Supporting Information

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[#,†,⁑]*

[#]Department of Computer Science, [*]Department of Electrical and Computer Engineering, [¶]Department of Chemistry, [§]Department of Physics and Astronomy, [⸗]Laboratory for Nanophotonics, Rice University, 6100 Main Street, Houston, Texas 77005, USA

[†]Department of Neuroscience, Baylor College of Medicine, Houston, Texas, 77030, USA

[‡]Department of Physics and Astronomy, University of Georgia, Athens, GA, 30602, USA

Corresponding Authors:
Ankit Patel abp4@rice.edu

Naomi Halas halas@rice.edu

[⁑]These authors contributed equally

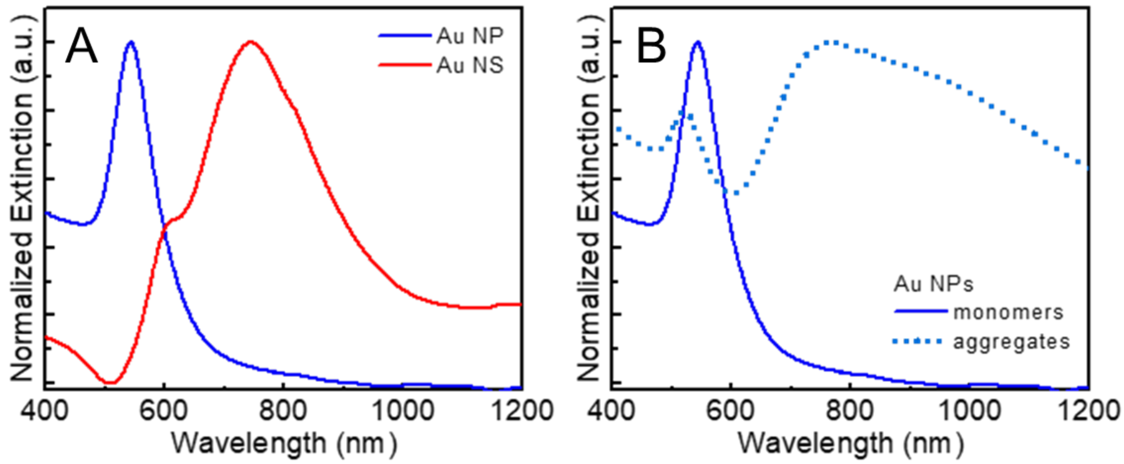

Figure S1. Optical properties of plasmonic nanoparticles. (A) Extinction spectra of (blue) Au NPs and (red) Au NS in water solution; **(B)** Extinction spectra of (line) Au NP in solution and (dash line) AuNPs on a quartz substrate.

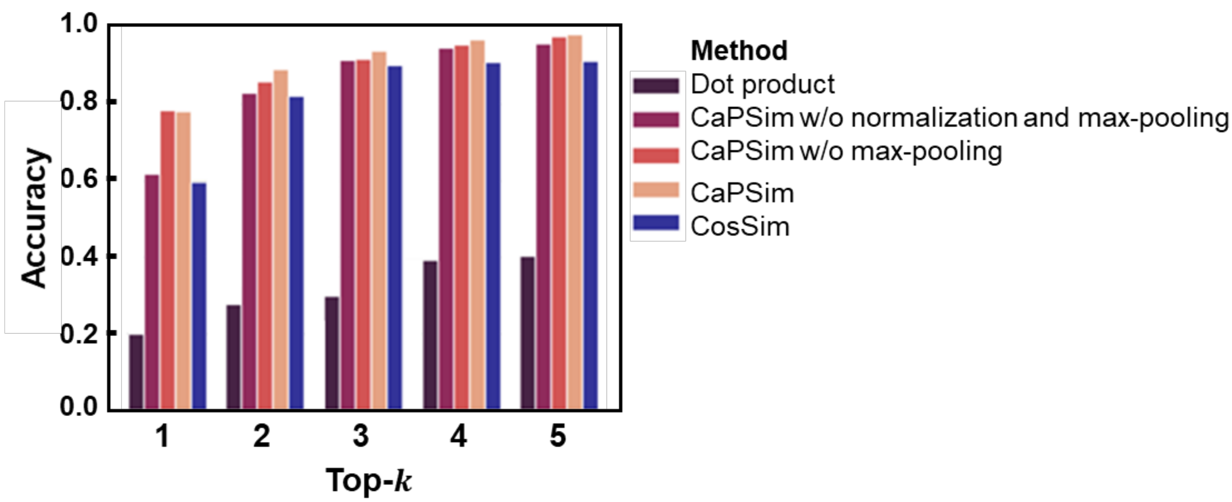

Figure S2. Ablation study results for CaPSim.

We performed a series of experiments where we started with a primitive version of CaPSim and added the designed steps back one by one. We evaluated these versions of CaPSim on the SERS data using the Au NS substrate. Compared to a plain dot product, adding the ability to extract the important peaks already boosts the spectral recognition performance to a level above CosSim, one of the best existing metrics. And each of the normalization and max-pooling steps can further improve the performance.

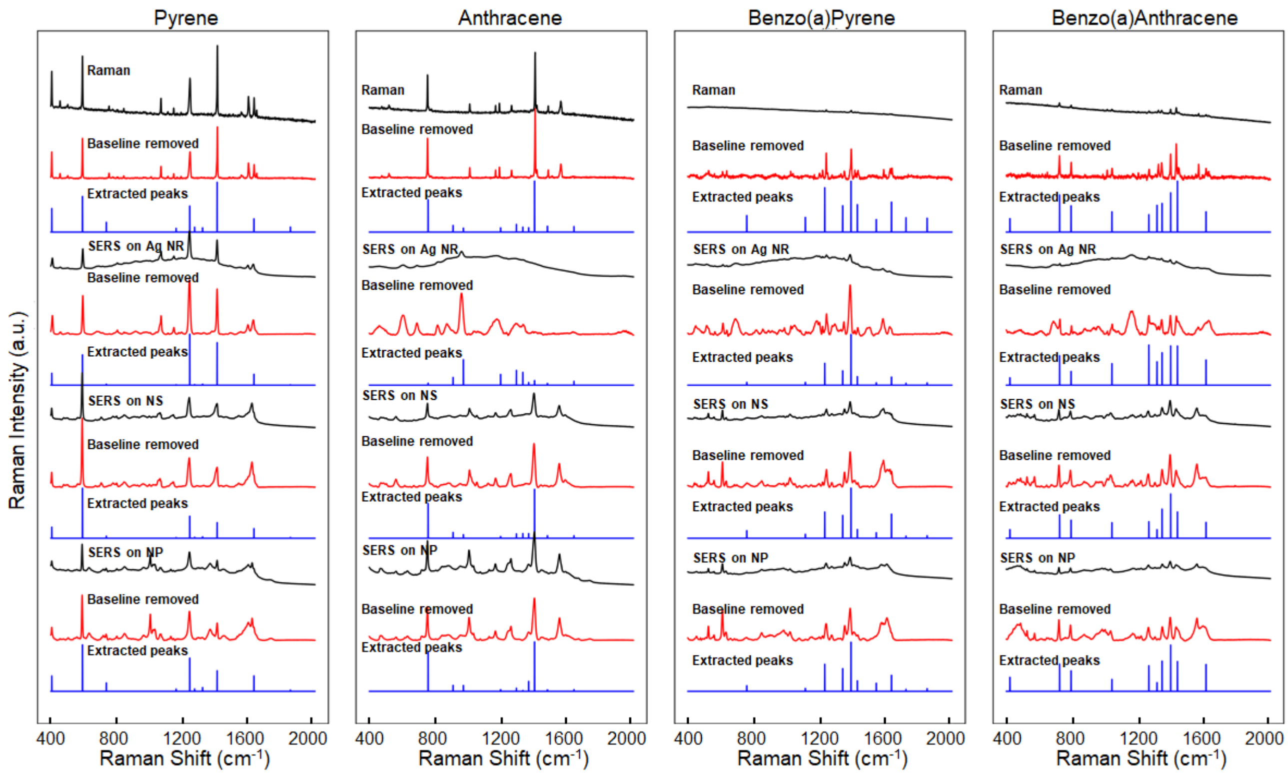

Figure S3. Processing results of the spectra on four of the PAHs we used. We removed the baseline for both SERS and Raman raw data and examined their characteristic peaks extracted by CaPE. The baseline removal can remove the slow-changing trend from all raw spectra data and leave peaks on. It also does not introduce spurious peaks.