AirNet: A Calibration Model for Low-Cost Air Monitoring Sensors Using Dual Sequence Encoder Networks

Haomin Yu,[1] **Qingyong Li,**[1][*] **Yangli-ao Geng,**[1] **Yingjun Zhang,**[1] **Zhi Wei**[2]

[1]Beijing Key Laboratory of Transportation Data Analysis and Mining, Beijing Jiaotong University, Beijing, China
[2]New Jersey Institute of Technology, Newark, NJ 07102, USA
{17112085, liqy, gengyla, zhangyj}@bjtu.edu.cn, zhiwei@njit.edu

Abstract

Air pollution monitoring has attracted much attention in recent years. However, accurate and high-resolution monitoring of atmospheric pollution remains challenging. There are two types of devices for air pollution monitoring, i.e., static stations and mobile stations. Static stations can provide accurate pollution measurements but their spatial distribution is sparse because of their high expense. In contrast, mobile stations offer an effective solution for dense placement by utilizing low-cost air monitoring sensors, whereas their measurements are less accurate. In this work, we propose a data-driven model based on deep neural networks, referred to as AirNet, for calibrating low-cost air monitoring sensors. Unlike traditional methods, which treat the calibration task as a point-to-point regression problem, we model it as a sequence-to-point mapping problem by introducing historical data sequences from both a mobile station (to be calibrated) and the referred static station. Specifically, AirNet first extracts an observation trend feature of the mobile station and a reference trend feature of the static station via dual encoder neural networks. Then, a social-based guidance mechanism is designed to select periodic and adjacent features. Finally, the features are fused and fed into a decoder to obtain a calibrated measurement. We evaluate the proposed method on two real-world datasets and compare it with six baselines. The experimental results demonstrate that our method yields the best performance.

Introduction

Air pollution has become a common problem in the process of urbanization and industrialization (Lin and Zhu 2018). It may lead to various respiratory problems including bronchitis, emphysema and asthma (Mabahwi, Leh, and Omar 2014). Thus, the demand for monitoring air quality is becoming increasingly essential. Generally, static stations provided by the government are equipped with instruments specialised for monitoring a number of air pollutants, such as carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), ozone (O_3) and particulate matter (PM10 and PM2.5) (Kumar et al. 2015). However, the spatial distributions of these static stations are sparse due to the high cost for installation and maintenance.
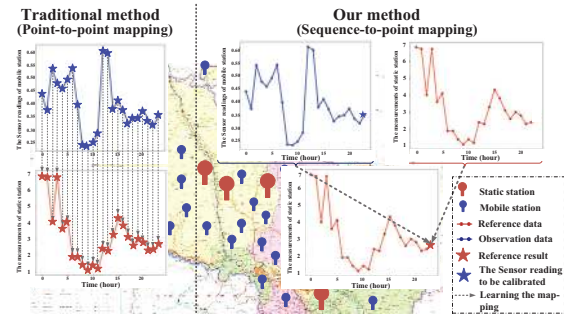
Figure 1: A sketch map of calibration task. The traditional methods formulate the calibration model as a point-to-point regression model. But our method introduces historical sequence data of static station and mobile station to formulate the calibration problem as sequence-to-point mapping problem.

To complement the accurate but costly static stations, the field has developed low-cost, mobile-enable air monitoring devices that are suitable for dense placement (Devarakonda et al. 2013; Piedrahita et al. 2014; Hasenfratz et al. 2015). Nevertheless, these devices are known to suffer from weak metrological characteristics, which makes their measurements less reliable (Spinelle et al. 2017). To improve accuracy of these devices, many scholars have developed low-cost calibration methods, which mainly formulate calibration task as a point-to-point regression problem (Spinelle et al. 2014; 2015; 2017; Saukh, Hasenfratz, and Thiele 2015; Lin et al. 2015; Barcelo-Ordinas et al. 2018; 2019). As shown in Fig.1, the point-to-point calibration methods are to establish a mapping from a sensor reading (to be calibrated) to a reference value. Spinelle et al. (Spinelle et al. 2014; 2015; 2017) utilized linear regression, multivariate linear regression and artificial neural networks to calibrate low-cost sensor. Saukh et al. (Saukh, Hasenfratz, and Thiele 2015) proposed using geometric mean regression for calibration. Their algorithm is especially optimized for calibrating large networks of heterogeneous sensors with different noise characteristic. Lin et al. (Lin et al. 2015) calibrated Aeroqual series 500 portable gas sensors by using ordinary least square

linear regression and major axis linear regression. Barcelo-Ordinas et al. (Barcelo-Ordinas et al. 2018) utilized multivariate estimation techniques to calibrate the gas sensor devices. Although various methods have been proposed to overcome the weaknesses of low-cost sensors, the calibration of low-cost gas sensors in actual deployment remains challenging due to several factors as summarized below.

- **Cross interference issue.** Every sensor faces the problem of cross interference. For example, the readings of a sensor for monitoring O_3 depends not only on the O_3 concentrations, but also on the other air pollutants and environmental factors (e.g. CO, temperature and humidity) (Barcelo-Ordinas et al. 2019; Anusankari et al. 2019; Lewis et al. 2016). This issue can lead to a relatively high degree of sensor response drift (Smith et al. 2019).

- **Numerical scale difference.** Different measurement modes bring on different numerical scales of the readings between static stations and mobile stations. For example, the measurement range of TGS 5042 A00 manufactured by Figaro for monitoring CO is 0-10000 μmol/mol (Spinelle et al. 2017). However, the measurement range of static station for CO is 0-150 mg/m^3. Furthermore, due to the the influence of the cross interference, the difference of numerical scales exhibits dynamic and complex nonlinearities. Such difference would increase the difficulty of calibration.

- **Limited available features.** In the actual deployment of sensors, the atmospheric environment is uncontrollable. There are many factors that affect the sensor readings, such as temperature, air pressure and sensor life. However, restricted by the equipment of mobile stations, only limited features are available, as described in Table 1.

To overcome the above challenges, we propose to exploit historical sequences of both the mobile station and the static station for optimizing the calibration task. Specifically, we convert the point-to-point regression problem into a sequence-to-point mapping problem, which will learn the mapping from sequence data captured from both the mobile station and the static station to a reference result. In particular, we propose AirNet, a low-cost air pollution monitoring sensor calibration model, to solve the sequence-to-point calibration problem. AirNet first extracts reference trend feature of the static station and observation trend feature of the mobile station via dual encoder neural networks. Next, to better integrate social related information, a social-based guidance mechanism is introduced to select periodic and adjacent similarity features. Finally, these features are fused and fed into a decoder to produce the calibration results. AirNet has demonstrated superior performance for low-cost sensor calibration. Our major contributions are summarized as follows:

- We formulate the low-cost sensor calibration task as a sequence-to-point mapping problem instead of the point-to-point regression adopted by the established studies. The new formulation empowers our model to exploit underlying temporal correlations between the source domain and the target domain, which is demonstrated to be useful in improving the calibration results.

- We propose dual sequence encoder networks for the sequence-to-point calibration task. The proposed networks allow to combine the sequence information of both a mobile station and the referred static station, which proves effective in withstanding numerical scale deviations.

- We design a social-based guidance mechanism to mine the periodic and adjacent correlation of pollutants underlying social regularities and further to assist the calibration task.

Problem Reformulation

In this section, we first introduce the initial features that are available from the air monitoring stations, including a mobile station to be calibrated and a referred static station, and then present our novel formulation for the monitoring sensor calibration problem.

Initial Features

Given a target pollutant (CO or O_3), its initial features include internal features associated with the aimed pollutant itself, and external features affecting sensor readings of the target pollutant.

The internal features consist of two parts. The first part is about the sensor readings of the aimed pollutant obtained from the mobile station.The second part is about the historical reference data of the aimed pollutant of the static station.

The external features, which refer to the features of other factors that affect the target pollutant sensor readings, are acquired by other sensors equipped in the mobile station. The external features comprise three parts.

- *Other gaseous pollutant measurements.* The mobile stations are equipped with monitoring sensors for other gaseous pollutants , including O_3 (or CO), NO_2, SO_2 and particulate matters (PM10 and PM2.5). Thus, we utilize the readings of these sensors to solve cross interference phenomenon up to a point.

- *Climate related indicators.* The mobile stations can also provide the temperature and humidity information, which is introduced into the initial features since it is related to the sensor sensitivity.

- *Empiric Features.* Some features derived from experience are added to our scenario to enhance our solution.

The key information and notations of initial features are presented in Table 1.

Problem Formulation

Linear regression has been commonly used for the calibration task in many existing studies (Spinelle et al. 2014; 2015; 2017). Basically, these regression methods treat the mobile value to be calibrated as independent variables, and the target static value as dependent variables. Suppose we observe data pairs at n time points and call them $(x_t^{\text{int}}, y_t^{\text{int}}), t = 1, ..., n$, where x_t^{int} is the mobile station reading, and y_t^{int} is the static station measurement. We can

Table 1: Key information and notations of initial features.

| Types of features	Notation	Feature source	Feature	Data descriptions
Internal features (IF)	$x^{\text{int}} \in \mathbb{R}$	Mobile station (Observation data)	Aimed pollutant	CO (or O_3)
	$y^{\text{int}} \in \mathbb{R}$	Static station (Reference data)	Aimed pollutant	CO (or O_3)
External features (EF)	$\boldsymbol{x}^{\text{ext}} \in \mathbb{R}^{l_{\text{ext}}}$	Mobile station (Observation data)	Other gaseous pollutants	O_3 (or CO), NO_2, SO_2, PM2.5, PM10
			Weather related data	Humidity, Temperature
			Empiric Features	PM05N, PM1N, PM25N, PM10N, PM_ZUFEN, PMN_SUM, PM10N_ratio, PM25N_ratio
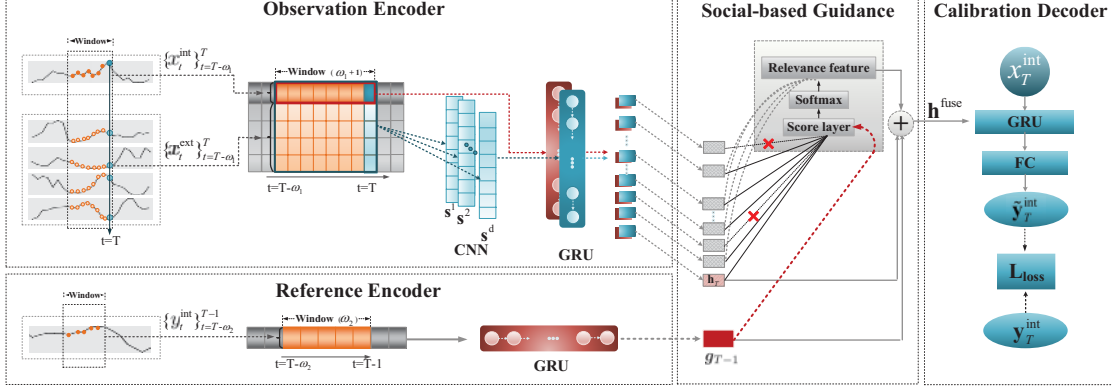
Figure 2: Framework of the proposed AirNet.

describe the underlying relationship between y_t^{int} and x_t^{int} with an error term ε_t by a linear regression (LR) model

$$y_t^{\text{int}} = \beta_0 + \beta x_t^{\text{int}} + \varepsilon_t. \qquad (1)$$

We can include external features as covariates ($\boldsymbol{x}^{\text{ext}}$) and extend the above LR model to a multiple linear regression (MLR) model

$$y_t^{\text{int}} = \beta_0 + \beta^{\text{int}} x_t^{\text{int}} + \boldsymbol{\beta}^{\text{ext}} \boldsymbol{x}_t^{\text{ext}} + \varepsilon_t. \qquad (2)$$

More generally, we are interested in learning a mapping function F

$$\tilde{y}_t^{\text{int}} = F_{\Theta}(x_t^{\text{int}}, \boldsymbol{x}_t^{\text{ext}}), \qquad (3)$$

where Θ is the collection of the F function parameters. We will estimate Θ by minimizing a certain loss function L

$$\begin{aligned}
\tilde{\Theta} &= \arg\min_{\Theta} \sum_t L(y_t^{\text{int}}, \tilde{y}_t^{\text{int}}) \\
&= \arg\min_{\Theta} \sum_t L(y_t^{\text{int}}, F_{\Theta}(x_t^{\text{int}}, \boldsymbol{x}_t^{\text{ext}})).
\end{aligned} \qquad (4)$$

The loss function can be, for example, Mean Square Error (MSE) , and Mean Absolute Error (MAE). Most existing methods differ at specifying different forms of F. For instance, F can be a linear function for regression methods, or more complex functions defined by artificial neural networks, decision trees, etc. We call all these existing approaches as point-to-point methods, which mean that when making calibration at a certain time point, we only consider the input and the output are the same time point.

However, we note that such a point-to-point regression may not be optimal because it ignores the (strong) temporal dependency among data. For example, concentrations of

atmospheric pollutants have a certain degree of periodicity and continuity affecting by social activities. To exploit the temporal dependency of data for improving calibration accuracy, we propose to integrate a sequence of historical mobile features within a time window of length ω_1, namely, $\{x_t^{\text{int}}\}_{t=T-\omega_1}^{T}$ and $\{\boldsymbol{x}_t^{\text{ext}}\}_{t=T-\omega_1}^{T}$. At the same time, to overcome the numerical scale difference, we also consider a sequence of historical static features within a time window of length ω_2 ($\{y_t^{\text{int}}\}_{t=T-\omega_2}^{T-1}$). In contrast to the point-to-point regression, we formulate the calibration task as a sequence-to-point problem. Formally, we are interested in learning a mapping function F

$$\tilde{y}_T^{\text{int}} = F_{\Theta}(\{x_t^{\text{int}}\}_{t=T-\omega_1}^{T}, \{\boldsymbol{x}_t^{\text{ext}}\}_{t=T-\omega_1}^{T}, \{y_t^{\text{int}}\}_{t=T-\omega_2}^{T-1}). \qquad (5)$$

Adding the two sequences of historical data would be helpful, but it is not trivial to model them. We expect the mapping function F be complex for the calibration task, so deep neural networks may be an ideal candidate for it. Therefore, we develop novel dual sequence encoder networks to accommodate the sequence data, as detailed in the following section.

The Model

As shown in Fig. 2, the proposed AirNet consists of four modules including an observation encoder, a reference encoder, a social-based guidance module and a calibration decoder. We detail each module as follows.

Observation Encoder

The observation encoder is in charge of extracting the observation information derived from the mobile station. This
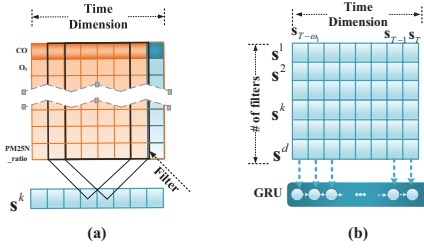
Figure 3: The operation of convolution layer. (a) The setting of the convolution filter height; (b) The concatenation operation after convolution.
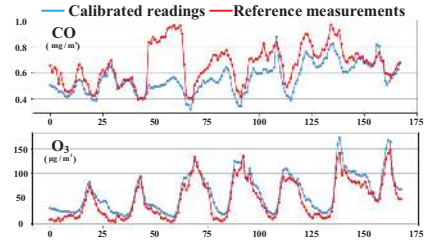

Figure 4: Calibrated readings without introducing the static station information. The horizontal axis represents time, and the vertical axis represents the concentration value.

module can capture observation trend and contribute to alleviate cross interference. To obtain the time series features of the sensor to be calibrated, we feed its mobile internal features $\{x_t^{\text{int}}\}_{t=T-\omega_1}^{T}$ into a gated recurrent unit (GRU) (Cho et al. 2014) :

$$\{\boldsymbol{h}_t^1\}_{t=T-\omega_1}^{T} = GRU^{\text{basic}}(\{x_t^{\text{int}}\}_{t=T-\omega_1}^{T}), \qquad (6)$$

where $\boldsymbol{h}_t^1$ denotes the hidden state at time t. The last hidden state $\boldsymbol{h}_T^1$, which we call the basic trend feature, encapsulates the trend information of mobile internal features. However, it does not take into account the cross interference challenge.

To handle the dynamic cross interference influence, we stitch the mobile internal and external features within a time window ω_1 into a matrix $\boldsymbol{X} \in \mathbb{R}^{(l_{\text{ext}}+1) \times (\omega_1+1)}$, where l_{ext} denotes the dimension of the mobile external features. Then, a convolution layer is applied on this matrix to learn the dynamic cross interference among features. Specifically, we utilize d filters with height h and width c to sweep through the input matrix $\boldsymbol{X}$. Note that h is set to $l_{ext}+1$, as shown in Fig. 3 (a), so as to model the cross interference among all mobile features. The process can be formulated as:

$$\boldsymbol{s}^k = \phi(\boldsymbol{W}_k * \boldsymbol{X} + \boldsymbol{b}_k), k = 1, \ldots, d, \qquad (7)$$

where $*$ represents the convolutional operation, and ϕ is an activation function (ReLU is utilized in our work). We treat $\{\boldsymbol{s}^k\}_{k=1}^{d}$ as row vectors and stack them into a matrix $\boldsymbol{S} \in \mathbb{R}^{d \times (\omega_1+1)}$ (See Fig. 3 (b)) . The t-th columns $\boldsymbol{s}_t$ of $\boldsymbol{S}$ encompasses the cross inference information of time t.

Then we send $\{\boldsymbol{s}_t\}_{t=T-\omega_1}^{T}$ into a GRU to code their temporal information:

$$\{\boldsymbol{h}_t^2\}_{t=T-\omega_1}^{T} = GRU^{\text{cross}}(\{\boldsymbol{s}_t\}_{t=T-\omega_1}^{T}), \qquad (8)$$

where $\boldsymbol{h}_t^2$ represents the hidden state at time step t. $\boldsymbol{h}_t^2$ encloses dynamic cross interference trend information, so we call it cross trend feature. Then, we concatenate the basic trend feature and the cross trend feature at time step t:

$$\boldsymbol{h}_t = [\boldsymbol{h}_t^1; \boldsymbol{h}_t^2], t = T - \omega_1, \ldots, T. \qquad (9)$$

where $\boldsymbol{h}_t$ represents final hidden state of moment t extracted by the observation encoder. Particularly, we call the hidden state of the last moment (i.e. $\boldsymbol{h}_T$) *observation trend feature*, which encapsulates the trend information of all observation

features. The observation trend feature contains not only the rising and falling trend of the aimed pollutant, but also the cross interference information. However, it cannot obtain the reliable numerical scale of the referred measurements. As a solution, we introduce the historical measurement sequence of the referred static station to acquire this information.

Reference Encoder

The reference encoder is responsible for encoding historical information of the target pollutant from the referred static station. As shown in Fig. 4, without the historical information of the static station (i.e. only the observation encoder is utilized), the calibrated readings inevitably occur numerical scale shift. Thus, we design a simple yet effective prediction strategy to obtain useful information of the referred static station containing reliable numerical scales.

Specifically, a historical sequence from the static station $\{y_t^{\text{int}}\}_{t=T-\omega_2}^{T-1}$ is fed into a GRU to obtain reference features:

$$\{\boldsymbol{g}_t\}_{t=T-\omega_2}^{T-1} = GRU^{\text{global}}(\{y_t^{\text{int}}\}_{t=T-\omega_2}^{T-1}), \qquad (10)$$

where $\boldsymbol{g}_t$ denotes the hidden state of moment t. We call the hidden state of the last moment (i.e. $\boldsymbol{g}_{T-1}$) *reference trend feature*, which encodes the dynamic scale information of the referred measurements. The reference trend feature can not only reveal the inherent regularity of the target pollutant, but also compensate for numerical scale difference.

Social-based Guidance Module

In this module, we first present the motivation and details of the social-guidance mechanism. Then, a feature fusion operation is introduced to balance extracted features.

Social-based Guidance Mechanism Social regularities generally have periodicity and continuity, such as going to work on weekdays, resting on weekends, sleeping at night and working during the day. Air pollutant concentrations may sometimes be affected by social regularities. As shown in the example in Fig. 5, the concentrations of O_3 at 22:00 demonstrate a strong correlation among weekdays, while there is no evident correlation between weekdays and weekends. Meanwhile, the concentrations of adjacent times also show similarity due to the continuity of social activities. However, periodic and adjacent similarity relationships do
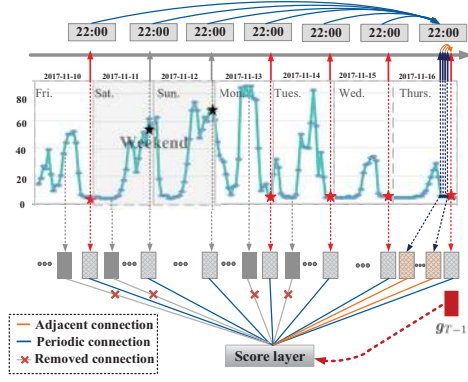
Figure 5: The illustration of social-based guidance mechanism.

not necessarily exist at all times. Thus, the proposed module is expected to be able to learn this information automatically.

To make full use of social knowledge, the periodic and adjacent similarities are treated as a priori strategy to filter the final hidden states obtained by observation encoder, as shown in Fig. 5. Then the reference trend feature is utilized to automatically learn the impact of the filtered hidden states. Specifically, the impact of the filtered hidden state at time t is calculated based upon the correlation with the reference trend feature $\boldsymbol{g}_{T-1}$:

$$\boldsymbol{I}_a = [T, T-1, \cdots, T-(l_a-1), T-l_a],$$
$$\boldsymbol{I}_p = [T, T-p, \cdots, T-(l_p-1)\cdot p, T-l_p\cdot p],$$
$$a_t = \sigma(\boldsymbol{v}^\top \tanh(\boldsymbol{W}\boldsymbol{g}_{T-1} + \boldsymbol{U}\boldsymbol{h}_t)), t \in \boldsymbol{I}_a \cup \boldsymbol{I}_p, \quad (11)$$
$$\boldsymbol{r} = \sum_{t\in\boldsymbol{I}_a} a_t\boldsymbol{h}_t + \sum_{t\in\boldsymbol{I}_p} a_t\boldsymbol{h}_t,$$

where $\boldsymbol{I}_a$ and $\boldsymbol{I}_p$ denote adjacent time slot set and periodic time slot set, respectively. p is periodic interval, which is set as 24 according to the 24-hour biological cycle. l_a and l_p represent the number of adjacent hidden states and the number of periodic hidden states (l_a is set as 6, and l_p is equal to $\lfloor\frac{\omega_1+1}{24}\rfloor$), respectively. $\boldsymbol{W}, \boldsymbol{U}, \boldsymbol{v}$ are the learnable parameters, σ is the softmax function, and $\boldsymbol{a}_t$ represents the impact score between reference trend feature and hidden state at time t. We call $\boldsymbol{r}$ *relevant feature* which encloses periodic information and adjacent information.

Feature Fusion The observation encoder can capture rising and falling trend information, but fails to obtain the correct numerical scale information. The reference encoder can compensate numerical scale difference, yet it is difficult to predict sudden changes of air pollutant concentrations. Furthermore, the social-based attention mechanism can mine potential information related to social regularities. For taking advantage of complementary information acquired by each module, the observation trend feature $\boldsymbol{h}_T$, the reference trend feature $\boldsymbol{g}_{T-1}$ and the relevant feature $\boldsymbol{r}$ are fused via a fully connected layer:

$$\boldsymbol{h}^{\text{fuse}} = FC^{\text{fuse}}([\boldsymbol{r}; \boldsymbol{g}_{T-1}; \boldsymbol{h}_T]) = \boldsymbol{W}_f[\boldsymbol{r}; \boldsymbol{g}_{T-1}; \boldsymbol{h}_T] + \boldsymbol{b}_f, \quad (12)$$

Table 2: The detailed settings of AirNet.

Module	Parameter	Value
Observation encoder	Filter size and number of CNN	16×5,100
	Stride of CNN	1
	Hidden dimension of GRU$^{\text{basic}}$	50
	Hidden dimension of GRU$^{\text{cross}}$	80
Reference encoder	Hidden dimension of GRU$^{\text{global}}$	50
Social-based guidance	Output dimension of FC$^{\text{fuse}}$	120
Calibration decoder	Output dimension of FC$^{\text{dec}}$	1

Table 3: The statistics of two datasets. Note that the number of instances is different because of missing values and error data.

Dataset	Lanzhou dataset	Fuzhou dataset
Target sensors	CO and O$_3$	CO and O$_3$
Time spans	2017/11/2-2018/10/2	2017/11/2-2018/10/2
Time intervals	1 hour	1 hour
# of instances	7674	7753

where $\boldsymbol{W}_f$ and $\boldsymbol{b}_f$ are the weight and bias to be learned.

Calibration Decoder

After the feature fusion, we feed the fused vector $\boldsymbol{h}^{\text{fuse}}$ along with the mobile internal feature x_T^{int} into a decoder to calibrate the readings of aimed sensor with:

$$\tilde{y}_T^{\text{int}} = FC^{\text{dec}}(f(\boldsymbol{h}^{\text{fuse}}, x_T^{\text{int}})) = \boldsymbol{w}_d^T f(\boldsymbol{h}^{\text{fuse}}, x_T^{\text{int}}) + b_d, \quad (13)$$

where f represents a GRU cell. $\boldsymbol{w}_d$ and b_d are learnable parameters.

Implementation Details

The proposed method is implemented with Pytorch, and the code is made publicly available[1]. The model is trained by minimizing the mean absolute loss function.

During the training procedure, the learning rate is set as 0.001 and the batch size is set as 128. The model parameters are optimized by Adam optimizer (Kingma and Ba 2014). To prevent overfitting, we further apply the dropout method with probability 0.2. The historical sequence window lengths of the mobile station and the static station (i.e. ω_1 and ω_2) are set to 167 and 24, respectively. For clarity, Table 2 summarizes other parameter settings in AirNet.

Experiments

In this section, we evaluate the performance of the proposed AirNet model. All experiments are conducted on a 64-bit Ubuntu 16.04 computer with 18 Intel 2.68GHz CPUs, 256 GB memory, and 8 NVIDIA TITAN X GPUs. All methods are repeated five times, the optimal result for each method is selected as its final result.

Setting

Datasets Experiments were carried out over two different datasets, as shown in Table 3. Meanwhile, we divide the total dataset by a ratio of 8:1:1 for training, validation and testing

---

[1]https://github.com/yuhaomin/AirNet

Table 4: The feature details of baselines and AirNet. IF_M, IF_S and EF_M denote mobile internal feature, static internal feature and mobile external features, respectively. CF and HF represent feature at current time (time to be calibrated) and historical sequence features, respectively.

Definition	Method	IF_M		IF_S		EF_M	
		CF	HF	CF	HF	CF	HF
Point-to-point	LR	✓	-	-	-	-	-
	MLR	✓	-	-	-	✓	-
	XGBoost	✓	-	-	-	✓	-
Sequence-to-point	ELC	✓	✓	-	-	-	-
	LSTNet	✓	✓	-	-	✓	✓
	DA-RNN	✓	✓	-	✓	✓	✓
	AirNet	✓	✓	-	✓	✓	✓

in chronological order. Here we briefly introduce these two datasets as below:

- **Lanzhou dataset.** Lanzhou is one of the most seriously air-polluted cities in China, due to its special valley topography and industrial structure (Tao et al. 2014). Note that the higher level of air pollution, the more serious the influence of the cross interference among sensor readings.

- **Fuzhou dataset.** Fuzhou has always been ranked ahead of China's cities in air quality, so the cross interference factors have less effects on the readings of the sensors in Fuzhou.

Evaluation Metrics   We evaluate the calibration quality of all the methods under test by two measures:

- Symmetric mean absolute percentage error (SMAPE)

$$\text{SMAPE} = \frac{100\%}{\tau} \sum_{j=1}^{\tau} \left(\frac{|y_{T,j}^{\text{int}} - \tilde{y}_{T,j}^{\text{int}}|}{(|y_{T,j}^{\text{int}}| + |\tilde{y}_{T,j}^{int}|)/2} \right), \quad (14)$$

- Mean absolute error (MAE)

$$\text{MAE} = \frac{1}{\tau} \sum_{j=1}^{\tau} |y_{T,j}^{\text{int}} - \tilde{y}_{T,j}^{\text{int}}|, \quad (15)$$

where τ refers to the number of testing sequences. $y_{T,j}^{\text{int}}$ and $\tilde{y}_{T,j}^{\text{int}}$ represent reference measurement and the calibration result for the j-th testing sequence, respectively.

Methods for Comparison

We compare AirNet with the following baselines.

- **LR**: Linear regression (LR) is a linear approach to model the relationship between a scalar response and a reference result, and it is widely applied in sensor calibration field (Spinelle et al. 2014; 2015; 2017).

- **MLR**: Multiple linear regression (MLR), which is commonly used for sensor calibration (Spinelle et al. 2014), attempts to learn the mapping from two or more explanatory variables to a reference result.

- **XGBoost**: XGBoost (Chen and Guestrin 2016) is an implementation of gradient boosted decision trees, which shows strong competitiveness in many competitions. The XGBoost of Scikit-Learn with default parameters is employed.

Table 5: Calibration results for the Lanzhou dataset and Fuzhou dataset.

Method	Lanzhou Dataset				Fuzhou Dataset			
	CO		O$_3$		CO		O$_3$	
	SMAPE	MAE	SMAPE	MAE	SMAPE	MAE	SMAPE	MAE
LR	46.47%	0.467	62.1%	15.069	24.15%	0.133	48.88%	26.099
MLR	29.9%	0.256	59.77%	13.469	18.56%	0.102	28.6%	14.174
XGBoost	27.14%	0.219	48.8%	12.411	20.03%	0.110	32.38%	16.42
ELC	32.54%	0.300	54.41%	13.931	23.05%	0.126	41.86%	20.739
LSTNet	34.36%	0.279	59.57%	13.926	19.48%	0.106	24.25%	11.429
DA-RNN	17.42%	0.131	38.16%	9.662	5.55%	0.029	14.66%	6.925
AirNet	**15.82%**	**0.122**	**28.26%**	**6.911**	**3.92%**	**0.021**	**10.06%**	**4.261**

- **ELC**: Ensemble of LSTM & CNN (ELC) (Zhao et al. 2018) leverages time-frequency features and global features extracted by LSTM.

- **LSTNet**: LSTNet (Lai et al. 2018) proposes a novel recurrent-skip component that leverages the periodic pattern to capture long-term dependencies.

- **DA-RNN**: DA-RNN (Qin et al. 2017) is a dual-staged attention model for time series prediction problem.

Note that ELC, LSTNet and DA-RNN were originally utilized to solve the time series prediction problem, and we apply them to our sensor calibration scenarios. Their parameters are set based on the recommendations in their original papers. The initial feature settings for all methods are shown in Table 4, where the initial feature setting for each baseline is set according to its own architecture. Furthermore, to fully study the performance gained from each component of AirNet, four other versions of AirNet are implemented:

- **AirNet-O**: An incomplete AirNet in which the observation encoder is removed. Note that AirNet-O turns the calibration task into a prediction task. After the observation decoder is removed, the guidance mechanism is not available. Thus, the guidance mechanism is removed.

- **AirNet-R**: An incomplete AirNet in which the reference encoder is removed. Namely, the reference trend feature is removed. Note that after the reference decoder is removed, the guidance mechanism is not available. Thus, the guidance mechanism is also removed.

- **AirNet-C**: An incomplete AirNet in which the convolution part is removed. That is, the cross trend feature is removed.

- **AirNet-G**: An incomplete AirNet in which the social-based guidance mechanism is removed.

Experimental Results

Table 5 and Table 6 present the numerical results of baselines and ablation studies, respectively. Figure 6 visualizes the results of two competitive baselines and AirNet. Next, we will analyze the experimental results in detail.

Comparison results with Baselines   The left part of Table 5 shows that our AirNet achieves the best performance for the Lanzhou dataset. The calibration results of most baseline methods are not ideal, demonstrating the challenge of calibration in heavy air pollution scenario. By comparison, the methods that utilize mobile external features generally
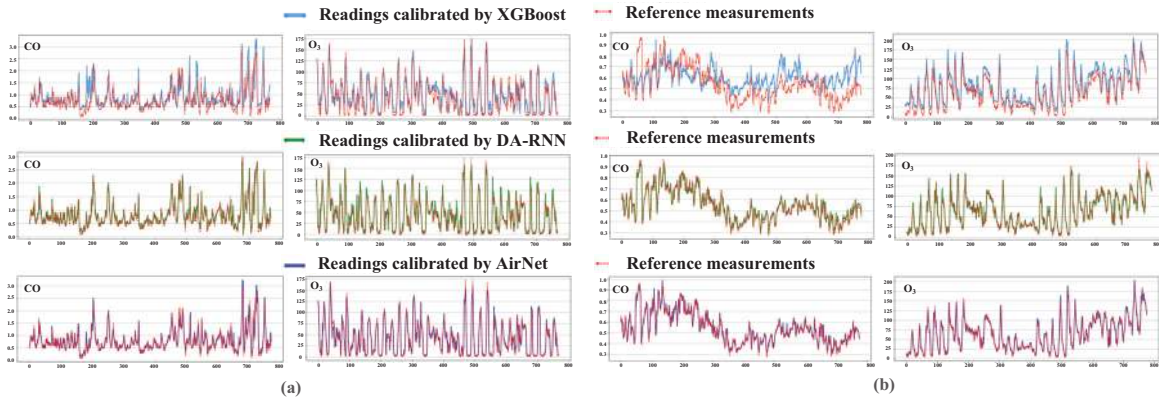
Figure 6: Calibration results for different methods applied to the (a) Lanzhou dataset and (b) Fuzhou dataset. (The horizontal axis represents time, and the longitudinal indicates the concentration value.)

Table 6: Ablation study results for the Lanzhou dataset and Fuzhou dataset.

Method	Lanzhou Dataset				Fuzhou Dataset			
	CO		O_3		CO		O_3	
	SMAPE	MAE	SMAPE	MAE	SMAPE	MAE	SMAPE	MAE
AirNet-O	17.02%	0.131	28.29%	7.908	5.29%	0.028	13.3%	6.25
AirNet-R	28.09%	0.225	41.91%	10.825	19.2%	0.106	23.06%	11.494
AirNet-C	16.09%	0.124	27.39%	7.489	4.23%	0.024	10.43%	4.391
AirNet-G	16.87%	0.128	**26.84**%	6.976	3.98%	**0.021**	11.2%	5.043
AirNet	**15.82**%	**0.122**	28.26%	**6.911**	**3.92**%	0.021	**10.06**%	**4.261**

obtain better calibration results than methods that do not introduce external features (LR and ELC). Among them, XG-Boost, which is suitable for modeling the nonlinear relationships, is superior to linear models such as LR and MLR. Besides, LSTNet does not perform well because it utilizes skip-links for all hidden states to capture long-term dependencies and periodic pattern, which may introduce much noise. DA-RNN, which introduces the features captured from both the mobile station and the static station, achieves competitive performance compared with AirNet. However, it cannot handle the sudden changes of pollutant concentrations very well, as shown in Fig. 6(a). The proposed AirNet obtains the best results, because it considers both information from two types of stations and the influence of social regularities.

The calibration performance for the Fuzhou dataset is shown in the right part of Table 5. The calibration results of Fuzhou are generally better than that of Lanzhou. This is because the sensor in Fuzhou is less affected by the cross interference phenomenon. More specifically, LR and ELC, which only take the mobile internal feature into account, do not perform well. In comparison, MLR, XGBoost and LST-Net achieve improved performance since they consider external features. As a point-to-point calibration method, although XGBoost achieves small calibration errors for CO, its deviation is actually large as shown in Fig. 6(b). DA-RNN is competitive compared with AirNet. On the whole, the results of AirNet are superior to the baseline methods, and AirNet appears suitable for different sensors in different regions.

Ablation Study   Table 6 lists ablation studies for the five configurations on the two datasets. Both AirNet-O and AirNet-R are used to test the importance of the reference encoder. Note that AirNet-O only retains the reference encoder but AirNet-R removes the reference encoder. By comparing the performance of two methods, it is clear that the prediction strategy designed to introduce historical static internal features plays a crucial role in our methods. AirNet-C, which removes the convolutional layer, has slightly worse performance. This is because it ignores the effect of cross interference on the sensor readings. For the test of the contribution of the social-based guidance mechanism, we can see from the results that AirNet works better than AirNet-G for calibrating CO sensor. However, the advantage of Air-Net is not obvious when calibrating O_3 for Lanzhou, since the O_3 sensor equipped in Lanzhou is subject to large external interference.

In summary, this ablation study justifies the efficacy of our architecture design. All components have contributed to the excellent performance of AirNet.

Conclusions

In this paper, we presented a new formulation (i.e. sequence-to-point mapping) for calibrating low-cost air monitoring sensors. With the new formulation, we proposed an end-to-end neural network model (AirNet) to solve the calibration problem. AirNet exploits dual sequence encoder networks to extract temporal information for both the static station and the mobile station, which proves useful in improving the calibration results. Furthermore, inspired by the temporal regularity of social activities, we designed a social-based guidance mechanism, which reveals the periodic and adjacent correlations among the features of different times and further enhances the calibration performance. The experimental results on two real-world datasets verified the effectiveness of the proposed formulation and method.

In future research, there are some topics that we will investigate. Firstly, we will study spatio-temporal calibration models by integrating multiple near mobile stations. Sec-

ondly, we will try to deploy AirNet on the web server to provide calibration service for low-cost sensors.

Acknowledgments

References

Anusankari, S.; Balaji Ganesh, A.; Subasri, R.; and Deepa, N. 2019. Optical determination of carbon dioxide and oxygen by a fluorescent membrane to evaluate the freshness of meat products. *Instrumentation Science & Technology* 1–26.

Barcelo-Ordinas, J. M.; Garcia-Vidal, J.; Doudou, M.; Rodrigo-Muñoz, S.; and Cerezo-Llavero, A. 2018. Calibrating low-cost air quality sensors using multiple arrays of sensors. In *Proceedings of the 2018 IEEE Wireless Communications and Networking Conference*, 1–6. IEEE.

Barcelo-Ordinas, J. M.; Doudou, M.; Garcia-Vidal, J.; and Badache, N. 2019. Self-calibration methods for uncontrolled environments in sensor networks: A reference survey. *Ad Hoc Networks* 88:142–159.

Chen, T., and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd Acm SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. ACM.

Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Devarakonda, S.; Sevusu, P.; Liu, H.; Liu, R.; Iftode, L.; and Nath, B. 2013. Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 15. ACM.

Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; and Thiele, L. 2015. Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing* 16:268–285.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kumar, P.; Morawska, L.; Martani, C.; Biskos, G.; Neophytou, M.; Di Sabatino, S.; Bell, M.; Norford, L.; and Britter, R. 2015. The rise of low-cost sensing for managing air pollution in cities. *Environment International* 75:199–205.

Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. 2018. Modeling long-and short-term temporal patterns with deep neural networks. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, 95–104. ACM.

Lewis, A. C.; Lee, J. D.; Edwards, P. M.; Shaw, M. D.; Evans, M. J.; Moller, S. J.; Smith, K. R.; Buckley, J. W.; Ellis, M.; Gillot, S. R.; et al. 2016. Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday discussions* 189:85–103.

Lin, B., and Zhu, J. 2018. Changes in urban air quality during urbanization in China. *Journal of Cleaner Production* 188:312–321.

Lin, C.; Gillespie, J.; Schuder, M.; Duberstein, W.; Beverland, I.; and Heal, M. 2015. Evaluation and calibration of aeroqual series 500 portable gas sensors for accurate measurement of ambient ozone and nitrogen dioxide. *Atmospheric Environment* 100:111–116.

Mabahwi, N. A. B.; Leh, O. L. H.; and Omar, D. 2014. Human health and wellbeing: Human health effect of air pollution. *Procedia-Social and Behavioral Sciences* 153:221–229.

Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R. P.; Lv, Q.; Hannigan, M.; et al. 2014. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques* 7(10):3325–3336.

Qin, Y.; Song, D.; Chen, H.; Cheng, W.; Jiang, G.; and Cottrell, G. 2017. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*.

Saukh, O.; Hasenfratz, D.; and Thiele, L. 2015. Reducing multi-hop calibration errors in large-scale mobile sensor networks. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 274–285. ACM.

Smith, K. R.; Edwards, P. M.; Ivatt, P. D.; Lee, J. D.; Squires, F.; Dai, C.; Peltier, R. E.; Evans, M. J.; Sun, Y.; and Lewis, A. C. 2019. An improved low-power measurement of ambient no 2 and o 3 combining electrochemical sensor clusters and machine learning. *Atmospheric Measurement Techniques* 1325–1336.

Spinelle, L.; Gerboles, M.; Villani, M. G.; Aleixandre, M.; and Bonavitacola, F. 2014. Calibration of a cluster of low-cost sensors for the measurement of air pollution in ambient air. *Sensors* 21–24.

Spinelle, L.; Gerboles, M.; Villani, M. G.; Aleixandre, M.; and Bonavitacola, F. 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. part a: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical* 215:249–257.

Spinelle, L.; Gerboles, M.; Villani, M. G.; Aleixandre, M.; and Bonavitacola, F. 2017. Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. part b: NO, CO and CO_2. *Sensors and Actuators B: Chemical* 238:706–715.

Tao, Y.; Mi, S.; Zhou, S.; Wang, S.; and Xie, X. 2014. Air pollution and hospital admissions for respiratory diseases in lanzhou, china. *Environmental Pollution* 185:196–201.

Zhao, Y.; Shen, Y.; Zhu, Y.; and Yao, J. 2018. Forecasting wavelet transformed time series with attentive neural networks. In *Proceedings of the 2018 IEEE International Conference on Data Mining*, 1452–1457. IEEE.