Deep Learning-Assisted Dynamic Mode Decomposition for Non-resonant Background Removal in CARS Spectroscopy

Adithya Ashok Chalain Valapil[1][0009−0004−3807−5144], Carl Messerschmidt[2][0009−0007−3056−6063], Maha Shadaydeh[1][0000−0001−6455−2400], Michael Schmitt[2][0000−0002−3807−3630], Jürgen Popp[2,3][0000−0003−4257−593X], and Joachim Denzler[1][0000−0002−3193−3300]

[1] Computer Vision Group, Friedrich-Schiller University, 07743, Jena, Germany
[2] Leibniz-Institute of Photonic Technology, 07745 Jena, Germany
[3] Institute of Physical Chemistry, Friedrich-Schiller University, 07743 Jena, Germany
{adithya.ashok,carl.messerschmidt,maha.shadaydeh,m.schmitt,juergen.popp, joachim.denzler}@uni-jena.de https://inf-cv.uni-jena.de

Abstract. Coherent Anti-Stokes Raman Spectroscopy (CARS) provides non-invasive, label-free chemical analysis at high spatial resolution, making it a powerful tool for biomedical and material imaging. However, their effectiveness is hindered by a dominant and unpredictable non-resonant background (NRB) that distorts meaningful spectral features. Existing NRB removal methods often require additional measurements or computationally intensive post-processing. In this work, we present a physics-informed framework that leverages the broadband, low-rank structure of the NRB using Dynamic Mode Decomposition (DMD) for unsupervised separation of resonant Raman modes from non-resonant contributions in the spectral domain. We further introduce **DA-DMD** - a Deep Learning-Assisted DMD approach, that uses an attention mechanism to adaptively weight DMD modes and a CNN with skip connection to enhance Raman signal reconstruction. Trained entirely on synthetic data, DA-DMD eliminates the need for experimental labels or calibration. We validate our methods on synthetic and real CARS measurements, demonstrating superior background suppression, fidelity preservation, and generalization compared to existing approaches. DA-DMD offers fast inference and improves robustness, positioning it as a practical tool for scalable chemical imaging in complex environments.

Keywords: Dynamic Mode Decomposition · Raman Spectroscopy · Channel Attention · Delay Embedding · Unsupervised Learning.

1 Introduction

Analyzing molecular composition is fundamental to fields such as biology, materials science, and medicine. Spectroscopy, particularly Raman scattering, enables label-free chemical imaging from spectral measurements by probing vibrational

modes of molecules [27, 33]. However, traditional Spontaneous Raman (SR) spectroscopy suffers from weak signal intensity, requiring long integration times that limit its use in real-time applications. Coherent Anti-Stokes Raman Spectroscopy (CARS) overcomes this limitation by using nonlinear optical interactions to coherently excite molecular vibrations, boosting signal strength by several orders of magnitude [14]. This makes CARS ideal for high-speed, chemically specific spectral analysis. However, CARS suffers from a critical challenge: the presence of a non-resonant background (NRB) - an intense, broadband signal arising from electronic nonlinearities rather than molecular vibrations. This NRB distorts spectral line shapes and conceals the very information CARS aims to reveal [37].

This study proposes decomposing the CARS spectra into two components, namely the NRB and the Raman signatures, using dynamic mode decomposition (**DMD**). The NRB spectrum is typically broad and featureless [23], implying slow changes over wavenumber or low-frequency dynamics, which can be separated from the Raman spectra that exhibit fast changes over wavenumber or high-frequency dynamics.

DMD is a data-driven method for extracting spatiotemporal patterns in sequential measurements [28]. This is achieved by decomposing these measurements into a set of modes, each of which is associated with a fixed oscillation frequency and a decay/growth rate over time [50]. DMD has been traditionally applied to spatiotemporal dynamic systems in diverse fields such as fluid mechanics [41], robotics [6], ecosystems [42] and neuroscience [8]. For the first time, in this study, we adapt DMD for spectral analysis by reinterpreting wavenumber progression as a pseudo-temporal coordinate. The Raman signature embedded in the CARS spectra causes strong correlation at adjacent wavenumbers of the CARS spectra due to the finite vibrational mode bandwidth, which has a peak centered at a wavenumber and influences the intensity values over a range of adjacent wavenumbers (lineshapes) [14, 17]. These correlations manifest as a sequence of local and fast intensity variations across the wavenumber axis, creating a sequential structure that we exploit for NRB removal.

We introduce two approaches using DMD. 1. **Unsupervised DMD**, which clusters the DMD modes into low-frequency (NRB) and high-frequency (Raman) components using spectral priors, while requiring no training data; 2. Deep Learning-Assisted DMD (**DA-DMD**), where we propose to enhance the modes selection by combining the DMD decomposition capability with channel attention [20, 46] and further refine the reconstructions via convolutional networks [36].

The efficacy of both approaches is validated through quantitative and qualitative analyses, employing both synthetic and experimental data sets. The results demonstrate that our unsupervised DMD approach not only matches but also surpasses conventional unsupervised methods in suppressing NRB. Additionally, DA-DMD attains state-of-the-art accuracy with expedited inference while training with considerably fewer samples compared to other deep-learning methods.

2 Related Works

Existing NRB removal strategies face critical trade-offs. There are experimental methods like interferometric CARS [38], frequency modulation CARS [16], and polarization CARS [13] for narrow-band operation, as well as time-resolved CARS [48] and Fourier transform CARS [35] for broadband mode. These methods suppress NRB at the source but require complex instrumentation and sacrifice signal strength.

Alternative approaches focus on the post-processing of measured spectra. There are widely used methods like time domain Kramers-Kronig (TDKK) [11,30] and Maximum Entropy Method (MEM) [45], which are computationally intensive. Therefore, the discrete Hilbert transform used in these methods was later replaced by a learned matrix approach in LeDHT [10]. Another approach uses singular value decomposition (SVD) and transformed basis vectors to establish the "factorized Kramers–Kronig and error correction" (fKK-EC) method [12]. These two methods significantly improved the speed of the phase retrieval task. However, these traditional methods require extra reference spectra, which can be time-consuming and may not always be available [25].

Conversely, deep learning methods offer a promising solution for NRB removal without requiring reference spectra. Methods like SpecNet [24, 44], long short-term memory (LSTM) [19], bi-directional LSTM (Bi-LSTM) [25], very deep convolutional autoencoders (VECTOR) [32, 49], generative adversarial networks (GAN) [31, 47] and convolutional neural networks with gated recurrent units (CNN+GRU) [47] can provide a more efficient and effective way to remove NRB and recover the underlying Raman spectra by training on synthetic spectra. We compare our DA-DMD approach with these methods.

Using multiscale frequency analysis, the wavelet prism signal decomposition technique [39] separates Raman features from NRB. This was recently extended to interpolated inverse discrete wavelet transforms that we refer to as IWT in our work [18]. We also compare our unsupervised DMD setup with this method.

3 Unsupervised Hankelized DMD for NRB Removal

3.1 DMD: Preliminaries

Let $\mathbf{x}_j \in \mathbb{R}^M$ denote sequential data with time index $j = 1, ..., N$, where M and N denote the dimension of each measurement and the number of measurement snapshots, respectively. DMD finds a best-fit linear operator $\mathbf{A} \in \mathbb{R}^{M \times M}$ such that $\mathbf{x}_{j+1} \approx \mathbf{A}\mathbf{x}_j$ holds for all pairs of measurements j [5]. Let $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \, \mathbf{x}_2 \cdots \mathbf{x}_{N-1} \end{bmatrix}$ and the time-shifted matrix $\mathbf{X}' = \begin{bmatrix} \mathbf{x}_2 \, \mathbf{x}_3 \ldots \mathbf{x}_N \end{bmatrix}$. This relationship can be described in a matrix form as $\mathbf{X}' \approx \mathbf{A}\mathbf{X}$. The best-fit linear operator is computed as $\mathbf{A} = \mathbf{X}'\mathbf{X}^\dagger$ where † is the Moore-Penrose pseudoinverse, which is computed using the singular value decomposition (SVD) of $\mathbf{X}$. The eigendecomposition of $\mathbf{A}$ yields the eigenvalues and eigenvectors (modes) that can be used for signal

reconstruction. Specifically, let ϕ_i and λ_i be the ith eigenvector-eigenvalue pair of $\mathbf{A}$, $\mathbf{x}_j$ at any time j can be reconstructed as

$$\mathbf{x}_j = \sum_{i=1}^{M} b_i \phi_i \lambda_i^j \tag{1}$$

where b_i denotes the weights of the modes in the initial state $\mathbf{x}_0$.

The eigenvalues λ_i are in general complex numbers that can be defined as $\lambda_i = e^{(\sigma_i + j2\pi f_i)\Delta(t)}$ where σ_i and f_i are respectively the growth/decay rate and the eigenfrequency for the i-the mode, and $\Delta(t)$ is the time interval between two consecutive snapshots of measurement.

3.2 Hankelized DMD of CARS Spectra

Time delay embedding is an established method for the geometric reconstruction of attractors for nonlinear systems [40, 43]. As traditional DMD assumes Markovian dynamics, where only the present step matters and lacks memory of past states, we employ delay embedding (Hankelization) [9, 51], which captures hidden correlations across wavenumber.

In our study, the dimension of the measurement at each wavenumber of the spectrum equals one. To match the definition of $\mathbf{x}_j$ in Section 3.1 and to be able to use DMD to decompose the spectrum into M modes, we use an M dimensional delay embedding vector to represent the wavenumbers. This forms a Hankel matrix $\mathbf{H}$ on input $\mathbf{x}_j$ as in Eq. 2 which embeds pseudo-temporal dynamics into the spectrum, allowing us to analyze the spectral evolution more comprehensively. The Hankel matrix $\mathbf{H}$ is constructed as follows:

$$\mathbf{H} = \begin{bmatrix} x_1 & x_2 & .. & x_L \\ x_2 & x_3 & .. & x_{L+1} \\ . & . & .. & . \\ . & . & .. & . \\ x_M & x_{M+1} & .. & x_N \end{bmatrix}. \tag{2}$$

Here, N refers to the length of original spectrum and $L = N - M + 1$. The embedding delay dimension M is calculated as delay or lag τ that encapsulates maximum information between the spectrum and its delayed version. It can be calculated using autocorrelation function (ACF) metric [7] with Eq. 9 in Appx. A. An example is shown for ethanol spectrum in Fig. 5. Here, the first major dip in the ACF curve at $\tau = 12$ is considered a good value for M in this case.

3.3 Mode Clustering for NRB Removal

After the decomposition of the spectrum into M modes, we separate these modes into two clusters; the first low-frequency cluster is used to reconstruct the NRB, while the second high-frequency cluster is used to reconstruct the Raman signatures. We show this interpretation using ethanol as an example

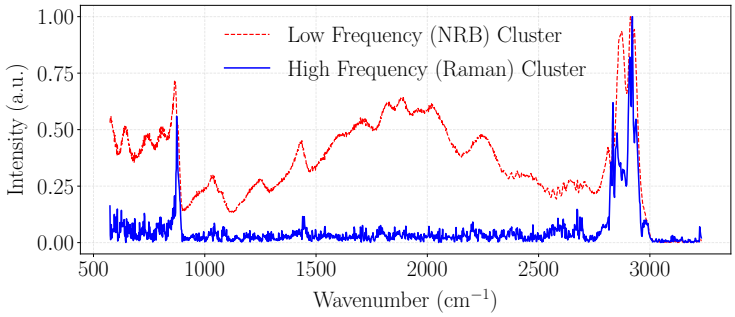

Fig. 1: **Clustering example.** Reconstruction of clusters for separating the Raman signatures (in blue) and NRB (in red) is shown. A 2-class clustering is applied on the frequencies of the eigenvalues λ_i to extract the Raman modes. Results are shown before the removal of very high-frequency modes in the Raman signature.

in Appx. B Fig. 6. The first mode resembles NRB spectral contribution, while the combination of the second and third modes shows major similarity with the Raman line shapes. Some of the missing peaks are often present at higher modes. To separate these M modes, we employ 2-class clustering [21] to group the eigenfrequencies of λ_i into two clusters, as illustrated in Fig. 1. The reconstructed cluster, according to Eq. 1, without the first mode, is expected to contain the Raman signatures. To refine results, we iteratively reapply DMD as suggested in [34] to the previous high-frequency cluster to eliminate the very high-frequency modes. Therefore, the cluster with the first mode is retained. The recursive process is terminated when a frequency above a threshold is eliminated [15, 29]. We learn this threshold from the Fourier transform of the spectrum. This process ultimately removes very high-frequency noises.

4 DA-DMD: Deep Learning-Assisted DMD

The main challenge in the above-discussed DMD method is the selection of relevant modes using clustering. Assuming the ideal selection of clusters, we still need parametrized smoothing functions like a recursive DMD process that sometimes even out relevant peaks. To address these limitations, we propose a deep learning-assisted DMD (DA-DMD) approach as shown in Fig. 2. We introduce an attention mechanism after Hankelized DMD to learn the relevance of each mode using synthetic training data, as each mode contributes differently to the final reconstructed spectrum. Once the importance of each mode is learned, a set of convolutional layers reduces multi-mode input to a single-channel representation corresponding to the clean spectrum.

A Squeeze-and-Excitation (SE) block is used for attention in our model [20]. It optimizes the feature representation by adaptively reweighing the extracted DMD modes based on the most relevant spectral components, while suppressing less significant ones. The squeeze operation condenses global information from

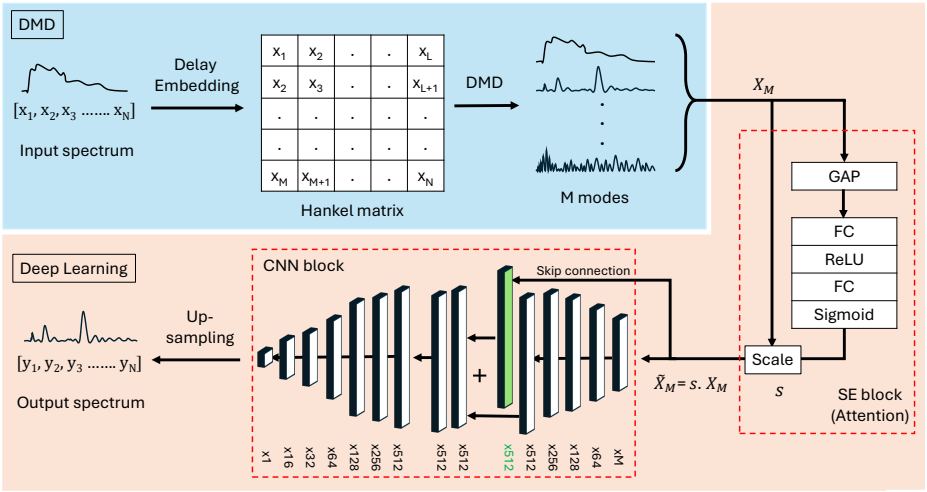

Fig. 2: **Flow diagram of DA-DMD**. Once the modes are extracted using Hankelized DMD, a Squeeze-and-Excitation (SE) block [20] is used for channel attention to extract the relevance of each mode (Eq. 4). A set of convolutional layers is then used to produce denoised spectra.

the extracted DMD modes by Global Average Pooling (GAP). For an input $X \in \mathbb{R}^{B\times M\times N}$ with batch size B, number of modes (equal to delay embedding) M and spectral length N, we obtain activation of mode m at spectral point i for batch b as $X_{b,m,i}$. The GAP reduces each mode to a single value as:

$$\bar{X}_m = \frac{1}{N}\sum_{i=1}^{N} X_{m,i}. \tag{3}$$

This reduces X to a compressed feature vector $\bar{X} \in \mathbb{R}^{B\times M}$, which represents the global contribution of mode m across the entire spectrum. Then comes the excitation step, where we pass $\bar{X}_m$ through a fully connected network or gated mechanism to determine the importance of each mode. The excitation mechanism consists of two fully connected layers with ReLU activation, followed by a sigmoid activation to obtain the mode attention vector as:

$$s = \sigma(W_2 \cdot ReLU(W_1 \cdot \bar{X})), \tag{4}$$

where $W_1 \in \mathbb{R}^{(M/k)\times M}$ reduces the dimension by a factor of k and $W_2 \in \mathbb{R}^{M\times(M/k)}$ re-expands it back to match M. The extracted DMD modes are scaled using learned importance weights, i.e., $\tilde{X}_m = s_m.X_m$.

Following the SE block's mode weighting, the architecture employs a CNN block with a set of convolutional layers [36] that expands the multi-channel input (corresponding to DMD modes) to capture fine characteristics, which then collapses into a single feature map (corresponding to Raman spectra). A skip connection is used from the SE block's output to a deeper convolutional layer

as shown in Fig.2. This preserves spatial information by bridging the shallow and deep representations during the forward pass in the network. The deeper layers are used to enhance gradient flow and enable iterative feature refinement. This design choice prevents feature dilution and ensures that the remaining CNN layers reconstruct a clean and accurate Raman spectrum from the learned representations. As the delay embedding step alters the results in reduced spectral length, a linear interpolation-based upsampling layer is added at the end to ensure that the final spectrum $y \in \mathbb{R}^N$ has the same length as the input.

5 Experiments and Results

Experiments are done on synthetic and real samples. We compare the proposed methods with several state-of-the-art techniques, including IWT, TDKK, and other deep learning (DA-DMD, VECTOR, GAN, CNN+GRU, Bi-LSTM, LSTM, SpecNet) methods. Both DMD and IWT are unsupervised methods and require no training. The results of each experiment are discussed mainly in comparison to those of DA-DMD.

We include both qualitative analysis and quantitative metric evaluation. Qualitatively, we discuss the characteristics of a good Raman reconstruction by the visual quality of peak occurrence and intensities. Quantitatively, we compare using two metrics: Mean Square Error (MSE) and Pearson's correlation (r), which are calculated as follows.

$$\text{MSE} = \sum_{i=1}^{K} (\hat{y}_i - y_i)^2, \quad r = \frac{\sum_{i=1}^{K} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{K} (\hat{y}_i - \bar{\hat{y}})^2} \sqrt{\sum_{i=1}^{K} (y_i - \bar{y})^2}}, \tag{5}$$

where $\hat{y}$ and y are predicted and ground truth respectively. The mean of K spectra is denoted by $\bar{y}$. The MSE is an error metric that is best when close to 0, while r shows the best similarity when close to 1.

5.1 Experiments on synthetic spectra

Synthetic data generation. We generate synthetic Raman-CARS pairs following [44], modeling Raman spectra as Lorentzian peaks(Eq. 6) with random NRB as noise (Eq. 7) among a fourth-degree polynomial and double sigmoid function. For each spectral pair, the function parameters are also randomly selected as specified below.

$$\text{Resonant: } \chi_r^{(3)}(\omega) = \sum_{p=1}^{P} \frac{E_p}{\Omega_p - \omega - i\gamma_p}, \tag{6}$$

where, $P \sim \mathcal{U}\{1, 25\}$ is number of peaks varied per sample, $E_p \in \mathcal{U}(0.01, 1.0)$ is the amplitude, $\Omega_p \in \mathcal{U}(0, 1)$ is the normalized resonance frequency, $\gamma_p \in \mathcal{U}(0.001, 0.02)$ is the linewidth and ω is the normalized Raman shift ranging 1000 points over [0, 1].

$$\text{Non-resonant: } \chi_{nrb}^{(3)}(\omega) = \begin{cases} \frac{1}{1+e^{-b_1(\omega - c_1)}} \cdot \frac{1}{1+e^{b_2(\omega - c_2)}} & \text{(Sigmoid)} \\ a\omega^4 + b\omega^3 + c\omega^2 + d\omega + e & \text{(Polynomial)} \end{cases}, \tag{7}$$

where, $b_1, b_2 \sim \mathcal{N}(10, 5)$ control the steepness of the rising and falling rates of the sigmoid functions, $c_1 \sim \mathcal{N}(0.2, 0.3)$ and $c_2 \sim \mathcal{N}(0.7, 0.3)$ determine the position of inflection points of the sigmoid function. The coefficients of polynomial function are $a, c \sim \mathcal{U}(-1, 1)$ and $b, d, e \sim \mathcal{U}(-10, 10)$.

$$\text{CARS spectrum: } I_{\text{CARS}}(\omega) \propto \left|\chi_r^{(3)}(\omega) + \chi_{\text{nrb}}\right|^2 , \tag{8}$$

to which a random noise $\sim \mathcal{U}(0.0005, 0.003)$ is also added to simulate high-frequency disturbances. The total spectrum is normalized to ensure that learning focuses on the spectral shape rather than absolute intensity.

DA-DMD architecture selection. Our DA-DMD model is implemented on PyTorch 2.2.1 with CUDA 12.1 and cuDNN 8.9.2. The architecture mainly comprises an SE block for reweighting the modes and a CNN block for supervision and spectral refinement. Ablation studies (Tab. 1) on synthetic spectra have been conducted to confirm the necessity of both of these blocks. We used 1700 spectra to train our model for each setup and then test our prediction on 300 samples. In the first setup (Without SE block), the SE channel attention block is removed by keeping only the CNN layers. In the second setup (Without CNN block), the result of the attention block is reduced to one prediction result. We see that the performance of our model without either SE or CNN block is worse than the complete DA-DMD model.

DA-DMD Model	MSE	Corr
Complete model	0.0005 ± 0.0008	0.9494 ± 0.1607
Without SE block	0.0050 ± 0.0025	0.6026 ± 0.1621
Without CNN block	0.0033 ± 0.0021	0.7664 ± 0.1580

Tab. 1: **Ablation Study.** Testing the impact of each of the SE and CNN blocks in the deep learning architecture of the DA-DMD model based on MSE and correlation (mean ± standard deviation) for 300 synthetic spectral predictions.

In contrast to our CNN block, other architectures like autoencoders were tested and failed. Various techniques like batch normalization, spatial attention, and multiple skip connections were tested to optimize the architecture, but they did not improve the performance significantly. We, therefore, came up with the architecture detailed in 4. It is trained for 50 epochs. We used Adam Optimizer with a learning rate of 0.001 during backpropagation [26]. The number of delay embedding is selected as $M = 12$ by estimating using ACF as discussed in Section 3.2. The reduction ratio $k = 4$ is optimized among common values used in the original SE block [20]. The Mean Square Error (MSE) is our loss function.

	DA-DMD	VECTOR	GAN	CNN+GRU	Bi-LSTM	LSTM	Spec Net
# Training samples	**2k**	200k	200k	200k	50k	50k	50k
Training time/epoch (s)	**6.6**	18.7	142	6667	1360	240	73
# Parameters	0.5M	111.8M	6.2M	84.0k	5.2k	**3.9k**	6.0M
Prediction time (ms)	6	7	**1**	713	168	103	71

Tab. 2: **Model evaluation.** Comparison of computational requirements in terms of the training (per epoch) and prediction time (per spectrum), number of samples, and number of parameters for all deep learning models.

Deep learning model evaluation. All deep learning models are trained on synthetic datasets. We use available pretrained models or training setup for VECTOR [4], GAN [2], CNN+GRU [2], Bi-LSTM [3], LSTM [3] and SpecNet [1].

The time-domain Kramers Kronig (TDKK) algorithm needs reference measurements and requires optimizations for phase and scale retrieval parameters. Another major drawback they have is the high computational time. Deep learning models overcome these limitations. The computational complexity of the deep learning models is compared in Tab. 2. All experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 1080 GPU (40 GB VRAM), Intel Core i5-11600 CPU, and 32 GB RAM, running Ubuntu 22.04. It is clear that DA-DMD has a significantly low training time. This is mainly because of the need for a significantly lower number of training data (2000 synthetic samples), as DMD largely supports the network with preprocessed or decomposed modes. The prediction time is also low compared to all other models (except GAN). VECTOR also performs closely well in the computational complexity test. Though all the Recurrent Neural Networks (RNN) based models (LSTM, Bi-LSTM, CNN+GRU) have fewer trainable parameters than DA-DMD, they demand high computational time.

Analysis of results. A comparison of all models is conducted with 1000 newly simulated spectra to assess the consistency of performance. All models' performance for a synthetic sample is shown in Fig. 3. The unsupervised DMD and IWT produce peaks at the correct location. The results of IWT show good peak recovery, but a slight phase shift often remains in the unsupervised setting. DMD also produces good performance, predicting peaks that correlate with Raman peaks. From Tab. 3, we can see that DA-DMD and VECTOR perform best considering all metrics. According to the previous study [47], GAN and CNN+GRU are expected to perform better for synthetic spectra. This performance discrepancy is due to the difference in hyperparameters for synthetic training data generation. Fig. 7 in Appx. C shows the results of a second test done using a synthetic data generator provided by [47]. Both GAN and CNN+GRU models perform best in this secondary test. Our DA-DMD and VECTOR show consistently good results across both tests. LSTM shows the worst performance among the deep learning models.

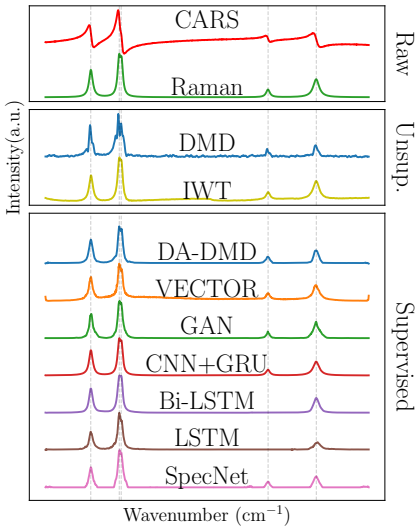

Fig. 3: Qualitative performance of **synthetic spectra** with reconstruction on one sample with all models.

Model	MSE	Corr
DMD	**0.0063** ±0.0132	**0.67** ±0.11
IWT	0.0109 ±0.0062	0.21 ±0.19
DA-DMD	**0.0006** ±0.0006	**0.98** ±0.01
VECTOR	0.0007 ±0.0010	0.97 ±0.02
GAN	0.0033 ±0.0073	0.88 ±0.14
CNN+GRU	0.0031 ±0.0017	0.90 ±0.09
Bi-LSTM	0.0017 ±0.0023	0.90 ±0.13
LSTM	0.0221 ±0.0430	0.78 ±0.31
SpecNet	0.0013 ±0.0016	0.94 ±0.05

Tab. 3: Quantitative performance using MSE and correlation is shown on **synthetic data** for the unsupervised DMD, IWT, and the seven supervised deep learning models (DA-DMD, VECTOR, GAN, CNN+GRU, Bi-LSTM, LSTM, SpecNet).

5.2 Experiments on real CARS spectra

There are two types of real samples used in our experiments. We use spectra measured using broadband CARS (BCARS) setup. Toluene CARS (3- & 2-color case as in [47]) and Raman spectra, along with the experimental results of existing deep learning models, are provided by Vernuccio et al. [47]. Ethanol CARS and Raman spectra are measured on BCARS, setup as in [22]. The TDKK reconstruction is obtained from the available CARS and NRB spectra. The deep learning models used in the previous section are used for prediction here. Both samples are homogeneous in nature, and the tests on them provide a good understanding of the performance in simple real-world scenarios. Fig. 4(a) and Fig. 4(b) show the reconstructed toluene and ethanol spectra, respectively.

Quantitative analysis. As illustrated in Tab. 4, our proposed DA-DMD achieves state-of-the-art performance on the fingerprint region of the ethanol spectrum, attaining the lowest MSE (0.0031) and highest correlation (0.8767) among all models, including TDKK, which uses extra reference measurements. Among the deep learning methods, ours shows a clear reduction in MSE and increase in correlation, simply from comparison to the second-best SpecNet model (MSE: 0.0043, correlation: 0.8367). DA-DMD also notably improves upon its unsupervised counterpart DMD (MSE: 0.0151, correlation: 0.3673), demonstrating the effectiveness of our domain adaptation framework. While the GAN and CNN+GRU models show moderate correlation scores (0.7818 - 0.8113), their MSE values (0.0181 - 0.0071) remain substantially higher than DA-DMD's. The

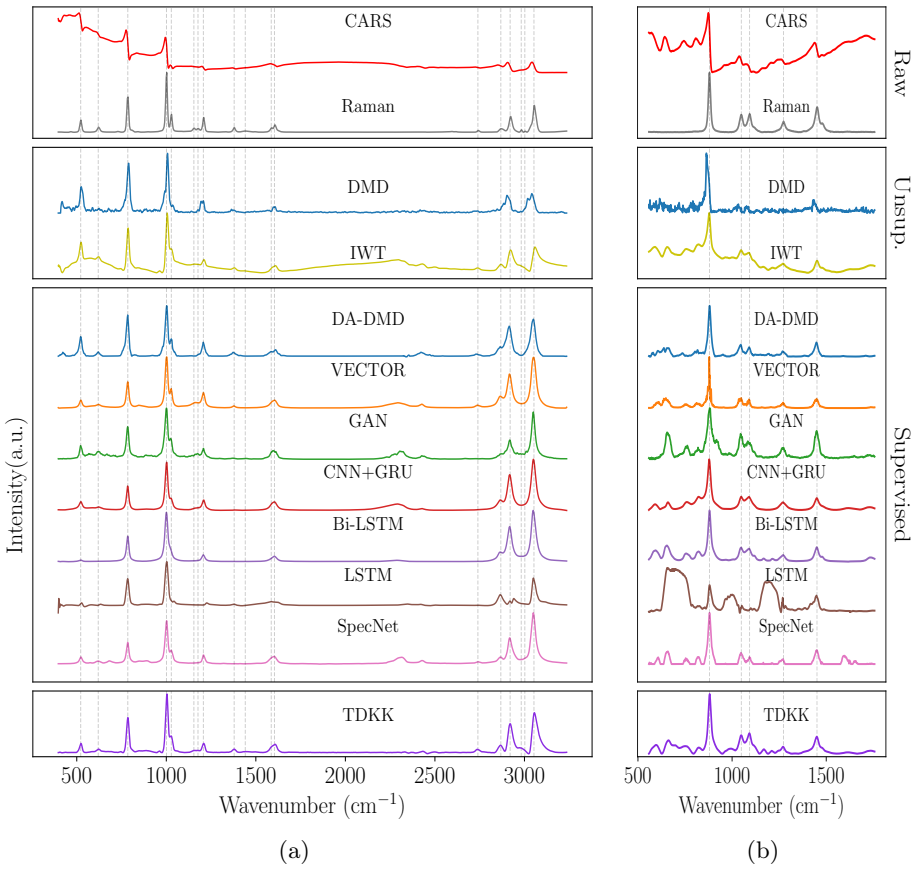

Fig. 4: Qualitative reconstruction performance on **real spectra**: (a) toluene and (b) ethanol fingerprint spectra. Dotted lines mark the original Raman peaks.

LSTM model fails completely with a negative correlation (-0.0054), highlighting the challenge of sequential modeling of this task.

For toluene analysis, DA-DMD delivers the highest correlation (0.8746) with competitive MSE (0.0060) just after GAN (0.0057), outperforming all other deep learning approaches, including Bi-LSTM (MSE: 0.0080, correlation: 0.8017) and SpecNet (MSE: 0.0063, correlation: 0.8043). While the TDKK baseline achieves marginally better MSE (0.0037) and correlation (0.8970), DA-DMD's performance remains within 0.0023 MSE units of this benchmark. The GAN model's strong MSE (0.0057) but lower correlation (0.8699) suggests that it captures magnitude variations better than spectral shapes, whereas DA-DMD balances both metrics effectively. Our method shows remarkable improvement over the original DMD implementation, increasing correlation by 18.8% (DMD: 0.7359 $\rightarrow$ DA-DMD: 0.8746) despite slightly higher MSE. LSTM shows the lowest performance among

	DMD	IWT	DA-DMD	VECTOR	GAN	CNN+GRU	Bi-LSTM	LSTM	Spec Net	TDKK
Tol - MSE	**0.0052**	0.0150	0.0060	0.0087	**0.0057**	0.0084	0.0080	0.0244	0.0063	0.0037
Tol - Corr	**0.7359**	0.7288	**0.8746**	0.8332	0.8699	0.8023	0.8017	0.8656	0.8043	0.8970
EtOH - MSE	**0.0151**	0.0502	**0.0031**	0.0047	0.0181	0.0071	0.0064	0.1267	0.0043	0.0050
EtOH - Corr	0.3673	**0.4270**	**0.8767**	0.7978	0.7818	0.8113	0.8245	-0.0054	0.8367	0.8753

Tab. 4: Quantitative performance on **real spectra** measured using MSE and correlation for all models on toluene (Tol) and ethanol (EtOH) reconstruction, qualitatively shown in Fig. 4.

the models when considering MSE. However, it shows a decent correlation of 0.8656 compared to other models. These results show the relevance of including multiple metrics for evaluation and the importance of qualitative analysis.

Qualitative analysis. For ethanol's fingerprint region (Fig. 4(b)), most deep learning models successfully capture all major spectral features but introduce high-frequency noise due to unsmoothed reconstructions. DA-DMD shows fewer artifacts and detects all peaks. Most methods struggle with minor false positives concentrated below 800 cm^{-1}. VECTOR, Bi-LSTM, and CNN+GRU also reconstruct good peak shapes at the correct positions. Notably, LSTM produces severe distortions, including artificial peaks at 700 cm^{-1} and 1200 cm^{-1} that obscure true spectral content. While unsupervised DMD and IWT avoid these artifacts, their reconstructions lack the sharpness required for precise intensity quantification, particularly in overlapping band regions.

As shown in Fig. 4(a), DA-DMD, Bi-LSTM, and GAN demonstrate superior performance in reconstructing toluene spectra, with both methods accurately resolving peak positions and relative intensities across the characteristic vibrational bands, especially in detecting peak 5. Very small single peaks (2, 9, 12) and double peaks (10, 11) are better detected in DA-DMD. Most of the models (except DMD and VECTOR) predict slight false variations in the silent region (1800 - 2800 cm^{-1}). The unsupervised DMD and IWT methods show promising baseline agreement but struggle with precise peak localization. DMD introduces positional errors, while IWT suffers from inconsistent phase correction artifacts.

6 Conclusion

In this study, we use DMD to decompose CARS spectra into different modes where low-frequency modes correspond to NRB, and high frequencies show Raman contribution. We then employ an unsupervised two-class clustering approach to remove NRB. Compared to state-of-the-art unsupervised methods such as IWT, DMD matches or outperforms IWT. To enhance the performance further, we then replace this mode clustering step with a supervised deep learning approach, DA-DMD, which improves mode selection. Compared to state-of-the-art deep learning methods, DA-DMD offers several advantages. 1. It provides low-latency

prediction (less than 10 ms per spectrum); 2. There is no need for reference measurements, and only minimal training on synthetic data is required, making DA-DMD easily adaptable to new samples and setups with minimal changes; 3. It accurately reconstructs the Raman signatures for both synthetic and real spectra. However, similar to other methods, our DMD-based models are sensitive to NRB magnitudes and peak overlap resolution. Future work will focus on: 1. Optimizing the deep learning architecture (layer depth and DMD mode count), informed by physical constraints in the loss function; 2. Synthesizing training data that better captures real-world spectral complexity; 3. Including comprehensive, comparable data from FT-Raman (ground truth) and BCARS spectra of various samples, to strengthen the broad applicability of the proposed methods.

Code availability

The DA-DMD model with related code and training data is available in GitHub https://github.com/spectra-analysis/DA_DMD.

Acknowledgment

This study was funded by the European Union's Horizon Europe research and innovation program for the project uCAIR with Grant Agreement No. 101135175.

References

1. Valensise et al., C.: SpecNet code, available at https://github.com/Valensicv/SpecNet
2. Vernuccio et al., F.: GAN and CNN+GRU code, available at https://github.com/crimson-project-eu/NRB_removal
3. Junjuri et al., R.: LSTM and Bi-LSTM code, available at https://github.com/Junjuri/Four-DL-models-comparison-for-evaluating-CARS
4. Wang et al., Z.: VECTOR code, available at https://github.com/villawang/VECTOR-CARS
5. Baddoo, P.J., Herrmann, B., McKeon, B.J., Nathan Kutz, J., Brunton, S.L.: Physics-informed dynamic mode decomposition. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **479**(2271) (2023). https://doi.org/10.1098/rspa.2022.0576
6. Berger, E., Sastuba, M., Vogt, D., Jung, B., Ben Amor, H.: Estimation of perturbations in robotic behavior using dynamic mode decomposition. Advanced Robotics **29**(5), 331–343 (2015). https://doi.org/10.1080/01691864.2014.981292
7. Box, G.E., Jenkins, G.M.: Time series analysis: forecasting and control. Holden-Day (1976)
8. Brunton, B.W., Johnson, L.A., Ojemann, J.G., Kutz, J.N.: Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. Journal of Neuroscience Methods **258**, 1–15 (2016). https://doi.org/10.1016/j.jneumeth.2015.10.010

9. Brunton, S.L., Brunton, B.W., Proctor, J.L., Kaiser, E., Kutz, J.N.: Chaos as an intermittently forced linear system. Nature Communications **8**(1) (2017). https://doi.org/10.1038/s41467-017-00030-8
10. Camp, C.H.: Raman signal extraction from cars spectra using a learned-matrix representation of the discrete hilbert transform. Optics Express **30**(15), 26057 (2022). https://doi.org/10.1364/oe.460543
11. Camp, C.H., Lee, Y.J., Cicerone, M.T.: Quantitative, comparable coherent anti-Stokes Raman scattering (CARS) spectroscopy: Correcting errors in phase retrieval. Journal of Raman Spectroscopy **47**(4), 408–415 (2016). https://doi.org/10.1002/jrs.4824
12. Camp, Jr, C.H., Bender, J.S., Lee, Y.J.: Real-time and high-throughput raman signal extraction and processing in CARS hyperspectral imaging. Opt. Express **28**(14), 20422–20437 (2020). https://doi.org/https://doi.org/10.1364/OE.397606
13. Cheng, J.X., Book, L.D., Xie, X.S.: Polarization coherent anti-stokes raman scattering microscopy. Optics Letters **26**(17), 1341 (2001). https://doi.org/10.1364/ol.26.001341
14. Cheng, J.X., Xie, X.S.: Coherent anti-stokes raman scattering microscopy:, instrumentation, theory, and applications. The Journal of Physical Chemistry B **108**(3), 827–840 (2003). https://doi.org/10.1021/jp035693v
15. Cooley, J.W., Lewis, P.A.W., Welch, P.D.: The fast fourier transform and its applications. IEEE Transactions on Education **12**(1), 27–34 (1969). https://doi.org/10.1109/te.1969.4320436
16. Ganikhanov, F., Evans, C.L., Saar, B.G., Xie, X.S.: High-sensitivity vibrational imaging with frequency modulation coherent anti-stokes raman scattering (fm cars) microscopy. Optics Letters **31**(12), 1872 (2006). https://doi.org/10.1364/ol.31.001872
17. Griffiths, P.R.: Introduction to the theory and instrumentation for vibrational spectroscopy (2001). https://doi.org/10.1002/0470027320.s8935
18. Härkönen, T., Vartiainen, E.: Interpolated inverse discrete wavelet transforms in additive and non-additive spectral background corrections. Optics Continuum **2**(5), 1068 (2023). https://doi.org/10.1364/optcon.488136
19. Houhou, R., Barman, P., Schmitt, M., Meyer, T., Popp, J., Bocklitz, T.: Deep learning as phase retrieval tool for CARS spectra. Optics Express **28**(14), 21002 (2020). https://doi.org/10.1364/OE.390413
20. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7132–7141 (2018). https://doi.org/10.1109/CVPR.2018.00745
21. Jin, X., Han, J.: K-Means clustering. In: Encyclopedia of Machine Learning, pp. 563–564. Springer US, Boston, MA (2011). https://doi.org/https://doi.org/10.1007/978-0-387-30164-8_425
22. Junjuri, R., Calvarese, M., Vafaeinezhad, M., Vernuccio, F., Ventura, M., Meyer-Zedler, T., Gavazzoni, B., Polli, D., Vanna, R., Bongarzone, I., Ghislanzoni, S., Negro, M., Popp, J., Bocklitz, T.: Estimation of biological variance in coherent raman microscopy data of two cell lines using chemometrics. The Analyst **149**(17), 4395–4406 (2024). https://doi.org/10.1039/d4an00648h
23. Junjuri, R., Meyer-Zedler, T., Popp, J., Bocklitz, T.: Investigating the effect of non-resonant background variation on the CARS data analysis of bacteria samples and classification using machine learning. Opt. Contin. **3**(11), 2244 (2024). https://doi.org/https://doi.org/10.1364/OPTCON.528930
24. Junjuri, R., Saghi, A., Lensu, L., Vartiainen, E.M.: Convolutional neural network-based retrieval of Raman signals from CARS spectra. Optics Continuum **1**(6), 1324 (2022). https://doi.org/10.1364/OPTCON.457365

25. Junjuri, R., Saghi, A., Lensu, L., Vartiainen, E.M.: Evaluating different deep learning models for efficient extraction of raman signals from cars spectra. Physical Chemistry Chemical Physics **25**(24), 16340–16353 (2023). https://doi.org/10.1039/d3cp01618h
26. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint (2014). https://doi.org/10.48550/ARXIV.1412.6980
27. Krafft, C., Schie, I.W., Meyer, T., Schmitt, M., Popp, J.: Developments in spontaneous and coherent raman scattering microscopic imaging for biomedical applications. Chemical Society Reviews **45**(7), 1819–1849 (2016). https://doi.org/10.1039/c5cs00564g
28. Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems. SIAM (2016). https://doi.org/10.1137/1.9781611974508
29. Liu, B., Wang, Y., Wang, W.: Spectrogram enhancement algorithm: a soft thresholding-based approach. Ultrasound Med. Biol. **25**(5), 839–846 (1999). https://doi.org/https://doi.org/10.1016/S0301-5629(99)00024-1
30. Liu, Y., Lee, Y.J., Cicerone, M.T.: Broadband CARS spectral phase retrieval using a time-domain Kramers–Kronig transform. Optics Letters **34**(9), 1363 (2009). https://doi.org/10.1364/OL.34.001363
31. Luo, Z., Xu, X., Lin, D., Qu, J., Lin, F., Li, J.: Removing non-resonant background of cars signal with generative adversarial network. Applied Physics Letters **124**(26) (2024). https://doi.org/10.1063/5.0201616
32. Muddiman, R., O' Dwyer, K., Camp, C.H., Hennelly, B.: Removing non-resonant background from broadband cars using a physics-informed neural network. Analytical Methods **15**(32), 4032–4043 (2023). https://doi.org/10.1039/d3ay01131c
33. Müller, M., Zumbusch, A.: Coherent anti-stokes raman scattering microscopy. Wiley **8**(15), 2156–2170 (2007). https://doi.org/10.1002/cphc.200700202
34. Noack, B.R., Stankiewicz, W., Morzyński, M., Schmid, P.J.: Recursive dynamic mode decomposition of transient and post-transient wake flows. Journal of Fluid Mechanics **809**, 843–872 (2016). https://doi.org/10.1017/jfm.2016.678
35. Ogilvie, J.P., Beaurepaire, E., Alexandrou, A., Joffre, M.: Fourier-transform coherent anti-stokes raman scattering microscopy. Optics Letters **31**(4), 480 (2006). https://doi.org/10.1364/ol.31.000480
36. O'Shea, K., Nash, R.: An introduction to convolutional neural networks. arXiv preprint (2015). https://doi.org/10.48550/ARXIV.1511.08458
37. Polli, D., Kumar, V., Valensise, C.M., Marangoni, M., Cerullo, G.: Broadband Coherent Raman Scattering Microscopy. Laser & Photonics Reviews **12**(9), 1800020 (2018). https://doi.org/10.1002/lpor.201800020
38. Potma, E.O., Evans, C.L., Xie, X.S.: Heterodyne coherent anti-stokes raman scattering (cars) imaging. Optics Letters **31**(2), 241 (2006). https://doi.org/10.1364/ol.31.000241
39. Saghi, A., Junjuri, R., Lensu, L., Vartiainen, E.M.: Semi-synthetic data generation to fine-tune a convolutional neural network for retrieving raman signals from cars spectra. Optics Continuum **1**(11), 2360 (2022). https://doi.org/10.1364/optcon.469753
40. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. Journal of Statistical Physics **65**(3–4), 579–616 (1991). https://doi.org/10.1007/bf01053745
41. Schmid, P.J., Li, L., Juniper, M.P., Pust, O.: Applications of the dynamic mode decomposition. Theoretical and Computational Fluid Dynamics **25**(1–4), 249–259 (2010). https://doi.org/10.1007/s00162-010-0203-9
42. Shadaydeh, M., Denzler, J., Migliavacca, M.: Physics informed modeling of ecosystem respiration via dynamic mode decomposition with control input. In:

ICLR Workshop on Machine Learning for Remote Sensing (ICLR-WS) (2024). https://doi.org/https://doi.org/10.48550/arXiv.2402.17625
43. Takens, F.: Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence, Warwick 1980, p. 366–381. Springer Berlin Heidelberg (1981). https://doi.org/10.1007/bfb0091924
44. Valensise, C.M., Giuseppi, A., Vernuccio, F., De la Cadena, A., Cerullo, G., Polli, D.: Removing non-resonant background from cars spectra via deep learning. APL Photonics **5**(6), 061305 (2020). https://doi.org/10.1063/5.0007821
45. Vartiainen, E.M., Rinia, H.A., Müller, M., Bonn, M.: Direct extraction of Raman line-shapes from congested CARS spectra. Optics Express **14**(8), 3622 (2006). https://doi.org/10.1364/OE.14.003622
46. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017). https://doi.org/https://doi.org/10.48550/arXiv.1706.03762
47. Vernuccio, F., Broggio, E., Sorrentino, S., Bresci, A., Junjuri, R., Ventura, M., Vanna, R., Bocklitz, T., Bregonzio, M., Cerullo, G., Rigneault, H., Polli, D.: Non-resonant background removal in broadband cars microscopy using deep-learning algorithms. Scientific Reports **14**(1) (2024). https://doi.org/10.1038/s41598-024-74912-5
48. Volkmer, A., Book, L.D., Xie, X.S.: Time-resolved coherent anti-stokes raman scattering microscopy: Imaging based on raman free induction decay. Applied Physics Letters **80**(9), 1505–1507 (2002). https://doi.org/10.1063/1.1456262
49. Wang, Z., O' Dwyer, K., Muddiman, R., Ward, T., Camp, C.H., Hennelly, B.M.: VECTOR: Very deep convolutional autoencoders for non-resonant background removal in broadband coherent anti-Stokes Raman scattering. Journal of Raman Spectroscopy **53**(6), 1081–1093 (2022). https://doi.org/10.1002/jrs.6335
50. Wu, Z., Brunton, S.L., Revzen, S.: Challenges in dynamic mode decomposition. Journal of The Royal Society Interface **18**(185) (2021). https://doi.org/10.1098/rsif.2021.0686
51. Yang, C., Mehouachi, F.B., Menendez, M., Jabari, S.E.: Urban traffic analysis and forecasting through shared koopman eigenmodes. arXiv preprint (2024). https://doi.org/10.48550/ARXIV.2409.04728

A Calculation of the delay embedding dimension M using the autocorrelation function

The autocorrelation function (ACF) is a statistical measure that quantifies the correlation between a series and its lagged values. Basically, it explains the average relationship between a data point in a series and its historical values. When analyzing sequential data, as in our case, the ACF is used to estimate the delay embedding dimension.

The normalized autocorrelation function (ACF) for a spectrum $\mathbf{x}$ is computed as:

$$\mathrm{ACF}(\tau) = \frac{\sum_{j=1}^{N-\tau}(x_j - \bar{x})(x_{j+\tau} - \bar{x})}{\sum_{j=1}^{N}(x_j - \bar{x})^2}, \tag{9}$$

where τ is the delay lag and $\bar{x}$ is the mean of $\mathbf{x}$. This quantifies self-similarity in spectral features across lags. An example $\mathrm{ACF}(\tau)$ function and the selection of the embedding dimension M as the lag (τ) of the first dip is shown in Fig. 5

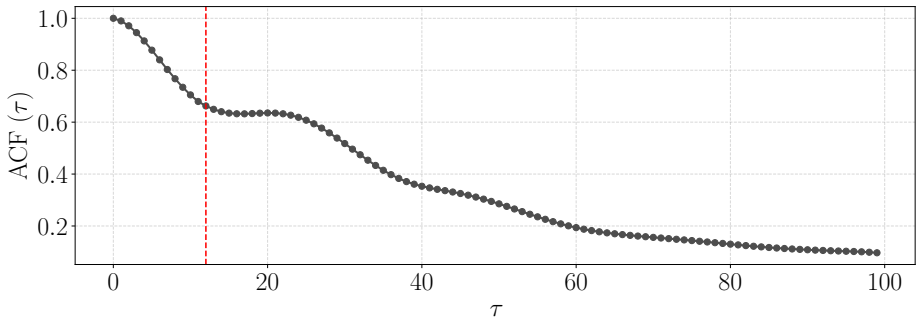

Fig. 5: Example usage of the autocorrelation function (ACF) to select delay embedding dimension. Here, we use an ethanol spectrum, and the ACF provides a clear dip around the delay lag $\tau = 12$ (red dotted vertical line).

B The DMD modes of the ethanol spectrum

Fig.6 visualizes seven modes out of twelve of the ethanol spectrum. The first real mode is assumed to present the NRB. Other complex modes are used to reconstruct the Raman spectrum. Certain smaller peaks with higher frequencies are missing in the second and third modes, and this is retrieved from other modes in the method discussed in the paper.

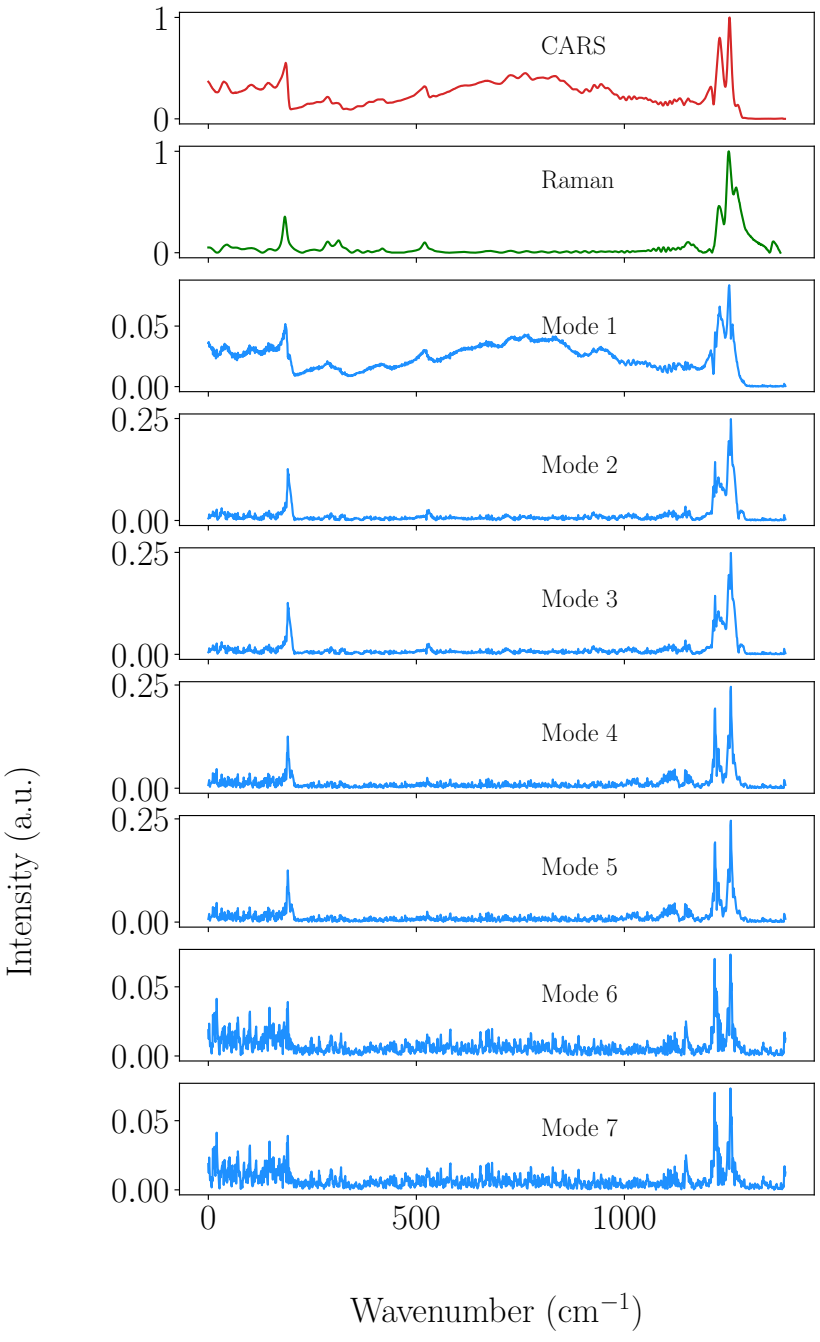

Fig. 6: The absolute values of the first seven DMD modes for ethanol spectra are plotted in blue. The corresponding CARS (red) and TDKK reconstructed Raman spectra (green) are also shown in the first two rows to highlight the similarity between the input CARS, output Raman/TDKK, and the DMD modes. The first mode is considered to be the NRB component, while the other complex conjugate modes are used to reconstruct the Raman spectrum.

C Deep learning model performance on an alternate set of synthetic spectra

We compare the performance of all deep learning models for a different set of synthetic spectra generated using different parameters as per [47]. Fig.7 shows the resulting MSE and correlation of 1000 spectra. In comparison to the results of Tab 3 in the paper, we noticed slightly different performance, which shows that the parameters used in the synthetic data generation influence the performance and also help verify the model performance consistency.

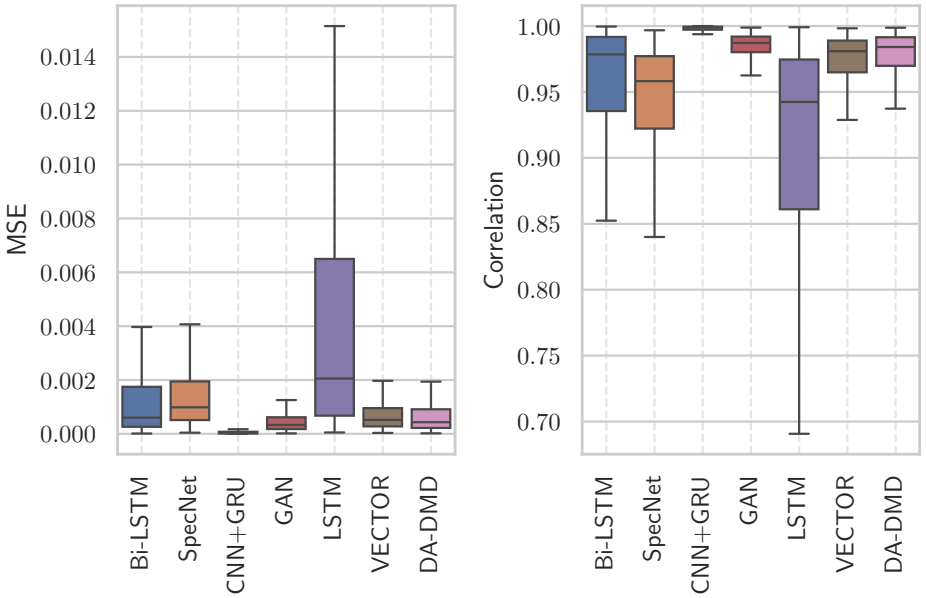

Fig. 7: Performance of seven deep learning models using correlation and MSE metrics on thousand synthetic datasets generated according to [47]. The plots show the mean and the skewness of these metrics.

We also tested metrics like the coefficient of determination (R^2) and Cosine Similarity (S_C), but they give very similar insight as correlation (r).