A Brain-Inspired Deep Separation Network for Single Channel Raman Spectra Unmixing

Gaoruishu Long[1], Jinchao Liu[1,*], Bo Liu[2,3], Jie Liu[1], Xiaolin Hu[4]
[1]College of Artificial Intelligence, Nankai University, Tianjin, China
[2]Guangdong Laboratory of Chemistry and Fine Chemical Industry, Guangdong, China
[3]Smekal Tech (Shantou) Ltd.
[4]Department of Computer Science and Technology, Tsinghua University, Beijing, China

***Abstract*—Raman spectra obtained in real world applications are often a noisy combination of several spectra of various substances in a tested sample. Unmixing such spectra into individual components corresponding to each of the substances is of great value and has been a longstanding challenge in Raman spectroscopy. Existing unmixing methods are predominantly designed to invert an overdetermined mixed model and therefore require multiple mixed spectra as input. However, open domain and/or non-cooperative detection applications in Raman spectroscopy such as controlled substance detection, call for single-channel solutions which can identify individual components from thousands of candidates by analyzing only a single noisy mixed spectrum. To our knowledge, sparse regression is the only existing solution which can cope with this scenario, yet it has very low tolerance to noises and can hardly be applicable in practice. To address these limitations, we introduce a novel neural approach for single-channel Raman spectrum unmixing inspired by speech separation. It aims at solving underdetermined systems and can decompose a noisy mixed spectrum from a library of thousands of components (substances). The core of our method is a deep separation neural network (RSSNet) which takes a mixed spectrum as input and outputs spectra of pure components. We created two synthetic datasets of single-channel Raman spectra unmixing and demonstrated feasibility and superiority of RSSNet on these datasets (outperform competing methods by >4dB). Furthermore, we verified that RSSNet, trained solely on synthetic data, can successfully unmix real-world mixed spectra of mixtures of mineral powders, exhibiting strong generalization. Our approach represents a new paradigm for Raman unmixing and enables new possibilities for fast detection of Raman mixtures.**

***Index Terms*—Blind Source Separation, Brain-Inspired Computing, Raman Spectroscopy, Signal Processing**

I. Introduction

Raman spectroscopy serves as a vital non-destructive tool for material identification, relying on the unique "molecular fingerprint" generated by vibrational modes [1]. However, in practical scenarios such as hazardous substance detection or pharmaceutical analysis, the observed spectra are often linear mixtures of multiple components. The objective of Raman unmixing is to blindly decompose a mixed spectrum $\mathbf{y}$ into its constituent pure spectra (*endmembers*) and their corresponding proportions (*abundances*). While effective in controlled settings, this inverse problem becomes non-trivial when the pure components are unknown (i.e., blind source separation).

Current unmixing paradigms face fundamental limitations in this context, as summarized in Table I. Standard geometrical and statistical approaches (e.g., NMF [12], VCA [13]) rely on statistical redundancy across multiple observations to solve the unmixing problem. Consequently, they fail in single-channel scenarios where the system becomes severely underdetermined. Similarly, while deep learning has shown potential in remote sensing [8]–[11] and was recently explicitly adapted for Raman hyperspectral imaging [7], these methods predominantly rely on spatial correlations and are typically constrained to resolving a limited number of endmembers in overdetermined settings. They performed extremely unsatisfactorily in scenarios involving a vast amount of (possible) pure components where ill-posed underdetermined systems need to be inverted, as we demonstrated in this study. In scenarios restricted to a single spectrum, sparse regression stands as the sole conventional candidate, bypassing the underdetermined limit via pre-defined dictionaries. However, a major drawback is that these methods are extremely sensitive to noise [14]–[17], making them impractical for real-world signals where low signal-to-noise ratios (SNR) degrade decomposition accuracy.

This limitation is critical in non-cooperative applications (e.g., controlled substance detection) where acquiring a full hyperspectral image is infeasible. Consequently, there is an urgent need for *single-channel Raman unmixing—identifying individual components from thousands of candidates using only a single, often noisy, input*. Motivated by blind source separation, speech separation in particular, we propose a neural network approach for single-channel Raman unmixing. Although speech processing deals with temporal audio signals, the fundamental principles of extracting meaningful individual components from complex mixtures are highly transferable to the challenges encountered in spectral analysis. Especially deep learning based speech separation methods have been proposed and showed great potential [18]–[21].

Our main contributions are summarized as follows:

- We present a novel neural separation-based paradigm inspired by speech separation for single-channel spectral unmixing in Raman spectroscopy.

* Corresponding author
This work was supported by the National Natural Science Foundation of China under Grant 62476144 and 62076140.

TABLE I
COMPARISON OF UNMIXING PARADIGMS FOR RAMAN SPECTROSCOPY AND SPECTRAL UNMIXING IN GENERAL. U.D. STANDS FOR UNDERDETERMINED, O.D. STANDS FOR OVERDETERMINED.

Unmixing Paradigms	Methods	Mixing Model	#Inputs	Noise Tolerance
Sparse regression-based	e.g. [2], [3]	U.D.	Single	Low
Geometrical and statistical-based	e.g. [4]–[6]	O.D.	Multiple	Low
Hybrid classical-learning-based	e.g. [7]–[11]	O.D.	Multiple	High
Neural separation-based	**RSSNet (ours)**	U.D.	Single	High

- We present the first neural network architecture, named RSSNet, which can decompose a single noisy Raman spectrum of a mixture from a library of thousands of substances with high accuracy.
- We validated the superiority of our method on both synthetic mineral datasets and collected real-world samples, and the extensive experiments demonstrate that RSSNet outperforms competing baselines by over 4 dB, exhibiting strong generalization where traditional methods fail.

The benchmark datasets used in this paper are available at https://github.com/Hydrochlor/RSSNet.

II. RELATED WORK

Spectral Unmixing Approaches. Existing methods largely fall into three categories. *Sparse regression* (e.g., SUnSAL [2]) relies on comprehensive spectral libraries to solve the unmixing problem. While applicable to single-spectrum inputs, these methods are highly sensitive to noise and dictionary completeness. *Geometrical and statistical methods* (e.g., NMF [12], VCA [13]) exploit the simplex structure of data but require multiple mixed observations (overdetermined systems) to estimate endmembers blindly, making them unsuitable for single-channel scenarios. Recently, *Hybrid classical-learning-based unmixing* (e.g., Autoencoders [7]) has emerged, utilizing neural networks to model spectral mixing. However, these models typically rely on spatial constraints within an image cube and struggle to decompose a single independent mixture into a vast number of potential components without spatial context.

Single-Channel Speech Separation. Our work draws inspiration from single-channel speech separation, often formulated as the "Cocktail Party Problem." Recent advancements have shifted from time-frequency masking to time-domain separation, yielding state-of-the-art performance. Architectures such as Conv-TasNet [19] and DPRNN [20] model raw waveforms directly, effectively handling long-range dependencies. TDANet [21] further improved efficiency using top-down attention mechanisms. These methods have demonstrated strong capabilities in disentangling overlapping signals from a single input, a characteristic we aim to transfer to spectral unmixing.

III. METHOD

Suppose that we have a mixed spectrum denoted as $\mathbf{y} \in \mathbb{R}^{1\times L}$, where L represents the length of the mixed spectrum. Under the linear mixing assumption, we have:

$$\mathbf{y} = \sum_{i=1}^{C} \alpha_i \mathbf{s}_i + \mathbf{e}, \tag{1}$$

where $\mathbf{s}_i \in \mathbb{R}^{1\times L}$ denotes the i-th individual substance spectrum, α_i denotes the proportion of $\mathbf{s}_i$, often referred to as abundance or concentration. $\mathbf{e} \in \mathbb{R}^{1\times L}$ denotes the noise signal. Our goal is to unmix $\mathbf{y}$ by directly inverting Eq. (1) using a neural network

$$\{\hat{\mathbf{s}}_i, i = 1, \cdots, C\} = \mathcal{U}_\phi(\mathbf{y}), \tag{2}$$

where $\mathcal{U}$ denotes a neural network with trainable weights ϕ to unmix $\mathbf{y}$. We use the scale-invariant source-to-noise ratio (SI-SNR) [22] to quantitatively measure the difference between the estimated individual spectra $\hat{\mathbf{s}}_i$ and the corresponding ground truth $\mathbf{s}_i$.

A. Overall architecture

In terms of overall pipeline, our method admits the popular encoder-separator-decoder (ESD) framework which has proved to be very effective in speech separation [21], [23], [24], as shown in Fig. 1. The hypothesis behind this ESD framework is that it allows learning a feature space where separation can be carried out with a better accuracy and efficiency.

B. Raman encoder and decoder

The encoder and decoder form the entrance and exit of our separation pipeline, respectively. Inspired by time-domain speech separation [23], we employ a lightweight encoder consisting of a 1-D convolution layer followed by Global Layer Normalization (GLN) and PReLU to project the mixed spectrum $\mathbf{y}$ into a latent feature representation $\mathbf{h} = \text{PReLU}(\text{GLN}(\text{conv1d}(\mathbf{y})))$, where $\mathbf{h} \in \mathbb{R}^{N\times L'}$. Symmetrically, the decoder reconstructs the estimated pure spectra $\hat{\mathbf{s}}_i$ from the masked features using a transposed convolution with identical kernel size and stride: $\hat{\mathbf{s}}_\mathbf{i} = \text{convtranspose1d}(\mathbf{h} \otimes \mathbf{M}_\mathbf{i})$.

We note that the encoder's kernel size is a critical hyperparameter for capturing sharp spectral peaks, as analyzed in our ablation study (See Table V).

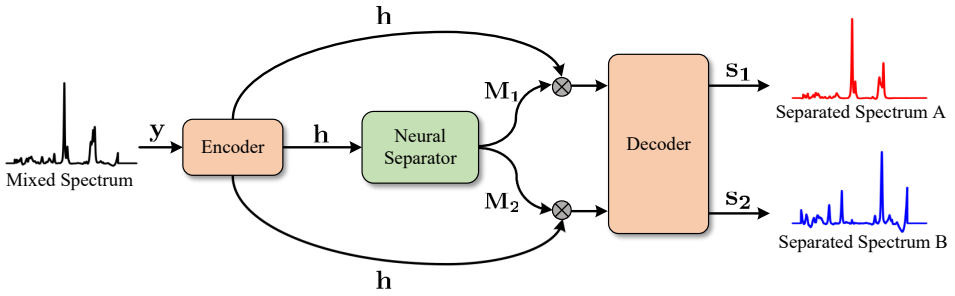

Fig. 1. Graphical illustration of single channel Raman spectra unmixing using a deep separation network. $\otimes$ denotes element-wise multiplication. The diagram assumes that there are only two distinct individual substance spectra in the mixed spectrum.

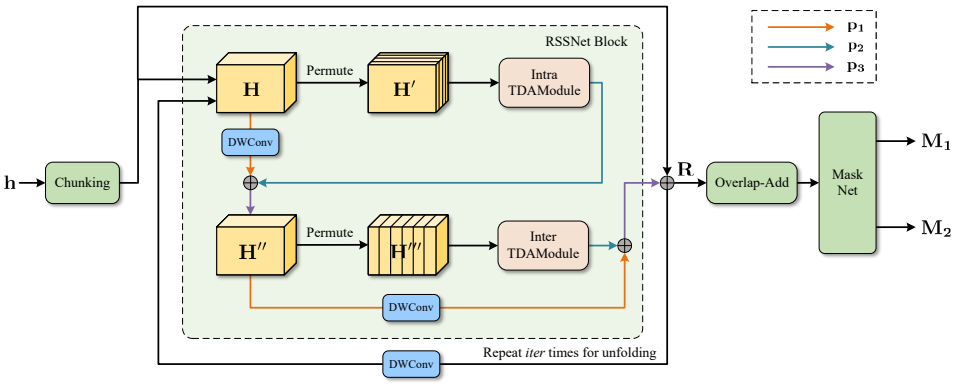

Fig. 2. The architecture of our separation network. $\oplus$ denotes element-wise addition, and *DWConv* denotes a depth-wise convolution layer. $\mathbf{p_1}$, $\mathbf{p_2}$ and $\mathbf{p_3}$ denote the paths for ablation study reported in Section IV-D. Here, separation network still assumes that there are only two distinct individual substance spectra in the mixed spectrum, taking $\mathbf{h}$ as input and finally obtains two masks, $\mathbf{M_1}$ and $\mathbf{M_2}$.

C. Neural separator

As illustrated in Fig. 2, the separator follows a dual-path design inspired by speech separation methods [20], [21]. It begins with a Chunking stage that segments the input feature into overlapping chunks to facilitate local-global modeling. These chunks are processed by stacked RSSNet Blocks, which incorporate Top-Down Attention (TDA) modules to capture multi-scale spectral dependencies. Subsequently, an Overlap-Add operation reconstructs the sequence, which is finally mapped to component-specific masks via the Mask Net.

Chunking. The input feature $\mathbf{h}$ is firstly segmented into T overlapping chunks of size K and stride $K/2$, forming a 3-D tensor $\mathbf{H} \in \mathbb{R}^{N\times K\times T}$.

RSSNet blocks. The core design of the neural separator followed the dual path structure proposed in DPRNN [20], within each path, a brain-inspired TDA module was adopted to enhance the ability of learning local and global dependencies. TDA module is a neural network designed with a top-down attention module, which modulates features of different scales top-down through the attention signal obtained from multi-scale features [21]. To be specific, when the RSSNet Block first receives the input $\mathbf{H}$, it performs a permute operation on it, reshaping the dimensions of $\mathbf{H}$ to $\mathbf{H}' \in \mathbb{R}^{T\times N\times K}$, indicating that it initially processes features along the dimension T. An Intra-TDA module then receives H' to learn local dependencies, which refers to the integration of features within the chunk.

The output of Intra-TDA module is added to $\mathbf{H}$ which passed through a depth-wise convolution layer, resulting in the input $\mathbf{H}''$ for the Inter-TDA module section. Notably, this part is more akin to a two-branch structure rather than a residual path, implying that the network's architecture incorporates two distinct pathways for feature propagation and modulation, differing from the conventional residual paths commonly observed in neural networks. We will further explain its role later. The Intra-TDA module can be represented mathematically as follows:

$$\mathbf{H}'' = f_1(\text{Permute}(\mathbf{H})) + \text{conv2d}(\mathbf{H}), \tag{3}$$

where $f_1(\cdot)$ denotes the network function fitted by Intra-TDA module.

The structure of the Inter-TDA module section is similar to that of the Intra-TDA module section. The input $\mathbf{H}''$ is first permuted, reshaping it to $\mathbf{H}''' \in \mathbb{R}^{K\times N\times T}$, which means this section processes features along the dimension K. Subsequently, it passed through Inter-TDA module, which learns global dependencies, specifically integrating features between chunks. Then, the output is obtained by adding it to $\mathbf{H}''$ passed through a convolution layer, resulting in the

final output of the RSSNet Block. Similarly, the mathematical expression can be written as:

$$\mathbf{H}'''' = f_2(\text{Permute}(\mathbf{H}'')) + \text{conv2d}(\mathbf{H}''), \tag{4}$$

where $\mathbf{H}''''$ denotes the final output of the block, $f_2(\cdot)$ denotes the network function fitted by Inter-TDA module.

Note that the depth-wise convolution layer on the path p_1 improved the performance (Table V). Intuitively speaking, learning local dependencies occurs along the dimension T, and global along the dimension K. Our depth-wise convolution layer *DWConv* here, due to its characteristic of performing separate convolution operations on each input channel, can be regarded as processing features along the dimension N. The *DWConv* path, based on the TCN concept, emulates the self-attention module in the transformer [25]. It allows for further extraction of relationships between feature sequences within each channel on top of local and global dependencies. As a result, it enhances the network's ability to model global correlations in spectral sequence.

Similarly to TDANet, we also consider the RSSNet Block as a recurrent model, and repeat the entire RSSNet Block $iter$ times. If we denote the input and output of the i-th iteration of the RSSNet Block as $\mathbf{R_i}$ and $\mathbf{R}'_\mathbf{i}$, we can obtain the output of the i-th iteration as follows:

$$\mathbf{R_{i+1}} = \text{conv2d}(\mathbf{R}'_\mathbf{i} + \mathbf{H}). \tag{5}$$

Typically, the convolution layer is set with a kernel size of 1×1. The addition of convolution layer is aimed at further integrating features from different scales. The output of the overall separation network is $\mathbf{R}'_{\mathbf{iter}}$.

Intra and inter TDA module. The TDA module mimics the top-down modulation in the biological visual and auditory cortex, where high-level semantic information guides the filtering of low-level sensory signals [21], as illustrated in Fig. 3. It integrates downsampling, global attention (GA) module, and local attention (LA) module to process sequential data effectively.

Initially, the input $\mathbf{L_0}$ undergoes feature extraction through a series of downsampling layers, progressively reducing the sequence length by half each layer while capturing multi-scale features. These features are aggregated to form global features $\mathbf{G}$ via average pooling and summation, ensuring a holistic representation of the input. The Global Attention (GA) module, inspired by the Transformer architecture [26], refines $\mathbf{G}$ by applying a multi-head attention (MHA) layer and a feed-forward network (FFN) layer, producing global top-down attention $\hat{\mathbf{G}}$. The global top-down attention $\hat{\mathbf{G}}$ is then interpolated and used to modulate the multi-scale features.

The local attention (LA) module gradually restores the modulated features back to the original resolution. By generating adaptive parameters ρ and $\mathbf{b}$ derived from adjacent feature scales as local top-down attention, the LA module progressively refines the feature representations, reconstructing fine-grained features.

Overlap-add and mask net. Overlap-add is the inverse operation of chunking. Define $\mathbf{R} = \mathbf{R}'_{\mathbf{iter}} \in \mathbb{R}^{N\times K\times T}$ as

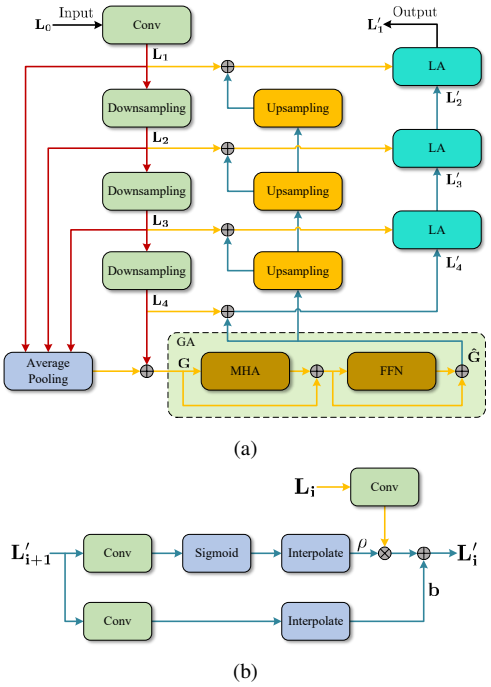

Fig. 3. **(a)** The architecture of the TDA module. $\oplus$ denotes element-wise addition. Here, we assume that the downsampling depth $S = 3$. The red, yellow, and blue arrows correspond to bottom-up connections, lateral connections, and top-down connections, respectively. **(b)** The internal structure of the LA layer. $\oplus$ denotes element-wise addition, $\otimes$ denotes element-wise product.

the input to the overlap-add section, it is responsible for transforming the divided chunks back into a feature sequence $\hat{\mathbf{R}} \in \mathbb{R}^{N\times L'}$. Then $\hat{\mathbf{R}}$ is used as the input to Mask Net, which generates masks $\mathbf{M_i} \in \mathbb{R}^{N\times L'}$ for C individual material spectra. Mask Net consists of a PReLU layer followed by a 1-D convolution layer with a kernel size of 1.

IV. Experiments

In our experiments, we set the feature dimension $N = 256$, encoder/decoder kernel size to 3 (stride 1), and chunk size $K = \sqrt{2L} \approx 45$ according to [20]. The separator comprises 6 stacked RSSNet blocks. Within the TDA modules, we configure the upsampling depth $S = 3$, use 8 attention heads (512-dim), and set the FFN hidden dimension to 1024. All depth-wise convolutions use 1×1 kernels and stride of 1, and the dropout rate is fixed at 0.1.

We use Adam as the optimizer with an initial learning rate of 1e-3. Gradient clipping is also applied during training, limiting the gradient's L2 norm to 5. Training ran for 100 epochs (*RRUFF-2Mix*) and 200 epochs (*UNIPR-2Mix*) on 2×GeForce RTX 3080 GPUs.

A. Datasets and evaluation metrics

We established two single-channel unmixing datasets, *RRUFF-2Mix* and *UNIPR-2Mix*, derived from RRUFF [32] and UNIPR [1] databases, respectively. All spectra were standardized to a length of 1,024. *RRUFF-2Mix* comprises

TABLE II
COMPARISON OF THE PROPOSED RSSNET AND COMPETING METHODS FOR RAMAN UNMIXING ON THE DATASETS RRUFF-2MIX AND UNIPR-2MIX. TF STANDS FOR TRANSFORMER AND NA STANDS FOR NOT APPLICABLE.

Methods	**RRUFF-2Mix**		**UNIPR-2Mix**		**Params** (M)
	SI-SNR (dB) ↑	SI-SNRi (dB) ↑	SI-SNR (dB) ↑	SI-SNRi (dB) ↑	
Sparse regression-based					
SUnSAL [2]	Failed	-	Failed	-	-
NNOMP [27]	Failed	-	Failed	-	-
Geometrical and statistical-based					
FastICA [28]	-20.18	-	-22.69	-	-
NMF [29]	-15.72	-	-22.97	-	-
VCA [13]	-18.41	-	-22.05	-	-
N-FINDR [30]	-17.84	-	-22.05	-	-
FCLS [4]	Failed	-	Failed	-	-
MCR-ALS [31]	Failed	-	Failed	-	-
KF-OSU [6]	-	-	-27.06	-	-
Hybrid classical-learning-based					
Unmixing-AE(Conv.) [7]	-21.33	-	-22.61	-	-
Unmixing-AE(Conv. TF) [7]	-20.56	-	-22.51	-	-
Unmixing-AE(Dense) [7]	-21.67	-	-22.86	-	-
Unmixing-AE(TF) [7]	-20.48	-	-22.86	-	-
mDAE [8]	-16.79	-	-22.04	-	-
SIDAEU [9]	-16.96	-	-22.14	-	-
CNNAEU [11]	-17.12	-	NA	-	-
Endnet [10]	Failed	-	Failed	-	-
Neural separation-based					
Conv-TasNet [19]	13.75 $\pm$0.38	15.12 $\pm$0.38	11.79 $\pm$0.12	12.21 $\pm$0.12	3.46
DPRNN [20]	16.73 $\pm$0.53	18.09 $\pm$0.53	12.44 $\pm$0.14	12.85 $\pm$0.14	38.98
AFRCNN [24]	15.90 $\pm$0.79	17.27 $\pm$0.79	12.06 $\pm$0.11	12.48 $\pm$0.11	15.73
TDANet [21]	11.74 $\pm$0.34	13.11 $\pm$0.34	9.89 $\pm$0.19	10.30 $\pm$0.19	**2.50**
RSSNet (ours)	**21.13** $\pm$0.81	**22.50** $\pm$0.81	**16.44** $\pm$0.42	**16.85** $\pm$0.42	5.25

Note: 'Failed' denotes failing to produce valid results, caused by either component misidentification (e.g. sparse regression) or non-convergence (e.g. Endnet).

60k/5k/5k samples for training/validation/testing, while the smaller *UNIPR-2Mix* contains 5k/1k/1k samples.

To generate synthetic mixtures, we randomly selected two pure spectra $\mathbf{s_1}$ and $\mathbf{s_2}$, and a mixing factor $\alpha \in [0, 1]$ to form clean mixed spectrum $\mathbf{m} = \alpha\mathbf{s_1} + (1-\alpha)\mathbf{s_2}$. To simulate real-world conditions where spectra are contaminated by sample matrix effects and detector fluctuations, we applied a robust noise protocol. Following [33]–[35], we added a mixture of Brown (red) and white noise at a 1:1 ratio to the clean spectra, with a Signal-to-Noise Ratio (SNR) varying between 10 and 20 dB.

To verify generalization, we acquired 21 real-world mixed spectra, covering a wide range of solid-solid, solid-liquid, and liquid-liquid mixtures. For the solid samples demonstrated in Fig. 4, we purchased powders of minerals such as Orpiment and Realgar. We first obtained pure spectral signatures, then physically mixed the powders, compressed them into pellets, and performed measurements. All real-world spectra were acquired using an 830nm laser transmission Raman spectrometer (Smekal, TRS3) and a backscatter probe.

In terms of evaluation metrics, we reported both SI-SNR and SI-SNR improvement (SI-SNRi) [36] for speech separation methods, and only SI-SNR for hyperspectral unmixing methods, as SI-SNRi is not applicable. For both evaluation metrics, higher values indicate better separation performance.

B. Compared methods

To justify the need for our proposed RSSNet, we validate all existing unmixing paradigms and show that none of them are applicable to our case as shown in Table II. Note that as discussed in the introduction, all existing unmixing methods except for sparse regression are designed to work with multiple mixed spectra, so we transformed our single-channel unmixing problem to multiple channel unmixing by stacking all the spectra to form an "hyperspectral image", so that these methods can be evaluated. This is only for the purpose of evaluation mathematically, and inapplicable in practice. Specifically, for geometrical and statistical methods (e.g., VCA, FastICA), we upsampled the spectral length to 2,048 for the *RRUFF-2Mix* dataset to satisfy the requirement that the spectral dimension must exceed the number of endmembers ($L > P$).

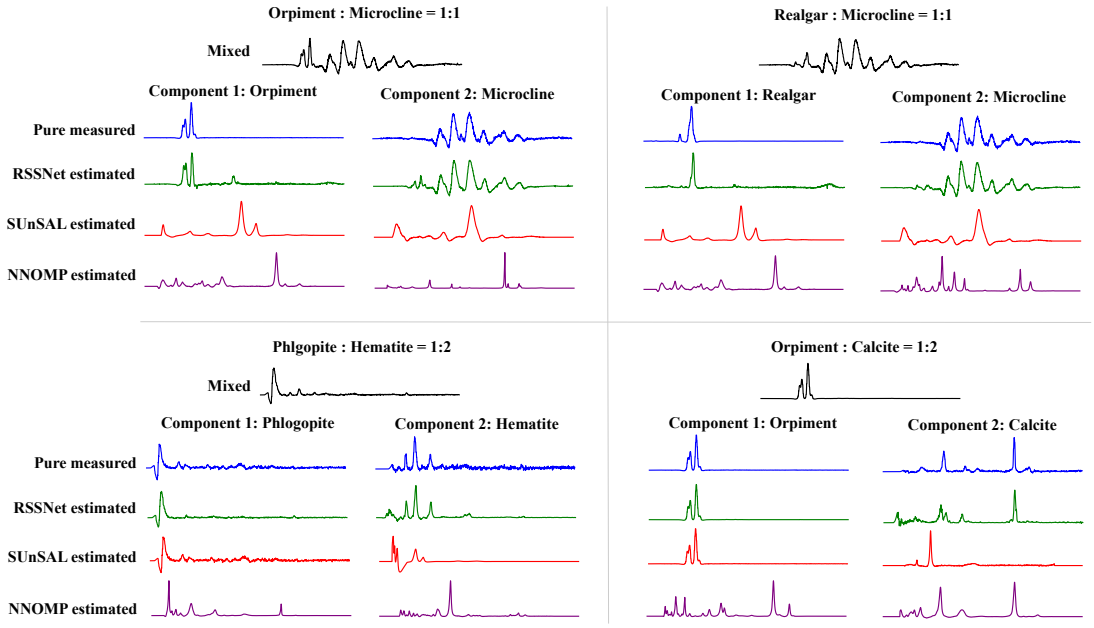

Fig. 4. Results of our RSSNet and existing methods unmixing real-world Raman spectra of mixtures of mineral powders Orpiment [As_2S_3], Microcline [$K(AlSi_3O_8)$], Realgar [AsS], Hematite [Fe_2O_3], Phlogopite [$KMg_3(AlSi_3O_{10})(OH)_2$] and Calcite [$Ca(CO_3)$]. The intensities of the pure spectra in the two mixed spectra above are comparable, and the characteristic peaks of both pure spectra are clearly visible in the mixed spectra. In contrast, the pure spectra in the two mixed spectra below differ significantly in magnitude; in the mixed spectra, only the peaks of the large one are clearly observable, while the other exhibits much lower amplitude and is nearly invisible. For visualization purposes, all spectra have been normalized. Please view at full resolution for better clarity.

TABLE III
COMPARISON OF NEURAL SEPARATION-BASED METHODS ON EXTENDED METRICS. FOR ALL METRICS, LOWER VALUES ARE BETTER (↓).

	RRUFF-2Mix			UNIPR-2Mix		
Method	**SID (↓)**	**SAD (↓)**	**RMSE (↓)**	**SID (↓)**	**SAD (↓)**	**RMSE (↓)**
AFRCNN	0.2275	0.2222	0.0068	0.3394	0.3084	0.0094
Conv-TasNet	0.2916	0.2992	0.0091	0.4405	0.3232	0.0098
DPRNN	0.1897	0.2175	0.0066	0.3286	0.3033	0.0092
TDANet	0.3368	0.3267	0.0099	0.4176	0.3798	0.0115
RSSNet (ours)	**0.1256**	**0.1432**	**0.0043**	**0.2351**	**0.2320**	**0.0070**

To demonstrate the superiority of the network architecture of RSSNet, we compared RSSNet with state-of-the-art speech separation networks Conv-TasNet [19], DPRNN [20], AFRCNN [24], TDANet [21] where we draw inspiration from.

C. Accuracy and robustness

On synthetic datasets. As detailed in Table II, RSSNet achieves state-of-the-art performance, surpassing baselines with an SI-SNRi of 22.50 dB (*RRUFF-2Mix*) and 16.85 dB (*UNIPR-2Mix*) using only 5.25M parameters. In stark contrast, traditional paradigms proved ineffective for this single-channel task. Sparse regression methods failed to identify the correct components from the spectral library, often selecting incorrect dictionary atoms due to their sensitivity to noise. This resulted in invalid reconstructed spectra as shown in Fig. 4. Meanwhile, geometrical and hybrid approaches produced unusable results (often negative SI-SNR) as they are fundamentally designed to require multiple mixed spectra. Also, we further evaluate shape recovery using Spectral Information Divergence (SID), Spectral Angle Distance (SAD), and RMSE. As shown in Table III, RSSNet outperforms baselines across all metrics on both datasets. This confirms RSSNet's ability to reconstruct precise spectral shapes while maintaining high noise separation performance—a key objective in Raman unmixing.

On real-world data. Qualitative results in Fig. 4 demonstrate that RSSNet accurately recovers pure spectral signatures, even in extreme scenarios where one component is spectrally submerged by the other. In contrast, existing methods SUnSAL and NNOMP both failed to unmix all four mixed spectra. In particular, SUnSAL outputted the same unmixed spectra for two different mixtures. For the other two mixtures, SUnSAL managed to find one correct component which dominated the mixed spectra, and it completely missed the other component almost invisible in the mixed spectra.

While Fig. 4 demonstrates qualitative feasibility on mineral powders, we conducted a rigorous quantitative evaluation on a larger dataset of 21 physically mixed samples. This

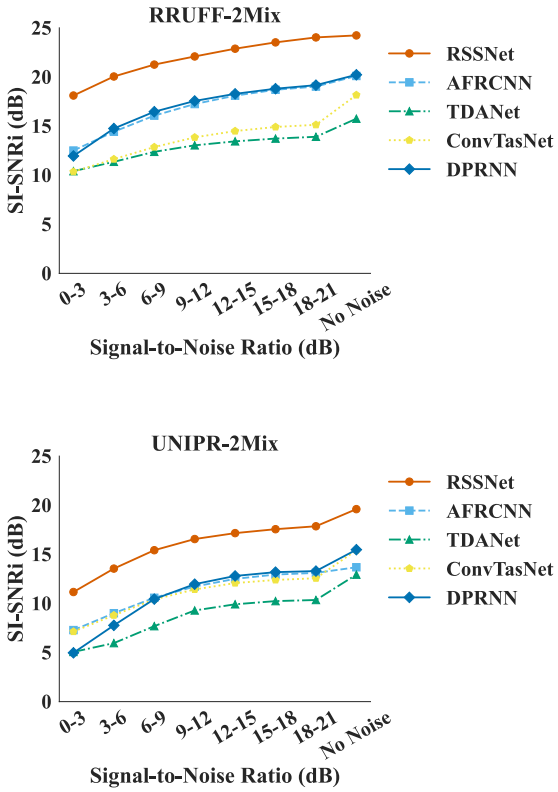

Fig. 5. Robustness against different levels of noises of RSSNet and competing methods.

dataset includes diverse scenarios such as solid-solid mixtures (e.g., Orpiment-Realgar), liquid-liquid mixtures (e.g., Ethanol-Methanol), and solid-liquid compositions. As shown in Table IV, RSSNet demonstrated superior sim-to-real generalization, achieving a Mean SI-SNR of 11.72 dB (Median: 11.22 dB) and securing the best performance in 16 out of 21 cases. In stark contrast, generic speech separation models struggled to bridge the domain gap: Conv-TasNet dropped to 6.31 dB, while DPRNN failed to yield meaningful results (0.09 dB mean). This highlights that our specialized architecture is essential for bridging the sim-to-real gap and handling the complex baselines and noise in physical spectra.

Robustness. We conducted experiments to investigate the performance of compared methods at different level of noises (Fig. 5). It is evident that our RSSNet significantly outperformed the competing networks in speech separation.

D. Ablation study

To validate the architectural design of RSSNet, we conducted ablation studies on the *RRUFF-2Mix* dataset, focusing on the Raman encoder's kernel size and the configuration of the depth-wise convolution (*DWConv*) in the separator.

Raman encoder kernel size. Raman spectra are characterized by sharp, distinct peaks where position and shape are critical. We hypothesized that a smaller kernel size is preferable to capture these fine-scale features. As shown in

TABLE IV
SUMMARY OF QUANTITATIVE SI-SNR RESULTS ON 21 REAL-WORLD MIXED SAMPLES. HIGHER VALUES INDICATE BETTER PERFORMANCE.

Method	**Mean** (dB)	**Median** (dB)
ConvTasNet	6.31	5.51
DPRNN	0.09	2.31
AFRCNN	9.32	10.38
TDANet	8.81	7.28
RSSNet (ours)	**11.72**	**11.22**

TABLE V
ABLATION STUDIES ON RSSNET COMPONENTS (RRUFF-2MIX). LEFT: ENCODER KERNEL SIZE. MIDDLE: DWCONV PATH POSITION (SEE FIG. 2). RIGHT: DWCONV KERNEL SIZE.

Encoder Kernel		**DWConv Path**		**DWConv Kernel**	
Size	SI-SNRi (dB)	Path	SI-SNRi (dB)	Size	SI-SNRi (dB)
2	19.53	None	22.29	$\mathbf{1 \times 1}$	**22.50**
3	**22.50**	$\mathbf{p_1}$	**22.50**	3×3	19.95
5	22.34	p_2	20.22	5×5	20.16
7	19.26	p_3	20.06	7×7	18.43
9	20.06	-	-	-	-

Table V, the network achieved the best performance (22.50 dB SI-SNRi) with a kernel size of 3. Performance gradually declined as the kernel size increased (dropping to 20.06 dB at size 9), confirming that extracting high-frequency local details is more important than a large initial receptive field for this task.

Depth-wise convolution strategy. We further investigated the role of the *DWConv* layer within the RSSNet Block. We evaluated three potential paths for this layer (p_1, p_2, p_3) as illustrated in Fig. 2. Placing the *DWConv* on path p_1 yielded the highest accuracy. Furthermore, we examined the kernel size for this specific layer. Results indicate that a 1×1 kernel significantly outperforms larger ones (e.g., 22.50 dB vs. 18.43 dB for 7×7). This 1×1 *DWConv* acts as a channel-mixing mechanism similar to Point-wise Convolution or the projection in self-attention, effectively modeling global correlations among feature channels. Since the TDA modules already handle temporal dependencies, using larger kernels here is redundant and degrades local sensitivity.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented a novel neural separation paradigm for single-channel Raman spectral unmixing and introduce RSSNet, a brain-inspired neural network successfully overcomes the limitations of traditional geometrical and statistical methods in underdetermined and noisy scenarios. Experimental results on both synthetic datasets and real-world mineral samples confirm that RSSNet achieves state-of-the-art performance in decomposing single mixed spectra, demonstrating strong robustness against spectral variations. Future work will focus on extending the framework to handle multi-component mixtures and investigating generalization to unseen substances in open-world scenarios.

REFERENCES

[1] J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, and S. J. Gibson, "Deep convolutional neural networks for raman spectrum recognition: a unified solution," *Analyst*, vol. 142, no. 21, pp. 4067–4074, 2017.

[2] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2010, pp. 1–4.

[3] M. Parente and M.-D. Iordache, "Sparse unmixing of hyperspectral data: The legacy of sunsal," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 21–24.

[4] D. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, 2001.

[5] M. Garrido, F. Rius, and M. Larrechi, "Multivariate curve resolution–alternating least squares (mcr-als) applied to spectroscopic data from monitoring chemical reactions processes," *Analytical and bioanalytical chemistry*, vol. 390, no. 8, pp. 2059–2066, 2008.

[6] H. Kouakou, J. H. de Morais Goulart, R. Vitale, T. Oberlin, D. Rousseau, C. Ruckebusch, and N. Dobigeon, "On-the-fly spectral unmixing based on kalman filtering," *Chemometrics and Intelligent Laboratory Systems*, p. 105252, 2024.

[7] D. Georgiev, Álvaro Fernández-Galiana, S. V. Pedersen, G. Papadopoulos, R. Xie, M. M. Stevens, and M. Barahona, "Hyperspectral unmixing for raman spectroscopy via physics-constrained autoencoders," *Proceedings of the National Academy of Sciences*, vol. 121, no. 45, p. e2407439121, 2024.

[8] R. Guo, W. Wang, and H. Qi, "Hyperspectral image unmixing using autoencoder cascade," in *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2015, pp. 1–4.

[9] F. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Neural network hyperspectral unmixing with spectral information divergence objective," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 755–758.

[10] S. Ozkan, B. Kaya, and G. B. Akar, "Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 482–496, 2019.

[11] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spectral–spatial hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 535–549, 2021.

[12] X. Liu, W. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 757–772, 2011.

[13] J. Nascimento and J. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[14] C. Li, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Sparse unmixing of hyperspectral data with noise level estimation," *Remote Sensing*, vol. 9, no. 11, 2017.

[15] X. Xu, Z. Shi, and B. Pan, "ℓ_0-based sparse hyperspectral unmixing using spectral information and a multi-objectives formulation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 141, pp. 46–58, 2018.

[16] M. Gong, H. Li, E. Luo, J. Liu, and J. Liu, "A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 2, pp. 234–248, 2017.

[17] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3890–3905, 2016.

[18] D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702–1726, 2018.

[19] Y. Luo and N. Mesgarani, "Conv-tasnet: Surpassing ideal time–frequency magnitude masking for speech separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 8, pp. 1256–1266, 2019.

[20] Y. Luo, Z. Chen, and T. Yoshioka, "Dual-path rnn: Efficient long sequence modeling for time-domain single-channel speech separation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 46–50.

[21] K. Li, R. Yang, and X. Hu, "An efficient encoder-decoder architecture with top-down attention for speech separation," in *The Eleventh International Conference on Learning Representations*, 2023.

[22] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 31–35.

[23] Y. Luo and N. Mesgarani, "Tasnet: Time-domain audio separation network for real-time, single-channel speech separation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 696–700.

[24] X. Hu, K. Li, W. Zhang, Y. Luo, J.-M. Lemercier, and T. Gerkmann, "Speech separation using an asynchronous fully recurrent convolutional neural network," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, 2021, pp. 22 509–22 522.

[25] D. Luo and X. Wang, "ModernTCN: A modern pure convolution structure for general time series analysis," in *The Twelfth International Conference on Learning Representations*, 2024.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30, 2017.

[27] M. Yaghoobi, D. Wu, R. J. Clewes, and M. E. Davies, "Fast sparse Raman spectral unmixing for chemical fingerprinting and quantification," in *Optics and Photonics for Counterterrorism, Crime Fighting, and Defence XII*, D. Burgess, G. Owen, H. Bouma, F. Carlysle-Davies, R. J. Stokes, and Y. Yitzhaky, Eds., vol. 9995, International Society for Optics and Photonics. SPIE, 2016, p. 99950E.

[28] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 07 1997.

[29] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[30] M. E. Winter, "N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *SPIE Proceedings*, vol. 3753. SPIE, October 1999, pp. 266–275.

[31] I. Matveeva, I. Bratchenko, Y. Khristoforova, L. Bratchenko, A. Moryatov, S. Kozlov, O. Kaganov, and V. Zakharov, "Multivariate curve resolution alternating least squares analysis of in vivo skin raman spectra," *Sensors*, vol. 22, no. 24, 2022.

[32] B. Lafuente, R. T. Downs, H. Yang, and N. Stone, *1. The power of databases: The RRUFF project*, Berlin, München, Boston, 2016, pp. 1–30.

[33] Y. Zhao, G. Che, and X. Zhao, "Adaptive noise removal for biological raman spectra with low snr," *Vibrational Spectroscopy*, vol. 123, p. 103441, 2022.

[34] Moester, M. J. B., Ariese, F., and de Boer, J. F., "Optimized signal-to-noise ratio with shot noise limited detection in stimulated raman scattering microscopy," *J. Eur. Opt. Soc.-Rapid Publ.*, vol. 10, p. 15022, 2015.

[35] R. Ranjan, G. Costa, M. A. Ferrara, M. Sansone, and L. Sirleto, "Noises investigations and image denoising in femtosecond stimulated raman scattering microscopy," *Journal of Biophotonics*, vol. 15, no. 6, p. e202100379, 2022.

[36] J. L. Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "Sdr – half-baked or well done?" in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 626–630.