Deep learning-driven super-resolution in Raman hyperspectral imaging: Efficient high-resolution reconstruction from low-resolution data

View Online   Export Citation   CrossMark

Md Inzamam Ul Haque,[1] (iD) Ariel Lebron,[2] (iD) Frances Joan D. Alvarez,[2] (iD) Jennifer F. Neal,[2] (iD) Marc Mamak,[2] (iD)
Debangshu Mukherjee,[3] (iD) Olga S. Ovchinnikova,[4,a] (iD) and Jacob D. Hinkle[5] (iD)

AFFILIATIONS

[1]The Bredesen Center, University of Tennessee, Knoxville, Tennessee 37996, USA
[2]Research and Development, The Procter and Gamble Company, Mason, Ohio 45040, USA
[3]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA
[4]Department of Materials Science and Engineering, University of Tennessee, Knoxville, Tennessee 37996, USA
[5]NVIDIA, Santa Clara, California 95051, USA

[a]Author to whom correspondence should be addressed: oovchinn@utk.edu

ABSTRACT

Deep learning (DL) has become an indispensable tool in hyperspectral data analysis, automatically extracting valuable features from complex, high-dimensional datasets. Super-resolution reconstruction, an essential aspect of hyperspectral data, involves enhancing spatial resolution, particularly relevant to low-resolution hyperspectral data. Yet, the pursuit of super-resolution in hyperspectral analysis is fraught with challenges, including acquiring ground truth high-resolution data for training, generalization, and scalability. The pressing issue of extended spectral acquisition times, notably for high-resolution scans, is a significant roadblock in hyperspectral imaging. Super-resolution methods offer a promising solution by providing higher spatial resolution data to expedite data collection and yield more efficient outcomes. This paper delves into a practical application of these concepts using Raman imaging, where spectral acquisition times can be prohibitively long. In this context, DL-based super-resolution models demonstrate their efficacy by predicting and reconstructing high-resolution Raman data from low-resolution input, eliminating the need for resource-intensive high-resolution scans. While previous work often relied on substantial high-resolution datasets, this study showcases the ability to achieve similar outcomes even with limited data, presenting a more practical and cost-effective approach. The results offer a glimpse into the transformative potential of this technology to streamline hyperspectral imaging applications by saving valuable time and resources through the successful generation of high-resolution data from low-resolution inputs.

In hyperspectral data analysis, deep learning (DL)[1] has become increasingly crucial in automatically extracting meaningful features from high-dimensional data. Hyperspectral data consist of a series of images acquired at a range of wavelengths, thus encoding a detailed spectral signature for each pixel in an image. Deep learning techniques have been utilized in multiple aspects of hyperspectral data analysis, including feature extraction,[2] super-resolution reconstruction,[3] classification,[4] spectral unmixing,[5] and transfer learning.[6] In the context of hyperspectral data analysis, super-resolution reconstruction refers to enhancing the spatial resolution of hyperspectral images. This can be particularly useful when we have low-resolution hyperspectral data, but data extraction requires finer details for machine learning applications like object detection or mineral identification. However,

super-resolution in hyperspectral data analysis comes with challenges, such as obtaining ground truth high-resolution hyperspectral data for training, generalization, and scalability. One major challenge in super-resolution for hyperspectral data is the high acquisition time required to collect ground truth high-resolution datasets. Super-resolution computational techniques offer a practical solution by reconstructing high-resolution information from low-resolution images acquired through shorter acquisition times.[7] This study focuses on hyperspectral data from Raman spectroscopy, a label-free, nondestructive molecular imaging technique with submicrometer resolution. However, high-resolution Raman scans are time-intensive, with spectral acquisition typically taking one second per sample point,[8] making high-throughput imaging applications prohibitively slow.[9] To address these

23 November 2024 10:53:34

limitations, we propose an approach using deep learning to predict high-resolution Raman data from low-resolution inputs, bypassing the need for extensive high-resolution scans. Our workflow leverages recent advances in single-image super-resolution (SISR) techniques,[10–15] adapted to reconstruct high-resolution data using a limited dataset of only 54 paired low- and high-resolution Raman images. This approach not only reduces the reliance on large paired datasets, which are challenging to obtain but also demonstrates the viability of super-resolution reconstruction on small, targeted datasets. Our method employs a residual network-based convolutional neural network (CNN), specifically SRResNet,[11] which has shown superior performance in spatial resolution enhancement in limited-data scenarios. Unlike previous studies, such as Horgan et al.,[7] which used transfer learning with extensive paired datasets, our approach directly trains on a small dataset, reducing the dependency on large training samples and addressing common challenges in transfer learning, like domain shifts and data distribution differences. Additionally, we implement innovative preprocessing steps, including the Savitzky–Golay filter for noise reduction and template matching for spatial localization within the low-resolution field of view, to enhance model performance and enable accurate extraction of high-resolution features. Our workflow (Fig. 1) provides a streamlined and adaptable framework that could be extended to various hyperspectral imaging applications, paving the way for efficient, time-saving solutions in fields requiring high-resolution data.

This study uses aqueous-based surfactant mixtures with varying amounts of polymer (samples A–F) for samples for Raman imaging. Low and high resolution Raman images are collected using a WiTec alpha 300R Confocal Raman microscope equipped with a 488 nm laser and Zeiss Epiplan-Neofluar DiC 20×/0.5NA objective or Zeiss EC Epiplan-Neofluar Dic 50×/0.8NA. High spatial resolution scans are collected with the following scanning parameters: 100 points per line and 100 lines per image. Images are $50 \times 50\ \mu m^2$ in size with 500 nm scanning resolution. Low spatial resolution scans are collected with the following scanning parameters: 100 points per line and 100 lines per image. Images are $200 \times 200\ \mu m^2$ in size with $2\ \mu m$ scanning resolution. Raman data are baseline corrected, and cosmic rays are removed. Raman images are generated around 2900 and $3200\ cm^{-1}$, corresponding to the CH and OH peak areas.

The step-by-step data analysis workflow is described here in detail. Raw MATLAB files in.mat format are provided for the six samples. The MATLAB files are converted to NumPy arrays for analysis in Python. Each array is sized (10 000, 954), containing 954 spectral components of 100×100 resolution images. For each sample, data from four different areas are provided. The dataset consists of 24 NumPy arrays of size (10 000, 954), both for low resolution and high resolution.

Preprocessing of data includes smoothing and normalizing the data. Smoothing Raman data helps to distinguish signals of interest from high-frequency noise. It also increases the signal-to-noise ratio and enhances the primary features in a dataset. The Savitzky–Golay filter,[16] also known as the Savgol filter, has been used to smooth the data. This filter is much faster compared to other smoothing filters, which can save significant time in preprocessing datasets. The effect of smoothing the data is shown in supplementary material Fig. S1, where Raman spectra are shown for a randomly selected spatial point from sample D, both for low resolution and high resolution, to illustrate the effect of smoothing better. Following the smoothing, MinMaxScaler from Scikit-learn[17] is used to normalize the data and get it into the (0–1) range.

Following preprocessing, dimension reduction techniques, such as principal component analysis (PCA)[18] and non-negative matrix factorization (NMF),[19] are applied. These techniques reduce data complexity by extracting critical information from Raman spectra, distinguishing signal from noise, and revealing hidden spectral features and variations. PCA and NMF also reduce data size while preserving essential information, showing which spectral regions are most informative, and can aid in developing quantitative analysis models. For data reduction using NMF, NMF is trained using the low-resolution data consisting of 24 NumPy arrays of size (10 000, 954) and gets the low-resolution components. Then, the trained NMF transforms the high-resolution data of 24 NumPy arrays of size (10 000, 954).

From here on, NMF components are referred to as images for simplicity. The high-resolution images are a subset or within the field of view of the low-resolution images. If a low-resolution image is upscaled 4× from (100, 100) to (400, 400) resolution, then the corresponding (100, 100) high-resolution image can be identified by scanning through the low-resolution image. The first step toward predicting the high-resolution images from low-resolution images is to get the exact spatial position (X, Y) of the high-resolution images present in the field of view of the upscaled low-resolution images. A template matching algorithm, "matchTemplate" from the OpenCV package, has been used to identify these spatial (X, Y) positions based on correlation using the upscaled low-resolution image as input and the original high-resolution image as the template.
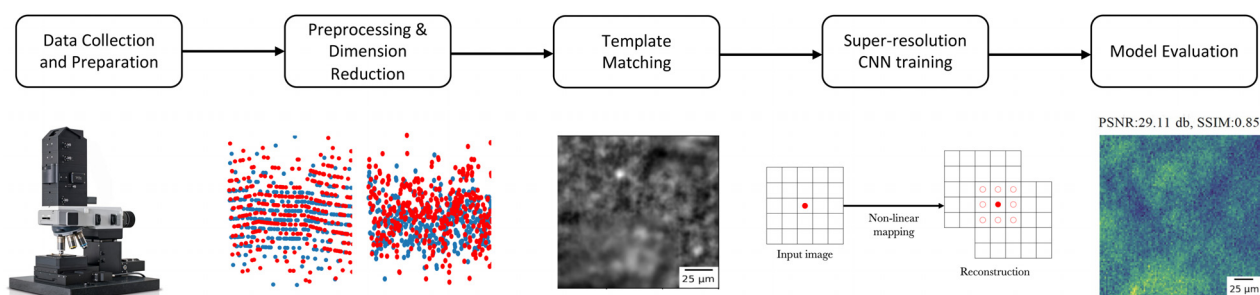

FIG. 1. Data analysis workflow.

The state-of-the-art super-resolution convolutional neural network (CNN) models, such as SRCNN,[11] SRResNet,[10] SRGAN,[10] and RCAN,[12] have been trained and tested with our small Raman dataset. The dataloader for training the CNN models contains the low-resolution images (100, 100) used as input images to the models, high-resolution images (100, 100) used as labels, and corresponding spatial (X, Y) coordinates retrieved before. The Raman dataset of 54 NMF components is divided into 40, 6, and 8 components for training, validation, and testing, respectively. The CNN models take the low-resolution image as input and upscale it $4\times$ to (400, 400) resolution. Adam optimizer is used with a learning rate of 0.0001 for training the SRResNet. The mean squared error (MSE) loss is used for backpropagation. After training for 1000 epochs, the training and validation losses decrease continuously, but the qualitative predicted image does not improve after 1000 epochs. To evaluate the model, image reconstruction quality metrics, such as peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM),[20] are used. PSNR is calculated as the ratio between the maximum possible signal power and the power of the noise that affects the quality of the image. The higher the PSNR, the better the quality of the image. SSIM is used for measuring the similarity between two images. In our case, the measurement or prediction of image quality is based on an initial uncompressed or distortion-free image as reference, the high-resolution image from the dataloader. After getting the upscaled image (400, 400) from the low-resolution image (100, 100) provided by the CNN models, the (X, Y) coordinates are used to extract the predicted high-resolution image of size (100, 100). Then, it is compared with the original high-resolution image of size (100, 100) using the above-mentioned metrics.

After applying both PCA and NMF techniques to our dataset, it was found qualitatively that NMF components more accurately recognize OH and CH bitmaps as seen in Fig. 2. PCA does not guarantee non-negativity in its components, which can be problematic for data where negative values have no physical meaning. Another reason to choose NMF over PCA is that some components of PCA have some vertical stripe artifacts, as seen in Fig. 2. Doing PCA, it was found that only the first two or three components from each sample area have an explained variance of greater than 1%. Hence, NMF components were chosen for the machine learning analysis. The NMF components are also checked in Fourier space to confirm if they have some signals or random noise (supplementary material Fig. S2 and supplementary material Fig. S3). Based on this criterion, only two or three components with good signals for low-resolution and high-resolution data were chosen from the 24 sample areas. Finally, 54 components were selected from the six samples for our deep-learning analysis for our low-resolution and high-resolution data.

The template matching algorithm results are shown in Fig. 3 for two components of sample A, where the objective was to locate the spatial coordinates of the high-resolution image as shown in the red boxes. The (X, Y) coordinates of the top left corner of the red boxes were identified by picking the most intensive point from the correlation result as shown in Figs. 3(c) and 3(f), respectively. The results for a few components were checked against the actual coordinates available for only a single sample. It was confirmed that this algorithm could successfully capture the (X, Y) coordinates. These spatial coordinates achieved by this step are essential after the super-resolution CNN model training step, which upscales the low-resolution image by
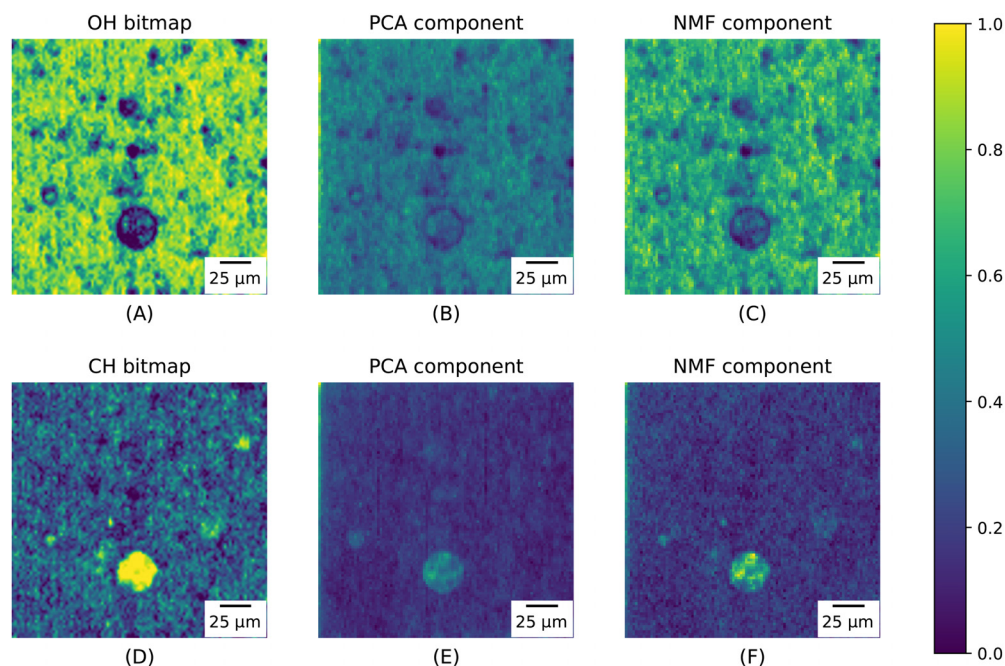

FIG. 2. High-resolution PCA and NMF components for sample A. OH and CH bitmaps for a certain area of sample A are shown here in (A) and (D), respectively. The corresponding PCA components are shown in (B) and (E), and NMF components are shown in (C) and (F). All the images are normalized in the (0–1) range, as shown in the color bar. The NMF components can be seen to recognize CH and OH bitmaps with more contrast more accurately.
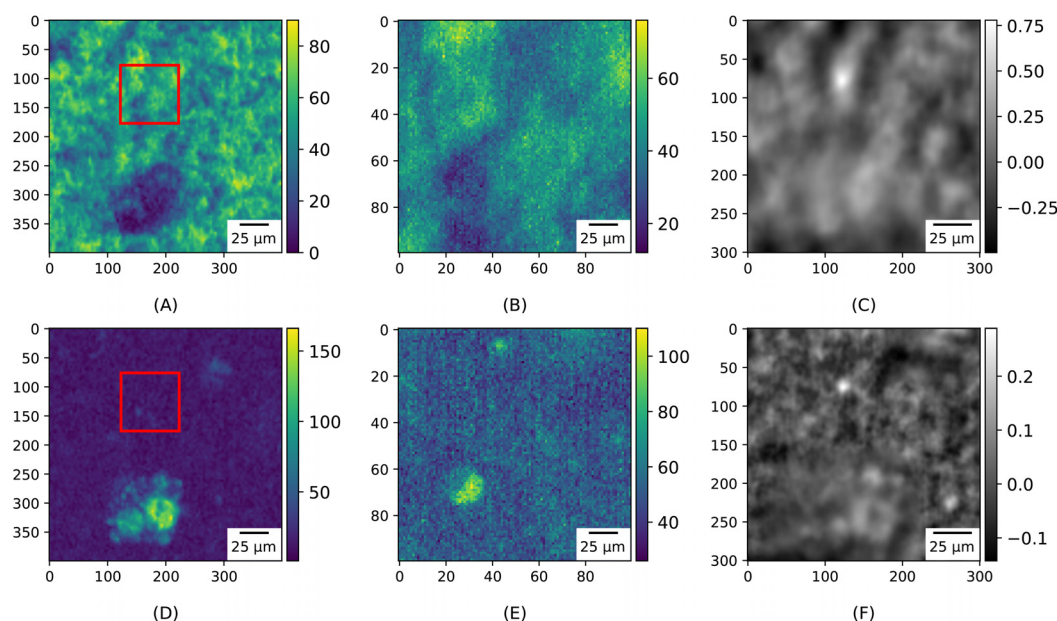
FIG. 3. Template matching is shown for two NMF components of sample A. (A) and (D) The two original upscaled low-resolution NMF components. (B) and (E) The corresponding original high-resolution NMF components that are used as templates here to scan through the low-resolution images in (A) and (D) and find the location of the high-resolution scans as shown in the red boxes. These red boxes are drawn using the maximum value of the correlation result as shown in (C) and (F).

4x. Then, the predicted high-resolution images are extracted using these spatial coordinates.

After training the super-resolution CNN models, the performance is evaluated for the training and test sets. Along with the prediction from the CNN model, prediction by linear interpolation is also computed. Tables I and II summarize the training and test set results, respectively, for the CNN models and linear interpolation. It indicates that the model maintains consistent performance across sets. The best quantitative results were achieved with the SRResNet model. SRResNet is a state-of-the-art CNN model with high upscaling factors (4X) as measured by peak signal-to-noise ratio (PSNR) and structural similarity (SSIM)[20] with a custom 16-block deep ResNet[21] optimized for mean squared error (MSE) loss.

Figure 4 shows the high-resolution NMF components prediction from low-resolution NMF components for two examples from the training set. As discussed, the high-resolution image can be found by scanning through the low-resolution image, and those coordinates are found by template matching. The red boxes in Figs. 4(a) and 4(e) are drawn from these coordinates, which depict the location of the original high-resolution components shown in Figs. 4(b) and 4(f). PSNR in decibels and SSIM are computed for both predictions. The calculated individual PSNR in (db)/SSIM is given on top of the predicted images.

Theoretically, a higher score is better for both PSNR and SSIM. The same results are shown for the test set in Fig. 5.

Qualitatively, the predictions look better with SRResNet for training and test sets. With the linear interpolation, the predictions are blurry or more smoothed. Qualitative results are essential since quantitative results are slightly biased. Although the PSNR or SSIM for prediction by SRResNet for the first training example is higher than the prediction by linear interpolation, for all other examples in both training and test set, the PSNR/SSIM are higher with linear interpolation. Higher PSNR means more noise is removed. However, as a least squares result, it is slightly biased toward over-smoothed or blurry results, i.e., an algorithm that eliminates not only the noise but also a part of the textures will have a good score. SSIM has been developed to have a quality reconstruction metric that also considers the similarity of the edges (high-frequency content) between the predicted and original image. To have a good SSIM measure, an algorithm needs to remove the noise while preserving the objects' edges. Although the prediction by SRResNet has less noise than the linear interpolation prediction, the texture might vary a bit, which explains the slightly lower SSIM scores. This is also reflected in Table II, where the PSNR and SSIM scores are lower for other CNN models than the linear interpolation model.

TABLE I. Linear interpolation and super-resolution CNN metrics for the training set.

Metric	Linear Interpolation	SRCNN	RCAN	SRGAN	SRResNet
PSNR	30.75	28.91	29.15	28.44	31.62
SSIM	0.92	0.91	0.90	0.91	0.93

TABLE II. Linear interpolation and super-resolution CNN metrics for the test set.

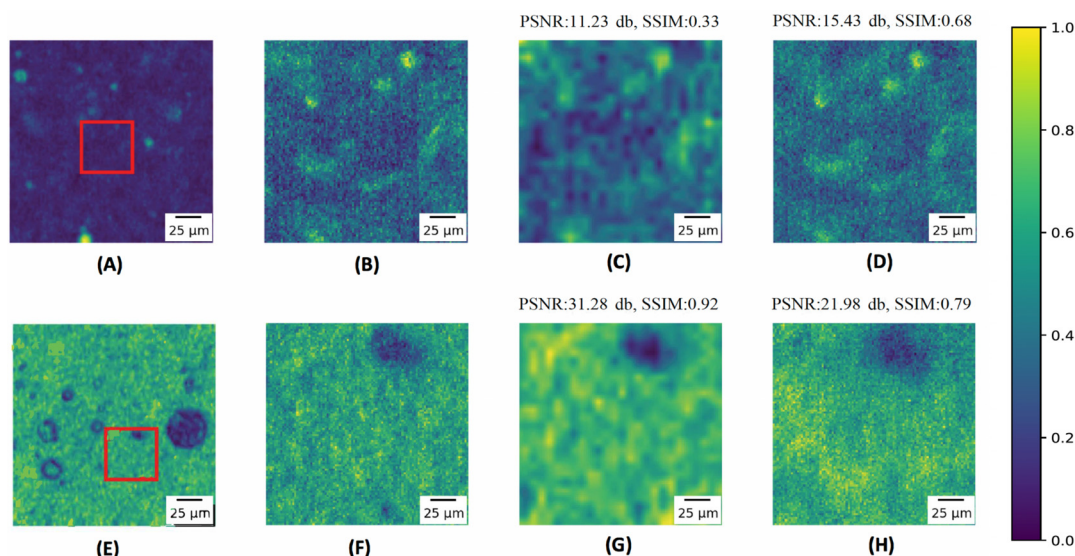| Metric	Linear Interpolation	SRCNN	RCAN	SRGAN	SRResNet
PSNR	27.25	25.82	25.73	26.34	28.93
SSIM	0.89	0.87	0.87	0.88	0.90

FIG. 4. High-resolution NMF components prediction from low-resolution NMF components using SRResNet for two examples from the training set. (A) and (E) The original low-resolution components. The red boxes depict the original high-resolution components' location in (B) and (F). Predictions achieved by linear interpolation are shown in (C) and (G). Finally, (D) and (H) The SRResNet prediction. The calculated individual PSNR in (db) and SSIM are only given on top of the predicted images.

In this study, we demonstrated the effectiveness of state-of-the-art super-resolution deep learning models in reconstructing high-resolution Raman images from low-resolution data, using a small, targeted dataset. Despite the limited data available, our approach achieved robust results by optimizing model training to maximize spatial resolution, thereby significantly reducing the need for time-intensive high-resolution scanning. This approach highlights the feasibility of deep learning in efficiently enhancing spatial resolution for Raman hyperspectral imaging, paving the way for faster and more resource-efficient imaging workflows. We found that using NMF components facilitated data complexity reduction, allowing the model to focus on primary spectral features and minimizing computational demands. However, we recognize that using normalized data directly could be an alternative, doing so without dimensionality reduction may compromise spectral feature preservation and computational efficiency. Potential improvements could involve expanding this work to a multi-
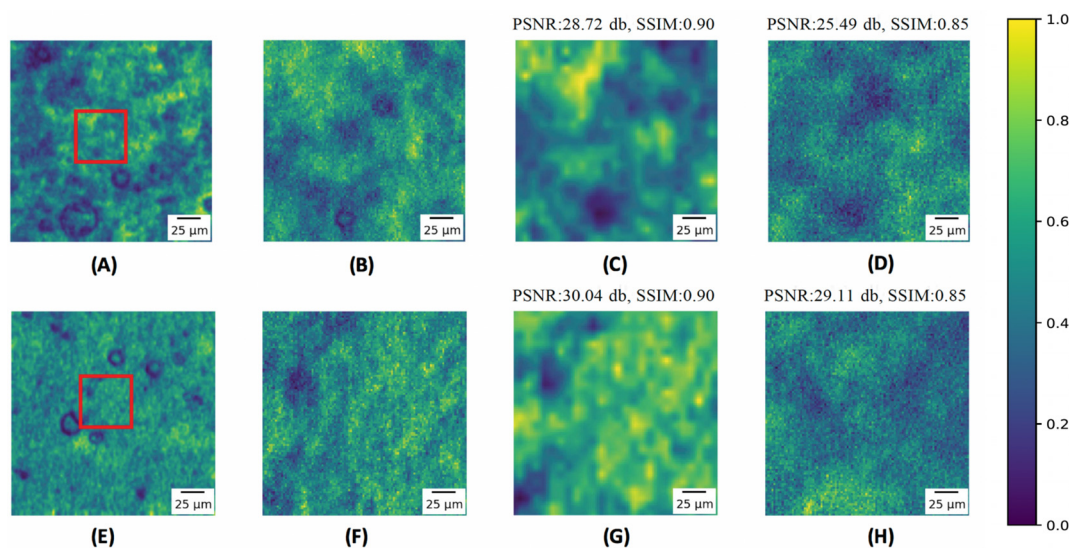

FIG. 5. High-resolution NMF components prediction from low-resolution NMF components using SRResNet for two examples from the test set. (A) and (E) The original low-resolution components. The red boxes depict the original high-resolution components' location in (B) and (F). Predictions achieved by linear interpolation are shown in (C) and (G). Finally, (D) and (H) the SRResNet prediction. The calculated individual PSNR in (db) and SSIM are only given on top of the predicted images.

channel CNN framework that incorporates multiple spectral components, further refining feature capture and predictive capability. Using larger datasets in future work might ensure robustness in terms of overfitting and generalizability. Additionally, while our primary focus is on super-resolution for Raman imaging, we anticipate that this workflow might be adaptable to other imaging modalities requiring high spatial resolution with minimal spectral loss. Our results affirm that super-resolution models can effectively reconstruct high-resolution data from low-resolution images, supporting high-throughput applications while maintaining spectral integrity. Ultimately, we aim to generalize these models for multimodal studies, enabling streamlined, time-efficient imaging across various hyperspectral data applications.

See the supplementary material for figures that are in a separate PDF file.

AUTHOR DECLARATIONS
Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Md Inzamam Ul Haque: Conceptualization (equal); Data curation (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Ariel Lebron:** Resources (equal); Validation (equal); Writing – review & editing (equal). **Frances Joan D. Alvarez:** Investigation (equal); Project administration (equal); Resources (equal); Validation (equal); Writing – review & editing (equal). **Jennifer F. Neal:** Resources (equal); Validation (equal); Writing – review & editing (equal). **Marc Mamak:** Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal). **Debangshu Mukherjee:** Supervision (equal); Visualization (equal); Writing – review & editing (equal). **Olga S. Ovchinnikova:** Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal). **Jacob D. Hinkle:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the Procter & Gamble Company.

REFERENCES

[1]Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
[2]W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," J. Sens. **2015**, 258619.
[3]Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," Neurocomputing **266**, 29–41 (2017).
[4]X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," IEEE Trans. Geosci. Remote Sensing **58**, 4604–4616 (2020).
[5]J. S. Bhatt and M. V. Joshi, "Deep learning in hyperspectral unmixing: A review," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020)* (IEEE, 2020), pp. 2189–2192.
[6]J. Lin, R. Ward, and Z. J. Wang, "Deep transfer learning for hyperspectral image classification," in IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), 2018.
[7]C. C. Horgan, M. Jensen, A. Nagelkerke, J.-P. St-Pierre, T. Vercauteren, M. M. Stevens, and M. S. Bergholt, "High-throughput molecular imaging via deep-learning-enabled Raman spectroscopy," Anal. Chem. **93**, 15850–15860 (2021).
[8]G. P. S. Smith, C. M. McGoverin, S. J. Fraser, and K. C. Gordon, "Raman imaging of drug delivery systems," Adv. Drug Delivery Rev. **89**, 21–41 (2015).
[9]C. J. Rowlands, S. Varma, W. Perkins, I. Leach, H. Williams, and I. Notingher, "Rapid acquisition of Raman spectral maps through minimal sampling: Applications in tissue imaging," J. Biophotonics **5**, 220–229 (2012).
[10]C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Honolulu, HI, 2017), pp. 105–114.
[11]C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," arXiv:1501.00092 (2015).
[12]Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," arXiv:1807.02758 (2018).
[13]W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," arXiv:1609.05158 (2016).
[14]X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, "ESRGAN: Enhanced super-resolution generative adversarial networks," arXiv:1809.00219 (2018).
[15]B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," arXiv:1707.02921v1 (2017).
[16]A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Anal. Chem. **36**, 1627–1639 (1964).
[17]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Ā. Duchesnay, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825–2830 arXiv:1201.0490 (2011).
[18]K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," London, Edinburgh, Dublin Philos. Mag. J. Sci. **2**, 559–572 (1901).
[19]D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems* (MIT Press, 2000), Vol. 13.
[20]Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," in *IEEE Transactions on Image Processing* (IEEE, 2004), Vol. 13, pp. 600–612.
[21]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.