RamPINN: Recovering Raman Spectra From Coherent Anti-Stokes Spectra Using Embedded Physics

Sai Karthikeya Vemuri **Adithya Ashok Chalain Valapil** **Tim Büchner** **Joachim Denzler**
Computer Vision Group, Friedrich Schiller University Jena, Germany

Abstract

Transferring the recent advancements in deep learning into scientific disciplines is hindered by the lack of the required large-scale datasets for training. We argue that in these knowledge-rich domains, the established body of scientific theory provides reliable inductive biases in the form of governing physical laws. We address the ill-posed inverse problem of recovering Raman spectra from noisy Coherent Anti-Stokes Raman Scattering (CARS) measurements, as the true Raman signal here is suppressed by a dominating non-resonant background. We propose RamPINN, a model that learns to recover Raman spectra from given CARS spectra. Our core methodological contribution is a physics-informed neural network that utilizes a dual-decoder architecture to disentangle resonant and non-resonant signals. This is done by enforcing the Kramers-Kronig causality relations via a differentiable Hilbert transform loss on the resonant and a smoothness prior on the non-resonant part of the signal. Trained entirely on synthetic data, RamPINN demonstrates strong zero-shot generalization to real-world experimental data, explicitly closing this gap and significantly outperforming existing baselines. Furthermore, we show that training with these physics-based losses alone, without access to any ground-truth Raman spectra, still yields competitive results. This work highlights a broader concept: formal scientific rules can act as a potent inductive bias, enabling robust, self-supervised learning in data-limited scientific domains[1].

[1]Project Page: `https://rampinn.github.io`

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

1 INTRODUCTION

Deep learning models require large datasets of paired inputs and ground-truth outputs to outperform existing established algorithmic approaches. This paradigm is further exaggerated in scientific disciplines where acquiring ground-truth data is experimentally expensive. Vibrational spectroscopy exemplifies this challenge (Ember et al. 2017; Ichimura et al. 2014; Zoladek et al. 2011; Schultz et al. 2024). While Raman spectroscopy provides high-fidelity molecular fingerprints (Raman et al. 1928), its long acquisition times limit its usage in in-situ applications (Wang et al. 2023). A high-speed alternative, Coherent Anti-Stokes Raman Scattering (CARS), produces spectra distorted by a non-resonant background (NRB) (Li et al. 2020; Polli et al. 2018). The NRB interferes coherently with the Raman signal, distorting spectral shapes and obscuring features, making its removal a challenging ill-posed inverse problem (Wang et al. 2024).

Our approach disentangles the Raman signal from the NRB, enforcing physical constraints on each component. We constrain the recovered Raman signal to satisfy the Kramers-Kronig (KK) relations (Kramers 1928; de Laer Kronig 1926; Lucarini et al. 2005), while modeling the NRB as a smooth component (Junjuri et al. 2024; Konorov et al. 2011). These physics-based priors, embedded into the learning objective, provide sufficient supervision to recover the true Raman spectrum from a single CARS measurement.

Traditional solvers (Lucarini et al. 2005; Härkönen et al. 2023; Camp 2022) face a circular dependency: applying the Kramers-Kronig relations requires an estimate of the non-resonant background, which itself is an unknown. A neural network resolves this dependency by disentangling the measured spectrum into its Raman and NRB components. The network is guided by a composite loss function that enforces the respective physical priors on each component. This reliance on physical laws, rather than paired data, enables training exclusively on synthetically generated spectra. We evaluate the model's zero-shot generalization on a public

benchmark of six chemically diverse molecules. This experimental setup simulates the data-scarce conditions common in this domain and validates the practical utility of our physics-informed approach.

We draw inspiration from the Physics-Informed Neural Network (PINN) paradigm (Karniadakis et al. 2021), but depart from its standard formulation in an essential way. Classical PINNs are typically developed for forward or inverse problems governed by partial differential equations, where physical knowledge enters through differential operators in the loss (Takamoto et al. 2023; Moseley et al. 2021; Haghighat et al. 2021; Cai et al. 2021; Jin et al. 2022; Cho et al. 2023; Krishnapriyan et al. 2021; Maddu et al. 2022; Vemuri et al. 2025; Stein et al. 2024). In contrast, CARS-to-Raman recovery is not naturally described by a governing differential equation. Instead, it is an ill-posed spectral inverse problem whose structure is determined by domain-specific physical relationships: the Kramers-Kronig (KK) relations linking the real and imaginary parts of the resonant susceptibility, and the smooth, slowly varying nature of the non-resonant background (NRB). We therefore extend the physics-informed learning paradigm beyond PDE-based constraints and introduce **RamPINN**, a model built around these spectroscopy-specific priors. Its dual-decoder architecture explicitly disentangles the resonant Raman component from the NRB, enabling us to impose distinct physical constraints on each branch: KK-consistency for the Raman component and smoothness regularization for the NRB component.

Our contributions are threefold: (i) We propose a RamPINN: A model for solving the ill-posed inverse problem of recovering Raman spectra by embedding known physics (the Kramers-Kronig relations and smoothness prior) as a differentiable, self-supervisory loss term. (ii) We demonstrate the robustness of this framework by showing it achieves state-of-the-art zero-shot generalization. Trained solely on synthetic data, our method successfully transfers to a benchmark of six diverse, real-world molecules, outperforming methods that rely on purely data-driven learning. (iii) We demonstrate the efficacy of the physical priors as a supervisory signal, showing that a variant of our model trained without ground truth Raman data remains competitive with fully supervised data-driven baselines.

2 RELATED WORK

Physics-informed Learning constrains neural network training by including known physics in the loss function (Lagaris et al. 1998). This idea was formalized by (Raissi et al. 2019; Karniadakis et al. 2021; Cuomo et al. 2022), who introduced Physics-Informed Neural Networks (PINNs), where differential equations guide the learning process. PINNs are used to solve PDEs and ODEs (Eivazi et al. 2022; Lin et al. 2021; Vemuri et al. 2025; Vemuri et al. 2023; Cho et al. 2023). For inverse problems, PINNs learn system parameters while enforcing physical equations and are used in geophysics, structural mechanics, and other fields (Vemuri et al. 2024; Rasht-Behesht et al. 2022; Stein et al. 2024; Difonzo et al. 2024). Beyond differential equations, domain-specific rules are integrated into the loss function for medical image registration (Arratia López et al. 2023) or geometry encoding (Gropp et al. 2020; Vemuri et al. 2026).

Deep Learning for Spectroscopy is applied to denoising (Tkachenko et al. 2013), background correction (Ling et al. 1985), and Raman peak extraction (Luo et al. 2021). Several works focused on recovering Raman spectra from corresponding CARS spectra. SpecNet (Valensise et al. 2020) introduced a CNN trained on synthetic CARS-Raman pairs to perform NRB removal, while VECTOR (Wang et al. 2021) used a deep autoencoder for end-to-end signal reconstruction. Further work regularizes the process via fingerprint weights and the CH-stretching region (Muddiman et al. 2023). Generative approaches (GAN and CNN+GRU) were tested for NRB removal (Vernuccio et al. 2024; Luo et al. 2024). Recurrent models (LSTM (Houhou et al. 2020), Bi-LSTM (Junjuri et al. 2023)) were explored to model spectral dependencies. Recent approaches, such as DA-DMD, have established a decomposition-based architecture to isolate Raman peaks from NRB using known physics (Chalain Valapil et al. 2026).

Traditionally, works utilize physical models to embed approximations of the Kramers-Kronig (KK) relations (Liu et al. 2009), which have been improved via learned Hilbert Kernels (Camp 2022; Camp et al. 2020). In signal extraction, methods based on wavelet transform (Härkönen et al. 2023; Saghi et al. 2022) were also used. However, these approaches are either limited to precomputed kernel operators or require knowledge specific to the measurement system.

In contrast, we formulate a novel physics-informed learning framework for CARS-to-Raman mapping. We compose a loss that explicitly enforces KK-consistency via a differentiable Hilbert transform and regularizes the non-resonant background using a smoothness prior. Therefore, the model disentangles resonant Raman features from background artifacts not only based on data but also using given rules. This is a novel deep-learning approach for CARS spectroscopy that leverages KK physics and NRB priors within a learnable loss formulation for supervised and self-supervised training.

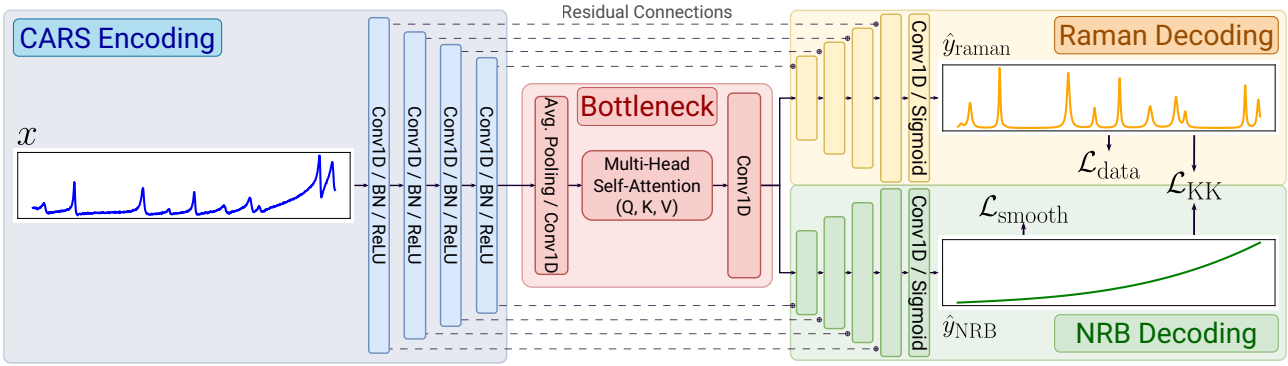


Figure 1: **RamPINN Architecture.** Our architecture builds upon recent advancements (Wang et al. 2021; Valensise et al. 2020; Vaswani et al. 2017; Ronneberger et al. 2015), with a key modification: a dual-branch decoder for reconstructing Raman and non-resonant background (NRB) signals. We incorporate physical constraints via losses $\mathcal{L}_{\text{KK}}$ and $\mathcal{L}_{\text{smooth}}$ on the predicted signals. Note that the encoder and decoder have identical dimensions (outlined in Section D), but are depicted here with different scales for clarity.

3 THEORY

We apply physics-informed learning (Raissi et al. 2019; Karniadakis et al. 2021) to map Coherent Anti-Stokes Raman Scattering (CARS) to Raman spectra (Raman et al. 1928). By embedding the problem's underlying physical laws, we improve model accuracy and robustness, particularly under data scarcity.

3.1 Physics-Informed Learning

Physics-informed learning is a paradigm that combines physical laws and neural networks, incorporating domain-specific knowledge into the learning process (Raissi et al. 2019; Karniadakis et al. 2021; Lagaris et al. 1997). This paradigm trains neural networks to satisfy the underlying physical equations governing a problem task while minimizing the data-driven loss function. These describe functional or differential relationships between input and output.

Let x be the input signal, $\hat{y}$ be the prediction, and f_θ be the neural network with trainable parameters θ. A physical model expresses a constraint $\mathcal{D}$:

$$\mathcal{D}(\hat{y}, x) = 0\,, \quad x \in \Omega\,, \tag{1}$$

where $\mathcal{D}$ is a known operator, such as a differential or integral transform within a domain Ω (Raissi et al. 2019). The physics-informed loss penalizes deviations from this constraint:

$$\mathcal{L}_{\text{phy}} = \frac{1}{|\Omega|} \int_\Omega \left|\mathcal{D}(f_\theta(x), u(x), x)\right|^2 dx\ . \tag{2}$$

In general this is combined with a data-driven loss term $\mathcal{L}_{\text{data}}$ (e.g., MSE) in a weighted objective:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{data}} \cdot \mathcal{L}_{\text{data}} + \lambda_{\text{phy}} \cdot \mathcal{L}_{\text{phy}}\ . \tag{3}$$

The physical constraint becomes part of the model's inductive bias, whose effect is controlled by the weighting parameter λ_{phy} (Vemuri et al. 2023; McClenny et al. 2023; Maddu et al. 2022; Wang et al. 2022).

3.2 CARS-to-Raman Spectra Mapping

This work aims to recover the Raman spectrum from a measured CARS signal. While spontaneous Raman spectroscopy directly captures molecular vibrations (Raman et al. 1928), it is limited by long acquisition times (Li et al. 2020; Wang et al. 2024). CARS is a faster alternative that uses nonlinear optical interactions to excite molecular vibrations more efficiently (Wang et al. 2024; Cheng et al. 2004). However, the measured CARS signal is not a clean Raman readout as it contains both resonant and non-resonant components that interfere nonlinearly. A more detailed discussion of this physical phenomenon is given in Section A.

The CARS intensity $I_{\text{CARS}}(\omega)$, as a function of wavenumber ω, is related to the third-order nonlinear susceptibility $\chi^{(3)}(\omega)$ by:

$$I_{\text{CARS}}(\omega) \propto \left|\chi^{(3)}(\omega)\right|^2 = \left|\chi_{\text{R}}(\omega) + \chi_{\text{NRB}}\right|^2\ , \tag{4}$$

where $\chi_{\text{R}}(\omega)$ is the complex-valued resonant component encoding molecular vibrational information, and χ_{NRB} is the real-valued non-resonant background (NRB), which is typically smooth and broad in real-world measurements (Junjuri et al. 2024; Konorov et al. 2011). While the NRB amplifies signal strength, it also distorts spectral shape, making Raman reconstruction from CARS a challenging problem.

To constrain this ill-posed problem, we leverage two well-established physical priors.

The Kramers-Kroning relation The resonant susceptibility χ_{R} satisfies the Kramers-Kronig (KK) relation (Lucarini et al. 2005; Kramers 1928; de Laer Kronig 1926), which links its real and imaginary parts through causality in linear response theory (Wang et al. 2021; Junjuri et al. 2023). This is further explained in Section B. Specifically,

$$\chi_{\mathrm{R}}(\omega) = \Re[\chi_{\mathrm{R}}(\omega)] \pm \Im[(\chi_{\mathrm{R}}(\omega))] \tag{5}$$

and

$$\Re[\chi_{\mathrm{R}}(\omega)] = \mathcal{H}(\Im[\chi_{\mathrm{R}}(\omega)]) \,, \tag{6}$$

where $\mathcal{H}(\cdot)$ is the Hilbert transform (Graf 2010). The imaginary part carries the true Raman spectrum (Müller et al. 2007). This relation provides a consistency constraint between what we want to recover ($\Im(\chi_{\mathrm{R}})$) and what we estimate from the observed signal (the real part $\Re(\chi_{\mathrm{R}})$).

To derive the specific loss function we use, we expand Equation (4) term (following Valensise et al. 2020),

$$|\chi_{\mathrm{NRB}} + \chi_{\mathrm{res}}|^2 = |\chi_{\mathrm{NRB}}|^2 + |\chi_{\mathrm{res}}|^2 + 2\,\Re\big(\chi_{\mathrm{NRB}}\chi_{\mathrm{res}}\big) \tag{7}$$

Here, we assume the resonant component is weak compared to the NRB $|\chi_{\mathrm{res}}| \ll |\chi_{\mathrm{NRB}}|$ (Valensise et al. 2020), such that we can neglect the purely resonant term $|\chi_{\mathrm{res}}|^2$ and obtain

$$I_{\mathrm{CARS}} \approx |\chi_{\mathrm{NRB}}|^2 + 2\,\Re\big(\chi_{\mathrm{NRB}}\chi_{\mathrm{res}}\big). \tag{8}$$

The non-resonant background is considered real-valued (instantaneous electronic response)

$$\chi_{\mathrm{NRB}} \in \mathbb{R}, \tag{9}$$

So the term in the RHS of 8 simplifies to

$$\Re\big(\chi_{\mathrm{NRB}}\chi_{\mathrm{res}}\big) = \chi_{\mathrm{NRB}}\,\Re[\chi_{\mathrm{res}}]. \tag{10}$$

Substituting this term into the 8 results in

$$I_{\mathrm{CARS}} \approx \chi_{\mathrm{NRB}}^2 + 2\,\chi_{\mathrm{NRB}}\,\Re[\chi_{\mathrm{res}}]. \tag{11}$$

As is standard in KK-based CARS processing (e.g., Camp 2022, Liu et al. 2009), one works with a normalized CARS signal obtained by dividing out the NRB envelope (the instrument is assumed to produce normalized spectra). We define:

$$x := \frac{I_{\mathrm{CARS}}}{\chi_{\mathrm{NRB}}}, \tag{12}$$

which yields

$$x \approx \chi_{\mathrm{NRB}} + 2\,\Re[\chi_{\mathrm{res}}]. \tag{13}$$

Raman spectra are interpreted on an overall scale (peak positions and shapes are important, not absolute magnitudes). The KK relation constrains the functional dependence of real and imaginary parts, not their absolute scale. For this reason, we use the KK-consistent imaginary component, :

$$\Im(\mathcal{H}(x - \hat{y}_{\mathrm{NRB}})) \tag{14}$$

as the physics-based target for the Raman branch.

NRB characteristics The background in our signal or NRB is smooth and lacks sharp peaks in real-world measurements (Junjuri et al. 2024; Konorov et al. 2011). This motivates a regularization term penalizing high curvature in the estimated background. These constraints directly integrate physical principles into the learning objective: one loss term promotes Kramers-Kronig (KK) consistency in the predicted Raman spectrum, while the other enforces smoothness in the NRB estimate (Junjuri et al. 2023; Vernuccio et al. 2024). Consequently, our approach is both data-driven and physics-grounded. This integration of physical priors fundamentally distinguishes RamPINN from previous, purely data-driven methods.

4 RAMPINN - METHODOLOGY

A CARS spectrum is represented as a function $x(\omega)$, where ω denotes the wavenumber, and the function value represents the measured relative intensity at that wavenumber. Similarly, the corresponding Raman spectrum and non-resonant background (NRB) are given by $y_{\mathrm{raman}}(\omega)$ and $y_{\mathrm{NRB}}(\omega)$, respectively.

The wavenumber (ω) is fixed across all samples during training and inference. We simplify notation by dropping the explicit dependence on ω, and refer to the spectra as vectors: x, y_{raman}, and y_{NRB}. All spectra are normalized, as we are concerned with relative spectral shapes (signatures) rather than absolute intensity values.

We aim to learn a mapping from the CARS spectrum to its underlying Raman and NRB components. A neural network f_θ, parameterized by θ, takes the normalized CARS spectrum x as input and reconstructs the Raman ($\hat{y}_{\mathrm{raman}}$) and NRB ($\hat{y}_{\mathrm{NRB}}$) spectra tuple:

$$f_\theta(x) = (\hat{y}_{\mathrm{raman}}, \hat{y}_{\mathrm{NRB}}) \ . \tag{15}$$

We aim to design a learning framework that enforces this decomposition in a way that respects the known physics of the CARS process. In the following sections, we describe how we achieve this using physics-informed loss functions based on the Kramers-Kronig relations and the smooth, peak-free nature of the NRB.

4.1 Model Architecture

Physics-informed learning is architecture-agnostic, yet recent advancements motivate our components (Valensise et al. 2020; Wang et al. 2021; Vaswani et al. 2017; Ronneberger et al. 2015). We use a 1D Convolutional U-Net (Ronneberger et al. 2015) tailored for spectral signal decomposition, shown in Figure 1. The model takes raw CARS spectra as input and outputs two components: the resonant Raman signal and the non-resonant background (NRB). The network employs a shared encoder and a dual-branch decoder for disentanglement, which are trained jointly under the physics-informed objective in Equation (18).

The encoder consists of four convolutional blocks with average pooling to preserve smooth signal features, progressively reducing temporal resolution while increasing feature depth. A self-attention block (Vaswani et al. 2017) captures long-range dependencies in the spectral domain, necessary for resolving overlapping peaks.

A dual-branch decoder reconstructs the Raman signal and NRB. Each branch comprises four upsampling blocks, a final convolution, and a sigmoid activation. To avoid checkerboard artifacts, we use upsampling followed by a one-dimensional convolutional layer rather than transposed convolutions (Höck et al. 2022; Büchner et al. 2023; Büchner et al. 2025; Kwarciak et al. 2024). Skip connections at each level concatenate encoder features with the corresponding decoder inputs, helping preserve fine-grained spectral structure throughout reconstruction.

Overall, this architecture is designed to decompose the input spectrum into physically meaningful components while maintaining flexibility for both supervised and physics-guided training. In Section D, we detail each component of the RamPINN model architecture. We also discuss the selection of our chosen backbone, comprising a U-Net (Ronneberger et al. 2015) with self-attention (Vaswani et al. 2017), over other existing backbone architectures.

4.2 Optimization

Let x be the measured CARS spectrum, and the goal is to predict $\hat{y} = (\hat{y}_{\text{raman}}, \hat{y}_{\text{NRB}})$, the Raman-resonant and non-resonant components, respectively, using a neural network $f_\theta(x)$. The Kramers-Kronig (KK) relations govern the underlying physical relationship (Lucarini et al. 2005; Kramers 1928; de Laer Kronig 1926; Guenther et al. 2004). A detailed description of Raman spectroscopy and Kramers-Kronig relationships is provided in Sections A and B, respectively.

Kramers-Kronig Regularization. Causality in optical response implies that the Raman component must be consistent with the imaginary part of the Hilbert transform of the residual signal (Liu et al. 2009; Lucarini et al. 2005). After subtracting the NRB, the residual signal should follow:

$$\mathcal{L}_{\text{KK}} = |\hat{y}_{\text{raman}} - \Im\left(\mathcal{H}(x - \hat{y}_{\text{NRB}})\right)|^2 \ . \tag{16}$$

Here, $\mathcal{H}(\cdot)$ denotes the differentiable Hilbert transform, which produces the analytic signal whose imaginary part corresponds to the KK-paired component. A complete formulation is provided in Section F.

NRB Regularization. The NRB is smooth and broad (Konorov et al. 2011; Wang et al. 2021). We enforce this behavior by regularizing the derivative of the predicted NRB signal, penalizing rapid changes (Puleio et al. 2023; Rosca et al. 2020):

$$\mathcal{L}_{\text{smooth}} = |\nabla \hat{y}_{\text{NRB}}|^2 \ . \tag{17}$$

Optimization Term. The loss is a weighted sum of a data fidelity term and the physics-based regularizers:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{data}}\mathcal{L}_{\text{data}} + \lambda_{\text{KK}}\mathcal{L}_{\text{KK}} + \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}} \ . \tag{18}$$

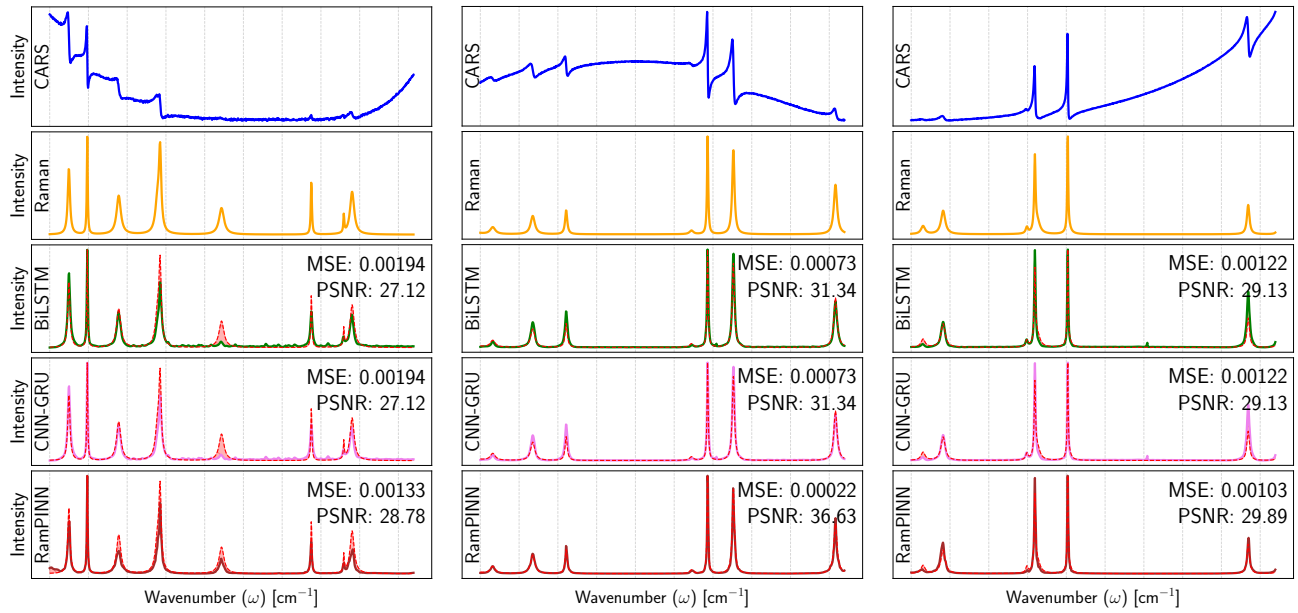
4.3 Implementation

We train RamPINN to learn the mapping from CARS spectra to their corresponding Raman and NRB components. The training algorithm, including the differentiable Hilbert transform for computing the Kramers-Kronig loss, is outlined in Section F.

Otherwise, we use standard deep learning training practices - Adam optimizer (Kingma et al. 2014) with a learning rate of 10^{-3} and early stopping based on validation loss. All models are implemented in PyTorch (Paszke 2019) with mini-batch training, gradient clipping, and loss normalization where appropriate. We fix the values of the regularization terms (obtained by an initial hyperparameter search) for RamPINN: $\lambda_{\text{smooth}} = 10$, $\lambda_{\text{data}} = 10$, and $\lambda_{\text{KK}} = 1$.

5 EXPERIMENTS AND RESULTS

We examine whether including physical knowledge during training enhances performance compared to data-driven baselines. Therefore, we compare RamPINN, supervised and self-supervised (by setting $\lambda_{\text{data}} = 0$), against state-of-the-art deep learning CARS-to-Raman approaches: SpecNet (Valensise et al. 2020), VECTOR (Wang et al. 2021), LSTM (Houhou et al. 2020), BiLSTM (Junjuri et al. 2023), GAN (Vernuccio et al.



(a) (b) (c)

Figure 2: **Qualitative Comparison of Raman Signal Extraction.** We visualize the Raman reconstruction of RamPINN, BiLSTM, and CNN-GRU on three synthetic samples (a-c). RamPINN outperforms the other methods, both qualitatively and quantitatively, as shown by the error values and lines (best viewed digitally). We provide additional plots in Section G.

2024), and CNN-GRU (Vernuccio et al. 2024). We follow their recommended training procedures. More details for these models are provided in Section E. All models have a comparable capacity, ensuring that any observed improvement can be attributed to the training methodology, specifically, the integration of physics-based constraints, rather than differences in model size. We also include traditional non-deep learning approaches for comparison: TDKK (Liu et al. 2009), LeDHT (Camp 2022), and IWT (Härkönen et al. 2023).

Due to the limited availability of real paired CARS and Raman spectra, we focus on synthetic training data. Our synthetic dataset mimics experimental conditions, described in Section C. We use 2000 synthetic samples, 1000 for training and 1000 for testing. All models are trained and evaluated (with an NVIDIA GTX 1080) on this common dataset and further tested on six public real-world examples for zero-shot signal extraction performance evaluation (Vernuccio et al. 2022a).

We evaluate performance using mean squared error (MSE) and Peak Signal-to-Noise ratio (PSNR (dB)) between predicted and ground truth spectra. To ensure robustness, we report results over 10 independent training runs. Quantitative results are given in Table 1. Beyond MSE/PSNR, we also report peak-aware metrics that assess whether the predicted spectra recover the correct peaks, how well their positions align, and how accurately their magnitudes are reproduced. Complete definitions and results (F1 for peak detection, normalized mean location error, and relative intensity error) are provided in Section J.

The supervised RamPINN outperforms all baselines by a large margin, achieving the lowest MSE and highest PSNR on the test set. This demonstrates the benefit of incorporating known physics into the learning process. Notably, the self-supervised RamPINN variant ($\lambda_{\text{data}} = 0$) remains competitive, outperforming most data-driven baselines that rely on ground truth data. This supports our hypothesis that leveraging physical knowledge improves performance, even with limited or no reconstruction regularization. This demonstrates that physical priors can effectively replace experimentally expensive paired data, enabling high-fidelity spectral recovery in data-limited settings.

The qualitative results for the top three performing models (RamPINN (supervised), BiLSTM (Junjuri et al. 2023), and CNN-GRU (Vernuccio et al. 2024)), as shown in Figure 2, further support these findings. Visually, RamPINN accurately reconstructs the Raman spectra from three synthetic CARS, especially recovering smaller peaks that other models miss. The Section G provides more examples. Although synthetic results are encouraging, zero-shot performance on real-world data is crucial for gauging practical applicability.

Table 1: **Quantitative Comparison of Raman Reconstruction Methods.** MSE and PSNR are computed between the predicted and ground truth Raman spectra on the test set. Lower MSE and higher PSNR indicate better performance. The top three models are ranked gold ● , silver ● , and bronze ● , respectively, based on MSE values (when two MSEs are the same, we use PSNR). All deep learning results are given over 10 independent runs. Traditional methods (denoted with †) estimate a single deterministic spectrum for comparison and require an NRB estimate.

Method	MSE ↓	PSNR (dB) ↑
TDKK†	0.0283	15.48
LeDHT†	0.0814	10.91
IWT†	0.0139	18.75
SpecNet	0.0064 ± 0.0003	21.91 ± 0.21
VECTOR	0.0205 ± 0.0002	16.88 ± 0.04
LSTM	0.0732 ± 0.1169	19.30 ± 8.77
BiLSTM	● 0.0007 ± 0.0002	● 31.57 ± 1.29
GAN	0.0088 ± 0.0118	22.74 ± 3.74
CNN-GRU	● 0.0019 ± 0.0004	● 27.38 ± 0.97
RamPINN	● 0.0006 ± 0.0001	● 33.83 ± 0.13
RamPINN (Self-sup)	0.0053 ± 0.0003	22.79 ± 0.28

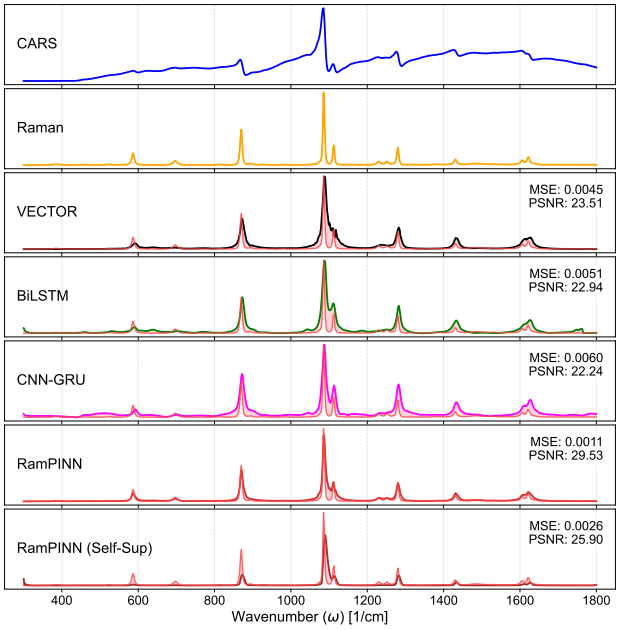


Figure 3: **Zero-shot Raman Spectra Extraction – Toluene.** Looking at input CARS, ground truth Raman, and reconstructions from baseline methods (VECTOR, BiLSTM, CNN-GRU), our RamPINN approach provides a more accurate zero-shot Raman spectra recovery, achieving the lowest MSE and highest PSNR.

5.1 Zero-shot Evaluation on Real-world Samples

To evaluate how well our models generalize beyond synthetic data, we conduct a zero-shot test: models trained entirely on synthetic spectra are applied directly to real CARS measurements, without any fine-tuning or adaptation. We use six publicly available samples, Acetone, DMSO, Ethanol, Isopropanol, Methanol, and Toluene with paired CARS and Raman spectra from (Vernuccio et al. 2022a; Vernuccio et al. 2022b).

We compare RamPINN predictions (supervised and self-supervised variants) with other baseline deep learning models against the ground-truth Raman spectra. For each model, we report the best-performing version. The goal is to assess how well each model handles real experimental signals without domain-specific calibration. Results are summarized in Table 2, with one representative prediction of Toluene shown in Figure 3. We can observe here that RamPINN reconstructs all peaks with accurate shape and intensity at correct locations. This is also seen for other samples for which qualitative visualizations are included in Section G.

These results indicate that sufficient synthetic data with self-supervision translates well to real-world applications. RamPINN consistently outperforms all purely data-driven baselines, demonstrating that embedding physical constraints leads to better generalization on real-world data. Notably, even the self-supervised RamPINN, trained without ground truth spectra, still generates competitive reconstructions, rivaling supervised baselines. This highlights the strength of physics-informed training, especially in settings where ground truth data are scarce or unavailable.

6 ABLATION STUDIES

We conduct ablations to assess the key parts of our method. Specifically, we: (1)Quantify the impact of our physics-informed KK loss and smoothness prior. (2)Analyze performance in relation to the amount of training data. (3)Test the model's robustness to violations of the non-resonant background (NRB) smoothness assumption. These studies probe the method's limitations and validate its suitability for practical application.

6.1 Effect of the Physics Loss Term

One of our contributions is the incorporation of the Kramers-Kronig (KK) loss ($\mathcal{L}_{\text{KK}}$), which isolates and quantifies its effect compared to purely data-driven approaches. Specifically, we vary the weight of the physics-based KK loss term, denoted by λ_{KK}, from zero to one in ten equidistant intervals.

As shown in Figure 4a, when $\lambda_{\text{KK}} = 0$, the model reduces to a data-driven approach without any physics

Table 2: **Real-world Zero-shot Evaluation.** Zero-shot evaluation results on six real-world CARS samples (Acetone, DMSO, Ethanol, Isopropanol, Methanol, Toluene) (Vernuccio et al. 2022a; Vernuccio et al. 2022b). RamPINN shows strong performance, consistently achieving the lowest Mean Squared Error (MSE) and highest Peak Signal-to-Noise Ratio (PSNR) across all samples, outperforming the baseline methods. The self-supervised RamPINN variant (Self-sup) also notably surpasses many baselines. The top three models are ranked gold ● , silver ● , and bronze ● , respectively, based on MSE values (when two MSEs are the same, we use PSNR).

	Acetone		DMSO		Ethanol		Isopropanol		Methanol		Toluene	
	MSE ↓	PSNR ↑	MSE ↓	PSNR ↑	MSE ↓	PSNR ↑	MSE ↓	PSNR ↑	MSE ↓	PSNR ↑	MSE ↓	PSNR ↑
TDKK	0.0283	15.48	0.0098	20.04	0.0140	18.30	0.0192	17.15	0.0064	22.06	0.0064	31.92
LeDHT	0.2663	5.74	0.0157	18.01	0.0351	14.53	0.0607	12.16	0.0315	15.01	0.1129	9.47
IWT	0.0078	21.02	0.0083	20.07	0.0149	18.26	0.0182	17.39	● 0.0031	24.97	0.0061	22.97
SpecNet	0.0108	19.65	0.0046	23.42	0.0036	24.43	0.0169	17.73	0.0041	23.88	0.0061	22.12
VECTOR	● 0.0041	23.88	0.0051	22.96	● 0.0033	24.84	0.0105	19.78	0.0039	24.12	0.0045	23.51
LSTM	0.0122	19.12	0.0101	19.95	0.0131	18.83	0.0165	17.82	0.0555	12.56	0.0075	21.25
BiLSTM	0.0043	23.67	0.0046	23.39	0.0041	23.92	● 0.0100	20.00	0.0037	24.30	0.0051	22.94
GAN	● 0.0035	24.53	● 0.0044	23.57	0.0046	23.38	0.0153	18.16	0.0037	24.27	● 0.0031	25.05
CNN-GRU	0.0105	19.80	● 0.0044	23.58	● 0.0033	24.85	0.0169	17.72	● 0.0037	24.32	0.0060	22.24
RamPINN	● **0.0011**	**29.69**	● **0.0013**	**28.98**	● **0.0008**	**30.87**	● **0.0026**	**25.80**	● **0.0010**	**30.14**	● **0.0011**	**29.53**
RamPINN (Self-sup)	0.0053	22.79	0.0076	21.20	0.0099	20.02	● 0.0035	24.56	0.0114	19.43	● 0.0026	25.90

constraint, and it shows weaker performance. As the value of λ_{KK} increases, the influence of physics during training increases, leading to better performance. Hence, we quantify how much the KK loss $\mathcal{L}_{\text{KK}}$ contributes to model performance.

6.2 Effect of Data

While we show that the Kramers-Kronig (KK) loss enhances standard data fidelity terms, it is crucial to test scaling the training data. We freeze the test set and vary the number of training samples from 0 to 1000, keeping all other settings fixed. This setup ensures systematic evaluation of the impact of supervision concerning generalization. Performance improves with more training data, as shown in Figure 4b. However, even in the fully self-supervised case, where RamPINN is trained without any data signal, the model recovers meaningful Raman spectra from CARS inputs. While not matching the accuracy of supervised models, these self-supervised predictions remain competitive, underscoring the strength of the physics-informed loss as an inductive bias. We observe diminishing returns with more data, suggesting that either the synthetic data lacks diversity or that the model has reached its representational capacity. The results show that the problem can be handled effectively with limited supervision.

6.3 Effect of Smoothness

We ablated the influence of λ_{smooth}. This prior is crucial for disentangling the signal, as it prevents the NRB decoder from fitting sharp Raman peaks to the synthetic data. While the KK-loss has a larger impact, the smoothness prior consistently improves performance as shown in Figure 4c.

6.4 Additional Robustness Studies

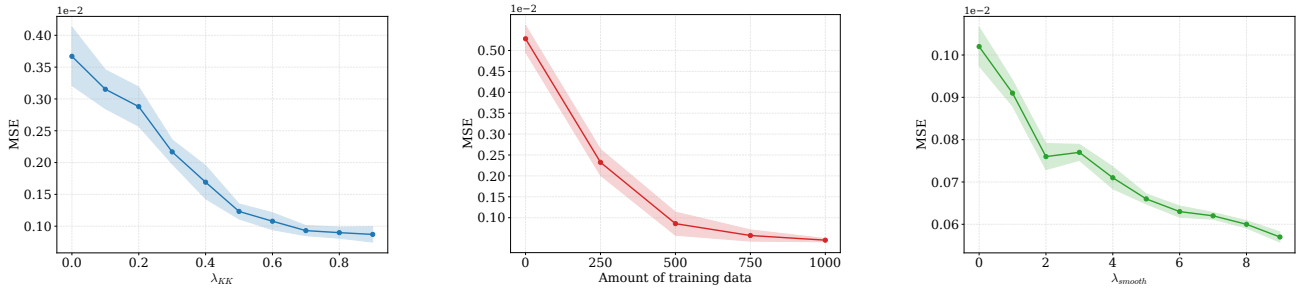
A key question is how robust RamPINN is to real-world experimental complexities, such as nonlinearities. To test these failure modes, we add artifacts. We conducted robustness studies, presented in full in Section I. We evaluated the model's performance against increasingly nonlinear NRB shapes and in the presence of sharp, peak-like artifacts designed to simulate instrumental noise. Our results consistently show that the model exhibits graceful degradation rather than catastrophic failure. RamPINN effectively learns to attribute non-physical artifacts to the NRB component, preserving the fidelity of the Raman reconstruction even under significant violations of assumptions.

7 LIMITATIONS & FUTURE WORK

Our work operates under the scientific constraint of data scarcity, using physical priors to enable zero-shot generalization from synthetic data. This framing naturally informs the primary avenues for future research.

Future work could explore adapting RamPINN for cases where it could learn from a large corpus of unpaired real-world CARS spectra while being fine-tuned on only partially available Raman spectra.

A natural extension of this work would be to design flexible background models, for example, using a learned function before capturing more complex NRB structures. Similarly, the current architecture assumes a fixed spectral resolution. Future work could address this by incorporating principles from Neural Operators (Kovachki et al. 2021), yielding a resolution-agnostic model that can adapt to data from different experimental setups without retraining.



(a) **Effect of** $\mathcal{L}_{\text{KK}}$ (b) **Effect of** $\mathcal{L}_{\text{data}}$ (c) **Effect of** $\mathcal{L}_{\text{smooth}}$

Figure 4: **Ablation Study Quantifying the Effects of Individual Loss Terms on Test MSE**. We observe that scaling our physics-based Kramers-Kronig regularization term, $\mathcal{L}_{\text{KK}}$, steadily decreases the test MSE with increasing influence. We also see similar effects with increasing the amount of training data ($\mathcal{L}_{\text{Data}}$) and the smoothness constraint imposed on NRB ($\mathcal{L}_{\text{smooth}}$).

8 CONCLUSION

In this work, we introduce RamPINN, a physics-informed neural network for reconstructing Raman spectra from CARS signals. By incorporating physical principles, specifically the Kramers-Kronig relations and NRB smoothness, as physics-based loss terms into the learning process, we demonstrate that RamPINN significantly enhances performance over purely data-driven baselines. While supervised training naturally yields the best results, the strong inductive bias induced by the physics terms provides competitive recovery even in the fully self-supervised setting.

Beyond spectroscopy, this work highlights a broader opportunity: many scientific domains possess deep, well-established knowledge that can be formally integrated into machine learning. Rather than replacing domain understanding, combining physical principles with modern learning methods offers a practical and interpretable path forward, especially in scientific settings where data is scarce or synthetic.

Sutton's "bitter lesson" suggests that, given enough data and computing power, learning outperforms inductive hand-crafted biases (Sutton 2019). However, in science-driven fields, data is often limited, and the inductive biases we embed are not heuristic approximations, but grounded in physical law. In these cases, physics-informed learning is not a compromise, but a necessary and principled foundation for generalization.

Acknowledgments

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

References

Arratia López, Pablo, Hernán Mella, Sergio Uribe, Daniel E. Hurtado, and Francisco Sahli Costabal (2023). "WarpPINN: Cine-MR image registration with physics-informed neural networks". In: *Medical Image Analysis* 89, p. 102925. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2023.102925.

Büchner, Tim, Christoph Anders, Orlando Guntinas-Lichius, and Joachim Denzler (2025). "Electromyography-Informed Facial Expression Reconstruction for Physiological-Based Synthesis and Analysis". In: *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 215–227. DOI: 10.1109/CVPR52734.2025.00029.

Büchner, Tim, Orlando Guntinas-Lichius, and Joachim Denzler (2023). "Improved obstructed facial feature reconstruction for emotion recognition with minimal change cyclegans". In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, pp. 262–274. DOI: 10.1007/978-3-031-45382-3_22.

Cai, Shengze, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis (2021). "Physics-Informed Neural Networks for Heat Transfer Problems". In: *Journal of Heat Transfer* 143.6, p. 060801. ISSN: 0022-1481. DOI: 10.1115/1.4050542.

Camp Jr, Charles H, John S Bender, and Young Jong Lee (2020). "Real-time and high-throughput Raman signal extraction and processing in CARS hyperspectral imaging". en. In: *Opt. Express* 28.14, pp. 20422–20437.

Camp, Charles H. (2022). "Raman signal extraction from CARS spectra using a learned-matrix representation of the discrete Hilbert transform". In: *Op-*

tics Express 30.15, p. 26057. ISSN: 1094-4087. DOI: 10.1364/oe.460543.

Chalain Valapil, Adithya Ashok, Carl Messerschmidt, Maha Shadaydeh, Michael Schmitt, Jürgen Popp, and Joachim Denzler (2026). "Deep Learning-Assisted Dynamic Mode Decomposition for Non-resonant Background Removal in CARS Spectroscopy". In: *Pattern Recognition (DAGM GCPR 2025)*. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, pp. 41–56. DOI: https://doi.org/10.1007/978-3-032-12840-9_4.

Cheng, J.-X. and X.S. Xie (2004). "Coherent Anti-Stokes Raman Scattering Microscopy". In: *The Journal of Physical Chemistry B* 108.3, pp. 827–840. DOI: 10.1021/jp035693v.

Cho, Junwoo, Seungtae Nam, Hyunmo Yang, Seok-Bae Yun, Youngjoon Hong, and Eunbyung Park (2023). "Separable Physics-Informed Neural Networks". In: *Advances in Neural Information Processing Systems*.

Cuomo, Salvatore, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli (2022). "Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next". In: *Journal of Scientific Computing* 92.3, p. 88. ISSN: 1573-7691. DOI: 10.1007/s10915-022-01939-z.

de Laer Kronig, Ralph (1926). "On the Theory of Dispersion of X-Rays". In: *Journal of The Optical Society of America* 12.6, pp. 547–557. DOI: 10.1364/JOSA.12.000547.

Difonzo, Fabio V., Luciano Lopez, and Sabrina F. Pellegrino (2024). "Physics informed neural networks for an inverse problem in peridynamic models". In: *Engineering with Computers*. ISSN: 1435-5663. DOI: 10.1007/s00366-024-01957-5.

Eivazi, Hamidreza, Mojtaba Tahani, Philipp Schlatter, and Ricardo Vinuesa (2022). "Physics-informed neural networks for solving Reynolds-averaged Navier-Stokes equations". In: *Physics of Fluids* 34 (7). ISSN: 10897666. DOI: 10.1063/5.0095270.

Ember, Katherine J I, Marieke A Hoeve, Sarah L McAughtrie, Mads S Bergholt, Benjamin J Dwyer, Molly M Stevens, Karen Faulds, Stuart J Forbes, and Colin J Campbell (2017). "Raman spectroscopy and regenerative medicine: a review". en. In: *NPJ Regen. Med.* 2.1.

Graf, Urs (2010). "Hilbert Transforms". In: *Introduction to Hyperfunctions and Their Integral Transforms: An Applied and Computational Approach*. Basel: Birkhäuser Basel, pp. 275–308. DOI: 10.1007/978-3-0346-0408-6_5.

Gropp, Amos, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman (2020). "Implicit geometric regularization for learning shapes". In: *Proceedings of the 37th International Conference on Machine Learning*. ICML'20. JMLR.org.

Guenther, Robert D., Duncan G. Steel, and Leopold Bayvel (2004). *Encyclopedia of Modern Optics*. Elsevier / Academic Press.

Haghighat, Ehsan, Maziar Raissi, Adrian Moure, Hector Gomez, and Ruben Juanes (2021). "A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics". In: *Computer Methods in Applied Mechanics and Engineering* 379, p. 113741. ISSN: 0045-7825. DOI: https://doi.org/10.1016/j.cma.2021.113741.

Härkönen, Teemu and Erik Vartiainen (2023). "Interpolated inverse discrete wavelet transforms in additive and non-additive spectral background corrections". In: *Optics Continuum* 2.5, p. 1068. ISSN: 2770-0208. DOI: 10.1364/optcon.488136.

Höck, Eva, Tim-Oliver Buchholz, Anselm Brachmann, Florian Jug, and Alexander Freytag (2022). "N2v2-fixing noise2void checkerboard artifacts with modified sampling strategies and a tweaked network architecture". In: *European Conference on Computer Vision*. Springer, pp. 503–518.

Houhou, Rola, Parijat Barman, Micheal Schmitt, Tobias Meyer, Juergen Popp, and Thomas Bocklitz (2020). "Deep learning as phase retrieval tool for CARS spectra". In: *Optics Express* 28.14, p. 21002. ISSN: 1094-4087. DOI: 10.1364/oe.390413.

Ichimura, Taro, Liang-da Chiu, Katsumasa Fujita, Satoshi Kawata, Tomonobu M Watanabe, Toshio Yanagida, and Hideaki Fujita (2014). "Visualizing cell state transition using Raman spectroscopy". In: *Plos one* 9.1, e84478.

Jin, Pengzhan, Shuai Meng, and Lu Lu (2022). "MIONet: Learning Multiple-Input Operators via Tensor Product". In: *SIAM Journal on Scientific Computing* 44.6, A3490–A3514. DOI: 10.1137/22M1477751.

Junjuri, Rajendhar, Tobias Meyer-Zedler, Jürgen Popp, and Thomas Bocklitz (2024). "Investigating the effect of non-resonant background variation on the CARS data analysis of bacteria samples and classification using machine learning". en. In: *Opt. Contin.* 3.11, p. 2244.

Junjuri, Rajendhar, Ali Saghi, Lasse Lensu, and Erik M. Vartiainen (2022). "Effect of non-resonant background on the extraction of Raman signals from CARS spectra using deep neural networks". In: *RSC Advances* 12.44, pp. 28755–28766. ISSN: 2046-2069. DOI: 10.1039/d2ra03983d.

– (2023). "Evaluating different deep learning models for efficient extraction of Raman signals from CARS spectra". In: *Physical Chemistry Chemical Physics* 25.24, pp. 16340–16353. ISSN: 1463-9084. DOI: 10.1039/d3cp01618h.

Karniadakis, George Em, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang (2021). "Physics-informed machine learning". In: *Nature Reviews Physics* 3.6, pp. 422–440. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00314-5.

Kingma, Diederik P and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Konorov, Stanislav O, Michael W Blades, and Robin F B Turner (2011). "Non-resonant background suppression by destructive interference in coherent anti-Stokes Raman scattering spectroscopy". en. In: *Opt. Express* 19.27, pp. 25925–25934.

Kovachki, Nikola B., Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar (2021). "Neural Operator: Learning Maps Between Function Spaces". In: *CoRR* abs/2108.08481.

Kramers, Hendrik Anthony (1928). *La diffusion de la lumière par les atomes*. Como.

Krishnapriyan, Aditi, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney (2021). "Characterizing possible failure modes in physics-informed neural networks". In: *Advances in neural information processing systems* 34, pp. 26548–26560.

Kwarciak, Kamil, Mateusz Daniol, Daria Hemmerling, and Marek Wodzinski (2024). "Unsupervised Skull Segmentation via Contrastive MR-to-CT Modality Translation". In: *Proceedings of the Asian Conference on Computer Vision*, pp. 160–174.

Lagaris, I.E., A. Likas, and D.I. Fotiadis (1998). "Artificial neural networks for solving ordinary and partial differential equations". In: *IEEE Transactions on Neural Networks* 9.5, pp. 987–1000. DOI: 10.1109/72.712178.

Lagaris, Likas, and Fotiadis (1997). "Artificial neural network methods in quantum mechanics". In: *Computer Physics Communications* 104.1, pp. 1–14. ISSN: 0010-4655. DOI: 10.1016/S0010-4655(97)00054-4.

Li, Shaowei, Yanping Li, Rongxing Yi, Liwei Liu, and Junle Qu (2020). "Coherent Anti-Stokes Raman Scattering Microscopy and Its Applications". In: *Frontiers in Physics* Volume 8 - 2020. ISSN: 2296-424X. DOI: 10.3389/fphy.2020.598420.

Lin, Chensen, Martin Maxey, Zhen Li, and George Em Karniadakis (2021). "A seamless multiscale operator neural network for inferring bubble dynamics". In: *Journal of Fluid Mechanics* 929, A18. DOI: 10.1017/jfm.2021.866.

Ling, Yong-Chien, Thomas J. Vickers, and Charles K. Mann (1985). "Background Correction in Raman Spectroscopic Determination of Dimethylsulfone, Sulfate, and Bisulfate". In: *Appl. Spectrosc.* 39.3, pp. 463–470.

Liu, Yuexin, Young Jong Lee, and Marcus T. Cicerone (2009). "Broadband CARS spectral phase retrieval using a time-domain Kramers–Kronig transform". In: *Opt. Lett.* 34.9, pp. 1363–1365. DOI: 10.1364/OL.34.001363.

Lucarini, Valerio, Jarkko J Saarinen, Kai-Erik Peiponen, and Erik M Vartiainen (2005). *Kramers-kronig relations in optical materials research*. en. 2005th ed. Springer Series in Optical Sciences. Springer.

Luo, Si-heng, Xin Wang, Gan-yu Chen, Yi Xie, Wenhan Zhang, Zhi-fan Zhou, Zhi-min Zhang, Bin Ren, Guo-kun Liu, and Zhong-qun Tian (2021). "Developing a Peak Extraction and Retention (PEER) Algorithm for Improving the Temporal Resolution of Raman Spectroscopy". In: *Analytical Chemistry* 93.24, pp. 8408–8413. DOI: 10.1021/acs.analchem.0c05391.

Luo, Ziyi, Xiangcong Xu, Danying Lin, Junle Qu, Fangrui Lin, and Jia Li (2024). "Removing non-resonant background of CARS signal with generative adversarial network". en. In: *Appl. Phys. Lett.* 124.26.

Maddu, Suryanarayana, Dominik Sturm, Christian L Müller, and Ivo F Sbalzarini (2022). "Inverse Dirichlet weighting enables reliable training of physics informed neural networks". In: *Machine Learning: Science and Technology* 3.1, p. 015026. DOI: 10.1088/2632-2153/ac3712.

McClenny, Levi D. and Ulisses M. Braga-Neto (2023). "Self-adaptive physics-informed neural networks". In: *Journal of Computational Physics* 474, p. 111722. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2022.111722.

Moseley, Ben, Andrew Markham, and Tarje Nissen-Meyer (2021). *Finite Basis Physics-Informed Neural Networks (FBPINNs): a scalable domain decomposition approach for solving differential equations*.

Muddiman, Ryan, Kevin O' Dwyer, Charles. H. Camp, and Bryan Hennelly (2023). "Removing non-resonant background from broadband CARS using a physics-informed neural network". In: *Analytical Methods* 15.32, pp. 4032–4043. ISSN: 1759-9679. DOI: 10.1039/d3ay01131c.

Müller, Michiel and Andreas Zumbusch (2007). "Coherent anti-Stokes Raman Scattering Microscopy". In: *ChemPhysChem* 8.15, pp. 2156–2170. ISSN: 1439-4235, 1439-7641. DOI: 10.1002/cphc.200700202.

Paszke, A (2019). "Pytorch: An imperative style, high-performance deep learning library". In: *arXiv preprint arXiv:1912.01703*.

Polli, Dario, Vikas Kumar, Carlo M. Valensise, Marco Marangoni, and Giulio Cerullo (July 2018). "Broadband Coherent Raman Scattering Microscopy". In: *Laser; Photonics Reviews* 12.9. ISSN: 1863-8899. DOI: 10.1002/lpor.201800020.

Puleio, Alessandro, Riccardo Rossi, and Pasqualino Gaudio (2023). "Calibration of spectra in presence of non-stationary background using unsupervised physics-informed deep learning". In: *Scientific Reports* 13.1. ISSN: 2045-2322. DOI: 10.1038/s41598-023-29371-9.

Raissi, M., P. Perdikaris, and G. E. Karniadakis (2019). "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *Journal of Computational Physics* 378, pp. 686–707. ISSN: 10902716. DOI: 10.1016/j.jcp.2018.10.045.

Raman, Chandrasekhara Venkata and Kariamanikkam Srinivasa Krishnan (1928). "A New Type of Secondary Radiation". In: *Nature* 121.3048, pp. 501–502. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/121501c0.

Rasht-Behesht, Majid, Christian Huber, Khemraj Shukla, and George Em Karniadakis (2022). "Physics-Informed Neural Networks (PINNs) for Wave Propagation and Full Waveform Inversions". In: *Journal of Geophysical Research: Solid Earth* 127.5, e2021JB023120. DOI: 10.1029/2021JB023120.

Rinia, Hilde A., Mischa Bonn, and Michiel Müller (2006). "Quantitative Multiplex CARS Spectroscopy in Congested Spectral Regions". In: *The Journal of Physical Chemistry B* 110.9, pp. 4472–4479. ISSN: 1520-5207. DOI: 10.1021/jp0564849.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, pp. 234–241.

Rosca, Mihaela, Theophane Weber, Arthur Gretton, and Shakir Mohamed (2020). "A case for new neural network smoothness constraints". In: *Proceedings on "I Can't Believe It's Not Better!" at NeurIPS Workshops*. Ed. by Jessica Zosa Forde, Francisco Ruiz, Melanie F. Pradier, and Aaron Schein. Vol. 137. Proceedings of Machine Learning Research. PMLR, pp. 21–32.

Saghi, Ali, Rajendhar Junjuri, Lasse Lensu, and Erik M. Vartiainen (2022). "Semi-synthetic data generation to fine-tune a convolutional neural network for retrieving Raman signals from CARS spectra". In: *Optics Continuum* 1.11, p. 2360. ISSN: 2770-0208. DOI: 10.1364/optcon.469753.

Schultz, Constanze, Tristan Wegner, Corinna Heusel, Tim Gallagher, Yanjun Zheng, Markus Werner, Seraphine V. Wegner, Tobias Meyer-Zedler, Oliver Werz, Michael Schmitt, Juergen Popp, and Frank Glorius (2024). "Alkyne-Tagged Imidazolium-Based Membrane Cholesterol Analogs for Raman Imaging Applications". In: *Chemical Science* 15.35, pp. 14323–14335. DOI: 10.1039/D4SC03155E.

Smith, D. Y. (1977). "Dispersion relations for complex reflectivities". In: *Journal of the Optical Society of America* 67.4, p. 570. ISSN: 0030-3941. DOI: 10.1364/josa.67.000570.

Stein, Gideon, Sai Karthikeya Vemuri, Yuanyuan Huang, Anne Ebeling, Nico Eisenhauer, Maha Shadaydeh, and Joachim Denzler (2024). "Investigating the Effects of Plant Diversity on Soil Thermal Diffusivity Using Physics- Informed Neural Networks". In: *ICLR Workshop on AI4DifferentialEquations In Science (ICLR-WS)*.

Sutton, Richard (2019). "The bitter lesson". In: *Incomplete Ideas (blog)* 13.1, p. 38.

Takamoto, Makoto, Timothy Praditia, Raphael Leiteritz, Dan MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert (2023). *PDEBENCH: An Extensive Benchmark for Scientific Machine Learning*.

Tkachenko, A., T. Van Reeth, V. Tsymbal, C. Aerts, O. Kochukhov, and J. Debosscher (2013). "Denoising spectroscopic data by means of the improved least-squares deconvolution method". In: *A&A* 560, A37. DOI: 10.1051/0004-6361/201322532.

Valensise, C. M., A. Giuseppi, F. Vernuccio, A. De la Cadena, G. Cerullo, and D. Polli (2020). "Removing non-resonant background from CARS spectra via deep learning". In: *APL Photonics* 5.6, p. 061305. ISSN: 2378-0967. DOI: 10.1063/5.0007821.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2017). *Attention Is All You Need*. DOI: 10.48550/arXiv.1706.03762.

Vemuri, Sai Karthikeya, Tim Büchner, and Joachim Denzler (2024). "Estimating Soil Hydraulic Parameters for Unsaturated Flow Using Physics-Informed Neural Networks". In: *Computational Science – ICCS 2024*. Ed. by Leonardo Franco, Clélia de Mulatier, Maciej Paszynski, Valeria V. Krzhizhanovskaya, Jack J. Dongarra, and Peter M. A. Sloot, pp. 338–351. DOI: 10.1007/978-3-031-63759-9_37.

Vemuri, Sai Karthikeya, Tim Büchner, Julia Niebling, and Joachim Denzler (2025). "Functional Tensor Decompositions for Physics-Informed Neural Networks". In: *Pattern Recognition*. Ed. by Apostolos Antonacopoulos, Subhasis Chaudhuri, Rama Chellappa, Cheng-Lin Liu, Saumik Bhattacharya, and Umapada Pal. Cham: Springer Nature Switzerland, pp. 32–46.

Vemuri, Sai Karthikeya and Joachim Denzler (2023). "Gradient Statistics-Based Multi-Objective Optimization in Physics-Informed Neural Networks". In: *Sensors* 23.21. ISSN: 1424-8220. DOI: 10.3390/s23218665.

Vemuri, Sai Karthikeya, Tim Büchner, and Joachim Denzler (2026). "F-INR: Functional Tensor Decomposition for Implicit Neural Representations". In: *Winter Conference on Applications of Computer Vision (WACV)*. (accepted). DOI: 10.48550/arXiv.2503.21507.

Vernuccio, Federico, Arianna Bresci, Benedetta Talone, Alejandro de la Cadena, Chiara Ceconello, Stefano Mantero, Cristina Sobacchi, Renzo Vanna, Giulio Cerullo, and Dario Polli (2022a). "Fingerprint multiplex CARS at high speed based on supercontinuum generation in bulk media and deep learning spectral denoising". In: *Optics Express* 30.17, p. 30135. ISSN: 1094-4087. DOI: 10.1364/oe.463032.

Vernuccio, Federico, Arianna Bresci, Benedetta Talone, Alejandro De la Cadena Pérez-Gallardo, Chiara Ceconello, Stefano Mantero, Cristina Sobacchi, Renzo Vanna, Giulio Cerullo, and Dario Polli (2022b). *Supplementary Material - Fingerprint Multiplex CARS at High Speed Based on Supercontinuum Generation in Bulk Media and Deep Learning Spectral Denoising*. DOI: 10.5281/ZENODO.6772179. (Visited on 05/02/2025).

Vernuccio, Federico, Elia Broggio, Salvatore Sorrentino, Arianna Bresci, Rajendhar Junjuri, Marco Ventura, Renzo Vanna, Thomas Bocklitz, Matteo Bregonzio, Giulio Cerullo, Hervé Rigneault, and Dario Polli (2024). "Non-resonant background removal in broadband CARS microscopy using deep-learning algorithms". In: *Scientific Reports* 14.1. ISSN: 2045-2322. DOI: 10.1038/s41598-024-74912-5.

Wang, Kaiqiang, Li Song, Chutian Wang, Zhenbo Ren, Guangyuan Zhao, Jiazhen Dou, Jianglei Di, George Barbastathis, Renjie Zhou, Jianlin Zhao, and Edmund Y. Lam (Jan. 2024). "On the use of deep learning for phase recovery". In: *Light: Science; Applications* 13.1. ISSN: 2047-7538. DOI: 10.1038/s41377-023-01340-x. URL: http://dx.doi.org/10.1038/s41377-023-01340-x.

Wang, Sifan, Xinling Yu, and Paris Perdikaris (2022). "When and why PINNs fail to train: A neural tangent kernel perspective". In: *Journal of Computational Physics* 449, p. 110768. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2021.110768.

Wang, Yumei, Liuru Fang, Yuhua Wang, and Zuzhao Xiong (Dec. 2023). "Current Trends of Raman Spectroscopy in Clinic Settings: Opportunities and Challenges". en. In: *Adv Sci (Weinh)* 11.7, e2300668.

Wang, Zhen, Xue Li, Yufeng Liu, Wenjun Zhang, and Xiaojun Xu (2021). "VECTOR: Very deep convolutional autoencoder for CARS spectrum denoising and background removal". In: *Optics Express* 29.4, pp. 6196–6210. DOI: 10.1002/jrs.6335.

Zoladek, Alina, Flavius C Pascut, Poulam Patel, and Ioan Notingher (2011). "Non-invasive time-course imaging of apoptotic cells by confocal Raman microspectroscopy". In: *Journal of Raman Spectroscopy* 42.3, pp. 251–258.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Material

A Raman Spectroscopy and CARS

Raman spectroscopy is a non-destructive chemical analysis technique widely used for molecular characterization, including in medical and biological applications. When light interacts with matter, most photons are elastically scattered (Rayleigh scattering), but a small fraction undergoes inelastic scattering, exchanging energy with molecular vibrational modes. This inelastic process produces a shift in photon energy that is characteristic of the molecule's internal structure (Raman et al. 1928).

The resulting Raman spectrum can be interpreted as a high-dimensional feature vector (intensity vs. frequency shift), which is fingerprint to a molecule. Each peak corresponds to a specific vibrational or rotational transition, enabling identification of chemical composition, molecular structure, orientation, and concentration. From a data perspective, Raman measurements are typically represented as one-dimensional spectra, where each sample consists of intensity values sampled over a discrete set of wavenumbers. The main drawback is that the acquisition time in Spontaneous Raman (SR) measurement is very high. This occurs because, naturally, these interactions are very low, and photon accumulation significant enough for quantification takes time.

In spontaneous Raman (SR) spectroscopy, a monochromatic laser illuminates the sample, and the scattered light is collected and spectrally resolved using a dispersive element (e.g., a grating spectrometer) and a detector such as a CCD. Because the Raman scattering cross-section is very small, only a tiny fraction of incident photons contribute to the signal. As a result, long acquisition times or signal averaging are required to achieve an acceptable signal-to-noise ratio (SNR). This leads to practical limitations such as low throughput, motion sensitivity, and challenges in real-time or in vivo measurements (Wang et al. 2023).

To overcome this, a faster approach called coherent anti-Stokes Raman scattering (CARS) was introduced (Polli et al. 2018; Cheng et al. 2004). In CARS, at least two synchronized laser beams are used: a pump beam at frequency ω_p and a Stokes beam at frequency ω_s. When the frequency difference $\omega_p - \omega_s$ matches a molecular vibrational resonance, a coherent vibrational excitation is generated in the sample. A third interaction with the pump field then produces an anti-Stokes signal at frequency $2\omega_p - \omega_s$. This uses a coherent process to stimulate the Raman process at a higher rate, such that more photons are released.

Unlike SR, where scattering events are independent and incoherent, CARS is a coherent process: the emitted photons are phase-matched and constructively interfere, leading to a signal that scales quadratically with the number of oscillators. This results in significantly higher signal intensity and enables much faster acquisition, often allowing imaging at video rates. From a measurement standpoint, CARS systems typically use scanning microscopy setups, where spectra are acquired either point-by-point (spectral scanning) or via multiplex detection schemes.

For machine learning applications, this distinction has important implications. SR data is generally linear in the molecular concentration and directly interpretable, whereas CARS signals are proportional to the square modulus of the third-order nonlinear susceptibility, $|\chi^{(3)}|^2$ as seen in Equation 19 of Appendix B. This introduces nonlinear mixing of resonant and non-resonant contributions, meaning that the measured spectrum is not a simple linear superposition of individual molecular signatures. Consequently, the inverse problem recovering chemically meaningful spectra from CARS measurements is more complex and often requires preprocessing, modeling, or learning-based approaches.

B Kramers-Kroning Principle

The Kramers-Kronig relations, or KK relations, are mathematical relations that connect real and imaginary parts of any complex function that is analytic in the upper half-plane.

In CARS, the NRB originates from electronic contributions to the third-order nonlinear susceptibility ($\chi^{(3)}$) (Lucarini et al. 2005; Kramers 1928; de Laer Kronig 1926; Guenther et al. 2004):

$$I_{\text{CARS}}(\omega) \propto \left|\chi^{(3)}(\omega)\right|^2 I_{pu}^2 I_s \,, \tag{19}$$

where I_{pu} and I_s are the pump and stokes laser intensity, respectively. The nonlinear susceptibility includes the Raman resonant part χ_r and the non-resonant part χ_{nr}, corresponding to electronic contributions and NRB, respectively. The non-resonant part is frequency independent. Excluding the third order in susceptibility notation, this is expressed as:

$$|\chi(\omega)|^2 = |\chi_r(\omega) + \chi_{nr}|^2 = |\chi_r(\omega)|^2 + 2\chi_{nr}\Re[\chi_r(\omega)] + |\chi_{nr}|^2 \,. \tag{20}$$

The non-resonant part arises due to the measurement setup, and the resonant part can be mathematically expressed as:

$$\chi_r(\omega) = \sum_r \frac{A_r}{\Omega_r - (\omega_{pu} - \omega_s) - i\gamma_r} \tag{21}$$

where $(\omega_{pu} - \omega_s)$ is the difference in pump and stokes frequency and A_r, Ω_r and γ_r are the amplitude, vibrational frequency and bandwidth of rth Raman mode respectively. We can relate the real and imaginary parts in Equation 21 as:

$$\chi_r(\omega) = |\chi_r(\omega)|\, e^{i\phi(\omega)} \,. \tag{22}$$

As $I_{\text{Raman}} \approx \Im[\chi_r(\omega)]$, we need to estimate the phase term (Rinia et al. 2006). From Equation 19 and Equation 20, the CARS measurement only has information on the squared modulus of susceptibility $I_{\text{CARS}} \approx |\chi(\omega)|^2$. To approximate NRB, a reference spectrum with Raman inactive substance with susceptibility $\chi_{nr}(\omega)$ is measured and applied such that $\chi_r(\omega) = \chi/\chi_{nr}$. Now let's represent the normalized intensity as $S(\omega)$. The phase information can be obtained from this using the Kramers-Kronig relations. It connects the real and imaginary parts as:

$$\begin{aligned} \Re[\chi_r(\omega)] &= \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{\Im[\chi_r(\omega')]}{\omega' - \omega} d\omega' \text{ and} \\ \Im[\chi_r(\omega)] &= \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{\Re[\chi_r(\omega')]}{\omega' - \omega} d\omega' \,. \end{aligned} \tag{23}$$

Here, the $\mathcal{P}$ represents the Cauchy principal value used to handle improper integrals with singularities by avoiding the singularity through a limiting process. Using the KK relation(Smith 1977), we can deduce the phase as:

$$\varphi(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{ln(\sqrt{S(\omega')})}{\omega' - \omega} d\omega' \,. \tag{24}$$

The Hilbert transform is a specific type of KK relation that applies to real-valued signals. For a real-valued function $x(t)$:

$$\mathcal{H}\{x(t)\} = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{x(t')}{t' - t} dt' \,. \tag{25}$$

Therefore, the discrete Hilbert transform on $ln(\sqrt{S(\omega)})$ will give the phase.

C Synthetic Data Generation

To train RamPINN, we generate synthetic CARS spectra by simulating both the resonant Raman signal and the non-resonant background (NRB). This synthetic dataset is constructed to closely mimic realistic spectral profiles observed in practical CARS experiments. Below, we describe the components and procedure used.

Resonant Raman Signal. Raman-active modes are modeled as Lorentzian-shaped peaks, following the standard formulation for the third-order nonlinear susceptibility as in Equation 21. In implementation, the Raman susceptibility $\chi^{(3)}$ is constructed as:

$$\chi_r^{(3)}(\omega) = \sum_{r=1}^{N} \frac{A_r}{\Omega_r - \omega - i\gamma_r}, \tag{26}$$

where:

- $N \sim \mathcal{U}\{1, 25\}$ is number of peaks varied per sample,
- $A_r \in \mathcal{U}(0.01, 1.0)$ is the amplitude (randomized),
- $\Omega_r \in \mathcal{U}(0, 1)$ is the normalized resonance frequency,
- $\gamma_r \in \mathcal{U}(0.001, 0.02)$ is the linewidth,
- ω is the normalized Raman shift ranging 1000 points over [0, 1].

The real and imaginary parts of $\chi^{(3)}$ are retained to compute the CARS intensity and use in the Kramers-Kronig loss. The signal is normalized such that the maximum absolute value of $\chi^{(3)}$ is 1.

Multiple such Lorentzian peaks are summed to construct a complete Raman spectrum. The real and imaginary parts of $\chi^{(3)}$ are retained to compute the CARS intensity and use in the Kramers-Kronig loss.

Non-Resonant Background (NRB). To simulate the NRB, we use two classes of functions commonly observed in experimental setups: Sigmoid-based and polynomial-shaped backgrounds. Each spectrum is randomly assigned one of these forms.

Sigmoid-Based NRB. The sigmoid NRB is generated using the following function, where we use a product of two sigmoids to better capture smooth peak-shaped NRB profiles:

$$\chi_{nrb}^{(3)}(\omega) = \frac{1}{1 + e^{-b_1(\omega - c_1)}} \cdot \frac{1}{1 + e^{b_2(\omega - c_2)}}, \tag{27}$$

where:

- $b_1, b_2 \sim \mathcal{N}(10, 5)$ (steepness),
- $c_1, c_2 \sim \mathcal{N}(0.2, 0.3)$ and $\mathcal{N}(0.7, 0.3)$ respectively (centers).

This form captures the smooth, sigmoidal nature of NRB often observed in real data, as suggested in prior studies (Junjuri et al. 2022).

Polynomial-Based NRB. Alternatively, NRB is modeled as a smooth polynomial function:

$$\chi_{nrb}^{(3)}(\omega) = a\omega^4 + b\omega^3 + c\omega^2 + d\omega + e\,, \tag{28}$$

with randomly sampled coefficients:

$$a, c \sim \mathcal{U}(-1, 1), \quad b, d, e \sim \mathcal{U}(-10, 10)\,. \tag{29}$$

These coefficients produce a range of smooth background shapes, consistent with observations from experimental NRB profiles.

CARS Spectrum Generation. The final CARS spectrum is computed as:

$$I_{\text{CARS}}(\omega) \propto \left|\chi_r^{(3)}(\omega) + \chi_{\text{nrb}}\right|^2, \tag{30}$$

where χ_{nrb} is treated as real-valued, either polynomial or sigmoid in form, depending on the random draw. The total spectrum is normalized to ensure that learning focuses on the spectral shape rather than absolute intensity. A random noise $\sim \mathcal{U}(0.0005, 0.003)$ is also added to simulate high frequency disturbances.

D RamPINN Model Architecture

The RamPINN model utilizes a 1D U-Net-like (Ronneberger et al. 2015) architecture adapted for spectral data processing. It takes a single-channel 1D CARS spectrum as input and outputs two separate 1D spectra corresponding to the estimated resonant Raman component and the Non-Resonant Background (NRB) component. The architecture consists of an encoder path, a bottleneck, and two parallel decoder paths. We implement the model in PyTorch (Paszke 2019).

Encoder. The encoder comprises four stages. Each stage consists of a `ConvBlock1D` module followed by an `AvgPool1d` operation with a kernel size of 2 (effectively downsampling by a factor of 2). The `ConvBlock1D` module contains a 1D Convolution (`nn.Conv1d`) with a kernel size of 5 and padding of 2 (maintaining sequence length within the block), followed by `BatchNorm1d` and a `ReLU` activation. The number of channels increases through the encoder: 1(input) $\rightarrow$ 64 $\rightarrow$ 128 $\rightarrow$ 256 $\rightarrow$ 512.

Bottleneck. The bottleneck connects the encoder and decoder paths at the lowest spatial resolution (highest feature abstraction). It consists of a `ConvBlock1D`, followed by a `SelfAttention1D` module, and another `ConvBlock1D`. The self-attention mechanism (SelfAttention1D) employs standard scaled dot-product attention, allowing the model to capture long-range dependencies within the compressed spectral representation (512 channels).

Decoders. Two identical, parallel decoder branches are used to reconstruct the resonant and NRB components separately. Each decoder consists of four stages, mirroring the encoder structure. Each stage uses an `UpBlock1D` module, which first performs linear upsampling (`nn.Upsample` with scale factor=2), followed by a 1D Convolution (kernel size 5, padding 2), `BatchNorm1d`, and `ReLU`. Skip connections are implemented by concatenating (`torch.cat`) the output of each decoder stage with the feature map from the corresponding encoder stage (after interpolating the decoder feature map to match the encoder feature map's spatial dimension using `F.interpolate`). The channel dimensions decrease through the decoder: 512 $\rightarrow$ 256 $\rightarrow$ 128 $\rightarrow$ 64 $\rightarrow$ 32.

Output Layers. Each decoder branch terminates with a final 1D Convolution (`nn.Conv1d` with kernel size 1) projecting the 32-channel feature map to a single channel. This output is passed through a `Sigmoid` activation function, scaling the predicted values between 0 and 1. Finally, both the resonant and NRB outputs are interpolated (`F.interpolate` with mode='linear') back to the original input sequence length to ensure compatibility with the ground truth spectra during training and evaluation.

This dual-decoder architecture with shared encoder and bottleneck allows the model to learn common features relevant to both components while dedicating specific pathways for reconstructing the distinct resonant and NRB signals, potentially leveraging shared information from the CARS input effectively.

D.1 Architecture Selection

RamPINN model mostly focus on physics loss but an optimal architecture is also trivial for performance. From Table 3 we see that the U-Net architecture outperforms all other forms of backends, and that the attention block positively impacts the fidelity of the reconstructed signal (PSNR).

E Baseline Implementations

To benchmark the performance of RamPINN, we implemented several state-of-the-art deep learning models from literature for CARS-to-Raman signal reconstruction.

RamPINN Architecture	Supervised	Self-Supervised
U-Net w/ Attention (Ours)	**0.0006 (33.83)**	**0.0053 (22.79)**
U-Net w/o Attention	0.0007 (31.56)	0.0058 (22.59)
ResNet	0.0016 (27.90)	0.0078 (21.06)
AutoEncoder	0.0025 (25.96)	0.0071 (21.47)
Fourier Neural Operator	0.0040 (23.97)	0.0135 (18.86)

Table 3: Comparison of different RamPINN backend architectures under supervised and self-supervised training.

E.1 SpecNet Model

SpecNet (Valensise et al. 2020) employs a CNN trained on synthetic CARS data to extract the imaginary part of the third-order susceptibility, corresponding to the Raman signal. While SpecNet performs well on synthetic datasets, its sensitivity to noise limits its applicability to real-world data, especially in scenarios with weak Raman peaks. Nonetheless, it provides a valuable benchmark for assessing the impact of noise on model performance. We trained the provided SpecNet till 10 epochs with a learning rate of 0.001, batch size 256 and loss metric as mean square error.

E.2 LSTM Model

(Houhou et al. 2020) introduced LSTM networks to process CARS spectra. Being a type of recurrent neural network, the LSTM layers model sequential dependencies across the spectrum. We trained the provided LSTM network till 30 epochs with a learning rate of 0.005, batch size 10 and loss metric as mean absolute error.

E.3 VECTOR: Very Deep Convolutional Autoencoder Model

The VECTOR model (Wang et al. 2021), is a deep convolutional autoencoder designed to retrieve Raman-like spectra from CARS measurements without requiring NRB reference data. The architecture comprises an encoder-decoder structure with skip connections to preserve essential spectral features. Trained on simulated noisy CARS spectra, VECTOR effectively learns to denoise and reconstruct the underlying Raman signal using a customized loss function consisting $L1$ penalty terms. Its performance surpasses earlier models like SpecNet and LSTM. We trained the provided VECTOR network till 40 epochs with a learning rate of 0.0001 and batch size 10.

E.4 BiLSTM Model

Building upon the LSTM approach, a Bi-LSTM architecture (Junjuri et al. 2023) for NRB removal in CARS spectra is explored. The BiLSTM processes the spectral data in both forward and backward directions, capturing comprehensive contextual information. This model exhibited superior performance in reconstructing Raman signals, particularly in regions with overlapping peaks or complex NRB structures. We trained the provided Bi-LSTM network till 10 epochs with a learning rate of 0.005, batch size 10 and loss metric as mean squared error.

E.5 GAN Model

A GAN model (Vernuccio et al. 2024) to extract Raman signals from CARS spectra in an adversarial fashion. The generator in the network has an encoder-decoder architecture with skip connections, and the discriminator is a CNN. We trained the provided GAN model for 1000 epochs with a learning rate of 0.0001 and batch size 8. A customized loss function was employed, utilizing a weighted Huber loss function based on whether peaks are classified as true-positive, false-positive, true-negative, or false-negative, as specified by the authors, for a chosen threshold.

E.6 CNN+GRU Hybrid Model

In the same paper (Vernuccio et al. 2024) proposed a CNN+GRU model that combines a set of convolutional layers and bidirectional gated recurrent units (GRU). This combination of these models helps to capture long-term

dependencies in data. We trained the provided CNN+GRU network till 50 epochs with a learning rate of 0.0005, batch size 256, and loss metric as Huber loss.

E.7 TDKK Method

Time-Domain Kramers-Kronig (TDKK) (Liu et al. 2009) is a traditional algorithm for NRB removal in broadband CARS microscopy, utilizing phase retrieval techniques. While effective, it requires additional NRB measurement with a buffer sample, and estimations of hyperparameters are necessary.

E.8 LeDHT Method

A learned-matrix representation of the discrete Hilbert transform (LeDHT) method (Camp 2022) introduces a learned matrix approach to the discrete Hilbert transform, enhancing the accuracy of Raman signal extraction by addressing distortions caused by NRB interference. It is computationally efficient and significantly enhances the performance of traditional phase retrieval techniques, such as the Kramers–Kronig relation. The approach involves training a matrix that approximates the Hilbert transform, which is then applied to new spectra via matrix multiplication.

E.9 IWT Method

IWP (Härkönen et al. 2023) introduces an approach that utilizes wavelet-based decomposition to model and correct additive or multiplicative spectral background errors. This method is implemented in MATLAB and can also be applied in an unsupervised setting, with the drawback of incomplete phase removal.

E.10 Model Card Overview

We provide a short overview of model parameters and differences for all baseline models and our proposed RamPINN architecture. The values are provided in Table 4. We estimate model size using each framework's built-in parameter-counting tools.

	SpecNet	VECTOR	BiLSTM	LSTM	CNN+GRU	GAN	RamPINN (Ours)
# Parameters	6.0M	111.8M	51.4k	3.9k	84.0k	6.2M	6.8M
Architecture	CNN	AE	BiLSTM	LSTM	CNN + BiGRU	Gen (AE) Disc (CNN)	AE + Attention
Loss Function	L2	L1	L2	L1	Huber	Adversarial	Physics-Informed
Framework	TensorFlow	PyTorch	TensorFlow	TensorFlow	TensorFlow	PyTorch	PyTorch

Table 4: Comparison of model parameters, architectures, and loss functions. "AE" denotes Autoencoder; "Gen" Generator; "Disc" Discriminator.

F Kramers–Kronig Implementation

Physical Intuition. The KK relations originate from causality in linear response theory: for any physical system, the imaginary (absorptive) part of the susceptibility is determined by the real (dispersive) part via the Hilbert transform. In the context of CARS spectroscopy, the measured signal combines a broadband non-resonant background (NRB) and a sharp resonant term. The NRB contributes primarily to the real part of the susceptibility. By estimating and removing the NRB, we isolate a residual that, under ideal conditions, should behave as the real part of a causal analytic signal. The imaginary part of this analytic continuation, obtained via the Hilbert transform, corresponds to the true Raman signature. The KK loss thus enforces this structure by minimizing the discrepancy between the predicted Raman component and the imaginary part of the analytic residual.

F.1 RamPINN Optimization Algorithm

Require: CARS spectra x, optional Raman labels y_{true}, learning rate η, number of iterations T
Ensure: Trained network $y_\theta(x) = (y_{\text{res}}, y_{\text{NRB}})$

for $t = 1$ **to** T **do**
  Sample mini-batch x
  $(y_{\text{res}}, y_{\text{NRB}}) \leftarrow y_\theta(x)$
  $r \leftarrow x - y_{\text{NRB}}$
  Hilbert Transform:
  $R_f \leftarrow \text{FFT}(r)$
  Construct H_f
  $\tilde{R}_f \leftarrow R_f \cdot H_f$
  $z \leftarrow \text{IFFT}(\tilde{R}_f)$
  $y_{\text{res}}^{(\text{KK})} \leftarrow \Im(z)$
  $\mathcal{L}_{\text{KK}} \leftarrow \|y_{\text{res}} - y_{\text{res}}^{(\text{KK})}\|^2$
  $\mathcal{L}_{\text{smooth}} \leftarrow \|\nabla y_{\text{NRB}}\|^2$
  if y_{true} is available **then**
    $\mathcal{L}_{\text{MSE}} \leftarrow \|y_{\text{res}} - y_{\text{true}}\|^2$
  else
    $\mathcal{L}_{\text{MSE}} \leftarrow 0$
  end if
  $\mathcal{L}_{\text{total}} \leftarrow \lambda_{\text{KK}} \cdot \mathcal{L}_{\text{KK}} + \lambda_{\text{smooth}} \cdot \mathcal{L}_{\text{smooth}} + \lambda_{\text{data}} \cdot \mathcal{L}_{\text{MSE}}$
  $\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}_{\text{total}}$
end for

G Additional Synthetic Samples

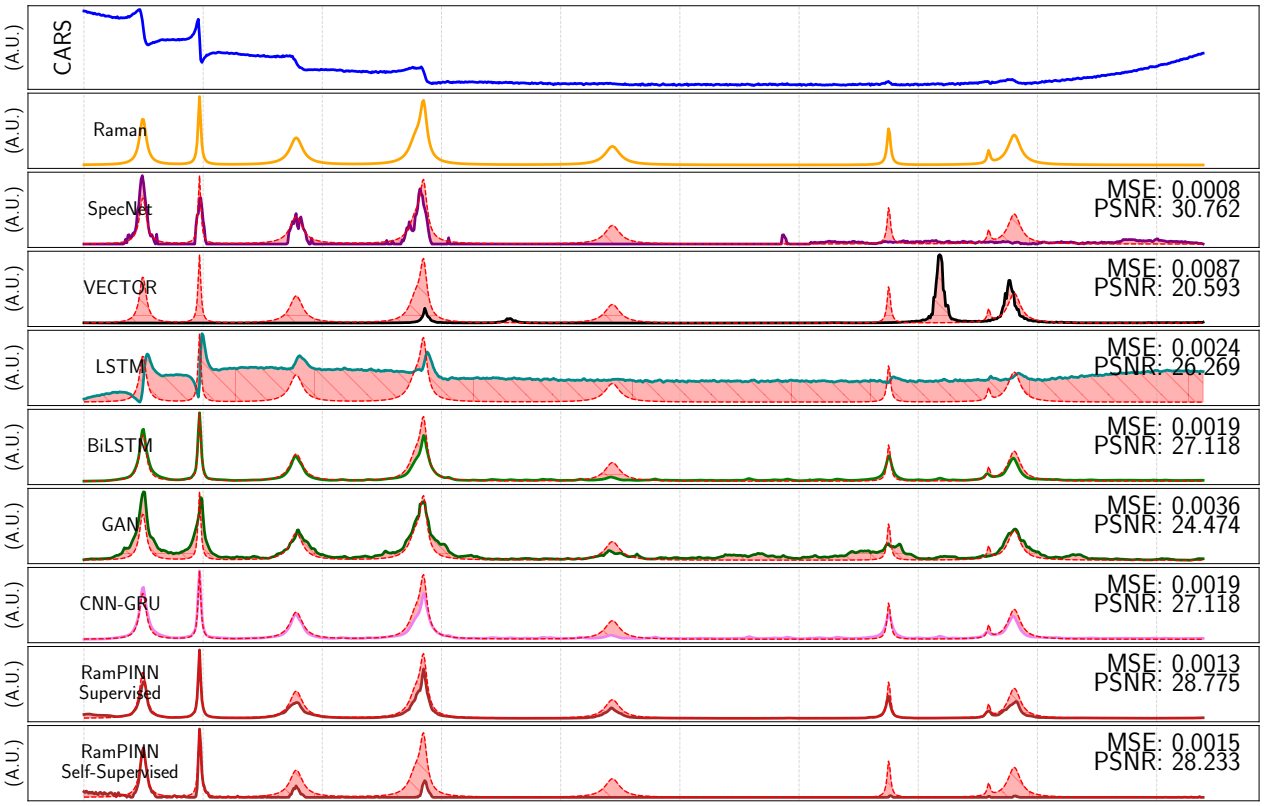


Figure 5: **Additional qualitative comparison of Raman reconstruction.**

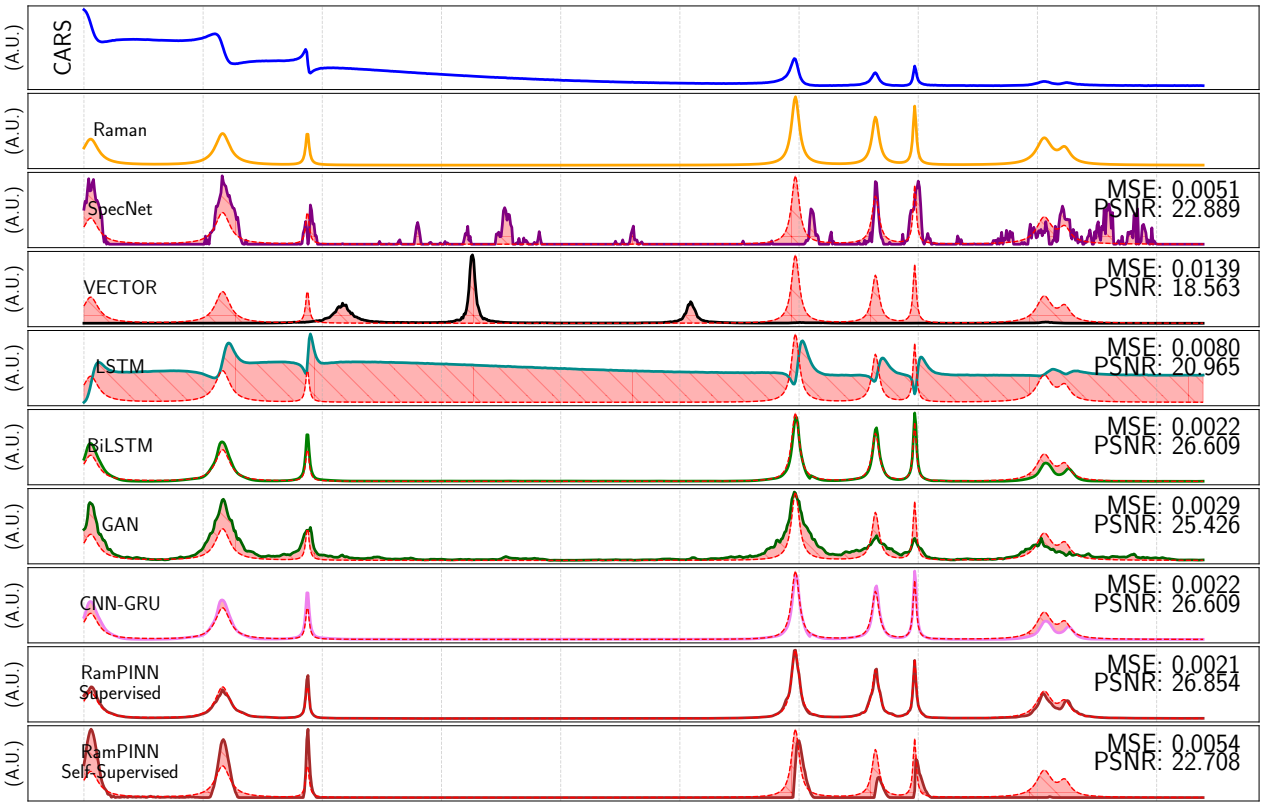


Figure 6: **Additional qualitative comparison of Raman reconstruction.**

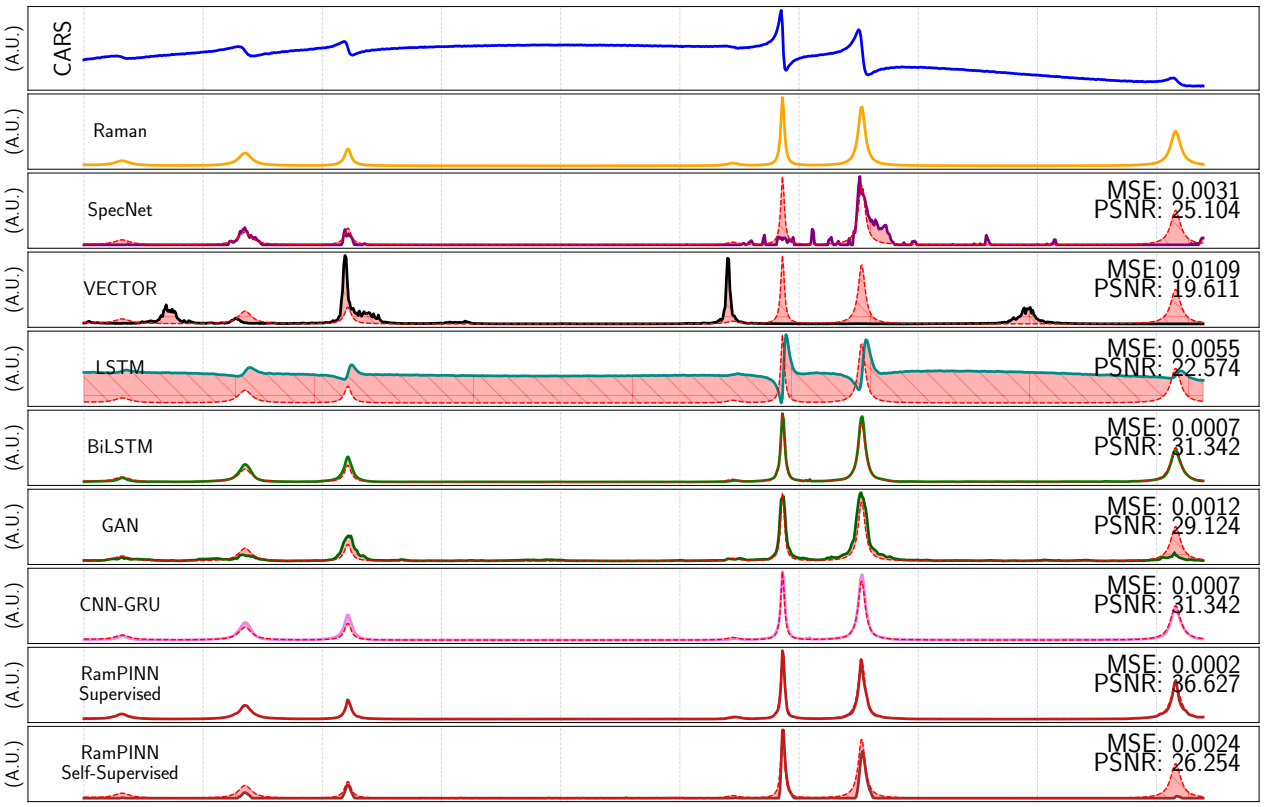


Figure 7: **Additional qualitative comparison of Raman reconstruction.**

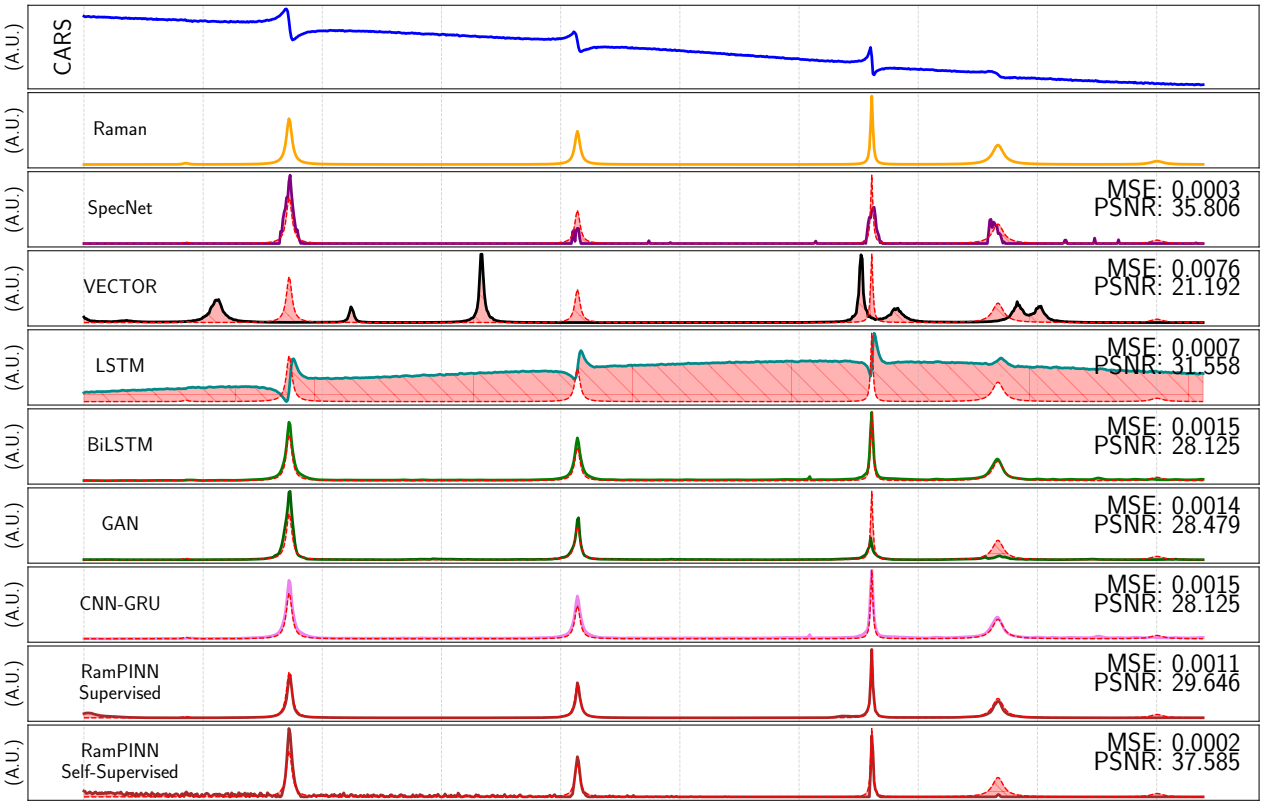


Figure 8: **Additional qualitative comparison of Raman reconstruction.**

H Additional Real-World Samples

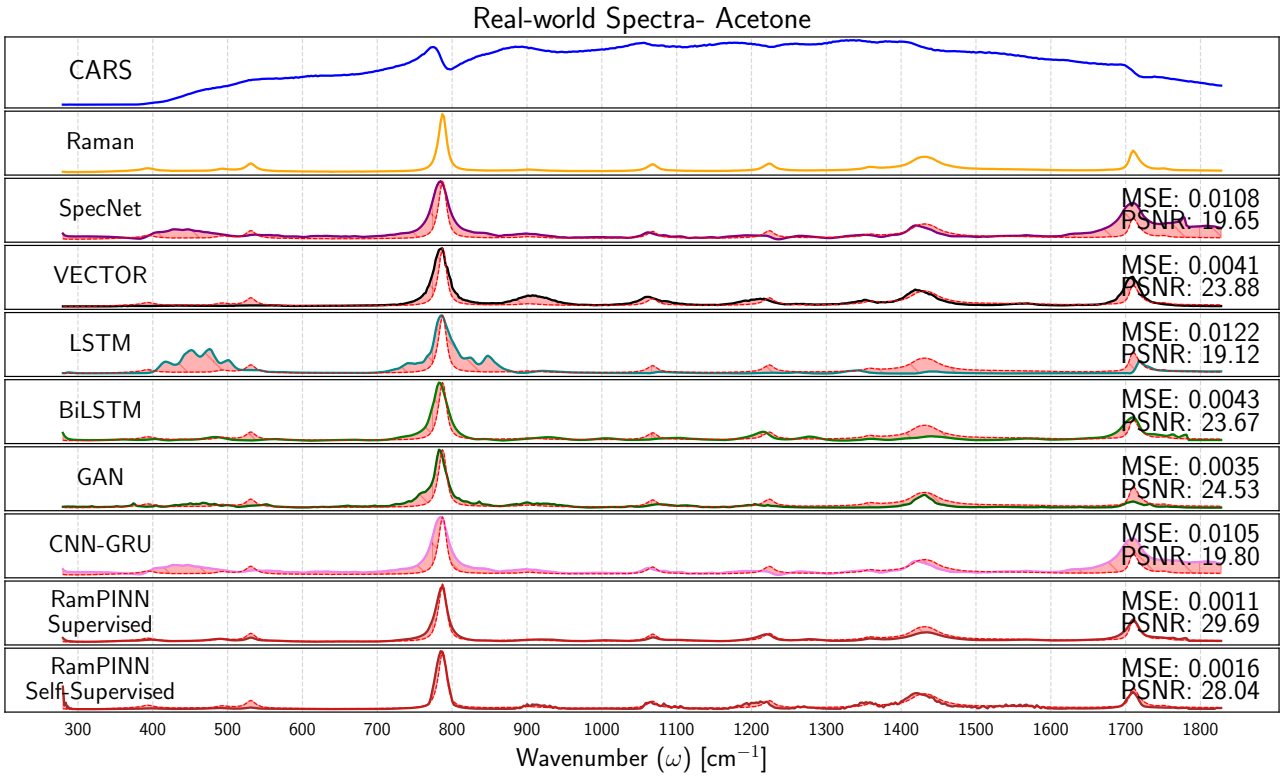


Figure 9: **Real-world sample – Acetone.**

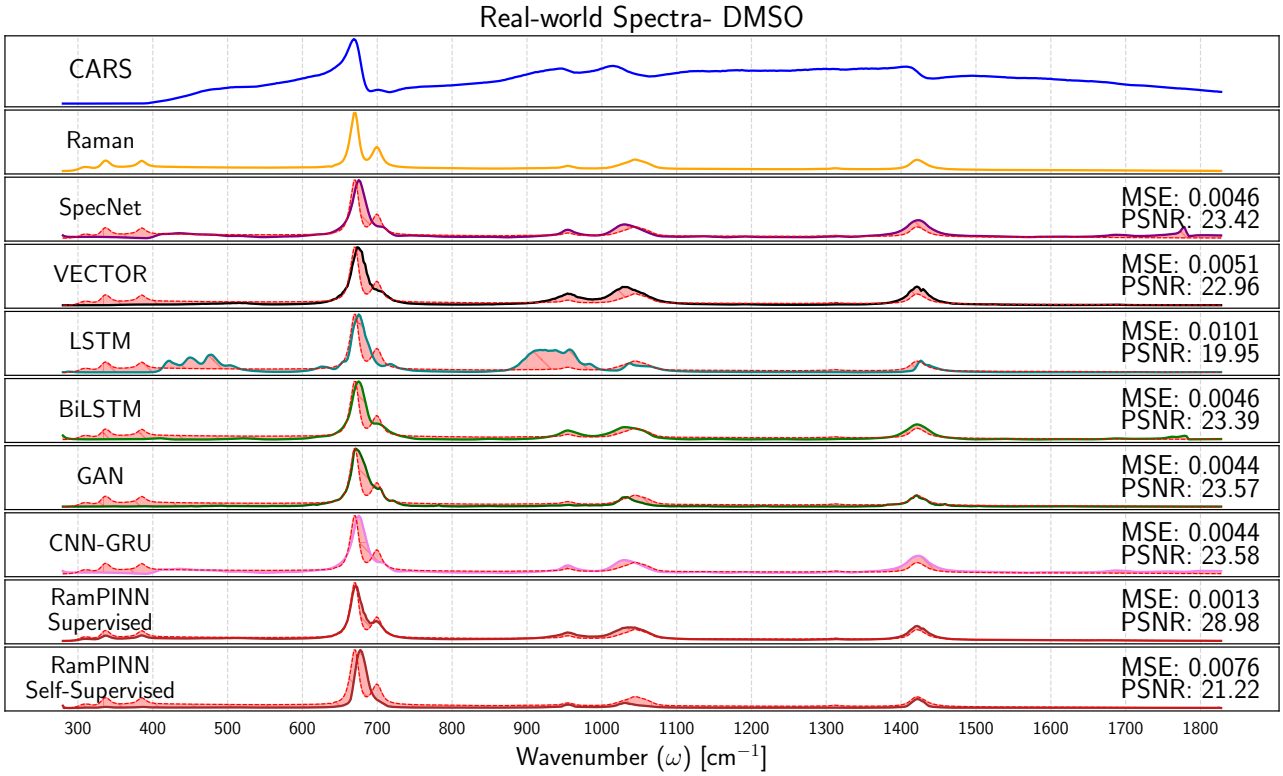


Figure 10: **Real-world sample – DMSO.**

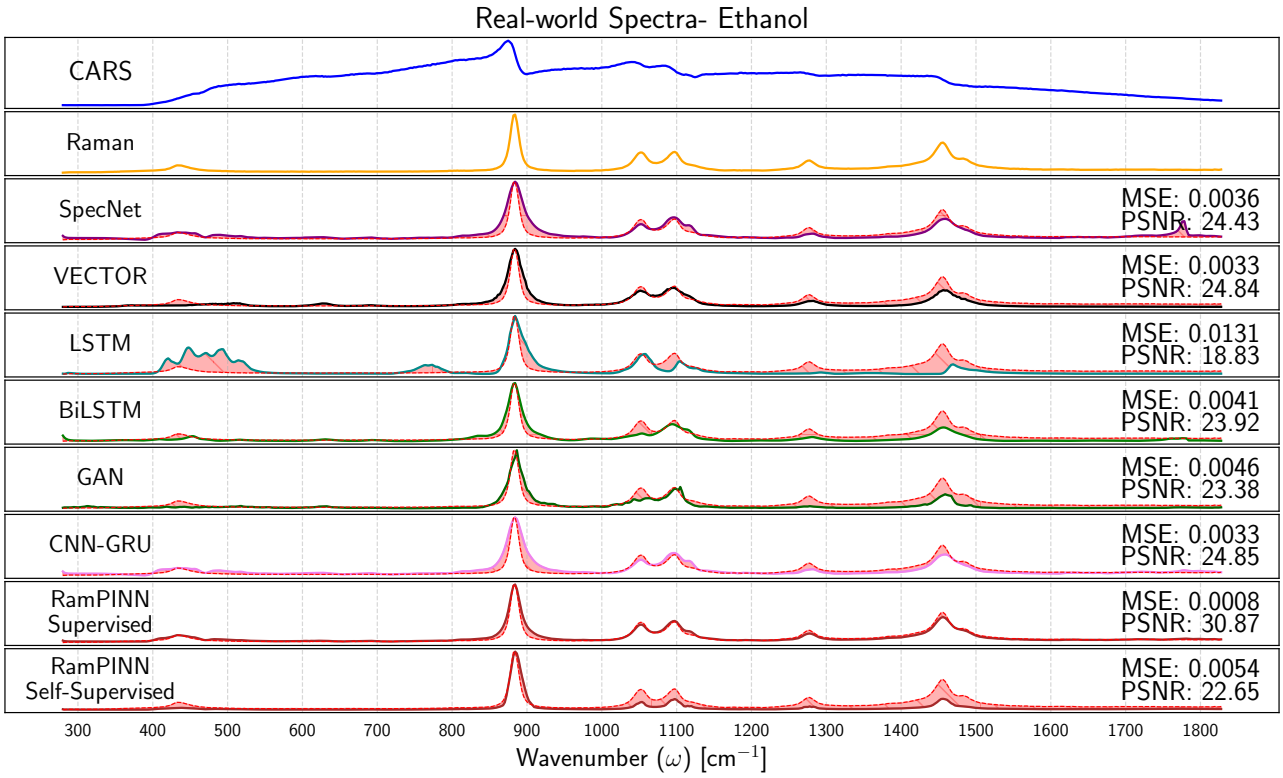


Figure 11: **Real-world sample – Ethanol.**

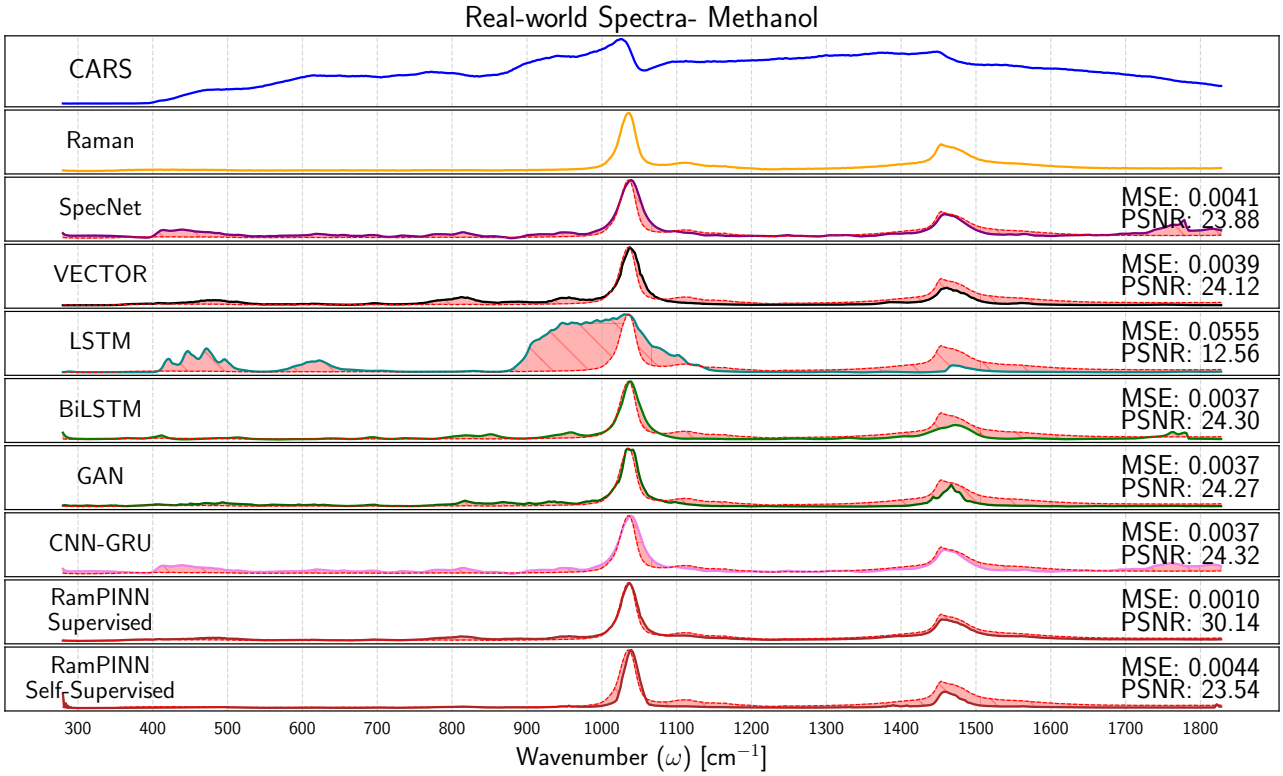


Figure 12: **Real-world sample – Methanol.**

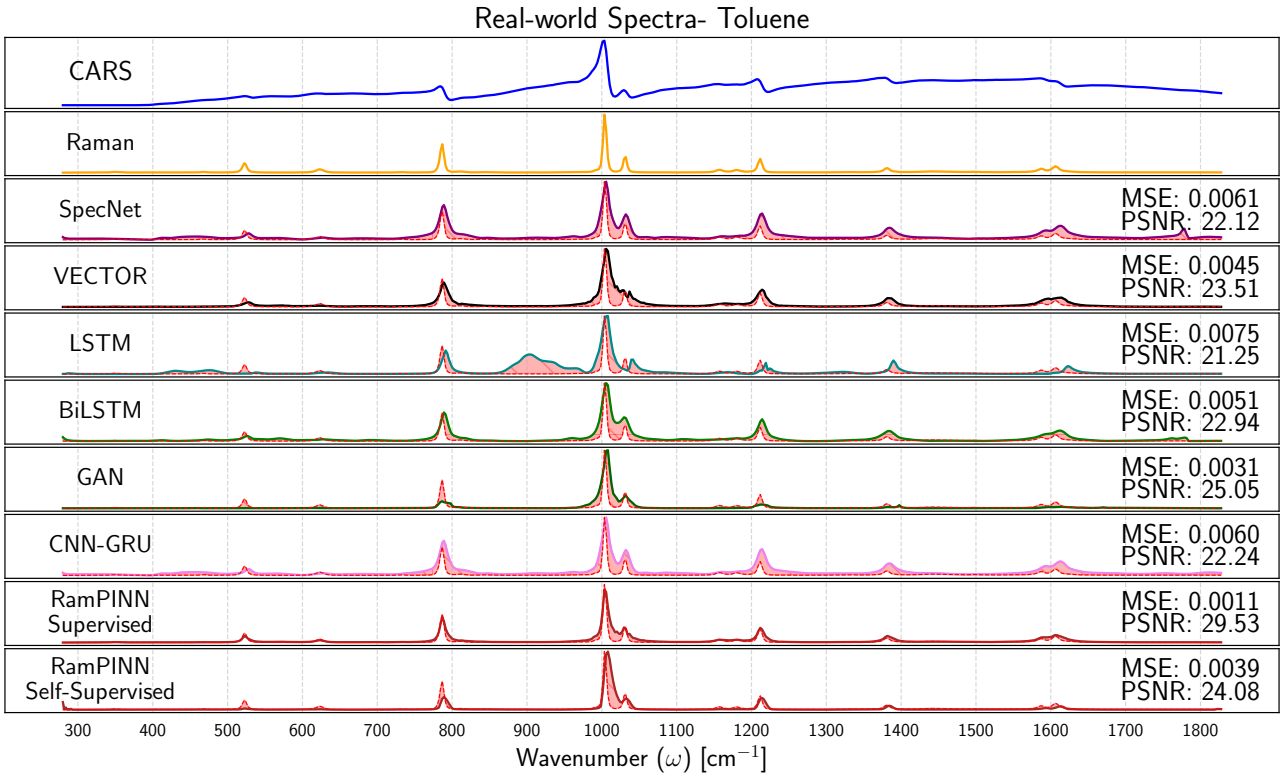


Figure 13: **Real-world sample – Toluene.**

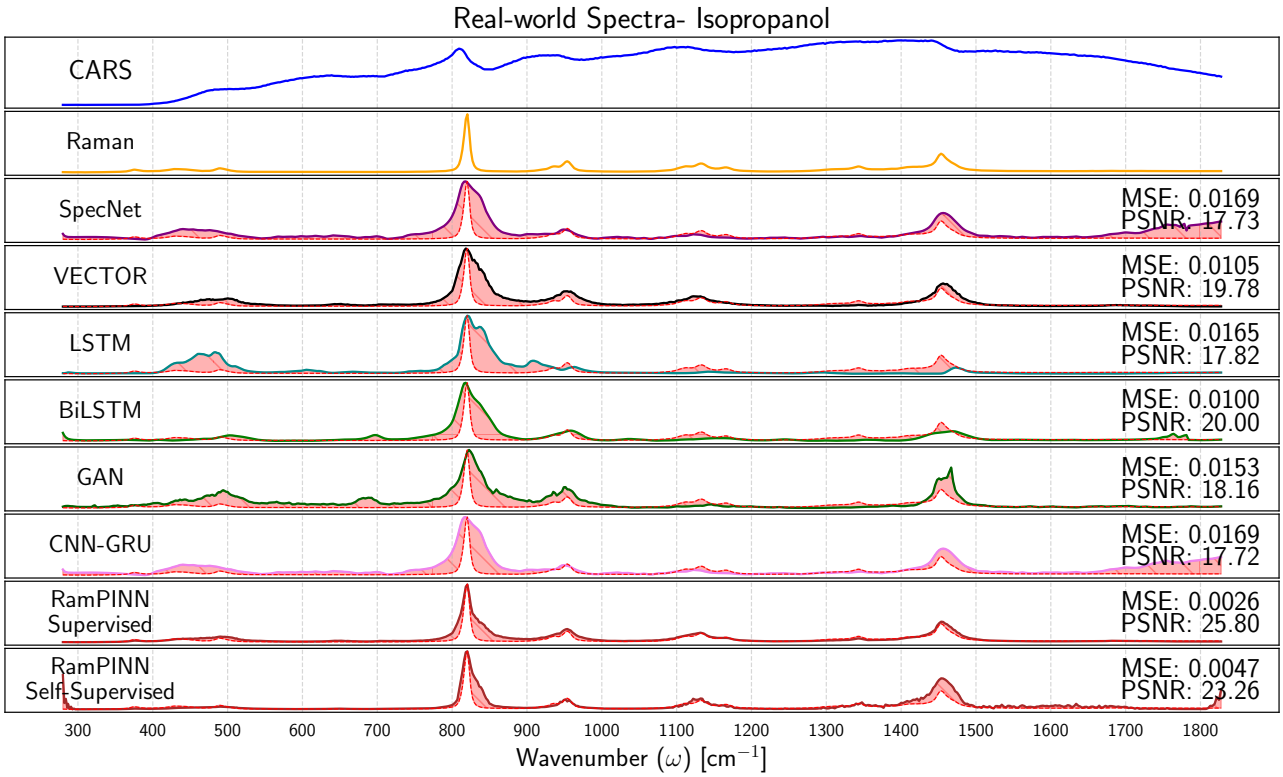


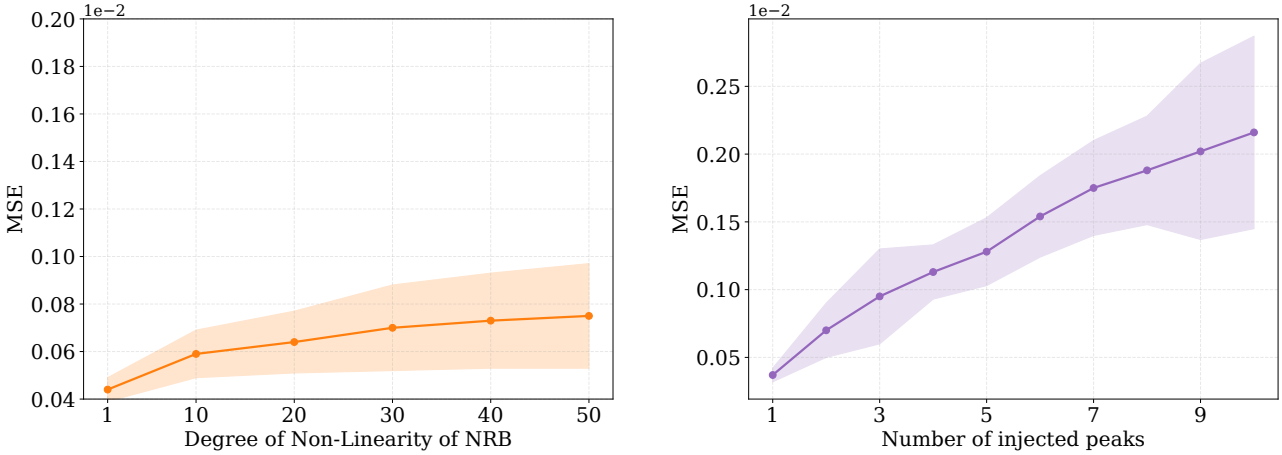
Figure 14: **Real-world sample – Isopropanol.**

I Robustness Studies

Our physics-informed objective is guided by the prior that the non-resonant background (NRB) is a smooth, peak-free component. While this is a common and valid approximation, real-world experimental conditions can introduce complexities. Therefore, we conducted two targeted experiments to assess the robustness of RamPINN when these core assumptions are violated.

Robustness to Increased Nonlinearity. To test the model's ability to handle more complex background shapes, we evaluated its performance on synthetic data with increasingly nonlinear NRB profiles. We generated these by increasing the maximum polynomial degree used in the simulation (see Appendix C for details). Figure 15a shows that the model's performance exhibits graceful degradation. Even with highly complex, nonlinear backgrounds, the reconstruction error remains low, demonstrating that the model is not brittle and can robustly handle a wide range of smooth background shapes. This indicates that the smoothness prior acts as a robust inductive bias, guiding the model to prefer smoother solutions rather than forcing it to fit only simple ones.

Robustness to Peak-like Artifacts. To simulate sharp instrumental artifacts or other features that violate the peak-free assumption, we injected a varying number of random Gaussian peaks into the input CARS spectra. These peaks, intentionally distinct from the Lorentzian shapes of true Raman signals, had randomly sampled heights. The results, presented in Figure 15b, show that RamPINN is remarkably resilient to such structured artifacts. The model effectively learns to attribute these sharp, non-physical features to the NRB component, thereby preserving the fidelity of the reconstructed Raman signal. This demonstrates the efficacy of our physics-architecture co-design: the network learns that physically-inconsistent features are best explained by the NRB decoder, allowing it to isolate the true Raman signal even in the presence of strong, peak-like interference.



(a) Robustness to increasing nonlinearity in NRB.

(b) Robustness to random peaks injected into CARS.

Figure 15: **Robustness of RamPINN**.

J Additional Metrics

MSE and PSNR summarize pointwise differences and are useful for broad comparisons, but they do not reflect spectroscopic goals. In Raman spectra tasks, what matters is whether the peaks are present, where they are, and how strong they are. Two typical failure modes illustrate the gap:

1. A small wavenumber shift of a narrow peak can be chemically significant yet only slightly affect MSE;
2. Smooth background deviations can inflate MSE while the peak set (locations and intensities) is essentially correct.

Setting and detection. All spectra are normalized in intensity. Positions are evaluated on a normalized axis $x = i/(L-1) \in [0,1]$. We detect peaks with `scipy.signal.find_peaks` on an optionally smoothed signal, using absolute (scale-free) thresholds under normalization: minimum height $h_{\min} \in [0,1]$, minimum prominence $p_{\min} \in [0,1]$, and a minimum separation $\delta \in (0,1]$ expressed as a fraction of the signal length. Predicted and true peaks are matched one-to-one by nearest neighbor within tolerance τ in normalized units: a match if $|\hat{x} - x| \leq \tau$. Therefore, alongside MSE/PSNR (main paper), we report peak-aware metrics that directly target detection, localization, and intensity fidelity.

Detection Metrics We count TP (predicted peaks matched within τ), FP (unmatched predicted peaks), and FN (unmatched true peaks), and compute

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}\,, \qquad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}\,, \tag{31}$$

$$\text{F1} = \frac{2\,\text{Precision}\,\text{Recall}}{\text{Precision} + \text{Recall}}\,. \tag{32}$$

This measures whether we found the right peaks.

Given N spectra, we provide (a) *macro* averages: mean $\pm$ std across spectra for each metric (ignoring NaNs for spectra with no matches in location/intensity metrics); (b) *micro* F1: sum $\text{TP}, \text{FP}, \text{FN}$ over all spectra first, then compute Precision, Recall, and F1 from these totals.

Defaults Parameters. Unless stated otherwise, we fix $\tau = 0.01$, $p_{\min} = 0.02$, $h_{\min} = 0$, $\delta = 0.01$, and a Savitzky–Golay window of 11. We keep $p_{\min}$, $h_{\min}$, δ fixed across methods for fairness, and vary only τ if needed to account for sampling density.

Table 5: **Detection Metrics (Macro and Micro) for the Test Synthetic Dataset.** Higher is better (↑). Values are mean ± std over spectra; best per column in **bold**. We observe that RamPINN outperforms in these metrics, demonstrating that it not only retrieves the Raman spectrum effectively but also accurately identifies the peaks.

Method	Macro Precision ↑	Macro Recall ↑	Macro F1 ↑	Micro Precision ↑	Micro Recall ↑	Micro F1 ↑
RamPINN	**0.983 ± 0.082**	**0.972 ± 0.092**	**0.960 ± 0.091**	**0.986**	**0.972**	**0.962**
RamPINN(self-sup)	0.925 ± 0.189	0.743 ± 0.225	0.797 ± 0.192	0.917	0.710	0.800
BiLSTM	0.852 ± 0.174	**0.972 ± 0.092**	0.896 ± 0.129	0.878	0.962	0.923
CNN-GRU	0.852 ± 0.174	0.945 ± 0.116	0.896 ± 0.129	0.878	0.940	0.923
SpecNet	0.260 ± 0.142	0.813 ± 0.196	0.370 ± 0.158	0.241	0.789	0.369
VECTOR	0.114 ± 0.150	0.131 ± 0.179	0.106 ± 0.120	0.116	0.127	0.121
GAN	0.029 ± 0.115	0.020 ± 0.077	0.020 ± 0.068	0.092	0.020	0.032

Mean Location Error (Normalized). Over matched pairs $(\hat{x}_j, x_j)$,

$$\text{MLE}_{\text{norm}} = \frac{1}{\text{TP}} \sum_{j=1}^{\text{TP}} |\hat{x}_j - x_j| \in [0,1]\,, \tag{33}$$

reported as NaN if $\text{TP} = 0$. This measures how well peak positions align.

Relative Intensity Error. Let A denote the per-peak amplitude. For each match,

$$r_j = \frac{|\hat{A}_j - A_j|}{\max\big(|A_j|, \varepsilon\big)} \,. \tag{34}$$

We report the mean and median of $\{r_j\}$; NaN if TP $= 0$. This measures how well peak magnitudes are recovered.

Table 6: **Localization and intensity errors on test synthetic data.** Lower is better (↓). Values are mean ± std; best per column in **bold**. We demonstrate that RamPINN outperforms classical methods in these metrics, indicating that RamPINN effectively aligns peak positions and magnitudes.

Method	Mean loc. err. (norm) ↓	Rel. int. err. (mean) ↓	Rel. int. err. (median) ↓	Pooled RIE (mean) ↓	Pooled RIE (median) ↓
RamPINN	**0.0004 ± 0.0004**	**0.399 ± 0.254**	**0.329 ± 0.181**	**0.429**	**0.309**
RamPINN(self-sup)	0.0009 ± 0.0008	0.449 ± 0.167	0.447 ± 0.196	0.455	0.453
BiLSTM	**0.0004 ± 0.0004**	0.674 ± 0.319	0.650 ± 0.319	0.705	0.641
CNN-GRU	0.0020 ± 0.0011	0.674 ± 0.319	0.650 ± 0.319	0.705	0.641
SpecNet	0.0006 ± 0.0004	0.599 ± 0.336	0.581 ± 0.341	0.599	0.525
VECTOR	0.0049 ± 0.0026	2.238 ± 3.733	2.122 ± 3.643	2.271	0.812
GAN	0.0046 ± 0.0024	0.857 ± 1.075	0.859 ± 1.075	0.824	0.816

Zero-shot Evaluation on Real Molecules. We applied our detection-focused metrics—precision, recall, F1, relative intensity error (mean; ↓), and mean location error (normalized; ↓)—to six real-world molecules (acetone, DMSO, ethanol, methanol, toluene, isopropanol); see Tables Tables 7 to 12. These metrics directly assess whether a model finds the *right peaks* and aligns them at the *right locations* with the *right intensities*, complementing global reconstruction scores like MSE/PSNR.

Takeaway. Not only does **RamPINN** reconstruct signals well (MSE/PSNR), it also *identifies* chemically plausible peaks and preserves their *alignment* and *relative intensities* under zero-shot distribution shift. This consistency across detection and alignment metrics evidences the strength of our physics-integrated approach.

Table 7: Acetone: per-model metrics. Arrows indicate direction; best per column in **bold**.

Method	Precision ↑	Recall ↑	F1 ↑	Rel. Int. Err. (mean) ↓	Mean Loc. Err. (norm) ↓
RamPINN	0.748	**0.875**	**0.778**	**0.251**	**0.0009**
RamPINN(self-sup)	0.489	**0.875**	0.538	0.600	0.0026
BiLSTM	0.417	0.625	0.500	0.332	0.0040
CNN-GRU	0.545	0.750	0.632	0.394	0.0036
VECTOR	**0.750**	0.750	0.750	1.137	0.0036
SpecNet	0.500	0.750	0.600	0.393	0.0036
GAN	0.133	0.500	0.211	0.631	0.0031
LSTM	0.364	0.500	0.421	0.535	0.0038

Table 8: DMSO: per-model metrics. Arrows indicate direction; best per column in **bold**.

Method	Precision ↑	Recall ↑	F1 ↑	Rel. Int. Err. (mean) ↓	Mean Loc. Err. (norm) ↓
RamPINN	**1.000**	**1.000**	**1.000**	0.093	0.0009
RamPINN(self-sup)	**1.000**	0.571	0.727	0.141	0.0038
BiLSTM	0.800	0.571	0.667	0.541	0.0031
CNN-GRU	0.667	0.571	0.615	0.615	0.0034
VECTOR	0.600	0.429	0.500	0.751	**0.0005**
SpecNet	0.571	0.571	0.571	0.615	0.0034
GAN	0.033	0.143	0.054	**0.090**	0.0031
LSTM	0.400	0.571	0.471	2.065	0.0031

Table 9: Ethanol: per-model metrics. Arrows indicate direction; best per column in **bold**.

Method	Precision ↑	Recall ↑	F1 ↑	Rel. Int. Err. (mean) ↓	Mean Loc. Err. (norm) ↓
RamPINN	**1.000**	**1.000**	**1.000**	**0.120**	**0.0003**
RamPINN(self-sup)	**1.000**	**1.000**	**1.000**	0.502	0.0005
BiLSTM	0.625	0.833	0.714	0.407	0.0009
CNN-GRU	0.667	**1.000**	0.800	0.229	0.0010
VECTOR	0.714	0.833	0.769	0.278	0.0006
SpecNet	0.667	**1.000**	0.800	0.229	0.0010
GAN	0.095	**1.000**	0.174	0.725	0.0036
LSTM	0.455	0.833	0.588	0.563	0.0052

Table 10: Methanol: per-model metrics. Arrows indicate direction; best per column in **bold**.

Method	Precision ↑	Recall ↑	F1 ↑	Rel. Int. Err. (mean) ↓	Mean Loc. Err. (norm) ↓
RamPINN	**0.500**	**0.667**	**0.571**	0.032	**0.0000**
RamPINN(self-sup)	0.333	**0.667**	0.444	0.185	0.0023
BiLSTM	0.125	0.333	0.182	**0.024**	**0.0000**
CNN-GRU	0.200	**0.667**	0.308	0.054	0.0023
VECTOR	0.400	**0.667**	0.500	0.161	0.0023
SpecNet	0.200	**0.667**	0.308	0.054	0.0023
GAN	0.020	0.333	0.037	0.935	0.0046
LSTM	0.250	**0.667**	0.364	0.369	0.0069

Table 11: Toluene: per-model metrics. Arrows indicate direction; best per column in **bold**.

Method	Precision ↑	Recall ↑	F1 ↑	Rel. Int. Err. (mean) ↓	Mean Loc. Err. (norm) ↓
RamPINN	**1.000**	0.827	**0.842**	**0.283**	**0.0002**
RamPINN(self-sup)	0.857	0.545	0.667	0.330	0.0020
BiLSTM	0.875	0.636	0.737	0.549	0.0020
CNN-GRU	0.889	0.727	0.800	0.568	0.0023
VECTOR	**1.000**	0.636	0.778	0.470	0.0024
SpecNet	0.800	0.727	0.762	0.565	0.0023
GAN	0.159	**0.909**	0.270	1.008	0.0050
LSTM	0.667	0.727	0.696	0.487	0.0059

Table 12: Isopropanol: per-model metrics. Arrows indicate direction; best per column in **bold**.

Method	Precision ↑	Recall ↑	F1 ↑	Rel. Int. Err. (mean) ↓	Mean Loc. Err. (norm) ↓
RamPINN	**1.000**	**0.667**	**0.800**	**0.159**	**0.0000**
RamPINN(self-sup)	0.375	**0.667**	0.480	0.397	0.0008
BiLSTM	0.500	0.444	0.471	0.445	0.0046
CNN-GRU	0.600	**0.667**	0.632	0.956	0.0036
VECTOR	0.833	0.556	0.667	0.283	0.0034
SpecNet	0.545	**0.667**	0.600	0.957	0.0036
GAN	0.125	0.556	0.204	0.899	0.0061
LSTM	0.444	0.444	0.444	1.524	0.0042