analytical chemistry

pubs.acs.org/ac Article

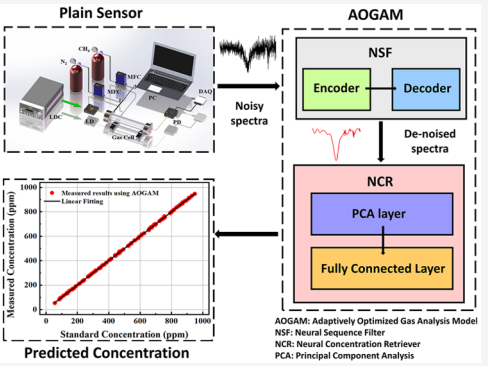
# Adaptively Optimized Gas Analysis Model with Deep Learning for Near-Infrared Methane Sensors

Jiachen Sun,[○] Linbo Tian,[○] Jun Chang,* Alexandre A. Kolomenskii, Hans A. Schuessler, Jinbao Xia, Chao Feng, and Sasa Zhang*

Cite This: *Anal. Chem.* 2022, 94, 2321−2332 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Noise significantly limits the accuracy and stability of retrieving gas concentration with the traditional direct absorption spectroscopy (DAS). Here, we developed an adaptively optimized gas analysis model (AOGAM) composed of a neural sequence filter (NSF) and a neural concentration retriever (NCR) based on deep learning algorithms for extraction of methane absorption information from the noisy transmission spectra and obtaining the corresponding concentrations from the denoised spectra. The model was trained on two data sets, including a computationally generated one and the experimental one. We have applied this model for retrieving methane concentration from its transmission spectra in the near-infrared (NIR) region. The NSF was implemented through an encoder−decoder structure enhanced by the attention mechanism, improving robustness under noisy conditions. Further, the NCR was employed based on a combination of a principal component analysis (PCA) layer, which focuses the algorithm on the most significant spectral components, and a fully connected layer for solving the nonlinear inversion problem of the determination of methane concentration from the denoised spectra without manual computation. Evaluation results show that the proposed NSF outperforms widely used digital filters as well as the state-of-the-art filtering algorithms, improving the signal-to-noise ratio by 7.3 dB, and the concentrations determined with the NCR are more accurate than those determined with the traditional DAS method. With the AOGAM enhancement, the optimized methane sensor features precision and stability in real-time measurements and achieves the minimum detectable column density of 1.40 ppm·m ($1\sigma$). The promising results of the present study demonstrate that the combination of deep learning and absorption spectroscopy provides a more effective, accurate, and stable solution for a gas monitoring system.






## 1. INTRODUCTION

Methane is an explosive and dangerous gas, which is often considered as target gas in applications of coal mine safety and natural gas pipeline leakage monitoring.[1−3] Meanwhile, methane is the second most important greenhouse gas, and its concentration has been increasing over the years with anthropogenic activities, whose subsequent effects such as climate change and temperature variability pose a threat to human survival.[4] Therefore, the test set's demand for methane sensors with high stability and sensitivity is gradually increasing. At present, methane sensors based on absorption spectroscopy are widely employed because of their high sensitivity, good selectivity, resistance to electromagnetic interference, long life, and contact-free measurements.[5−7]

Several high-sensitivity gas detection techniques based on absorption spectroscopy have become mature, such as cavity-enhanced absorption spectroscopy (CEAS),[8] cavity ring-down spectroscopy (CRDS),[9] Fourier transform infrared (FTIR) spectroscopy,[10] photoacoustic spectroscopy (PAS),[11] quartz-enhanced photoacoustic spectroscopy (QEPAS),[12−14] light-induced thermoelastic spectroscopy (LITES),[15] vernier frequency comb spectroscopies,[16] and so on. However, the stability and sensitivity of these techniques are greatly affected by factors such as temperature, humidity, vibration, and noise in harsh environments. In contrast, the tunable diode laser absorption spectroscopy (TDLAS)[17] technique has higher robustness to various natural factors. The TDLAS technique can be divided into direct absorption spectroscopy (DAS)[18] and wavelength modulation spectroscopy (WMS).[19] The WMS is often used in trace gas detection since by increasing the detection frequency it is possible to greatly reduce low-frequency noise ($1/f$ noise) and improve detection sensitivity. However, high cost, relatively complex system, and nonlinear detection at higher concentration levels are also its intrinsic problems. As a calibration-free detection technique, the DAS

**Received:** November 22, 2021
**Accepted:** January 5, 2022
**Published:** January 18, 2022

can directly obtain the concentration of absorbed gas by fitting the absorbance curves. The DAS is the most popular technique in methane detection applications with the advantages of easy miniaturization, low cost, and the ability of full-range concentration measurements.

The main factor restricting the sensitivity and stability of the DAS technique is the various noises contained in the sensor system, including thermal noise (Gaussian white noise), shot noise, interference fringe noise, flicker ($1/f$) noise, and so on.[20,21] At present, software-based filtering technology is extensively used because of its simple implementation and low cost. The filtering algorithms often used by researchers include Savitzky−Golay (SG) filtering algorithm,[22,23] Kalman filtering (KF) algorithm,[24,25] and the dual-optimized back-propagation adaptive Kalman filtering (BP-KF) algorithm.[26,27] In addition, in chemometrics, the orthogonal signal correction (OSC) and its variants are also used as preprocessing tools to remove systematic noise such as baseline variation and multiplicative scatter effects, in particular, the direct orthogonal signal correction (DOSC) is widely used.[28] The core idea of SG filtering is to filter the data in the window by weighting. However, it is particularly dependent on stable innovation or residual sequences. Thus, the variation of noise intensity in the data window will result in biased estimation.[23] The KF memorizes all past information, inferring the result by iteration, and is easily affected by time-varying noise, resulting in deviation of predictions.[29] Although BP-KF is designed to correct the erratic state values in nonlinear systems through neural networks, it is still challenging to evaluate the accurate noise covariance in dynamic detection, and the present model does not provide a good solution.[26] The basic idea of OSC is to filter out the information not related to the property $Y$ to be measured in the original spectral matrix $X$ by means of mathematical orthogonality, but the extremely low signal-to-noise ratio (SNR) or nonorthogonal noise will lead to degradation of the filtering effect. On the other hand, the capability of artificial neural networks to expressively represent complex data inspired us to readdress its application in solving filtering problems. Recurrent neural networks (RNNs), such as long short-term memory (LSTM)[30] and gated recurrent units (GRUs),[31] as deep learning models specialized in processing sequential data, have performed well in the fields of optical wireless communication[32,33] and optical sensing measurements[34,35] in recent years. Benefiting from the ability to automatically extract information from data through computational and statistical methods, deep learning is gradually becoming a promising research direction of data filtering in absorption spectroscopy.[36−38]

Another factor restricting the detection accuracy of DAS is the lack of an adequate absorption model. The current concentration retrieval method in absorption spectroscopy relies on employing the Voigt profile to fit the absorbance curve. However, it has been proven that only the Voigt profile fitting is still insufficient, and the Dicke narrowing and speed-dependent effects should also be considered.[39] In addition, more studies have demonstrated that taking the global spectrum into account can achieve more accurate concentration retrieval than referring only to the absorption peak.[18,40] Therefore, it is necessary to develop data-driven methods that more accurately derive the concentration. Deep learning enables computers to learn from experiences without actually modeling the physical rules that govern the spectral absorption process of gas molecules, which provides us a new solution for nonlinear inverse problems when relations between dependent and independent variables are not clear. Machine learning algorithms such as multivariate linear regression and back-propagation neural networks are applied in the earliest attempts to retrieve species concentrations and isotopic abundance,[40] while the extrapolation power for such simple models trained over small data sets remains doubtful. A one-dimensional convolutional neural network (1D-CNN) pre-trained through transfer learning is applied to establish an inversion model of transmitted signals to methane concentrations in the latest research.[41] Although a satisfactory concentration retrieval accuracy is obtained, the end-to-end model greatly reduces the interpretability of its intrinsic mechanisms and impedes subsequent improvements.

Based on the above considerations, we developed an adaptively optimized gas analysis model (AOGAM) composed of the neural sequence filter (NSF) and the neural concentration retriever (NCR). On the one hand, the proposed NSF is an encoder−decoder structured filter composed of double-layer bidirectional long short-term memory (LSTM) to mine the specific pattern between the noisy spectrum and the corresponding denoised spectrum. The encoder facilitates the automatic feature extraction from the original noise spectrum and converts it into an abstract vector according to specific rules. Subsequently, the attention mechanism in the decoder references to "attending to" certain parts of the input abstract vector for generating the context vector, which is predicted as the noise reduction spectrum by the LSTM in the decoder. We discussed the detailed structure of our NSF in Section 3.2.1. The proposed NSF outperforms other filters mentioned in this paper on both simulated and experimental test sets. In Section 4, we evaluate the performance of several filters in detail and analyze and discuss the reasons why the proposed neural sequence filter is better. On the other hand, we also design an NCR composed of a principal component analysis (PCA) layer based on singular value decomposition (SVD) and a fully connected (FC) layer to independently establish the mapping function between the transmission spectrum and concentration, which effectively solves the nonlinear ill-posed inverse problem and indirectly avoids the interference caused by manual concentration retrieval. In Section 4.3, we compare the concentration retrieval performance of the plain methane sensor (without noise filtering) with that obtained using AOGAM. This comparison shows that the implementation of the filtering with the NSF and the concentration retrieval by the NCR plug-in substantially improves the accuracy of the result. The conclusions are made in Section 5.

## 2. SENSOR CONFIGURATION

**2.1. Line Selection.** The choice of methane absorption line should ensure high sensitivity while also considering the cost since near-infrared (NIR) lasers are significantly less expensive than mid-infrared lasers. Consequently, we choose the R(3) absorption line near 1653.7 nm labeled by the blue box in Figure 1. The interference of water vapor and carbon dioxide in the air with the methane detection at this absorption line is evaluated in the Supporting Information. The simulation parameters in the illustration in Figure 1 and the experimental conditions described in this paper are as follows: temperature $T$ = 296 K, pressure $P$ = 1 atm, and optical path length $L$ = 130 cm.


https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

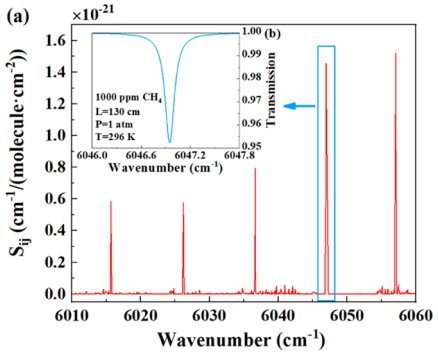



**Figure 1.** Selection of the absorption line of $CH_4$ in the spectral range of 6010−6060 $cm^{-1}$ based on the HITRAN database. Inset: transmission spectra of 1000 ppm $CH_4$ with $L$ = 130 cm, $P$ = 1 atm, and $T$ = 296 K.

**2.2. DAS Experimental Setup.** Figure 2 illustrates the DAS experimental setup. The distributed feedback laser diode

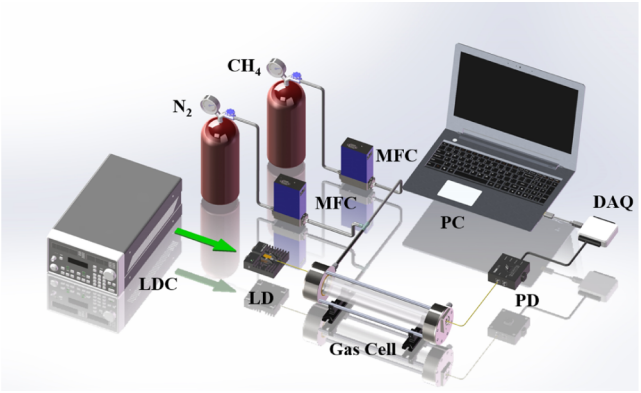



**Figure 2.** DAS experimental setup. MFC: mass flow controller; PC: personal computer; DAQ: data acquisition; PD: photodetector; DFB-LD: distributed feedback laser diode; LDC: laser diode controller.

(DFB-LD) (SWLD-165310S22-01, Allwave Devices Inc., China) with a center wavelength near 1653 nm was employed. The DFB-LD was installed on a 14-Pin Butterfly Laser Diode Mount and controlled with a laser diode controller (LDC) (LDC501, Stanford Research Systems Inc.). The driving signal is a 0.9 Hz sawtooth wave generated by the data acquisition (DAQ) card (PCI-6221, National Instruments Corporation). The output beam of the pigtailed butterfly DFB-LD was fiber-coupled to a multipass Herriott gas cell. The output light intensity from the Herriott gas cell was converted to a voltage signal with a photodetector (PD) module (PDB450C, Thorlabs) and collected with the DAQ card. Finally, the subsequent processing was completed using the personal computer (PC). The methane gas and high-purity nitrogen were mixed with different flow ratios to form 100 groups of mixed gases with different methane concentrations from 0 to 1000 ppm, and the errors were less than 2%. Among them, the precise control of the flow was completed by two mass flow controllers (MFCs, CS200A, Beijing Sevenstar Flow Co., Ltd., China). In more detail, the gas mixing and error evaluation of methane concentration can be found in the Supporting Information.

## 3. METHODOLOGY

**3.1. Data set.** Obtaining a large amount of data needed for neural network training by performing experiments is an extremely time-consuming and laborious task for DAS. Therefore, we generate the simulated set of methane transmission spectra at experimental conditions, which contains 1000 methane transmission spectra composed of 1111 sampling points with a concentration ranging from 0 to 1000 ppm. We have also generated the noisy transmission spectra, each composed of the pure transmission spectrum, Gaussian noise (with the 0 mean value and a variance of $9.71886 \times 10^{-5}$), and interference noise (simulated by an Airy function),[42] to achieve consistency with the real signal and thereby improve the generalization ability of the NSF. (The Supporting Information provides details on how the noisy transmission spectra data set was simulated.) We regard the generated noisy transmission spectra as the input and the denoised ones as the ground truth for the training of the NSF. At the same time, the concentration values corresponding to each transmission spectrum constitute a concentration set, which is used as the ground truth of the gas concentration for the training of the NCR.

In addition, we experimentally collected 100 methane transmission spectra with concentrations from 0 to 1000 ppm to form an experimental data set to evaluate the filtering performance of the designed filter in the practical setting. In addition, we also evaluated the performance enhancement of the conventional methane sensor complemented by the NSF and NCR.

With respect to the training strategy, we first split the simulated data set into the training set and the test set according to the proportion of 90 and 10%. We then use 10-fold cross-validation on the training set to divide the training set into a training set and a validation set for hyperparameter selection and architecture optimization, while the test set is used to verify the filtering performance and generalization of the final model. The more detailed model tuning strategy is discussed in the Supporting Information.

For a fair comparison, the proposed NSF and BP-KF are trained over the same simulation training set. Meanwhile, to keep the correctness of the implementation of the BP-KF, we no longer adjust it but use the original architecture form.[26] Similarly, the S-G filter is also optimized on the simulated training set to determine the optimal window size and the order of the fitting polynomial.[23] In addition, all filters are evaluated and verified on a simulated test set and experimental set.

**3.2. Model Architecture.** As a competitive neural network model, our AOGAM model includes the NSF of the encoder−decoder structure and the NCR composed of a PCA layer and a fully connected layer. The encoder first encodes the noisy transmission spectra into abstract vectors according to specific rules and redistributes the attention weights through the attention layer to generate the context vectors. Subsequently, for a given context vector, the decoder outputs the denoised spectra one element at a time. After that, the denoised spectra are transformed from high-dimensional spectral signals to a low-dimensional space through a PCA layer based on singular value decomposition (SVD) and converted to predicted concentrations through a single fully connected layer. We choose the LSTM rather than the GRU as the recurrent unit because the former performed better in our scenario. It is


https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

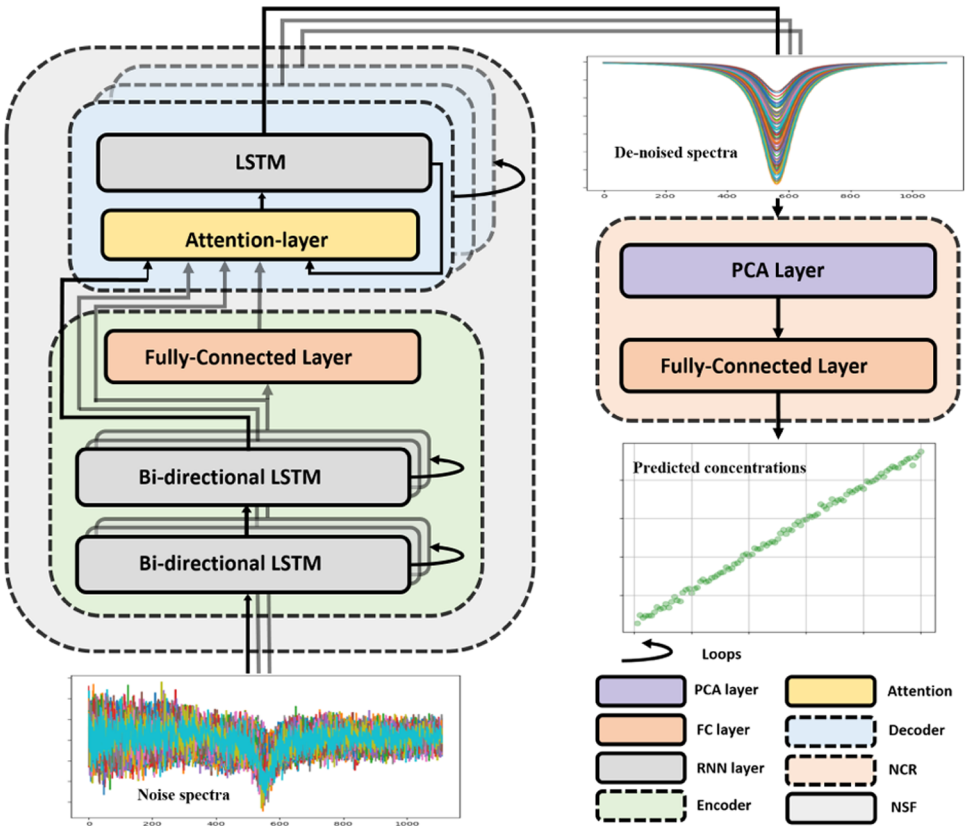



**Figure 3.** Complete model and the data processing flow chart. The noisy spectra are first filtered by the encoder−decoder structured NSF to generate the denoised spectra, and then the NCR predicts the corresponding concentration values. PCA: principal component analysis; FC: fully connected; RNN: recurrent neural network; NCR: neural concentration retriever; NSF: neural sequence filter.

generally considered that the GRU and LSTM have similar performance, and the internal structure of GRU is simpler, resulting in faster convergence. However, from the perspective of the model, LSTM with more complex parameters has better flexibility (the experimental results also prove that LSTM has more stable filtering performance than GRU) and can achieve a higher SNR than GRU. The LSTM unit computes the new memory content without any separate control of the amount of information flowing from the previous time step.[31] Therefore, the LSTM is more suitable for information extraction on the complex and changeable absorption spectra data streams. The complete model and the data processing flow chart are shown in Figure 3.

*3.2.1. Architecture of the Neural Sequence Filter. 3.2.1.1. Encoder.* In this paper, the encoder is designed to be stacked by a two-layered bidirectional LSTM and a fully connected layer.

Considering the noisy transmission spectra data set $x \in \mathbb{R}^{m\times d_x}$ that is composed of $m$ single spectra, $x^{(i)} = (x^{(i)\langle 1\rangle}, ..., x^{(i)\langle t\rangle}, ..., x^{(i)\langle dx\rangle})$, where $m$ is the number of spectra, $d_x$ is the number of the sampling points, $i$ denotes the index of the spectrum in the data set, and $t$ is the position index of a sampling point in the spectrum. The LSTM block receives the previous cell state $c^{\langle t-1\rangle} \in \mathbb{R}^{d_e\times 1}$, the previous hidden state $h^{\langle t-1\rangle} \in \mathbb{R}^{d_e\times 1}$, and the input $x^{(i)\langle t\rangle} \in \mathbb{R}^{1\times 1}$ at the current position. The LSTM block removes or adds information by carefully regulated structures called gates and follows the below equations:

$$f^{\langle t\rangle} = \sigma(W_f\cdot[h^{\langle t-1\rangle},\ x^{(i)\langle t\rangle}] + b_f) \quad (1)$$

$$i^{\langle t\rangle} = \sigma(W_i\cdot[h^{\langle t-1\rangle},\ x^{(i)\langle t\rangle}] + b_i) \quad (2)$$

$$\tilde{c}^{\langle t\rangle} = \tanh(W_c\cdot[h^{\langle t-1\rangle},\ x^{(i)\langle t\rangle}] + b_c) \quad (3)$$

$$c^{\langle t\rangle} = f^{\langle t\rangle} * c^{\langle t-1\rangle} + i^{\langle t\rangle} * \tilde{c}^{\langle t\rangle} \quad (4)$$

$$o^{\langle t\rangle} = \sigma(W_o\cdot[h^{\langle t-1\rangle},\ x^{(i)\langle t\rangle}] + b_o) \quad (5)$$

$$h^{\langle t\rangle} = o^{\langle t\rangle} * \tanh(c^{\langle t\rangle}) \quad (6)$$

where $f^{\langle t\rangle}$, $i^{\langle t\rangle}$, and $o^{\langle t\rangle}$ are the gates that respectively control the LSTM block to achieve the functions of forgetting, updating, and outputting. $W_f \in \mathbb{R}^{d_e\times d_x}$, $W_i \in \mathbb{R}^{d_e\times d_x}$, $W_C \in \mathbb{R}^{d_e\times d_x}$, and $W_o \in \mathbb{R}^{d_e\times d_x}$ are the trainable parameters in the LSMT block; $b_f \in \mathbb{R}^{d_e\times 1}$, $b_i \in \mathbb{R}^{d_e\times 1}$, $b_C \in \mathbb{R}^{d_e\times 1}$ and $b_o \in \mathbb{R}^{d_e\times 1}$ are the biases; $\sigma$ and tanh stand for the sigmoid and the tanh activation function, respectively; $\tilde{c}^{\langle t\rangle}$ denotes the candidate vector; · stands for the dot product; and * represents the element-wise multiplication. In our case, the number of the spectra $m = 1000$, the dimension of the hidden state in encoder LSTM $d_e = 256$, and the dimension of the input spectrum $d_x = 1111$.

Recall that the two-layered bidirectional LSTM outputs hidden state in two directions at each layer (as shown in Figure S2), which we denote $\overrightarrow{h}_l^{(i)\langle t\rangle}$ and $\overleftarrow{h}_l^{(i)\langle t\rangle}$ to indicate the hidden states from two directions by $l$th layer at position $t$,


https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

respectively. Therefore, the final output of the two-layered bidirectional LSTM is

$$h^{(i)\langle t\rangle} = \text{concatenate}(h_2{}^{(i)\langle t\rangle},\ h_2{}^{(i)\langle t\rangle}) \tag{7}$$

The purpose of the fully connected layer is to mix the output of the last position of each layer and generate a temporary vector $h_{\text{Tem}}^{(i)} \in \mathbb{R}^{d_d\times 1}$ whose dimension is compressed from $d_e$ to the dimension of the hidden state in the decoder

$$h_{\text{Tem}}^{(i)} = \tanh(h^{(i)\langle d_x\rangle}\cdot W_d + b_d) \tag{8}$$

where $W_d \in \mathbb{R}^{(d_x\cdot n_l\cdot n_d)\times d_d}$ denotes the trainable parameters and $b_d \in \mathbb{R}^{d_d\times 1}$ stands for the bias in the fully connected layer; $n_l$ and $n_d$ represent the number of layers and the number of directions, respectively, and $d_d = 512$ indicates the dimension of the hidden state in the decoder LSTM. The $h_{\text{Tem}}^{(i)}$ will be fed into the decoder along with the two-layered bidirectional LSTM outputs $h^{(i)} = (h^{(i)\langle 1\rangle}, ..., h^{(i)\langle d_x\rangle})$ as the output of the encoder. Note that the fully connected layer also receives the cell state outputted from the two-layered bidirectional LSTM to generate $c_{\text{Tem}}^{(i)} \in \mathbb{R}^{d_d\times 1}$, which will be considered as the initial state values in the decoder LSTM block along with the $h_{\text{Tem}}^{(i)}$.

*3.2.1.2. Decoder.* The decoder is composed of a single-layer unidirectional LSTM and an attention layer. Compared with directly decoding the encoded abstract vector, we found that adding an attention mechanism can enable the decoder to search clues at other positions of the input vector when predicting the specific position of the denoised spectrum so as to achieve better decoding. The schematic diagram of the decoder is shown in Figure S3.

With respect to the position $t'$ of the denoised spectrum predicted by the decoder, the attention layer receives the previous hidden state $\hbar^{(i)\langle t'-1\rangle} \in \mathbb{R}^{d_d\times 1}$ of the decoder LSTM (the initial state values of decoder LSTM are $h_{\text{Tem}}^{(i)}$ and $c_{\text{Tem}}^{(i)}$) and the outputs $h^{(i)} = (h^{(i)\langle 1\rangle}, ..., h^{(i)\langle d_x\rangle})$ of the encoder. The energy value $e^{(i)\langle t'-1,\ t\rangle} \in \mathbb{R}^{d_a\times 1}$ contributed by the output at position $t$ of the encoder $h^{(i)\langle t\rangle}$ to the predicted position $t'$ of the decoder can be expressed as

$$e^{(i)\langle t'-1,\ t\rangle} = \tanh(\text{concatenate}(\hbar^{(i)\langle t'-1\rangle},\ h^{(i)\langle t\rangle})\cdot W_e + b_e) \tag{9}$$

The Softmax layer then converts the energy values into the corresponding attention weights

$$\alpha^{(i)\langle t'-1,\ t\rangle} = \frac{\exp(e^{(i)\langle t'-1,\ t\rangle}\cdot W_\alpha{}^{\langle t\rangle})}{\sum_t^{d_x}\exp(e^{(i)\langle t'-1,\ t\rangle}\cdot W_\alpha{}^{\langle t\rangle})} \tag{10}$$

In the above equations, $W_e \in \mathbb{R}^{(2\cdot d_d+2d_e)\times d_a}$ and $b_e \in \mathbb{R}^{d_a\times 1}$ are the trainable weights and bias in the energy layer, respectively. In our case, $d_a = 128$ and $W_\alpha$ are the trainable parameters in the Softmax layer. The attention weights tell the decoder how much importance it should attach to each position from the noisy spectrum. We will see in Section 4.2 that for a specific position in the noisy spectrum, the decoder will refer to different parts of the whole spectrum to varying extents when denoising.

The attention layer outputs the context vector by weighted sum and the outputs of the encoder

$$\text{context}^{\langle t'\rangle} = \sum_t^{d_x}\alpha^{(i)\langle t'-1,\ t\rangle}\cdot h^{(i)\langle t\rangle} \tag{11}$$

The decoder LSTM iteratively accepts the context vector and predicts each spectral component for the complete denoised spectrum $\hat{x}^{(i)} = (\hat{x}^{(i)\langle 1\rangle}, ..., \hat{x}^{(i)\langle d_x\rangle})$.

*3.2.2. Neural Concentration Retriever.* The PCA layer compresses the dimension of the original spectrum while preserving the information to the greatest extent, thereby greatly reducing the functional complexity of mapping the denoised spectrum to its corresponding concentration. In addition, compared with employing only the Voigt line shape to fit the absorbance curve and using absorption peak to retrieve concentration, we expect that the neural network can take the whole dimensionally reduced spectrum into account to deduce more accurate concentration retrieval results. Indeed, considering only the absorption peak inevitably introduces deviations of the result due to peak value fluctuations.

The denoised spectrum output by the NSF first passes through a PCA layer based on the SVD. The SVD of $\hat{x} = \{\hat{x}^{(1)}, ...\ \hat{x}^{(i)}, ..., \hat{x}^{(m)}\} \in \mathbb{R}^{m\times d_x}$ is

$$\text{SVD}(\hat{x}_{m,\ d_x}) = U_{m,m}{}^*\Sigma_{m,d_x}{}^*V^{\text{T}}_{d_x,d_x} \tag{12}$$

where $U_{m,m}$ is the left singular value matrix, $\sum_{m,d_x}$ is a diagonal matrix with diagonal elements having singular values, and $V^{\text{T}}_{d_x,d_x}$ is the right singular value matrix.

Then, the predicted denoised spectrum is approximated by

$$\tilde{x}_{m,\ d_x} \approx U_{m,k}{}^*\Sigma_{k,k}{}^*V^{\text{T}}_{k,d_x} \tag{13}$$

where $k$ is the dimension after the reduction. The first $k$ rows of $V^{\text{T}}$ are used as the dimension reduction spectrum, and the new bases of the reduced dimension transformed space are expressed as

$$d\hat{x}_{m,k} = \hat{x}_{m,\ d_x}{}^*V_{d_{x,}\ k} \tag{14}$$

The prediction of concentration is given by the fully connected layer

$$\hat{y} = ReLU(d\hat{x}\cdot W_r + b_r) \tag{15}$$

where $W_r \in \mathbb{R}^{k\times 1}$ is the trainable parameter and $b_r$ is the bias. We will see the influence of different $k$ values on spectral information in Section 4.3.

*3.2.3. Loss Function.* As an integral, the AOGAM is optimized on the elastic mean squared error (MSE) loss function through the learning algorithm (Adam)

$$\begin{aligned}\mathcal{L} &= L_1 + L_2\\ &= \alpha^*\frac{1}{m}\sum_{i=0}^{m}(\hat{x}^{(i)} - x^{(i)})^2 + \beta^*\frac{1}{m}\sum_{i=0}^{m}(\hat{y}^{(i)}, y^{(i)})^2\end{aligned} \tag{16}$$

where $x^{(i)}$ and $y^{(i)}$ are the $i$th denoised spectrum and the corresponding methane concentration, respectively, and $\alpha$ and $\beta$ are the elasticity factor. The $L_1$ and $L_2$ denote subloss functions of filtering regression and concentration regression, respectively. The sum of the elasticity factors is 1, which ensures that the model is trained on two tasks simultaneously and helps to adjust the trade-off between the filtering and concentration inversion. As a result, fast and accurate filtering and concentration retrieval performance are achieved.


https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

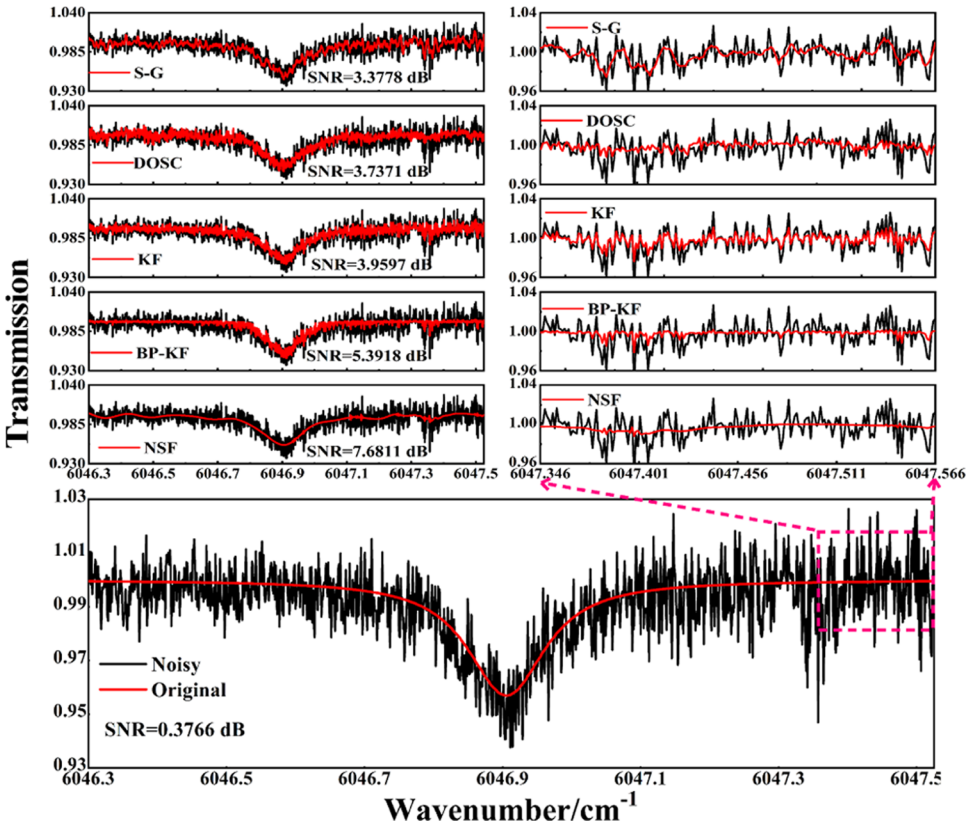



**Figure 4.** Effect of several filtering algorithms (for the background comprised of Gaussian white and interference noises).

## 4. EXPERIMENTAL RESULTS AND ANALYSIS

**4.1. Filtering Performance Comparison and Analysis of Simulated Data.** As an important part of the proposed model, the filtering performance of the NSF is compared with several common filters on the simulated data set. According to ref 23, we optimized the window length of the S-G filter and the order of the fitting polynomial. To ensure the accuracy of implementation, we established the state space equation of the Kalman filter based on the methane transmission spectrum and the model architecture based on the back-propagation neural network according to ref 27. The covariance matrices of the process noise and the measurement noise were obtained by our simulation and from the experimental transmission spectra. The DOSC was chosen as the representative of the OSC variants and trained on the same data set. The number of components regarding the DOSC has been carefully selected. Although it was suggested[28] that one or two components were sufficient, we found in practice that a small number of components could hardly filter out noise on our data set. When this number reaches 896, the performance is the best, and the filtering performance starts to decline when it is less than or more than this optimal number.

Since our model is trained on the simulation data set, we first evaluate and compare the filtering performance of several algorithms on the simulated test set. We refer to the simulated methane transmission spectrum of 200 ppm as the representative sample to intuitively show the effects of different filtering algorithms, as shown in Figure 4. Compared with the original signal (we refer to the simulated pure signal with no noise as the original signal), the SNR of the denoised signal filtered by the S-G filter is the lowest among all of the filters. The DOSC can only remove the noise projection in the direction that is orthogonal to the concentration matrix, but there are still some noise residues. When the background noise of our spectral matrix is too large, the spectral loading matrices corresponding to the first several main components usually are not the concentration matrix information but the spectral signals that are independent of the concentration matrix. When the spectral matrix filters out this part of the information, the useful absorption information will be distorted. In addition, the noise in the spectra is not completely white noise that is orthogonal to the concentration matrix. Therefore, even if the orthogonal components are completely filtered by DOSC, there will still be a large amount of noise residue. Although the state-of-the-art BP-KF significantly improves the filtering effect in the nonabsorption region compared with the conventional KF while still preserving the essential KF results, still a lot of noise residues is observed in the absorption area, resulting in the variation of the absorption peak, which will undoubtedly have a significant negative impact on the stability and accuracy of the final concentration determination. The reason for the low performance of KF is that the established space state equations cannot accurately represent the physical model corresponding to the application scenario. In addition, the KF "remembers" all of the previous information by iteration, which will result in biased estimation due to time-varying noise. Therefore, the performance of the KF suffers from the estimation gap of noise parameter between the preset noise and that of the real measurement. On the other hand, although the BP-KF is designed to optimize the state parameters to compensate for the deficiency of KF in the nonlinear system, its generalization ability is insufficient in different systems, which has been proven by its good denoising performance in the nonabsorption area (stable area) while rendering the weak


https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

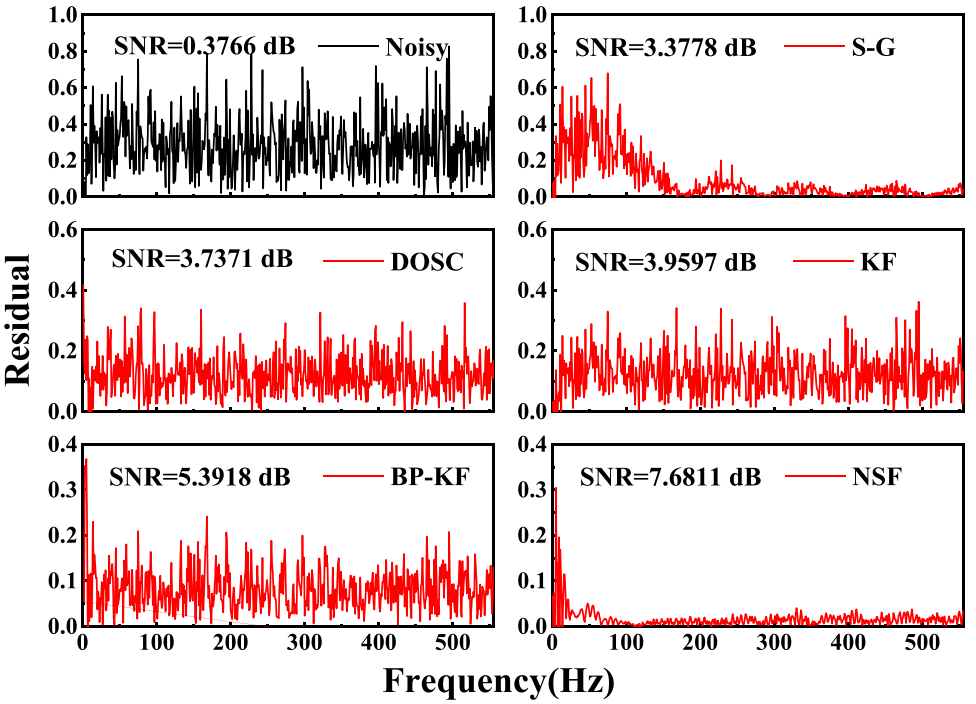



**Figure 5.** Frequency domain distributions of the original transmission spectra and filtered transmission spectra.

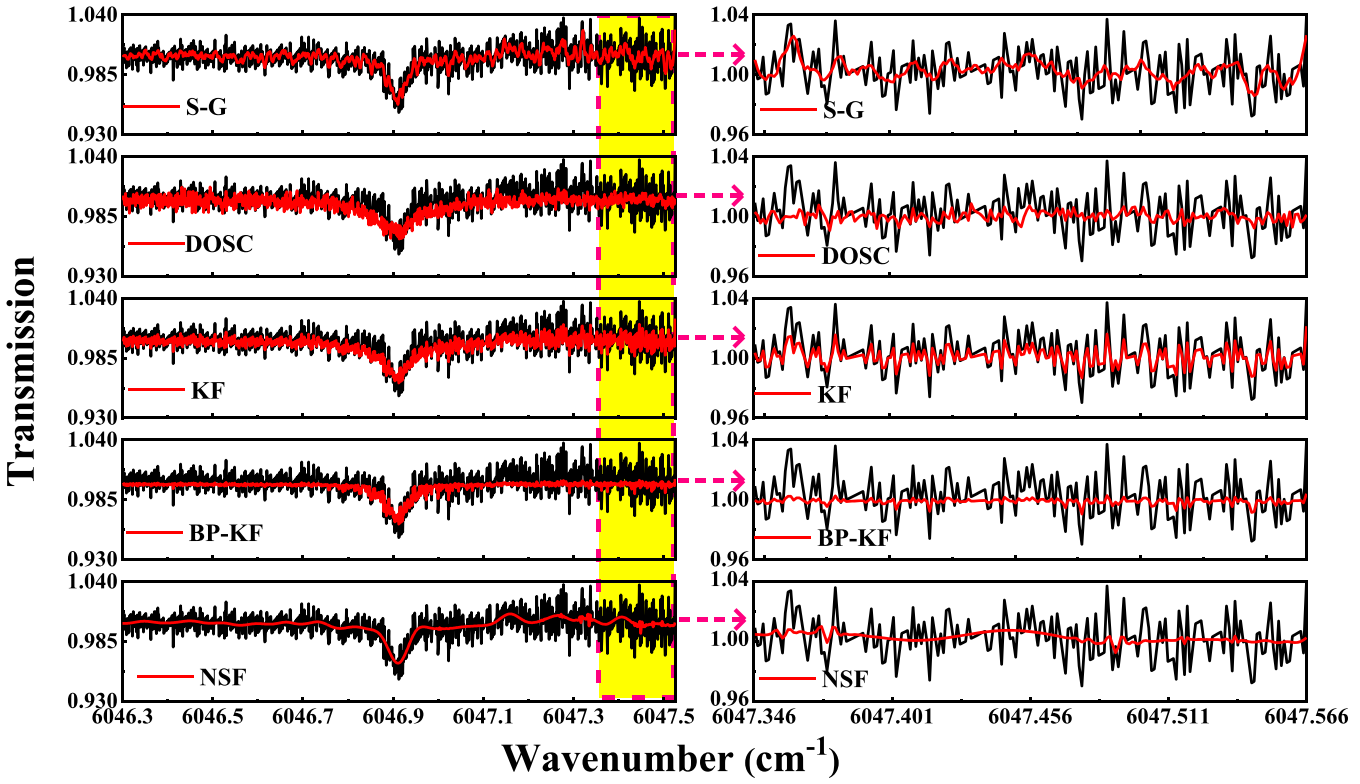



**Figure 6.** Effects of various filtering algorithms on processing experimental data.

performance in the absorption area (area with drastic fluctuation). In contrast, the NSF proposed in this paper shows significant smoothing and denoising effects in both absorption and nonabsorption regions. Unlike the KF-based algorithms, the NSF combines the information from both directions as the reference rather than predicting a specific component only through the previous information; therefore, its prediction of the original signal waveform is the closest to the original pure signal. When predicting a specific point, the attention mechanism achieves the most reasonable prediction by distributing the attention weights of the whole spectrum. Specifically, the SNR of the denoised spectrum obtained with the NSF is the highest compared to other filtering algorithms, reaching an improvement by 7.3045 dB compared to the



https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

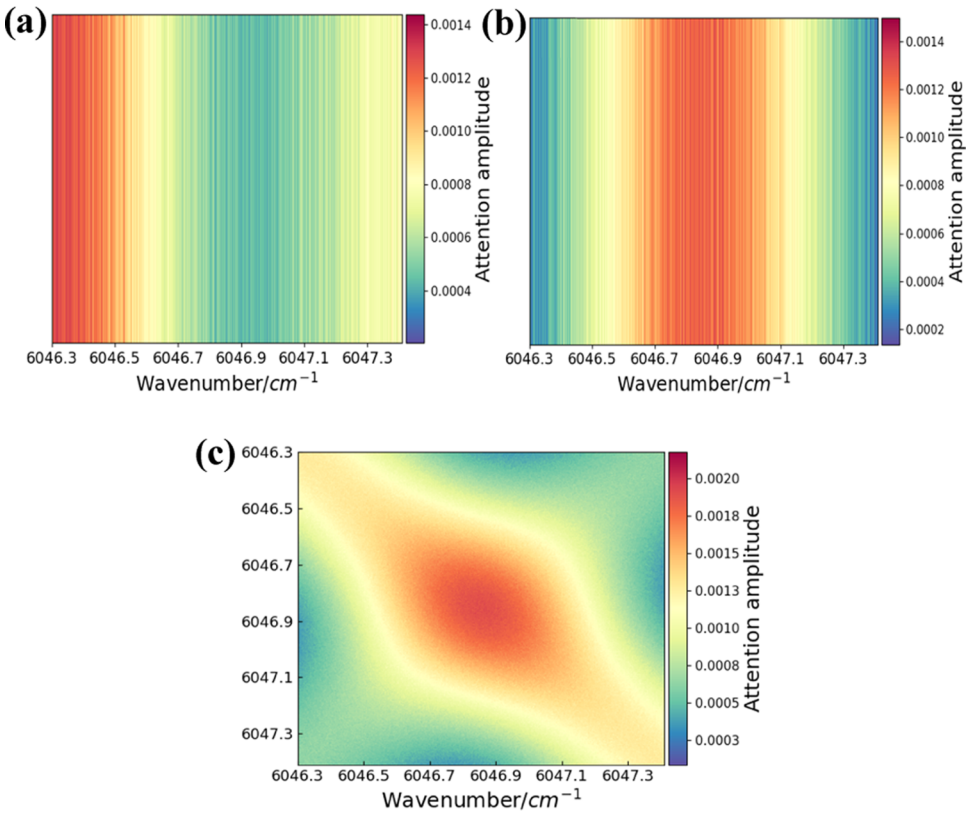



**Figure 7.** Attention maps for different positions of the filtered spectrum. The attention maps are generated through the attention layer. (a) Attention map for the position in the nonabsorption region with a wavenumber of 6046.37 cm$^{-1}$. (b) Attention map for the position in the nonabsorption region with a wavenumber of 6046.89 cm$^{-1}$. (c) Global attention map of all positions of the spectrum with respect to the particular positions.

original noisy signal (from 0.3766 to 7.6811 dB). The obtained SNR differences are statistically important ($P < 0.005$), and thus the NSF filter significantly outperforms other filters.

Furthermore, we also characterize the filtering performance from the perspective of the frequency domain. The distribution of frequency residuals after filtering by different filters obtained by fast Fourier transform is shown in Figure 5.

The filtering result of KF shows that it only somewhat suppresses the noise amplitude of the full-frequency band compared with the noisy spectrum. In contrast, DOSC, S-G, and the BP-KF are more focused on filtering high-frequency components while also suppressing the noise in the full band, which is consistent with the smoothing effect of the nonabsorption region in Figure 4. However, the BP-KF has large residuals at low-frequency regions, which is also consistent with the phenomenon that a large amount of noise remains in the absorption region. On the other hand, the NSF maintains the lowest residual even in the low-frequency region, which is the key to filtering out noise and retaining absorption information.

**4.2. Comparison of the Filtering Performance and Analysis of Experimental Data.** In addition to the simulated test set, we also apply the filters to the experimental data set to evaluate the generalization ability of the filters under practical application conditions. From the collected 100 experimental spectra, we selected the methane transmission spectrum with the preset concentration of 200 ppm as an example. The effects of the different filters are shown in Figure 6. As expected, the NSF still shows the best filtering performance even though it is trained over simulated data. The S-G filter still performs worst, followed by the DOSC, KF, and the BP-KF. The spectral frequency response after DOSC filtering is similar to that of S-G, and the filtering performance is very limited, which is mainly because the noise condition is complicated and is only partially orthogonal to the target property. The BP-KF accepts single input at each position and predicts the corresponding filtered output, which is only affected by all previous states. In contrast, our NSF takes the whole spectrum into account, which means it comprehensively evaluates the impact of other points based on the learned noise-related pattern with the help of an attention mechanism to predict a specific position of the spectrum. It still outperforms all other listed filters in processing the experimental data, and Wilcoxon signed-rank tests are statistically significant at $P < 0.005$ for all test data, which proves the feasibility and generalization power of the model in the practical application scenario. The NSF algorithm perceives the change during the transition from the nonabsorption to the absorption region, whereas the transition is not smooth under BP-KF processing.

Additionally, the attention mechanism and the encoder−decoder structure help to overcome the dilemma of training on the super-long sequence. Even though LSTM reduces the effects of vanishing and exploding gradients for very long sequences, it does not eliminate them entirely. While LSTM has difficulty processing highly complex feature representations to generate accurate outputs, the attention mechanism counters this problem by finding the relevance between the output data it is generating and the encoder's hidden state representations. The attention structure creates a context


https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

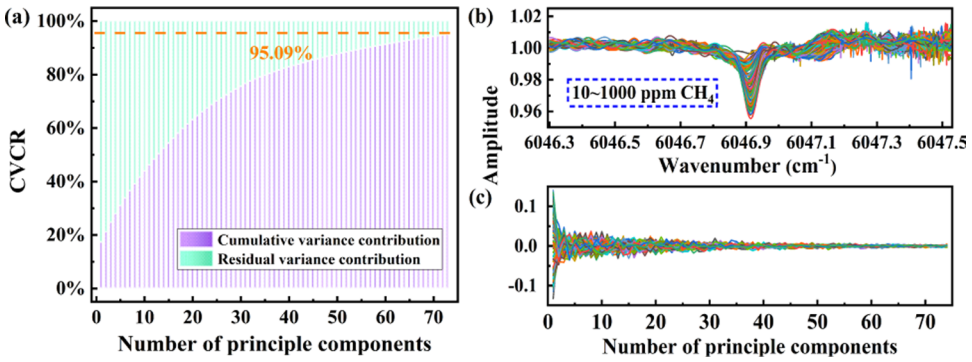



**Figure 8.** (a) Relationship between cumulative variance contribution and the number of principal components. (b) One hundred experimental spectra filtered by the NSF. (c) Dimension-reduced experimental spectra compressed by the PCA layer.

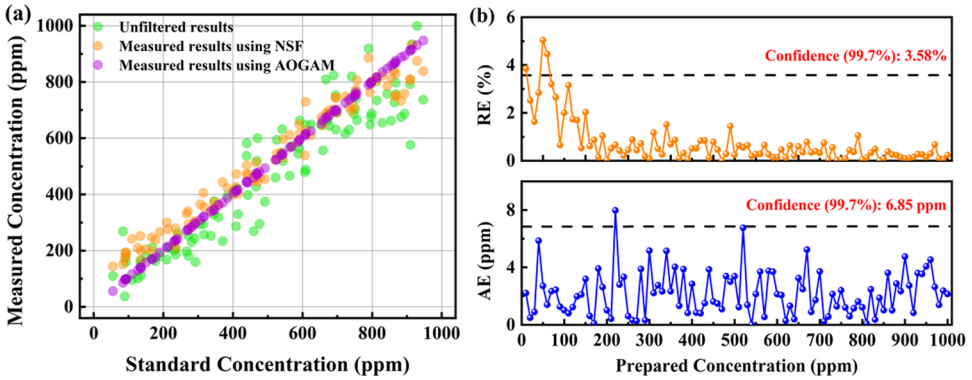



**Figure 9.** (a) Comparison of coefficients of determination for results measured from the plain methane sensor, filtered by the NSF and AOGAM. (b) Error analysis of the measured results using AOGAM: absolute error (AE) and relative error (RE) vs concentration.

vector from encoder hidden states. It then allows the decoder to use the context vector to generate a more precise and relevant output.

By observing the attention map, we can intuitively understand the fine-grained information flow between the input noisy spectrum and the output denoised spectrum. Figure 7 shows the attention maps of the NSF in predicting different positions of the denoised spectrum. For each position of the spectrum, the attention mechanism helps the NSF to refer different positions to various extents when filtering different regions of the noisy spectrum, as shown in Figure 7c. For instance, with respect to filtering the position with a wavenumber of 6046.37 $cm^{-1}$ in the nonabsorption region, the NSF considers other positions in the nonabsorption region and ignores the absorption region, as shown in Figure 7a. In contrast, the NSF is more attending to the spectrum near the absorption parts when generating the output of the position with a wavenumber of 6046.89 $cm^{-1}$, as shown in Figure 7b. Assigning different attention weights to various locations makes the NSF better adapt to changing and complex application scenes.

**4.3. Assessment of Concentration Retriever Performance.** In addition to analyzing the filtering effect, we also evaluate and discuss the concentration retrieval performance of the NCR, which is another key part of AOGAM.

Referring to the components of the spectrum as dimensions, each spectrum is represented by a high-dimensional vector composed of 1111 dimensions, which leads to hardly manageable calculations. Therefore, by adding a layer composed of PCA operation, the dimension of the denoised spectrum that is filtered by the NSF is reduced while maximizing the retention of spectral information. The high-dimensional space of the original dimension $d_x$ is compressed, and the original spectrum is projected to the new low-dimensional subspace composed of $k$ sets of mutually orthogonal eigenvectors, in which the cumulative variance contributions of first $k$ singular values corresponding to the orthogonal basis $\lambda$ follow the formula

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{d}\lambda_i} \geq t \tag{17}$$

As shown in Figure 8a, the cumulative variance contribution rate (CVCR) of the first 74 principal components has reached 95%, and thus, it is reasonable to represent the spectrum with the original 1111 dimensions by the new vector of 74 dimensions. The filtered 100 experimental spectra are shown in Figure 8b, and the new vectors (displayed in two-dimensional space) after dimensionality reduction through the PCA layer are shown in Figure 8c.

We used the concentrations preset by two MFCs as the standard concentration to evaluate the measurement accuracy of the methane sensor. We compared three cases of concentration retrieval results: (1) the measured concentration obtained from the original experimental transmission spectra without any processing, (2) the measured concentration obtained from the transmission spectra filtered by the NSF, and (3) the concentration results predicted by the complete AOGAM. The concentration retrieval layer is composed of two fully connected layers. It receives the experimental data from the dimensionality-reducing PCA layer and outputs the prediction of methane concentration.


https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

The results demonstrate the benefits of introducing the NCR. As seen in Figure 9a, the impact of different processing stages on the experimental results is clear: the methane sensor with the enhancement of AOGAM achieved the highest coefficient of determination ($R^2$ = 0.99989) and is followed by the methane sensor assisted only by the NSF ($R^2$ = 0.95122), while the methane sensor without any optimizing process was the most unreliable ($R^2$ = 0.85738); the differences are statistically significant at $P < 0.005$. Unlike the AOGAM, the other two methods must convert the measured transmission spectra into absorbance spectra, then subsequently perform the Voigt profile fitting on the absorbance spectra, and finally calculate the concentration corresponding to the peak, which will inevitably introduce interferential errors.[18,40] The experimental results demonstrate that the method of extracting only the maximum absorbance value brings greater fluctuation and unreliability to the measured methane concentrations, especially in the dynamic measurements with time-varying noise. In contrast, the NCR that extracts information from the whole spectrum (dimensionally reduced spectrum) constructs afresh the mapping function between the full spectrum and its corresponding concentration, thus avoiding distortions introduced by processing with prior knowledge. In addition, the PCA layer greatly reduces the complexity of the NCR and the difficulty of concentration determination, furnishing also higher generalization ability and robustness to complex environments. Table 1 (column 1) summarizes the coefficients of determination for the three cases.

**Table 1. Comparison for $R^2$, RE, and AE among Three Situations**

| | $R^2$ | RE (%) (mean) | AE (ppm) (mean) |
|---|---|---|---|
| plain sensor | 0.85738 | 19.25 | 75.84 |
| NSF | 0.95122 | 17.40 | 43.87 |
| AOGAM | 0.99989 | 0.73 | 2.19 |

Meanwhile, we also evaluate the precision of the concentration determination in three cases. We conducted a systematic error analysis by comparing the two statistical indicators of relative error (RE) and absolute error (AE), as shown in Figure 9b. The AE and the RE of the methane sensor assisted by AOGAM are under 6.85 ppm and 3.58% with a confidence of 99.7%, respectively. Although the precision deteriorates when the concentration is extremely low, the RE remains below 5%, and the overall performance can meet the needs of most practical applications. Table 1 (the second column and the third column) shows the mean error analysis results for the three cases. The performance of the AOGAM-assisted methane sensor has been significantly improved compared with the other two cases.

To assess the long-term detection stability of the optimized methane sensor, the custom methane gas with a concentration of 460 ppm is injected into the gas cell, and real-time detection is carried out for 1000 s. The optimized methane sensor shows good stability. The standard deviation is 1.08 ppm, which is significantly better than the performance of the plain methane sensor (the standard deviation is 11.11 ppm), as shown in Figure 10a. At the same time, the statistical histogram can be well fitted by a Gaussian distribution with the half width at half-maximum (HWHM) of 1.31 ppm, as shown in Figure 10b. The results show that the proposed optimization model can effectively improve the stability of the methane sensor in real-time detection. Since the minimum detection limit of the AOGAM-assisted methane sensor is 1.08 ppm ($1\sigma$) at room temperature and the optical path length is about 1.3 m, a minimum detectable column density of 1.08 ppm × 1.3 m = 1.40 ppm·m is achieved.

In Table 2, we compared our results with three other near-infrared methane sensors that have been reported and employed the R(3) absorption line near 1653.7 nm. We indicate the detectable column density, which is the most important evaluation parameter, and it can be clearly seen that the AOGAM-enhanced methane sensor proposed in this paper has the lowest detectable column density. The bold value in Table 2 represents the optimal detectable column density. Moreover, our sensor employing the AOGAM shows ultrahigh filtering as well as concentration retrieval performance, thus improving the DAS technique and realizing the full-range concentration detection at a lower cost and simpler system

**Table 2. Comparison of Our Sensor with the Three Other Reported Sensors**

| technique | detectable range | optical path length | detectable column density (ppm·m) | refs |
|---|---|---|---|---|
| WMS | (0−5) × $10^4$ ppm | 40 cm | ∼11.8 | 43 |
| DAS + WMS | full range | 10 cm | ∼6.48 | 44 |
| WMS | unavailable | 52.2 m | ∼5.22 | 45 |
| DAS | full range | 130 cm | **∼1.40** | this paper |

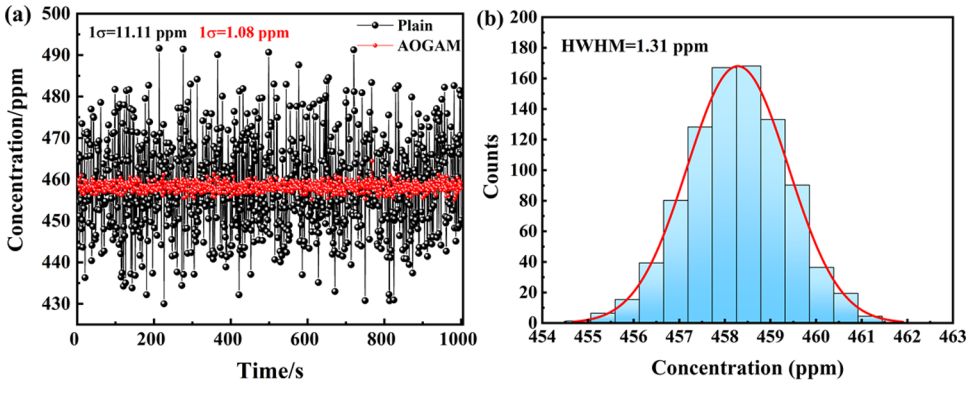



**Figure 10.** (a) Long-term measurement and (b) statistical histogram of methane concentration.



https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

compared with the three other sensors employing the WMS technique.

In addition, to make the conclusion of our AOGAM even more justified, we also made a special trip to a coal mine in Yan'an city, Shaanxi Province, China, for 1-day underground in situ measurement. We compared the methane concentration detection results of the plain methane sensor and that of the AOGAM-enhanced methane sensor. The methane concentrations from the detection with the AOGAM-enhanced methane sensor showed more stability and less variance ($P < 0.001$) than those of the plain sensor, as shown in Figure 11.

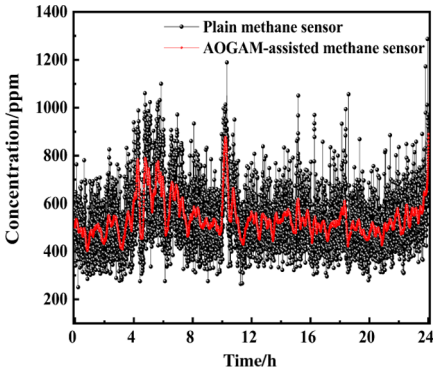



**Figure 11.** In situ measurement results of the plain methane sensor and the AOGAM-assisted methane sensor.

## 5. CONCLUSIONS

In this study, we developed an adaptively optimized gas analysis model (AOGAM) for the determination of the gas concentration from spectral measurements. This model is composed of a neural sequence filter (NSF) and a neural concentration retriever (NCR) based on the deep learning algorithms for extraction of methane absorption information from the noisy transmission spectra and characterization of the corresponding concentrations from the denoised spectra, respectively. The proposed NSF outperforms other widely used filters on both simulated and experimental test sets and shows high robustness in noisy conditions. We explicate the underlying mechanism of the global filtering of the NSF through the attention map generated during the filtering process, which increases the interpretability of our NSF. Furthermore, the NCR provides a new solution for the nonlinear inversion problem of the determination of methane concentration from the spectra by automatically mapping the accurate corresponding concentrations based on the denoised spectra without manual computation. The evaluation results demonstrate that the AOGAM-assisted methane sensor has improved characteristics in terms of the accuracy of gas concentration determination and stability of real-time measurements and achieved a detection column density of 1.40 ppm·m, which is the lowest among various NIR methane sensors reported in the literature. The proposed optimized methane sensor is suitable for the realistic conditions of different applications, such as for the needs of coal mines and chemical industry gas detection and other in situ measurement scenarios.

At the same time, we note that although AOGAM can significantly improve the concentration retrieval performance of a plain methane sensor, it is difficult to gain further progress for methane sensor parameters based solely on optimization algorithms because of inherently low absorption cross sections and the overtone nature of the absorption bands in the NIR spectral region. In contrast, the molecular fingerprint absorption in the mid-infrared region corresponds to the fundamental absorption bands of most gases, which have orders of magnitude higher absorption cross sections, making it easier to reach lower detection limits. Extending our technique to the mid-infrared will be our next research goal. In addition, current gas sensing based on absorption spectroscopy still mainly focuses on single components and has difficulty dealing with the multicomponent gas mixtures with overlapping absorption spectra. In this respect, it is important to fully investigate the possibilities of the deep learning approach as applied to the simultaneous detection of gas mixtures. Therefore, we will continue working in this direction in our future research.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.1c05059.

Evaluation and analysis of cross-interference; introduction to the process of generating simulated data set; error analysis of gas concentration configuration with MFC; model tuning strategy; and full codes implementation for all algorithms (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

**Jun Chang** − *School of Information Science and Engineering, Shandong University, Qingdao 266237, China; Shandong Provincial Key Laboratory of Laser Technology and Application and Key Laboratory of Education Ministry for Laser and Infrared System Integration Technology, Shandong University, Qingdao 266237, China;* orcid.org/0000-0003-1318-2352; Email: changjun@sdu.edu.cn

**Sasa Zhang** − *School of Information Science and Engineering, Shandong University, Qingdao 266237, China; Shandong Provincial Key Laboratory of Laser Technology and Application and Key Laboratory of Education Ministry for Laser and Infrared System Integration Technology, Shandong University, Qingdao 266237, China*; Email: sasazhang@sdu.edu.cn

### Authors

**Jiachen Sun** − *School of Information Science and Engineering, Shandong University, Qingdao 266237, China; Shandong Provincial Key Laboratory of Laser Technology and Application, Shandong University, Qingdao 266237, China*

**Linbo Tian** − *Shandong Provincial Key Laboratory of Laser Technology and Application and Key Laboratory of Education Ministry for Laser and Infrared System Integration Technology, Shandong University, Qingdao 266237, China*

**Alexandre A. Kolomenskii** − *Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, United States*

**Hans A. Schuessler** − *Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, United States*

**Jinbao Xia** − *State Key Laboratory of Crystal Materials, Shandong University, Jinan 250100, China; The State Key Laboratory of Applied Optics, Changchun 130000, China*




https://doi.org/10.1021/acs.analchem.1c05059
Anal. Chem. 2022, 94, 2321−2332

**Chao Feng** − *Center for Optics Research and Engineering, Shandong University, Qingdao 266237, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.1c05059

### Author Contributions

○J.S. and L.T. contributed equally to this work; J.S.: conceptualization, investigation, writing—original draft, methodology, and software; L.T.: conceptualization, investigation, writing—original draft, methodology, and software; J.C.: supervision, funding acquisition, and project administration; A.A.K.: writing—review and editing; H.A.S.: funding acquisition and writing—review and editing; J.X.: writing—review and editing and funding acquisition; C.F.: writing—review and editing and software; and S.Z.: supervision, funding acquisition, and project administration.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (61475085), the Key Research and Development Program of Shandong Province (no. 2020CXGC010104), Qingdao science and technology demonstration and guidance project (no. 21-1-4-sf-1-nsh), Open Fund of State Key Laboratory of Applied Optics (SKLA02020001A12), and the Robert A. Welch Foundation (Grant no. A1546).

## ■ REFERENCES

(1) Wittstock, V.; Scholz, L.; Bierer, B.; Perez, A. O.; Wöllenstein, J.; Palzer, S. *Sens. Actuators, B* **2017**, *247*, 930−939.
(2) Shi, L.; Wang, J.; Zhang, G.; Cheng, X.; Zhao, X. *Process Saf. Environ. Prot.* **2017**, *107*, 317−333.
(3) Iwaszenko, S.; Kalisz, P.; Słota, M.; Rudzki, A. *Remote Sens.* **2021**, *13*, No. 510.
(4) Cui, X.; Dong, F.; Zhang, Z.; Sun, P.; Xia, H.; Fertein, E.; Chen, W. *Atmos. Environ.* **2018**, *189*, 125−132.
(5) Dang, J.; Kong, L.; Zheng, C.; Wang, Y.; Sun, Y.; Yu, H. *Opt. Lasers Eng.* **2019**, *123*, 1−7.
(6) Wei, T.; Wu, H.; Dong, L.; Cui, R.; Jia, S. *Opt. Express* **2021**, *29*, 12357−12364.
(7) Yang, H.; Bu, X.; Cao, Y.; Song, Y. *Infrared Phys. Technol.* **2021**, *114*, No. 103670.
(8) Fiedler, S. E.; Hese, A.; Ruth, A. A. *Chem. Phys. Lett.* **2003**, *371*, 284−294.
(9) Berden, G.; Peeters, R.; Meijer, G. *Int. Rev. Phys. Chem.* **2000**, *19*, 565−607.
(10) Smith, B. C. *Fundamentals of Fourier Transform Infrared Spectroscopy*; CRC Press, 2011.
(11) Zhang, Q.; Chang, J.; Wang, Q.; Wang, Z.; Wang, F.; Qin, Z. *Sensors* **2018**, *18*, No. 42.
(12) Zhang, Q.; Chang, J.; Cong, Z.; Sun, J.; Wang, Z. *IEEE Photonics J.* **2018**, *10*, No. 6804308.
(13) Wu, H.; Dong, L.; Zheng, H.; Yu, Y.; Ma, W.; Zhang, L.; Yin, W.; Xiao, L.; Jia, S.; Tittel, F. K. *Nat. Commun.* **2017**, *8*, No. 15331.
(14) Zifarelli, A.; Giglio, M.; Menduni, G.; Sampaolo, A.; Patimisco, P.; Passaro, V. M.; Wu, H.; Dong, L.; Spagnolo, V. *Anal. Chem.* **2020**, *92*, 11035−11043.
(15) Wei, T.; Zifarelli, A.; Dello Russo, S.; Wu, H.; Menduni, G.; Patimisco, P.; Sampaolo, A.; Spagnolo, V.; Dong, L. *Appl. Phys. Rev.* **2021**, *8*, No. 041409.
(16) Zhu, F.; Bounds, J.; Bicer, A.; Strohaber, J.; Kolomenskii, A. A.; Gohle, C.; Amani, M.; Schuessler, H. A. *Opt. Express* **2014**, *22*, 23026−23033.
(17) Cui, R.; Dong, L.; Wu, H.; Ma, W.; Xiao, L.; Jia, S.; Chen, W.; Tittel, F. K. *Anal. Chem.* **2020**, *92*, 13034−13041.
(18) Raza, M.; Ma, L.; Yao, C.; Yang, M.; Wang, Z.; Wang, Q.; Kan, R.; Ren, W. *Opt. Laser Technol.* **2020**, *130*, No. 106344.
(19) Rieker, G. B.; Jeffries, J. B.; Hanson, R. K. *Appl. Opt.* **2009**, *48*, 5546−5560.
(20) Lins, B.; Zinn, P.; Engelbrecht, R.; Schmauss, B. *Appl. Phys. B* **2010**, *100*, 367−376.
(21) McManus, J. B.; Zahniser, M. S.; Nelson, D. D.; Shorter, J. H.; Herndon, S. C.; Jervis, D.; Agnese, M.; McGovern, R.; Yacovitch, T. I.; Roscioli, J. R. *Appl. Phys. B* **2015**, *119*, 203−218.
(22) Kireev, S. V.; Kondrashov, A. A.; Shnyrev, S. L. *Laser Phys. Lett.* **2018**, *15*, No. 105701.
(23) Zhang, T.; Kang, J.; Meng, D.; Wang, H.; Mu, Z.; Zhou, M.; Zhang, X.; Chen, C. *Sensors* **2018**, *18*, No. 4295.
(24) Zheng, F.; Qiu, X.; Shao, L.; Feng, S.; Cheng, T.; He, X.; He, Q.; Li, C.; Kan, R.; Fittschen, C. *Opt. Laser Technol.* **2020**, *124*, No. 105963.
(25) Xia, J.; Feng, C.; Zhu, F.; Zhang, S.; Kolomenskii, A. A.; Dong, J.; Schuessler, H. A. *Measurement* **2021**, *168*, No. 108467.
(26) Zhou, S.; Liu, N.; Shen, C.; Zhang, L.; He, T.; Yu, B.; Li, J. *Spectrochim. Acta, Part A* **2019**, *223*, No. 117332.
(27) Zhou, S.; Shen, C.; Zhang, L.; Liu, N.; He, T.; Yu, B.; Li, J. *Opt. Express* **2019**, *27*, 31874−31888.
(28) Westerhuis, J. A.; De Jong, S.; Smilde, A. K. *Chemom. Intell. Lab. Syst.* **2001**, *56*, 13−25.
(29) Yan, F.; Li, S.; Zhang, E.; Guo, J.; Chen, Q. *Neurocomputing* **2021**, *446*, 130−144.
(30) Hochreiter, S.; Schmidhuber, J. *Neural Comput.* **1997**, *9*, 1735−1780.
(31) Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. 2014, arXiv:1412.3555. arXiv.org e-Print archive. https://arxiv.org/abs/1412.3555.
(32) Lee, J.; Song, T.; He, J.; Kandeepan, S.; Wang, K. *Opt. Express* **2021**, *29*, 26165−26182.
(33) He, J.; Lee, J.; Song, T.; Li, H.; Kandeepan, S.; Wang, K. *Opt. Lett.* **2019**, *44*, 3745−3748.
(34) Datta, A.; Raj, V.; Sankar, V.; Kalyani, S.; Srinivasan, B. *Opt. Express* **2021**, *29*, 26745−26764.
(35) Xu, Z.; Sun, C.; Ji, T.; Manton, J. H.; Shieh, W. *J. Lightwave Technol.* **2021**, *39*, 475−480.
(36) Zhao, Z.; Wu, Z.; Zheng, Y.; Ma, P. *ISPRS J. Photogramm. Remote Sens.* **2021**, *180*, 227−237.
(37) Li, D.; Zhou, J.; Liu, Y. *Mech. Syst. Signal Process.* **2021**, *147*, No. 107057.
(38) Azar, J.; Makhoul, A.; Couturier, R.; Demerjian, J. *Comput. Electr. Eng.* **2021**, *92*, No. 107065.
(39) Li, J.; Du, Y.; Ding, Y.; Peng, Z. *J. Quant. Spectrosc. Radiat. Transfer* **2021**, No. 107790.
(40) Niu, M. S.; Han, P. G.; Song, L. K.; Hao, D. Z.; Zhang, J. H.; Ma, L. *Opt. Express* **2017**, *25*, A896−A905.
(41) Tian, L.; Sun, J.; Chang, J.; Xia, J.; Zhang, Z.; Kolomenskii, A. A.; Schuessler, H. A.; Zhang, S. *Measurement* **2021**, No. 109739.
(42) Ma, Y.; Li, H.; Yu, Y.; Yao, Y.; Fang, Y.; Zhou, B.; Huang, J.; Min, J.; Liang, K. *J. Mod. Opt.* **2013**, *60*, 1967−1975.
(43) Li, B.; Zheng, C.; Liu, H.; He, Q.; Ye, W.; Zhang, Y.; Pan, J.; Wang, Y. *Sens. Actuators, B* **2016**, *225*, 188−198.
(44) Zhang, L.; Pang, T.; Zhang, Z.; Sun, P.; Xia, H.; Wu, B.; Guo, Q.; Sigrist, M. W.; Shu, C. *Sens. Actuators, B* **2021**, *334*, No. 129680.
(45) Zhang, L.; Zhang, Z.; Sun, P.; Pang, T.; Xia, H.; Cui, X.; Guo, Q.; Sigrist, M. W.; Shu, C.; Shu, Z. *Spectrochim. Acta, Part A* **2020**, *239*, No. 118495.