Environmental Health Sciences of the National Institutes of Health (P42ES027725−01A), the Welch Foundation grants C-1220 (NJH) and C-1222 (PN), and the Carl and Lillian Illig Fellowship (Smalley-Curl Institute, H20398−239440).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Julia Dolive, Mattie Kumela, and Beniam Kumela for help with the Raman measurements.

ABBREVIATIONS

CPs	characteristic peaks
CaPE	characteristic peak extraction
CaPSim	characteristic peak similarity
CosineSim	cosine similarity
Pearson Corr	Pearson correlation
EuclideanDist	Euclidean distance
PYR	pyrene
ANTH	anthracene
B[a]P	benzo[a]pyrene
B[a]A	benzo[a]anthracene
PHE	phenanthrene
Au NPs	gold colloids
Au NSs	gold nanoshells
Ag NRs	silver nanorods
PLL	poly-L-lysine
SEM	scanning electron microscopy
SERS	surface-enhanced Raman spectroscopy

REFERENCES

(1) Kneipp, K.; Kneipp, H.; Itzkan, I.; Dasari, R. R.; Feld, M. S. Ultrasensitive chemical analysis by Raman spectroscopy. *Chem. Rev.* **1999**, *99*, 2957.

(2) Langer, J.; de Aberasturi, D. J.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguie, B.; Baumberg, J. J.; Bazan, G. C.; Bell, S. E. J.; Boisen, A.; Brolo, A. G.; et al. Present and Future of Surface-Enhanced Raman Scattering. *ACS Nano* **2020**, *14*, 28−117.

(3) Ding, S. Y.; You, E. M.; Tian, Z. Q.; Moskovits, M. Electromagnetic theories of surface-enhanced Raman spectroscopy. *Chem. Soc. Rev.* **2017**, *46*, 4042−4076.

(4) Schie, I. W.; Ruger, J.; Mondol, A. S.; Ramoji, A.; Neugebauer, U.; Krafft, C.; Popp, J. High-Throughput Screening Raman Spectroscopy Platform for Label-Free Cellomics. *Anal. Chem.* **2018**, *90*, 2023−2030.

(5) Wang, Y. L.; Zhao, C.; Wang, J. J.; Luo, X.; Xie, L. J.; Zhan, S. J.; Kim, J. M.; Wang, X. Z.; Liu, X. J.; Ying, Y. B. Wearable plasmonic-metasurface sensor for noninvasive and universal molecular fingerprint detection on biointerfaces. *Sci. Adv.* **2021**, *7*, No. eabe4553.

(6) Peron, O.; Rinnert, E.; Lehaitre, M.; Crassous, P.; Compere, C. Detection of polycyclic aromatic hydrocarbon (PAH) compounds in artificial sea-water using surface-enhanced Raman scattering (SERS). *Talanta* **2009**, *79*, 199−204.

(7) Xu, Y.; Kutsanedzie, F. Y. H.; Hassan, M.; Zhu, J. J.; Ahmad, W.; Li, H. H.; Chen, Q. S. Mesoporous silica supported orderly-spaced gold nanoparticles SERS-based sensor for pesticides detection in food. *Food Chem.* **2020**, *315*, 126300.

(8) Lenzi, E.; Henriksen-Lacey, M.; Molina, B.; Langer, J.; de Albuquerque, C. D. L.; Jimenez de Aberasturi, D.; Liz-Marzan, L. M. Combination of Live Cell Surface-Enhanced Raman Scattering Imaging with Chemometrics to Study Intracellular Nanoparticle Dynamics. *ACS Sens.* **2022**, *7*, 1747−1756.

(9) Safir, F.; Vu, N.; Tadesse, L. F.; Firouzi, K.; Banaei, N.; Jeffrey, S. S.; Saleh, A. A. E.; Khuri-Yakub, B. T.; Dionne, J. A. Combining Acoustic Bioprinting with AI-Assisted Raman Spectroscopy for High-Throughput Identification of Bacteria in Blood. *Nano Lett.* **2023**, *23*, 2065−2073.

(10) Jackson, J. B.; Halas, N. J. Surface-enhanced Raman scattering on tunable plasmonic nanoparticle substrates. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17930−17935.

(11) Witkowska, E.; Nicinski, K.; Korsak, D.; Szymborski, T.; Kaminska, A. Sources of variability in SERS spectra of bacteria: comprehensive analysis of interactions between selected bacteria and plasmonic nanostructures. *Anal. Bioanal. Chem.* **2019**, *411*, 2001−2017.

(12) Rastogi, R.; Dogbe Foli, E. A.; Vincent, R.; Adam, P.-M.; Krishnamoorthy, S. Engineering Electromagnetic Hot-Spots in Nanoparticle Cluster Arrays on Reflective Substrates for Highly Sensitive Detection of (Bio)molecular Analytes. *ACS Appl. Mater. Interfaces* **2021**, *13*, 32653−32661.

(13) Khlebtsov, N. G.; Lin, L.; Khlebtsov, B. N.; Ye, J. Gap-enhanced Raman tags: fabrication, optical properties, and theranostic applications. *Theranostics* **2020**, *10*, 2067−2094.

(14) Borah, R.; Ninakanti, R.; Bals, S.; Verbruggen, S. W. Plasmon resonance of gold and silver nanoparticle arrays in the Kretschmann

(attenuated total reflectance) vs. direct incidence configuration. *Sci. Rep.* **2022**, *12*, 15738.
(15) Park, J. E.; Lee, Y.; Nam, J. M. Precisely Shaped, Uniformly Formed Gold Nanocubes with Ultrahigh Reproducibility in Single-Particle Scattering and Surface-Enhanced Raman Scattering. *Nano Lett.* **2018**, *18*, 6475−6482.
(16) Bi, C. X.; Song, Y. H.; Zhao, H. Y.; Liu, G. Q. Shape controlled synthesis of concave octahedral Au@AuAg nanoparticles to improve their surface-enhanced Raman scattering performance. *RSC Adv.* **2022**, *12*, 19571−19578.
(17) Wei, X. Y.; Fan, Q. K.; Liu, H. P.; Bai, Y. C.; Zhang, L.; Zheng, H. Q.; Yin, Y. D.; Gao, C. B. Holey Au-Ag alloy nanoplates with built-in hotspots for surface-enhanced Raman scattering. *Nanoscale* **2016**, *8*, 15689−15695.
(18) Huang, H. J.; Shiao, M. H.; Lin, Y. W.; Lin, B. J.; Su, J.; Lin, Y. S.; Chang, H. W. Au@Ag Dendritic Nanoforests for Surface-Enhanced Raman Scattering Sensing. *Nanomaterials* **2021**, *11*, 1736.
(19) Zhao, Y. X.; Kang, H. S.; Zhao, W. Q.; Chen, Y. L.; Ma, L.; Ding, S. J.; Chen, X. B.; Wang, Q. Q. Dual Plasmon Resonances and Tunable Electric Field in Structure-Adjustable Au Nanoflowers for Improved SERS and Photocatalysis. *Nanomaterials* **2021**, *11*, 2176.
(20) Wustholz, K. L.; Henry, A. I.; McMahon, J. M.; Freeman, R. G.; Valley, N.; Piotti, M. E.; Natan, M. J.; Schatz, G. C.; Van Duyne, R. P. Structure-Activity Relationships in Gold Nanoparticle Dimers and Trimers for Surface-Enhanced Raman Spectroscopy. *J. Am. Chem. Soc.* **2010**, *132*, 10903−10910.
(21) Arbuz, A.; Sultangaziyev, A.; Rapikov, A.; Kunushpayeva, Z.; Bukasov, R. How gap distance between gold nanoparticles in dimers and trimers on metallic and non-metallic SERS substrates can impact signal enhancement. *Nanoscale Adv.* **2021**, *4*, 268−280.
(22) Hu, O.; Sellers, C.; Kwon, J. S. I.; Wu, G. Integration of surface-enhanced Raman spectroscopy (SERS) and machine learning tools for coffee beverage classification. *Chem. Eng.* **2022**, *3*, 100020.
(23) Leong, Y. X.; Lee, Y. H.; Koh, C. S. L.; Phan-Quang, G. C.; Han, X. M.; Phang, I. Y.; Ling, X. Y. Surface-Enhanced Raman Scattering (SERS) Taster: A Machine-Learning-Driven Multireceptor Platform for Multiplex Profiling of Wine Flavors. *Nano Lett.* **2021**, *21*, 2642−2649.
(24) Kim, N.; Thomas, M. R.; Bergholt, M. S.; Pence, I. J.; Seong, H.; Charchar, P.; Todorova, N.; Nagelkerke, A.; Belessiotis-Richards, A.; Payne, D. J.; Gelmi, A.; Yarovsky, I.; Stevens, M. M. Surface enhanced Raman scattering artificial nose for high dimensionality fingerprinting. *Nat. Commun.* **2020**, *11*, 207.
(25) Lussier, F.; Thibault, V.; Charron, B.; Wallace, G. Q.; Masson, J. F. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrAC-Trends Anal. Chem.* **2020**, *124*, 115796.
(26) Yang, S. W.; Xie, Y. H.; Liu, J. Z.; Zhao, S.; Jin, S. Z.; Zhang, D.; Chen, Q.; Huang, J.; Liang, P. Raman spectral classification algorithm of cephalosporin based on VGGNeXt. *Analyst* **2022**, *147*, 5486−5494.
(27) Yang, Y. J.; Li, H.; Jones, L.; Murray, J.; Haverstick, J.; Naikare, H. K.; Mosley, Y. Y. C.; Tripp, R. A.; Ai, B.; Zhao, Y. P. Rapid Detection of SARS-CoV-2 RNA in Human Nasopharyngeal Specimens Using Surface-Enhanced Raman Spectroscopy and Deep Learning Algorithms. *ACS Sens.* **2023**, *8*, 297−307.
(28) Yang, Y. J.; Xu, B. B.; Haverstick, J.; Ibtehaz, N.; Muszynski, A.; Chen, X. Y.; Chowdhury, M. E. H.; Zughaier, S. M.; Zhao, Y. P. Differentiation and classification of bacterial endotoxins based on surface enhanced Raman scattering and advanced machine learning. *Nanoscale* **2022**, *14*, 8806−8817.
(29) Yang, Y. J.; Xu, B. B.; Murray, J.; Haverstick, J.; Chen, X. Y.; Tripp, R. A.; Zhao, Y. P. Rapid and quantitative detection of respiratory viruses using surface-enhanced Raman spectroscopy and machine learning. *Biosens. Bioelectron.* **2022**, *217*, 114721.
(30) Samuel, A. Z.; Mukojima, R.; Horii, S.; Ando, M.; Egashira, S.; Nakashima, T.; Iwatsuki, M.; Takeyama, H. On Selecting a Suitable Spectral Matching Method for Automated Analytical Applications of Raman Spectroscopy. *ACS Omega* **2021**, *6*, 2060−2065.
(31) Pozzi, F.; Porcinai, S.; Lombardi, J. R.; Leona, M. Statistical methods and library search approaches for fast and reliable identification of dyes using surface-enhanced Raman spectroscopy (SERS). *Anal. Methods.* **2013**, *5*, 4205−4212.
(32) Pezzotti, G.; Boschetto, F.; Ohgitani, E.; Fujita, Y.; Shin-Ya, M.; Adachi, T.; Yamamoto, T.; Kanamura, N.; Marin, E.; Zhu, W. L.; Nishimura, I.; Mazda, O. Raman Molecular Fingerprints of SARS-CoV-2 British Variant and the Concept of Raman Barcode. *Adv. Sci.* **2022**, *9*, 2103287.
(33) Caldarone, A.; Piccotti, F.; Morasso, C.; Truffi, M.; Sottotetti, F.; Guerra, C.; Albasini, S.; Agozzino, M.; Villani, L.; Corsi, F. Raman analysis of microcalcifications in male breast cancer. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy* **2021**, *263*, 120185.
(34) Pezzotti, G.; Ohgitani, E.; Fujita, Y.; Imamura, H.; Shin-Ya, M.; Adachi, T.; Yamamoto, T.; Kanamura, N.; Marin, E.; Zhu, W. L.; Nishimura, I.; Mazda, O. Raman Fingerprints of the SARS-CoV-2 Delta Variant and Mechanisms of Its Instantaneous Inactivation by Silicon Nitride Bioceramics. *ACS Infect. Dis.* **2022**, *8*, 1563−1581.
(35) Miller, E. A.; Yamahara, K. M.; French, C.; Spingarn, N.; Birch, J. M.; Van Houtan, K. S. A Raman spectral reference library of potential anthropogenic and biological ocean polymers. *Scientific Data* **2022**, *9*, 780.
(36) Salim, Y. S.; Rashid, N. A.; Halim, S. I. A.; Chan, C. H.; Ong, C. H.; Harun, M. K. Fourier transform infrared (FTIR) authentication and batch-to-batch consistency for different types of paints using benchtop and handheld FTIR spectrophotometers for oil and gas industry. *Polym. Eng. Sci.* **2021**, *61*, 2757−2770.
(37) Mandende, H.; Ndou, C.; Mothupi, T. Hyperspectral core scanner: An effective mineral mapping tool for apatite in the Upper Zone, northern limb, Bushveld Complex. *J. South. Afr. Inst. Min. Metall.* **2023**, *123*, 81−92.
(38) Bell, D. C.; Boehm, R. C.; Feldhausen, J.; Heyne, J. S. A Data Set Comparison Method Using Noise Statistics Applied to VUV Spectrum Match Determinations. *Anal. Chem.* **2022**, *94*, 14861−14868.
(39) Rodriguez, J. D.; Westenberger, B. J.; Buhse, L. F.; Kauffman, J. F. Standardization of Raman spectra for transfer of spectral libraries across different instruments. *Analyst* **2011**, *136*, 4232−4240.
(40) Bajomo, M. M.; Ju, Y.; Zhou, J.; Elefterescu, S.; Farr, C.; Zhao, Y.; Neumann, O.; Nordlander, P.; Patel, A.; Halas, N. J. Computational chromatography: A machine learning strategy for demixing individual chemical components in complex mixtures. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2211406119.
(41) Kortli, Y.; Jridi, M.; Al Falou, A.; Atri, M. Face Recognition Systems: A Survey. *Sensors* **2020**, *20*, 342.
(42) Driskell, J. D.; Shanmukh, S.; Liu, Y.; Chaney, S. B.; Tang, X. J.; Zhao, Y. P.; Dluhy, R. A. The use of aligned silver nanorod Arrays prepared by oblique angle deposition as surface enhanced Raman scattering substrates. *J. Phys. Chem. C* **2008**, *112*, 895−901.
(43) Ding, S. Y.; Yi, J.; Li, J. F.; Ren, B.; Wu, D. Y.; Panneerselvam, R.; Tian, Z. Q. Nanostructure-based plasmon-enhanced Raman spectroscopy for surface analysis of materials. *Nat. Rev. Mater.* **2016**, *1*, 16021.
(44) Dahan, K. A.; Li, Y.; Xu, J.; Kan, C. X. Recent progress of gold nanostructures and their applications. *Phys. Chem. Chem. Phys.* **2023**, *25*, 18545−18576.
(45) Chaney, S. B.; Shanmukh, S.; Dluhy, R. A.; Zhao, Y. P. Aligned silver nanorod arrays produce high sensitivity surface-enhanced Raman spectroscopy substrates. *Appl. Phys. Lett.* **2005**, *87*, 031908.
(46) Liu, Y. J.; Chu, H. Y.; Zhao, Y. P. Silver Nanorod Array Substrates Fabricated by Oblique Angle Deposition: Morphological, Optical, and SERS Characterizations. *J. Phys. Chem. C* **2010**, *114*, 8176−8183.
(47) Lau, W. F.; Bai, F.; Huang, Z. F. Ballistic glancing angle deposition of inclined Ag nanorods limited by adatom diffusion. *Nanotechnology* **2013**, *24*, 465707.

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261

(48) Song, C. Y.; Chen, J.; Zhao, Y. P.; Wang, L. H. Gold-modified silver nanorod arrays for SERS-based immunoassays with improved sensitivity. *J. Mater. Chem. B* **2014**, *2*, 7488−7494.
(49) Chaney, S. B.; Shanmukh, S.; Dluhy, R. A.; Zhao, Y. P. Aligned silver nanorod arrays produce high sensitivity surface-enhanced Raman spectroscopy substrates. *Appl. Phys. Lett.* **2005**, *87*, 031908.
(50) Zhang, Z. M.; Chen, S.; Liang, Y. Z. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst* **2010**, *135*, 1138−1146.
(51) O'Bannon, E.; Williams, Q. Vibrational spectra of four polycyclic aromatic hydrocarbons under high pressure: implications for stabilities of PAHs during accretion. *Phys. Chem. Miner.* **2016**, *43*, 181−208.
(52) Chen, J.; Huang, Y. W.; Zhao, Y. P. Characterization of polycyclic aromatic hydrocarbons using Raman and surface-enhanced Raman spectroscopy. *J. Raman Spectrosc.* **2015**, *46*, 64−69.
(53) Xiao, L. L.; Zhang, M.; Liu, Z.; Bian, W. W.; Zhang, X. L.; Zhan, J. H. Hydrophobic silver nanowire membrane for swabbing extraction and in situ SERS detection of polycyclic aromatic hydrocarbons on toys. *Anal. Methods* **2017**, *9*, 1816−1824.
(54) Colangeli, L.; Mennella, V.; Baratta, G. A.; Bussoletti, E.; Strazzulla, G. RAMAN AND INFRARED-SPECTRA OF POLYCYCLIC AROMATIC HYDROCARBON MOLECULES OF POSSIBLE ASTROPHYSICAL INTEREST. *Astrophys. J.* **1992**, *396*, 369−377.
(55) Negri, P.; Marotta, N. E.; Bottomley, L. A.; Dluhy, R. A. Removal of Surface Contamination and Self-Assembled Monolayers (SAMs) from Silver (Ag) Nanorod Substrates by Plasma Cleaning with Argon. *Appl. Spectrosc.* **2011**, *65*, 66−74.
(56) Nuntawong, N.; Eiamchai, P.; Wong-ek, B.; Horprathum, M.; Limwichean, K.; Patthanasettakul, V.; Chindaudom, P. Shelf time effect on SERS effectiveness of silver nanorod prepared by OAD technique. *Vacuum* **2013**, *88*, 23−27.
(57) Subr, M.; Petr, M.; Peksa, V.; Kylian, O.; Hanus, J.; Prochazka, M. Ag Nanorod Arrays for SERS: Aspects of Spectral Reproducibility, Surface Contamination, and Spectral Sensitivity. *J. Nanomater.* **2015**, *2015*, 729231.
(58) Bachenheimer, L.; Scherzer, R.; Elliott, P.; Stagon, S.; Gasparov, L.; Huang, H. C. Degradation Mechanism of Ag Nanorods for Surface Enhanced Raman Spectroscopy. *Sci. Rep.* **2017**, *7*, 16282.
(59) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84−90.
(60) Zhao, W. J.; Tan, S. L.; Li, P.; *Ieee In SONG: Approximate Nearest Neighbor Search on GPU*; IEEE 36th Int. Conf. Data Eng. (ICDE): Dallas, TX, Apr 20−24; Dallas, TX, 2020; pp 1033−1044.
(61) Wang, M. Z.; Xu, X. L.; Yue, Q.; Wang, Y. X. A Comprehensive Survey and Experimental Comparison of Graph-Based Approximate Nearest Neighbor Search. *Proc. VLDB Endow.* **2021**, *14*, 1964−1978.
(62) Gao, J. H.; Ji, W. X.; Chang, F. L.; Han, S. Y.; Wei, B. X.; Liu, Z. M.; Wang, Y. Z. A Systematic Survey of General Sparse Matrix-matrix Multiplication. *ACM Comput. Surv.* **2023**, *55*, 55.
(63) Kundu, J.; Le, F.; Nordlander, P.; Halas, N. J. Surface-enhanced infrared absorption (SEIRA) spectroscopy on nanoshell aggregate substrates. *Chem. Phys. Lett.* **2008**, *452*, 115−119.
(64) Le, F.; Brandl, D. W.; Urzhumov, Y. A.; Wang, H.; Kundu, J.; Halas, N. J.; Aizpurua, J.; Nordlander, P. Metallic nanoparticle arrays: A common substrate for both surface-enhanced Raman scattering and surface-enhanced infrared absorption. *ACS Nano* **2008**, *2*, 707−718.
(65) Wang, H.; Kundu, J.; Halas, N. J. Plasmonic nanoshell arrays combine surface-enhanced vibrational spectroscopies on a single substrate. *Angew. Chem., Int. Ed.* **2007**, *46*, 9040−9044.
(66) Oldenburg, S. J.; Averitt, R. D.; Westcott, S. L.; Halas, N. J. Nanoengineering of optical resonances. *Chem. Phys. Lett.* **1998**, *288*, 243−247.
(67) Abell, J. L.; Lee, J.; Zhao, Q.; Szu, H.; Zhao, Y. P. Differentiating intrinsic SERS spectra from a mixture by sampling induced composition gradient and independent component analysis. *Analyst* **2012**, *137*, 73−76.
(68) Fenn, M. B.; Pappu, V.; Georgeiv, P. G.; Pardalos, P. M. Raman spectroscopy utilizing Fisher-based feature selection combined with Support Vector Machines for the characterization of breast cell lines. *J. Raman Spectrosc.* **2013**, *44*, 939−948.

https://doi.org/10.1021/acsnano.3c05510
ACS Nano 2023, 17, 21251−21261