Check for updates

# End-to-end methane gas detection algorithm based on transformer and multi-layer perceptron

CHANG LIU,[1,2,†] GANG WANG,[1,2,†] CHEN ZHANG,[1,2]
PIETRO PATIMISCO,[3,4] RUYUE CUI,[1,2] CHAOFAN FENG,[1,2]
ANGELO SAMPAOLO,[3,4] VINCENZO SPAGNOLO,[1,3,4,5] iD
LEI DONG,[1,2,3,6] iD AND HONGPENG WU[1,2,3,7]

[1] *State Key Laboratory of Quantum Optics and Quantum Optics Devices, Institute of Laser Spectroscopy, Shanxi University, Taiyuan 030006, China*
[2] *Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China*
[3] *PolySense Lab—Dipartimento Interateneo di Fisica, Politecnico and University of Bari, Bari, Italy*
[4] *PolySense Innovations Srl, Bari, Italy*
[5] *vincenzoluigi.spagnolo@poliba.it*
[6] *donglei@sxu.edu.cn*
[7] *wuhp@sxu.edu.cn*
[†] These authors contributed equally to this work.

**Abstract:** In this paper, an end-to-end methane gas detection algorithm based on transformer and multi-layer perceptron (MLP) for tunable diode laser absorption spectroscopy (TDLAS) is presented. It consists of a Transformer-based U-shaped Neural Network (TUNN) filtering algorithm and a concentration prediction network (CPN) based on MLP. This algorithm employs an end-to-end architectural design to extract information from noisy transmission spectra of methane and derive the $CH_4$ concentrations from denoised spectra, without intermediate steps. The results demonstrate the superiority of the proposed TUNN filtering algorithm over other typically employed digital filters. For concentration prediction, the determination coefficient ($R^2$) reached 99.7%. Even at low concentrations, $R^2$ remained notably high, reaching up to 89%. The proposed algorithm results in a more efficient, convenient, and accurate spectral data processing for TDLAS-based gas sensors.

## 1. Introduction

Methane, an odorless and colorless gas, is highly flammable and explosive. It has found wide-ranging applications as a fuel in industrial and urban areas. Apart from natural sources such as wetlands, biogas, thawing permafrost, and methane hydrates, human activities are the primary contributors to methane emissions, particularly arising from industries involved in natural gas and coal production, agricultural practices, landfill sites, and biomass combustion [1–6]. Over the past three decades, the emission and concentration of $CH_4$ in the atmosphere have been steadily increasing, making it the second-largest greenhouse gas after $CO_2$. Despite the concentration of methane in ambient air (∼1.9 ppm) being approximately 200 times lower than that of $CO_2$ (> 400 ppm), its global warming potential is 25 times that of $CO_2$. Methane contributes to about 15% of the expected global warming, significantly impacting global climate change [7]. Moreover, in the industrial domain, methane explosions in mines and natural gas system leaks pose threats to human safety [8]. Consequently, conducting research and monitoring methane concentrations with high sensitivity, precision, and robustness is essential for both the domains of climate change research and the industrial sector.

Tunable diode laser absorption spectroscopy (TDLAS) is a spectroscopic measurement technique that uses lasers as light source to retrieve the absorption spectra of gases in real-time [9–11]. Due to its advantages of high sensitivity and selectivity, TDLAS has found wide-ranging

applications in different fields, such air pollutants detection [12] and gas leak detection [13]. However, the ultimate sensitivity as well as the resolution of TDLAS measurements can be compromised by several sources of noise. These can be divided in two categories: interference noise, caused by spectral and optical interferences; and electronic noise, coming from detector, laser, and other electronic devices [14–18]. Therefore, the use of noise filtering on measured spectral data is mandatory to improve the ultimate performance of a TDLAS-based gas sensor.

In recent years, software-based filtering methods have rapidly developed due to their simplicity, minimal equipment dependence, and cost-effectiveness. The Savitzky-Golay (S-G) smoothing filter proposed by Li et al. [19] is particularly attractive, as it only requires the specification of two parameters: the width of the smoothing window and the order of the smoothing polynomial. However, its effectiveness of filtering results strictly correlated to both parameters [20,21]. Kalman Filter (KF) is widely used in linear systems [22,23] but suffers from significant biases when used in nonlinear signal processing. Dual-optimized Back Propagation (BP) neural networks with variance compensation, known as BP-KF, optimize the parameters of the Kalman filter using BP neural networks and eliminate variations in dynamic system parameters through variance compensation [24,25]. However, it still necessitates precise state-space equations and an assessment of noise contributions. A shallow neural network (SNN) utilizing a multi-layer perceptron (MLP) has been introduced for spectral filtering and denoising, yielding commendable outcomes [26–28]. However, the static weights of the fully connected layer in SNN post-training limit its adaptability to capture the variable significance of individual sampling points across diverse samples. In summary, although these methods have demonstrated good performance in filtering, their applicability remains significantly limited. To address the aforementioned issue of fixed weights for each sampling point, we introduce an innovative filtering algorithm named transformer-based U-shaped neural network (TUNN). This approach dynamically assigns varying weights to individual sampling points, aiming to obviate the intricate manual parameter evaluation process and enhance the efficacy of spectral filtering.

The rapid development of deep neural networks applied on direct absorption spectroscopy techniques have shown the ability to analyze and extract useful information from complex nonlinear data, learning stable mapping relationships [29–36]. They have been extensively applied for different purposes, such as multi-component gas identification [29,30], concentration retrieval [24,31], and hyperspectral image classification [32,33].

The accurate prediction of the absorption spectra of the target substance as a function of gas conditions is subject to an accurate knowledge of the spectral parameters (such as line intensities, collision broadening, and narrowing) as well as of the line shape function for reconstructing the absorption profile. Relying on pure Voigt profiles is not the best choice because, for example, it overlooks the influences of Dicke narrowing effects and velocity-changing effects from collisions. To comprehensively consider the impact of all sampling points and eliminate prior knowledge of spectral parameters, we devise a robust concentration prediction network (CPN) based on MLP.

To address all these issues, we propose an end-to-end methane gas detection algorithm based on transformer and MLP to extract information accurately and efficiently from noisy spectral data and retrieve the methane concentration. Leveraging the advantages of the end-to-end architecture, the algorithm eliminates the need for complex calculations and enhance the accuracy and reliability of methane concentration retrieval from the absorption spectra.

The paper is structured as follows. In Section 2, a detailed description of the algorithm architecture, including the TUNN filtering algorithm and the CPN based on MLP, is provided. Section 3 introduces the TDLAS-based methane sensor and the dataset acquisition procedure required for algorithm training. In Section 4, the algorithm is applied to the dataset, together with the most recent filtering algorithms, and the performance compared in terms of accuracy in concentration retrieval. Finally, Section 5 provides insights into future perspectives.

## 2. Algorithm architecture

### 2.1. TUNN—transformer-based U-shaped neural network filtering algorithm

The filtering network proposed in this study is based on the U-shaped network [36], which employs a typical encoder-decoder structure consisting of an encoder, a decoder, and an MLP, as depicted in Fig. 1. In the TUNN, multiple transformer blocks are incorporated, due to the inherent capacity of a transformers to effectively capture long-range dependencies within the dataset by using global self-attention mechanisms. The algorithm analyzes the spectral data at a global level, enabling it to consider dependencies and relationships across all data points. This enhances the algorithm's ability to discern critical features within the spectra, resulting in more accurate and reliable filtering of methane absorption information, even in the presence of noise. Before the first transformer block, a $1 \times 3$ convolutional layer is applied, projecting the dimensionality of the spectral data containing noisy samples from $T \times 1$ to $T \times 8$, where $T = 1050$ is the total number of sampling points in the dataset. This enriches the features of each sampling point in the spectral data. The embedded data is subsequently processed through four encoder blocks for feature extraction, each comprising a transformer block and a downsampling layer. The decoder exhibits the same structural symmetry as the encoder part, in the form of a series of transformer blocks and upsampling layers. The main objective of the decoder is to progressively restore the high-dimensional features to the filtered sequence. Finally, the data pass through a series of convolutional layers to restore its dimensions back to the original data dimension, to obtain the denoised spectral data. In addition, to capture multi-scale and multi-level information in the network, skip connections [37] to integrate deep-level and shallow-level information are used. The intricate details of each module within the TUNN are elaborated as follows.
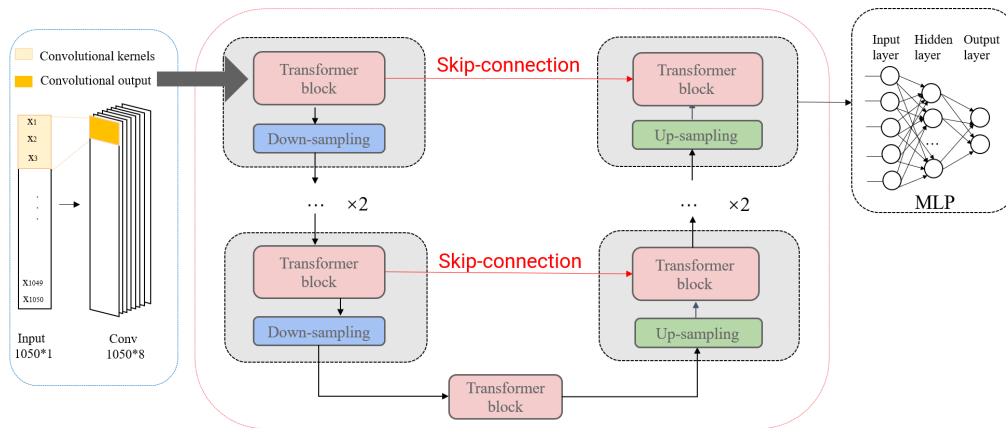


**Fig. 1.** The complete architecture of the TUNN.

Firstly, we introduce the transformer block, comprising a self-attention (SA) layer and a feed-forward network (FFN) layer, as illustrated in Fig. 2. Specific details of the SA layer on the transformer model refer to the work by Vaswani et al. [38]. In the TUNN, the SA layer enables each sample point within the sequence to efficiently attend to information from all other sample points in the sequence, including itself. In particular, for spectral data comprising 1050 sample points, the SA layer calculates the dot-product similarity between a specific sample point and all other sample points, including itself. Subsequently, the SA layer applies a softmax transformation to derive the corresponding attention weights. These weights are utilized to perform a weighted sum, yielding a new sample point at that position that effectively incorporates information from all other sample points. To enhance representational capacity and generalization capabilities,

linear transformations are introduced to project the input data $X$ into distinct spaces denoted as $Q(Query)$, $K(Key)$, and $V(Value)$. For the input sequence $X = [x_1, \ldots, x_i, \ldots, x_T] \in \mathbb{R}^{T \times d}$, where $T$ represents the number of sample points and d denotes the dimension of each sample point, the SA layer is computed as follows:

$$Self - Attention = SA(X)$$

$$= softmax(\frac{QK^T}{\sqrt{d_k}})V$$

$$= softmax(\frac{XW_Q W_K^T X^T}{\sqrt{d_k}})XW_V \tag{1}$$

where $W_Q \in \mathbb{R}^{d \times d_Q}$, $W_K \in \mathbb{R}^{d \times d_k}$, $W_V \in \mathbb{R}^{d \times d_v}$ are trainable parameters of weight matrices. The *softmax* calculation formula for the $i$-th element $x_i$ is:

$$softmax(x_i) = \frac{exp(x_i)}{\sum_{j=1}^{T} exp(x_j)} \tag{2}$$
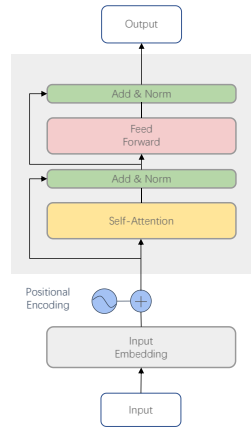


**Fig. 2.** The transformer block.

The FFN layer projects data into a higher-dimensional space to learn more abstract features. Meanwhile, an activation function to enhance the expressive capabilities of each sampling point is applied. This process allows for the establishment of intricate associations between different positions, ultimately facilitating the more effective capture of contextual information. Given an input sequence $X \in \mathbb{R}^{T \times d}$, the output is computed using the following formula:

$$FFN(X) = GELU(XW_1 + b_1)W_2 + b_2 \tag{3}$$

where $W_1 \in \mathbb{R}^{d \times 4d}$, $W_2 \in \mathbb{R}^{4d \times d}$, $b_1 \in \mathbb{R}^{T \times 4d}$, $b_2 \in \mathbb{R}^{T \times d}$ are trainable weight matrices and biases respectively. Additionally, the Gaussian error linear unit (GELU) is utilized as the activation function.

Based on the above formulas, for the input of noisy spectral dataset sample $X \in \mathbb{R}^{T \times d_1}$, the calculation for the $l$-th layer of the transformer ($l = 1, \ldots, K$) follows the formulae ($X^0 = X$):

$$\hat{X}^l = LN(SA(X^{l-1}) + X^{l-1})$$

$$X^l = LN(FFN(\hat{X}^l) + \hat{X}^l) \tag{4}$$

where $LN(X) = \frac{X - E[X]}{\sqrt{Var[X] + \varepsilon}}$, $E[X]$ denotes the mean value of $X$, $Var[X]$ represents the variance of $X$, and $\varepsilon$ denotes an infinitesimal value to prevent division by zero errors during computation.

It is important to highlight that, in implementing skip connections during the decoding stage, we utilized a concatenation approach, as illustrated in the following formula:

$$X^{decoder} = [X^{encoder}; X^{up}] \tag{5}$$

where $X^{encoder}$ represents the output of the corresponding transformer encoder module, $X^{up}$ denotes the output of the up-sampling layer, $X^{decoder}$ signifies the input of the corresponding transformer decoder module. The symbol $[\cdot; \cdot]$ denotes the concatenation operation.

Firstly, the spectral data set $X \in \mathbb{R}^{1050 \times 8}$ is extracted by transformer block in the encoder part. Subsequently, the down-sampling layer performs convolution on the deep-level features extracted by the transformer block, reducing the sequence length by half while doubling the number of feature dimensions. This process is repeated four times. As a result, the final output of the encoder is $X \in \mathbb{R}^{65 \times 128}$.

The decoder, which is symmetrical to the encoder structure, is also constructed based on transformer blocks. Differing from the encoder, the decoder employs deconvolution on the data before passing it through the transformer, reducing feature dimensions by half while doubling the number of sequence length. In the final stage of decoding, the high-dimensional features are effectively restored, gradually reconstructing the original denoised sequence, resulting in $X \in \mathbb{R}^{1050 \times 40}$.

Furthermore, to provide multi-scale and multi-level information for the sequence reconstruction in the decoder, TUNN incorporates the skip-connection, which concatenating data from the decoder with data of the same sequence length from the encoder facilitating the transfer of encoder's corresponding scale feature information into the deconvolution process, thereby reducing the loss of features during downsampling compression. As such, the decoder is equipped to achieve a more proficient restoration of the original sequence.

## 2.2. CPN—concentration prediction network

The CPN has been applicated widely in various fields due to its simple structure and powerful ability to fit nonlinear functions. Specifically, the MLP consists of an input layer, one or more hidden layers, and an output layer. The neurons in the hidden layers are typically equipped with nonlinear activation functions, enabling the network to capture complex relationships within the data. During the forward propagation of the MLP, input data is fed through the input layer and propagated through the hidden layers towards the output layer. The weights of neurons in each layer are trained and updated using backpropagation and optimization learning algorithms to minimize the loss function, enabling the network to achieve optimal generalization effect. To provide a concise explanation of the calculation principle of MLP, we will consider an example of an MLP network with a single hidden layer, and the corresponding formulae are as follows:

$$H = \sigma(XW^{(1)} + b^{(1)}) \tag{6}$$

$$O = \sigma(HW^{(2)} + b^{(2)}) \tag{7}$$

where $X \in \mathbb{R}^{n \times d}$ represents the input data with n samples, each sample having $d$ input features. The variable $H \in \mathbb{R}^{n \times h}$ corresponds to the output of the hidden layer, and $O \in \mathbb{R}^{n \times 1}$ represents the final output of network. $W^{(1)} \in \mathbb{R}^{d \times h}$ and $b^{(1)} \in \mathbb{R}^{n \times h}$ represent the weight and bias of the hidden layer respectively, $W^{(2)} \in \mathbb{R}^{h \times 1}$ and $b^{(2)} \in \mathbb{R}^{n \times 1}$ are the weight and bias of the output layer respectively. The symbol $\sigma$ denotes a non-linear activation function.

For the CPN, the input data consists of frequency spectrum data $X \in \mathbb{R}^{1050 \times 1}$ that has undergone denoising through TUNN. The output data is the final concentration information. To sum up, the CPN consists of one input layer, seven hidden layers, and one output layer. The precise number of neurons in each layer is illustrated in Fig. 3, while the activation function employed is rectified linear unit (ReLU).
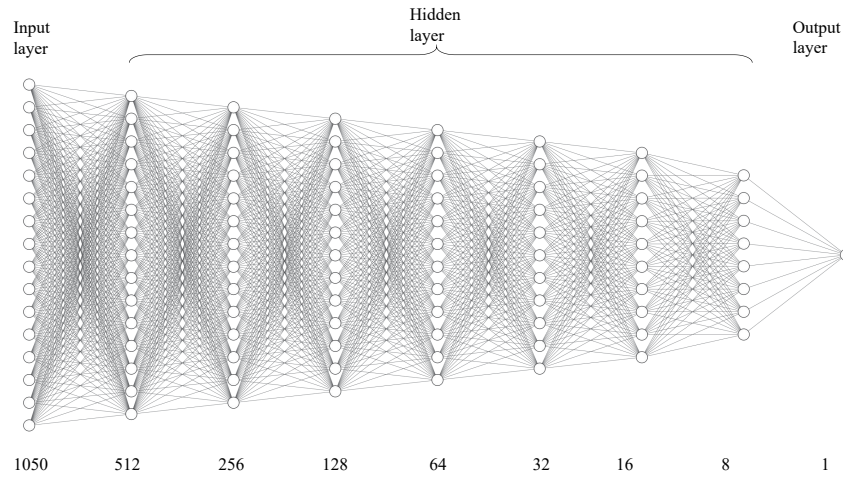
**Fig. 3.** The specific architecture of the CPN.

## 2.3.  Implementation details

**Table 1.  The detailed information of the TUNN**

| Input -1050 × 1 | |
|---|---|
| Input projection layer:  Conv3 - 1050 × 8 | |
| **Encoder** | **Decoder** |
| Transformer_0 - 1050 × 8 | Upsample_0 - 131 × 64 |
| Downsample_0 - 525 × 16 | Skip-connection0 ([Transformer_3;Upsample_0]) |
| | Transformer_5 - 131 × 128 |
| Transformer_1 - 525 × 16 | Upsample_1 - 262 × 64 |
| Downsample_1 - 262 × 32 | Skip-connection1 ([Transformer_2;Upsample_1]) |
| | Transformer_6 - 262 × 96 |
| Transformer_2 - 262 × 32 | Upsample_2 - 525 × 48 |
| Downsample_2 - 131 × 64 | Skip-connection2 ([Transformer_1;Upsample_2]) |
| | Transformer_7 - 525 × 64 |
| Transformer_3 - 131 × 64 | Upsample_3 - 1050 × 32 |
| Downsample_3 - 65 × 128 | Skip-connection3 ([Transformer_0;Upsample_3]) |
| | Transformer_8 - 1050 × 40 |
| Transformer_4 - 65 × 128 | Conv3 - 1050 × 20 |
| | Conv3 - 1050 × 10 |
| | Conv3 - 1050 × 1 |

The TUNN and the CPN were implemented using the PyTorch deep learning framework. For each part of the TUNN, each encoder and decoder consist of transformer blocks with one layer. Within these blocks, we utilized 8 attention heads. In the input projection layer, we used 8 convolution kernels (the dimensions of (1, 3), the stride of 1). The convolution kernel dimensions of the downsampling layer are $1 \times 2$ with a stride of 2. Furthermore, similar to the downsampling layers, all convolution kernels in the upsampling layers also utilized $1 \times 2$ kernels with a stride of 2. The final component, responsible for mapping the high-dimensional transformed spectrum

back to a 1-dimensional feature, is comprised of three convolutional layers. Additionally, to mitigate overfitting, we applied dropout with a rate of 0.1 within the transformer. The detailed information of the TUNN is provided in Table 1. For the CPN, specific details can be found in Section 2.2.

During the algorithm training process, we used a batch size of 16, Adam optimizer with a learning rate of 0.001, and performed 100 epochs (shown in Table 2). All training and testing are conducted on an NVIDIA 3060 GPU.

**Table 2. Training parameters**

| Batch Size | Epochs | Optimizer | Learning Rate |
|------------|--------|-----------|---------------|
| 16 | 100 | Adam | 0.001 |

## 3. Sensing system

### 3.1. Sensing system configuration

According to the HITRAN, methane exhibits a strong R(3) absorption line near 6050 cm$^{-1}$, depicted in Fig. 4, spectrally free from air background absorption interferences. The schematic diagram of the methane sensing system is shown in Fig. 5.
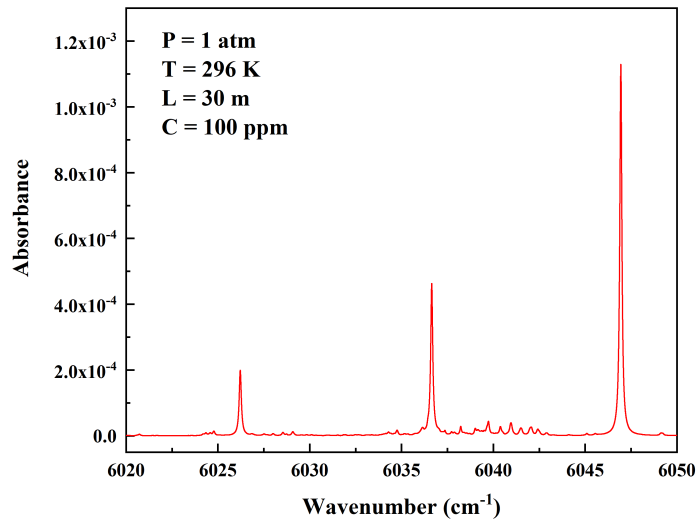


**Fig. 4.** The simulated absorption spectrum of 100ppm methane obtained based on the HITRAN database (T=297 K, P=1 atm, L=30 m, v=6020-6050 cm$^{-1}$).

A continuous-wave fiber-coupled 1.65 $\mu$m DFB laser (TR-1654-DFB-B) was employed as the laser source to scan the selected absorption line of methane. The laser was driven by a custom-designed control electronics unit (CEU) to regulate the laser temperature and current: the laser temperature was fixed at 25 °C, while the current was scanned from 85 mA to 115 mA by applying a slow 40 mHz ramp signal to the laser current driver. The laser beam emitted by the DFB laser was coupled into an optical fiber and entered a multi-pass gas cell with an optical path length of 30 m [4,39]. The pressure inside the multi-pass cell was fixed at 1 atm. An InGaAs photodetector (PDA10D-EC, Thorlabs, USA) generated a voltage signal proportional to the light intensity exiting from the multi-pass, which was in turn collected by a data acquisition card (USB-6361, NI), and transmitted to a PC for subsequent processing. Mixtures with different methane concentrations were obtained by diluting a cylinder with certified concentration of 5000
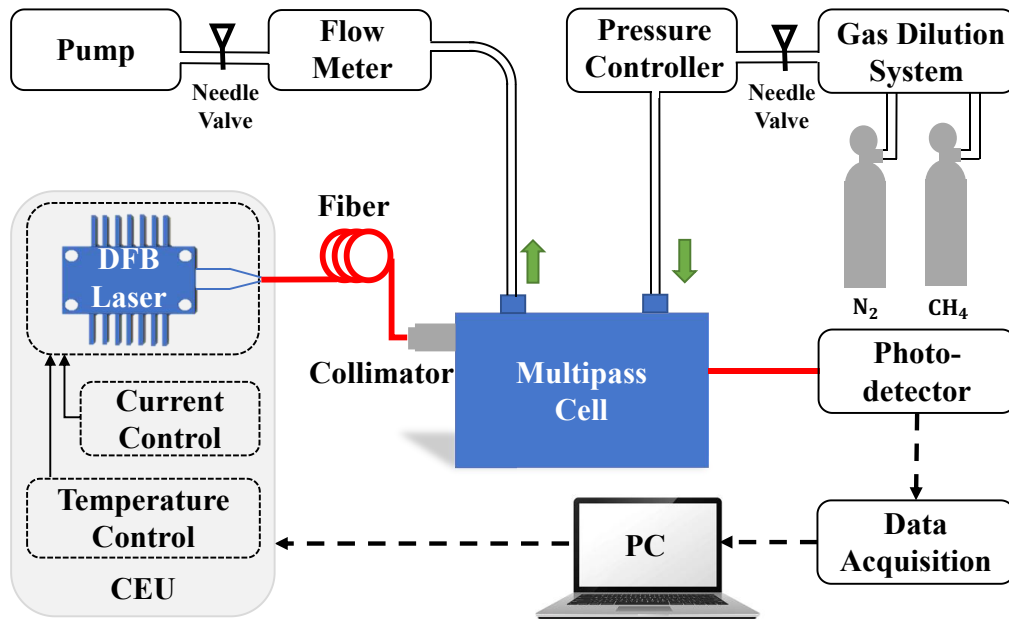
**Fig. 5.** Schematic diagram of the methane sensing system.

ppm of methane in $N_2$ with a cylinder of high-purity nitrogen, by using a gas blender (GB100, MCQ instruments, Italy).

### 3.2. Data set acquisition

To evaluate the performance of the proposed algorithm, a set of 20 scans of the selected methane absorption line were collected at different $CH_4$ concentrations ranging from 5-100 ppm at steps of 5 ppm, by using the sensing system described in the previous section. For each $CH_4$ concentration, the spectral scan was collected for 20 cycles. To further increase the prediction accuracy at low concentrations, scans were also collected for 1, 2, 3 and 4 ppm. At the end, the experimental dataset comprised 480 sets of spectral data for validation. Although deep learning typically would require a larger amount of labeled data for effective training, the acquisition of a large amount of experimental data is resource-intensive and time-consuming. Thus, simulated spectral scans at the experimental conditions were generated from 1 to 100 ppm of methane concentrations, at steps of 1 ppm, to enrich the dataset for algorithm training. Gaussian noise with a mean of 0 and a standard deviation of 0.00276 was added to simulated spectra. For each concentration of simulated spectra, 20 cycles were generated considering the Gaussian noise distribution. Considering $CH_4$ concentration of 25 ppm as representative, a comparison between the simulated scan and the experimental one is depicted in Fig. 6, with a resulting standard deviation of $1.71 \times 10^{-5}$.
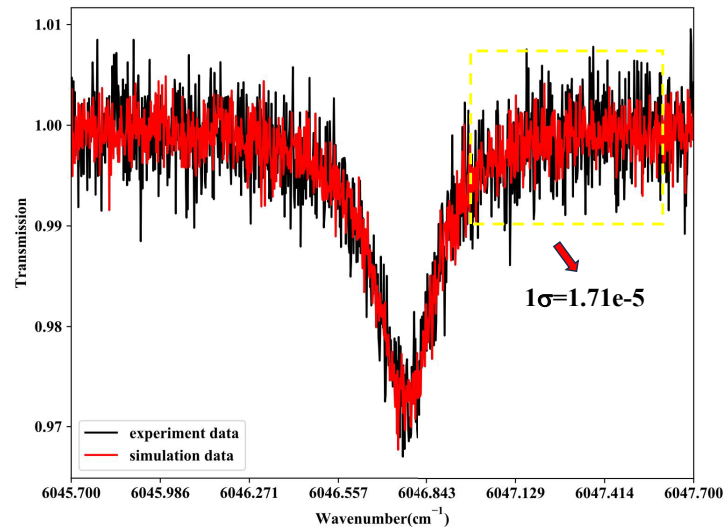
**Fig. 6.** Comparison between simulated and experimental data for a methane concentration of 25 ppm in $N_2$.

## 4.    Results and analysis

### 4.1.    Comparison and analysis of filtering performance

The filtering performance of TUNN with the popular denoising, including adaptive S-G filtering, KF and BP-KF, were compared on simulated data and experimental data. To ensure equitable comparisons, each algorithm was optimized to achieve the best filtering results. Specifically, for the adaptive S-G filter, the optimization was reached by fine-tuning both the polynomial order and window size [19]. For KF and BP-KF, we followed guidelines provided in [21] to discern the suitable system state equation. Additionally, we integrated a carefully designed BP neural network to further refine the system state equation and enhance the filtering results. Considering the simulated data referred to methane concentration of 5 ppm $N_2$ as representative, the results of the employed denoising algorithms is reported in Fig. 7.
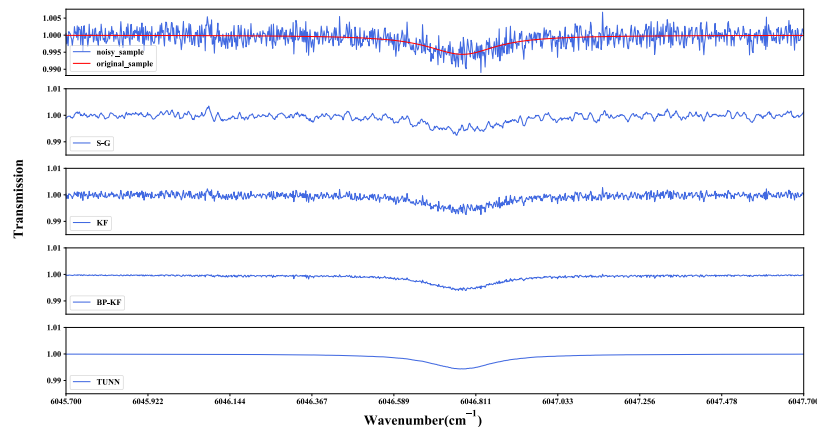


**Fig. 7.** Performance of different filtering algorithms on simulated data of 5ppm $CH_4$ concentration in $N_2$.

The filtering capability of TUNN is significantly better than alternative methods. Indeed, the denoised signal achieved through the implementation of TUNN closely resembles the pristine, noise-free signal. The results are summarized in Table 3.

**Table 3. SNR and RMSE of various filtering algorithms on simulated data of 5ppm**

| Filter Methods | SNR | RMSE |
|---|---|---|
| Noise Signal | 14.48 | $2.08 \times 10^{-3}$ |
| S-G | 17.97 | $9.20 \times 10^{-4}$ |
| KF | 18.28 | $9.10 \times 10^{-4}$ |
| BP-KF | 29.71 | $3.37 \times 10^{-4}$ |
| **TUNN** | **44.45** | $\mathbf{6.12 \times 10^{-5}}$ |

The signal-to-noise ratio (SNR) increases up to 44.45 dB, which is more than 3 times higher than the SNR of original noisy data, with a root mean square error (RMSE) reduced from $2.08 \times 10^{-3}$ to $6.12 \times 10^{-5}$. The adaptive S-G filtering exhibits the poorest filtering performance with a slight improvement of the SNR (3.49 dB), and the RMSE reduced to $9.20 \times 10^{-4}$. With KF, the SNR improves from 14.48 dB to 18.28 dB, while the RMSE metric does not demonstrate a remarkable improvement. Compared to KF, the BP-KF, which integrates a BP neural network for training optimization, shows some improvement in filtering performance. The filtered spectral signal exhibits greater smoothness in contrast to the S-G filtering and KF algorithm, with an enhancement of SNR to 29.71 dB and a reduced RMSE of $3.37 \times 10^{-4}$. However, it is evident that a significant amount of noise was still not-filtered throughout the transmission spectrum.

To fully analyze the filtering performance of TUNN on simulated data, the average SNR, mean absolute error (MAE) and RMSE of simulated data were calculated employing various filtering methods, for all datasets, from 1-100 ppm $CH_4$ in $N_2$. The results are shown in Table 4. Compared with BP-KF which also incorporates a neural network, TUNN yields to an enhancement of the average SNR by approximately 16.74 dB. Moreover, it significantly reduces the average MAE from $1.66 \times 10^{-3}$ to $6.88 \times 10^{-5}$, as well as the average RMSE from $2.07 \times 10^{-3}$ to $1.28 \times 10^{-4}$. The high performance of TUNN can be ascribed to several key factors. Firstly, the transformer block within TUNN possesses the capability to process high-sampling-rate data in parallel. Moreover, the self-attention mechanism inherent in the transformer block facilitates robust attention interactions between each sampling point and all other sampling points across the transmission spectrum. For each sampling point in the input spectral data, we compute its dot product with the other elements and then normalize it using the softmax function to obtain weights. These weights measure the relevance of the current element with respect to the others. This enables the entire network to effectively prioritize the crucial sampling points information necessary for accurate denoising of the spectrum. Furthermore, the incorporation of skip connections empowers TUNN to adeptly amalgamate information from various levels in the encoder stage into the decoder stage. Consequently, all the above aid the network in accurately reconstructing the denoised transmission spectrum.

To verify the performance of the filtering model in practical application scenarios, the TUNN algorithm was applied on the experimental dataset. As an illustrative example, we selected the transmission spectrum with a concentration of 5 ppm $CH_4$ in $N_2$ to demonstrate the filtering performance of various filtering methods, as depicted in Fig. 8. It is worth noticing that the outcomes obtained with the S-G filter and KF are notably unsatisfactory.

Substantial noise persists even after filtration, accompanied by persistent distortions induced by the noise. These factors negatively affect the accuracy and stability of concentration prediction. Although the transmission spectrum filtered by BP-KF has eliminated most of the noise compared with S-G filter and KF, the observed bias in the critical absorption peak position is still significant,

**Table 4. Average SNR, MAE and RMSE of different filtering algorithms on the simulated dataset**

| Filter Methods | SNR (mean) | MAE (mean) | RMSE (mean) |
|---|---|---|---|
| Noise Signal | 27.15 | $1.66 \times 10^{-3}$ | $2.07 \times 10^{-3}$ |
| S-G | 32.77 | $7.53 \times 10^{-4}$ | $9.44 \times 10^{-4}$ |
| KF | 33.00 | $7.26 \times 10^{-4}$ | $9.09 \times 10^{-4}$ |
| BP-KF | 39.69 | $4.53 \times 10^{-4}$ | $8.13 \times 10^{-4}$ |
| **TUNN** | **43.89** | $\mathbf{6.88 \times 10^{-5}}$ | $\mathbf{1.28 \times 10^{-4}}$ |
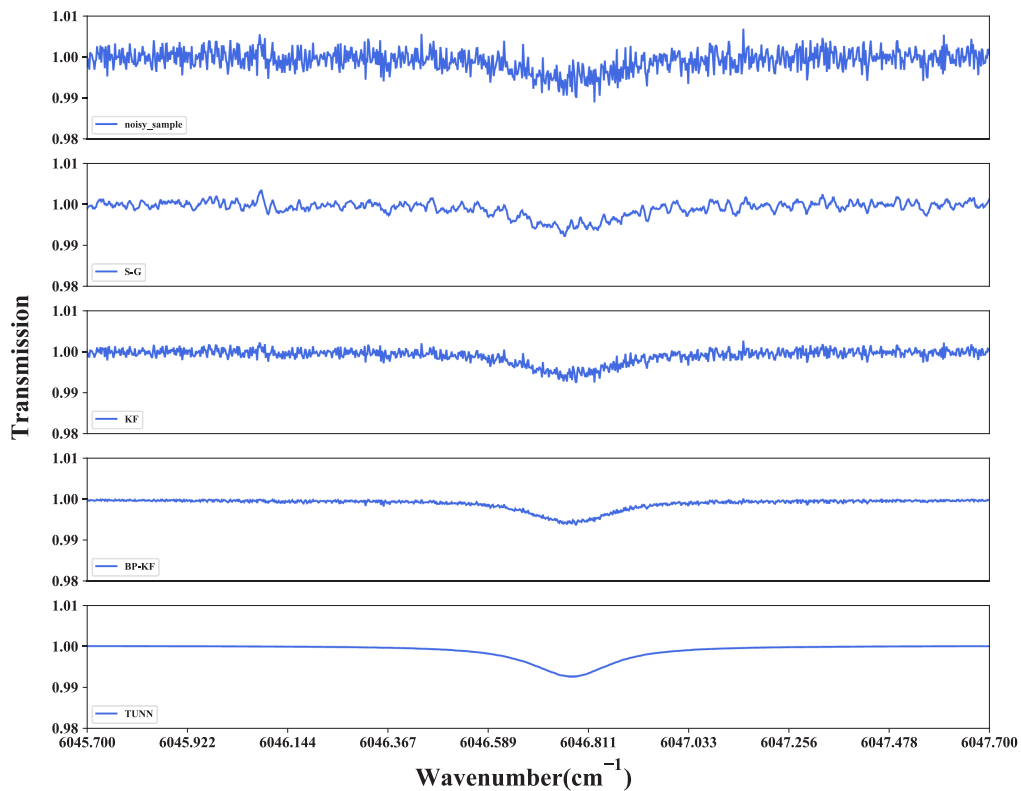


**Fig. 8.** Performance of different filtering algorithms on experimental data of 5ppm $CH_4$ in $N_2$.

which is a result of the oversimplified structure of the BP neural network. Compared with all preceding algorithms, TUNN demonstrates the best performance following the interaction of transformer attention and the fusion of multi-scale features within the multi-layer U-shaped network. The transmission spectrum data filtered by TUNN is characterized by its exceptional smoothness, and the positions of absorption peaks are notably more accurate. The availability of such smoothly denoised data proves highly advantageous for subsequent concentration prediction.

Similarly, for a quantitative analysis of TUNN's denoising performance, we have computed the average SNR, MAE, and RMSE across 480 sets of experimental data spectra ranging from 1 to 100 ppm $CH_4$ in $N_2$, as displayed in Table 5. Our proposed TUNN outperforms other filtering methods on the experimental data, exhibiting the highest average SNR, along with substantially reduced MAE and RMSE compared to the original noisy signal. Figure 9 illustrates TUNN's noise filtering capabilities across 24 randomly selected data sets with varying concentrations.
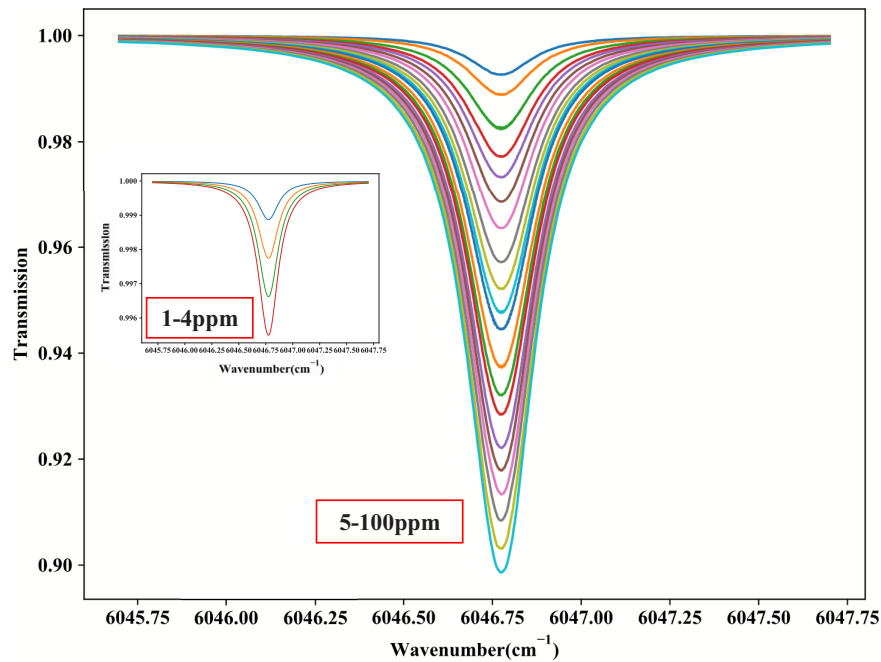


**Fig. 9.** Experimental transmission spectrum after TUNN filtering.

**Table 5. Average SNR, MAE and RMSE of different filtering algorithms on the experimental dataset**

| Filter Methods | SNR (mean) | MAE (mean) | RMSE (mean) |
|---|---|---|---|
| Noise Signal | 27.66 | $1.67 \times 10^{-3}$ | $2.08 \times 10^{-3}$ |
| S-G | 33.54 | $7.47 \times 10^{-4}$ | $9.41 \times 10^{-4}$ |
| KF | 33.61 | $7.26 \times 10^{-4}$ | $9.09 \times 10^{-4}$ |
| BP-KF | 40.00 | $4.72 \times 10^{-4}$ | $8.48 \times 10^{-4}$ |
| **TUNN** | **43.86** | $\mathbf{6.88 \times 10^{-5}}$ | $\mathbf{8.41 \times 10^{-4}}$ |

## 4.2. Performance comparison and analysis of concentration prediction

In order to assess the applicability and reliability of the proposed algorithm for concentration measurements, we conducted experiments using 480 sets of gas samples with different methane

concentrations. We compared the concentrations extracted from the sensor spectra with the results of concentration prediction. The comparative performance of concentration prediction is visualized in Fig. 10.
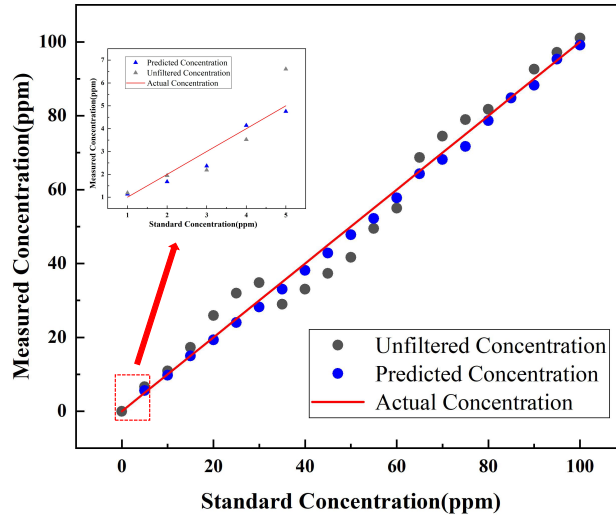


**Fig. 10.** Comparison of $R^2$ between the results obtained after TUNN+CPN filtering and the measured results from methane sensing system.

From this figure, it is evident that the utilization of the TUNN+CPN algorithm results in a significantly improved linear relationship between the predicted concentration and the real concentration values, with respect to that obtained by the standard method. Additionally, for a quantitative assessment of the concentration retrieval performance of TUNN+CPN, we computed the determination coefficient ($R^2$), the mean relative error (RE), and the mean absolute error (AE), as presented in Table 6. Specifically, upon employing the TUNN+CPN method, in the 5-100 ppm $CH_4$ in $N_2$ range, the $R^2$ of the methane sensor increases to 99.7%, while the average RE decreases by 0.066, and the average AE decreases by 2.259. Meanwhile, we conducted a comprehensive comparison of concentration determination precision in seven cases under low SNR conditions, as illustrated in Table 7. Within the 1-4 ppm range, the $R^2$ of the methane sensor improved to 89.1% when employing the TUNN and CPN. This enhancement was accompanied by a decrease in the average RE by 0.016 and a reduction in the AE by 0.075. Additionally, we observed that performing concentration calculations using traditional methods on the basis of filtering leads to a certain level of performance improvement. Moreover, the greater the filtering effectiveness, the more pronounced the enhancement. Furthermore, our findings indicate that relying solely on CPN for noise data prediction without incorporating filtering algorithms leads to a noticeable decline in predictive performance. This phenomenon is attributed to substantial noise at low concentrations, indicating that autonomous prediction using only MLP may not adequately align with the true concentration values. Therefore, noise filtration proves to be crucial for accurate concentration prediction.

Compared with the standard method, the performance has undergone significant improvement. We attribute the success of TUNN+CPN to two key factors. Firstly, TUNN effectively filters and denoises the transmission spectrum, even in the presence of substantial noise. Secondly, MLP has the capability to comprehensively consider the characteristics of all sampling points within the transmission spectrum, enabling to establish a robust concentration determination function that closely fits with the corresponding concentration values. In contrast, the standard method relies solely on absorption peak information to predict corresponding concentrations, making its

**Table 6. Comparison of $R^2$, AE, and RE between common sensors and TUNN+CPN after applying the algorithm**

| Concentration | 5-100 ppm | | | 1-4 ppm | | |
|---|---|---|---|---|---|---|
| | $R^2$ | RE (mean) | AE (mean) | $R^2$ | RE (mean) | AE (mean) |
| Methane Sensing System | 0.984 | 0.103 | 3.430 | 0.815 | 0.512 | 0.383 |
| **TUNN + CPN** | **0.997** | **0.037** | **1.171** | **0.891** | **0.136** | **0.308** |

**Table 7. Comparison of $R^2$, AE, and RE after applying different filtering algorithms in low SNR**

| Concentration | 1-4 ppm | | |
|---|---|---|---|
| | $R^2$ | RE (mean) | AE (mean) |
| Methane Sensing System | 0.815 | 0.512 | 0.383 |
| S-G | 0.829 | 0.415 | 0.364 |
| KF | 0.831 | 0.396 | 0.325 |
| BP-KF | 0.852 | 0.217 | 0.316 |
| CPN | 0.807 | 0.528 | 0.395 |
| TUNN | 0.862 | 0.164 | 0.313 |
| **TUNN + CPN** | **0.891** | **0.136** | **0.308** |

performance susceptible to external factors such as noise fluctuations. Thereby, TUNN+CPN shows superior performance thanks to their collaborative interaction.

## 5.  Conclusion

In this study, an end-to-end transformer-based methane spectrum filtering and concentration prediction algorithm was proposed, which achieves accurate and stable concentration measurement from methane absorption spectra. This algorithm comprises two primary components: TUNN for filtering noise and CPN for concentration prediction. When applied to transmission spectra with noise, the algorithm effectively eliminates most of the noise, extracts pertinent information, and subsequently correctly quantifies the methane concentration. Compared with other popular filtering algorithms such as S-G, KF, and BP-KF, the proposed TUNN exhibits stronger performance on both simulated and experimental datasets. The SNR was significantly improved, and it maintained stability even at lower concentrations. The noise-filtered methane spectrum was subsequently processed through the CPN to yield precise concentration predictions. Therefore, based on methane sensing system, our algorithm attained improved concentration detection accuracy and sustained greater stability even in environments characterized by low SNR. Note that, although the proposed algorithm significantly improves the concentration retrieval performance of methane sensing system, it is limited to detecting a single gas component and cannot identify or measure the concentrations of multi-component gases. In future studies, algorithms that support the gas classification and concentration measurement of multi-component gas sensor will be developed.

**Disclosures.** The authors declare no conflicts of interests.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

## References

1. M. Saunois, A. R. Stavert, B. Poulter, *et al.*, "The global methane budget 2000–2017," Earth Syst. Sci. Data **12**(3), 1561–1623 (2020).
2. E. G. Nisbet, E. J. Dlugokencky, and P. Bousquet, "Methane on the rise—again," Science **343**(6170), 493–495 (2014).
3. S. Iwaszenko, P. Kalisz, M. Słota, *et al.*, "Detection of natural gas leakages using a laser-based methane sensor and uav," Remote Sens. **13**(3), 510 (2021).
4. R. Y. Cui, L. Dong, H. P. Wu, *et al.*, "Three-dimensional printed miniature fiber-coupled multipass cells with dense spot patterns for ppb-level methane detection using a near-ir diode laser," Anal. Chem. **92**(19), 13034–13041 (2020).
5. H. P. Wu, L. Dong, X. K. Yin, *et al.*, "Atmospheric ch4 measurement near a landfill using an icl-based qepas sensor with vt relaxation self-calibration," Sens. Actuators, B **297**, 126753 (2019).
6. H. P. Wu, L. Dong, H. D. Zheng, *et al.*, "Beat frequency quartz-enhanced photoacoustic spectroscopy for fast and calibration-free continuous trace-gas monitoring," Nat. Commun. **8**(1), 15331 (2017).
7. K. Liu, L. Wang, T. Tan, *et al.*, "Highly sensitive detection of methane by near-infrared laser absorption spectroscopy using a compact dense-pattern multipass cell," Sens. Actuators, B **220**, 1000–1005 (2015).
8. V. Wittstock, L. Scholz, B. Bierer, *et al.*, "Design of a led-based sensor for monitoring the lower explosion limit of methane," Sens. Actuators, B **247**, 930–939 (2017).
9. L. Dong, C. G. Li, N. P. Sanchez, *et al.*, "Compact ch4 sensor system based on a continuous-wave, low power consumption, room temperature interband cascade laser," Appl. Phys. Lett. **108**(1), 011106 (2016).
10. K. Krzempek, R. Lewicki, L. Nähle, *et al.*, "Continuous wave, distributed feedback diode laser based sensor for trace-gas detection of ethane," Appl. Phys. B **106**(2), 251–255 (2012).
11. S. Z. Li, Y. P. Yuan, Z. J. Shang, *et al.*, "Ppb-level nh3 photoacoustic sensor combining a hammer-shaped tuning fork and a 9.55 $\mu$m quantum cascade laser," Photoacoustics **33**, 100557 (2023).
12. L. G. Shao, B. Fang, F. Zheng, *et al.*, "Simultaneous detection of atmospheric co and ch4 based on tdlas using a single 2.3 $\mu$m dfb laser," Spectrochim. Acta, Part A **222**, 117118 (2019).
13. H. B. Lu, C. T. Zheng, L. Zhang, *et al.*, "A remote sensor system based on tdlas technique for ammonia leakage monitoring," Sensors **21**(7), 2448 (2021).
14. A. Janani and M. Sasikala, "Investigation of different approaches for noise reduction in functional near-infrared spectroscopy signals for brain–computer interface applications," Neural Comput. Appl. **28**(10), 2889–2903 (2017).
15. B. Lins, P. Zinn, R. Engelbrecht, *et al.*, "Simulation-based comparison of noise effects in wavelength modulation spectroscopy and direct absorption tdlas," Appl. Phys. B **100**(2), 367–376 (2010).
16. P. Werle, R. Mücke, and F. Slemr, "The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (tdlas)," Appl. Phys. B **57**(2), 131–139 (1993).
17. P. Werle and F. Slemr, "Signal-to-noise ratio analysis in laser absorption spectrometers using optical multipass cells," Appl. Opt. **30**(4), 430–434 (1991).
18. H. P. Wu, A. Sampaolo, and L. Dong, "Quartz enhanced photoacoustic H2S gas sensor based on a fiber-amplifier source and a custom tuning fork with large prong spacing," Appl. Phys. Lett. **107**(11), 111104 (2015).
19. J. S. Li, H. Deng, P. F. Li, *et al.*, "Real-time infrared gas detection based on an adaptive savitzky–golay algorithm," Appl. Phys. B **120**(2), 207–216 (2015).
20. B. Zimmermann and A. Kohler, "Optimizing savitzky–golay parameters for improving spectral resolution and quantification in infrared spectroscopy," Appl. Spectrosc. **67**(8), 892–902 (2013).
21. T. Y. Zhang, J. W. Kang, D. Z. Meng, *et al.*, "Mathematical methods and algorithms for improving near-infrared tunable diode-laser absorption spectroscopy," Sensors **18**(12), 4295 (2018).
22. F. Zheng, X. B. Qiu, L. G. Shao, *et al.*, "Measurement of nitric oxide from cigarette burning using tdlas based on quantum cascade laser," Opt. Laser Technol. **124**, 105963 (2020).
23. H. T. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," Automatica **78**, 202–210 (2017).
24. S. Zhou, C. Y. Shen, L. Zhang, *et al.*, "Dual-optimized adaptive kalman filtering algorithm based on bp neural network and variance compensation for laser absorption spectroscopy," Opt. Express **27**(22), 31874–31888 (2019).
25. S. Zhou, N. W. Liu, C. Y. Shen, *et al.*, "An adaptive kalman filtering algorithm based on back-propagation (bp) neural network applied for simultaneously detection of exhaled co and n2o," Spectrochim. Acta, Part A **223**, 117332 (2019).
26. X. N. Liu, S. D. Qiao, G. W. Han, *et al.*, "Highly sensitive HF detection based on absorption enhanced light-induced thermoelastic spectroscopy with a quartz tuning fork of receive and shallow neural network fitting," Photoacoustics **28**, 100422 (2022).
27. Y. F. Ma, T. T. Liang, S. D. Qiao, *et al.*, "Highly Sensitive and fast hydrogen detection based on light-induced thermoelastic spectroscopy," Ultrafast Sci. **3**, 0024 (2023).
28. X. N. Liu and Y. F. Ma, "New temperature measurement method based on light-induced thermoelastic spectroscopy," Opt. Lett. **48**(21), 5687–5690 (2023).
29. X. F. Pan, Z. Zhang, H. Zhang, *et al.*, "A fast and robust mixture gases identification and concentration detection algorithm based on attention mechanism equipped recurrent neural network with double loss function," Sens. Actuators, B **342**, 129982 (2021).
30. Y. T. Yang, J. C. Jiang, J. F. Zeng, *et al.*, "Ch4, c2h6, and co2 multi-gas sensing based on portable mid-infrared spectroscopy and pca-bp algorithm," Sensors **23**(3), 1413 (2023).

31. Z. W. Liu, C. T. Zheng, T. Y. Zhang, *et al.*, "High-precision methane isotopic abundance analysis using near-infrared absorption spectroscopy at 100 torr," Analyst **146**(2), 698–705 (2021).
32. Y. C. Kim, H. G. Yu, J. H. Lee, *et al.*, "Hazardous gas detection for ftir-based hyperspectral imaging system using dnn and cnn," Proc. SPIE **10433**, 1043317 (2017).
33. J. Nalepa, M. Myller, and M. Kawulok, "Transfer learning for segmenting dimensionally reduced hyperspectral images," IEEE Geosci. Remote Sensing Lett. **17**(7), 1228–1232 (2020).
34. Q. S. Wen, T. Zhou, C. L. Zhang, *et al.*, "Transformers in time series: A survey," in *32th International Joint Conference on Artificial Intelligence (IJCAI)* (2023), pp. 6778–6786.
35. C. Zuo, J. M. Qian, S. J. Feng, *et al.*, "Deep learning in optical metrology: a review," Light: Sci. Appl. **11**(1), 39 (2022).
36. E. D'Andrea, S. Pagnotta, E. Grifoni, *et al.*, "An artificial neural network approach to laser-induced breakdown spectroscopy quantitative analysis," Spectrochimica Acta Part B: Atomic Spectroscopy **99**, 52–58 (2014).
37. M. Drozdzal, E. Vorontsov, G. Chartrand, *et al.*, "The importance of skip connections in biomedical image segmentation," in *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis* (2016), pp. 179–187.
38. A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," in *31th Advances in Neural Information Processing Systems (NeurIPS)* (2017), pp. 5998–6008.
39. B. Sun, P. Patimisco, A. Sampaolo, *et al.*, "Light-induced thermoelastic sensor for ppb-level $h_2s$ detection in a $sf_6$ gas matrices exploiting a mini-multi-pass cell and quartz tuning fork photodetector," Photoacoustics **33**, 100553 (2023).