CatUA: Catalyzing Urban Air Quality Intelligence Through Mobile Crowd-Sensing

Nan Zhou, Yuxuan Liu, Haoyang Wang, Fanhang Man, Jingao Xu, *Member, IEEE*, Fan Dang, *Senior Member, IEEE*, Chaopeng Hong, Yunhao Liu, *Fellow, IEEE*, Xiao-Ping Zhang, *Fellow, IEEE*, Yali Song, Qiuhua Wang, and Xinlei Chen, *Member, IEEE*

***Abstract*—Mobile air pollution sensing methods have emerged to collect air quality data with improved spatial and temporal resolutions. However, existing methodologies struggle to effectively process spatially mixed gas samples due to the highly dynamic fluctuations experienced by sensors, resulting in significant measurement deviations. We identify an opportunity to address this issue by exploring potential patterns within sensor measurements. To this end, we propose CatUA, a novel city-scale fine-grained air quality estimation system designed to deliver accurate mobile air quality data. First, we design AirBERT, a representation learning model specifically aimed at discerning mixed gas concentrations from sensor data. Second, we implement a Prompt-informed Training Strategy that leverages extensive unlabeled and minimal labeled city-scale data to enhance the performance of CatUA. Notably, the Auto-Prompt mechanism allows CatUA to conveniently acquire new knowledge tailored to specific downstream tasks. To ensure the practicality of CatUA, we have invested considerable effort in developing the software stack on our meticulously crafted Sensing Front-end, which has successfully gathered city-scale air quality data for over 1,200 hours. Experiments conducted on the collected data demonstrate that CatUA reduces sensing errors by 96.9% with a latency of only 44.9 ms, outperforming the state-of-the-art baseline by 42.6%.**

Received 6 November 2024; revised 5 March 2025; accepted 7 April 2025. Date of publication 11 April 2025; date of current version 6 August 2025. This work was supported in part by National Key R&D program of China under Grant 2022YFC3300703, in part by the Yunnan Forestry and Grassland Science and Technology Innovation Joint Special Project 202404CB090017, in part by the Natural Science Foundation of China under Grant 62371269, in part by Guangdong Innovative and Entrepreneurial Research Team Program under Grant 2021ZT09L197, in part by Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan under Grant JC20220011, and in part by Meituan Academy of Robotics Shenzhen. An earlier version of this paper was presented in part at the Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services (ACM MobiSys 2024), [DOI: 10.1145/3643832.3661872]. Recommended for acceptance by C.H. Liu. *(Nan Zhou, Yuxuan Liu, and Haoyang Wang are co-first authors.) (Corresponding author: Xinlei Chen.)*

Nan Zhou, Yuxuan Liu, Haoyang Wang, Fanhang Man, Chaopeng Hong, Xiao-Ping Zhang, and Xinlei Chen are with Shenzhen International Graduate School, Tsinghua University, Beijing 100190, China (e-mail: zhoun24@mails.tsinghua.edu.cn; yuxuan-l21@mails.tsinghua.edu.cn; haoyang-22@mails.tsinghua.edu.cn; mfh21@mails.tsinghua.edu.cn; hongco@sz.tsinghua.edu.cn; xiaoping.zhang@sz.tsinghua.edu.cn; chen.xinlei@sz.tsinghua.edu.cn).

Jingao Xu and Yunhao Liu are with the School of Software and BNRist, Tsinghua University, Beijing 100190, China (e-mail: xujingao13@gmail.com; yunhao@greenorbs.com).

Fan Dang is with the School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China (e-mail: dangfan@bjtu.edu.cn).

Yali Song is with the College of Soil and Water Conservation, Southwest Forestry University, Kunming 650233, China (e-mail: songyali@swfu.edu.cn).

Qiuhua Wang is with the College of Civil Engineering, Southwest Forestry University, Kunming 650233, China (e-mail: qhwang2010@swfu.edu.cn).

Digital Object Identifier 10.1109/TMC.2025.3560120

***Index Terms*—AQI monitoring, mobile crowd-sensing, self-supervised learning, prompt training.**

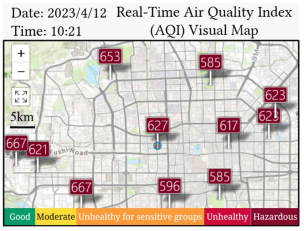
I. Introduction

In THE quest to safeguard public health, monitoring air quality emerges as a crucial undertaking. The World Health Organization has attributed over 7 million premature deaths worldwide to deteriorating air quality-related diseases [2]. Typically, city-scale air quality monitoring heavily relies on meteorological stations. However, these stations, limited in number and fixed in location, can only offer a coarse-grained view of urban air quality as depicted in Fig. 1(a). This limitation is particularly concerning given that major air pollutants (e.g., PM, VOCs, NO_x, CO) demonstrate significant dispersion effects over merely hundred-meter-scale distances. As a result, people lack access to fine-grained air quality information in their immediate living and working environments [3].

To enhance the granularity of air quality monitoring, a naive approach is to extensively deploy static air quality sensing nodes across the city (Fig. 1(b)). However, the high costs of deploying and maintaining such a widespread sensing system render it impractical for large-scale implementation [8], [9], [10]. In contrast, a more promising solution lies in mobile crowd-sensing (MCS), which utilizes crowd-sourced vehicles (e.g., taxis) equipped with sensing nodes (Fig. 1(b)) [11]. These vehicles continuously gather and report air pollutant concentration data and their sampling locations while moving, offering city-wide coverage and fine-grained measurements.

Albeit inspiring, vehicle mobility presents notable challenges for MCS in precise city-scale air quality monitoring. Commercial sensors, with a gas collection and response time of 30∼150 s,[1] may result in spatially mixed gas samples as vehicles move around 300 m (i.e., 40 km/h velocity). This issue is compounded of sensors response, gas mixing process and continuously creating blended measurements from adjacent locations with low accuracy. As depicted in Fig. 2, measurements from sensors on moving vehicles are average 170% less accurate than those from stationary ones. Therefore, it is imperative to

[1]We have analyzed the response time of air pollution sensors manufactured by 5 popular enterprises on the market, including City Tech [12], Alphasense [13], Figaro [14], Membrapor [15], and SGX Sensortech [16].



(a) The AQI of a major city in 2023 indicates hazardous effects on human bodies. There is only 1 official monitoring station per 468km².

(b) Static AQI systems: #1 AQMesh [4], #2 Eclipse [5], #3 ECOMSMART [6]. Mobile AQI systems: #4 ECOMTREK [7].

Fig. 1. Means of urban air quality monitoring: monitoring stations, static sensing, and mobile sensing.

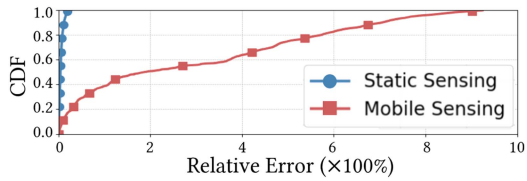


Fig. 2. The average relative errors for static and mobile sensing are 4% and 174%, respectively.

estimate accurate and fine-grained air quality data from those spatially mixed gas measurement time series [17].

Nowadays, Bidirectional Encoder Representations from Transformers (BERT) has proven effective in time series analysis, thanks to its bidirectional context interpretation. In this work, we aim to explore the potential of BERT to accurately derive true pollutant concentrations from spatially mixed gas measurements. Specifically, BERT's ability to contextualize data points makes it well-suited for unraveling the intricate data in urban air quality monitoring. However, translating this insight into a practical city-wide MCS system is non-trivial and faces two challenges: *(C1) Distinctive characteristics of spatially mixed gas measurements:* Unlike typical time series, gas measurements exhibit pronounced non-stationary and non-Markovian characteristics, posing challenges in isolating authentic gas concentration from interference introduced by other gas samples. Existing approaches frequently overlook the mixing effect of gas samples from diverse sensing positions or focus solely on scenarios with low sensor mobility [18]. *(C2) The network for city-scale air quality estimation is hard to converge with limited labeled data:* Initially, acquiring labeled data is challenging, as mobile sensors only sporadically encounter sparsely scattered stations which provide ground truth. Consequently, plenty of sensor readings lack ground truth, negatively impacting the performance of learning-based methods that heavily rely on labeled data. Furthermore, attaining a generalized model for city-scale deployment necessitates the collection of labeled data across diverse usage scenarios in a city [19], [20], [21]. This is particularly challenging due to the geographical variation of air pollution, exacerbating the situation.

To address these challenges, this work proposes CatUA, the first city-scale air quality estimation system with high-frequency sampling mobile sensors. *(S1) To tackle (C1):* we design a 3-block AirBERT model based on correlation analysis to remove the mixing influence of other gas samples on the true concentration, which enables BERT to process the spatially mixed gas measurements. AirBERT, thoughtfully crafted, is lightweight and can be deployed on edge devices for real-time estimation. *(S2) To tackle (C2):* we propose a Prompt-informed Training Strategy that significantly minimizes the reliance on labeled data. Initially, we design a pretraining phase based on a masked language model (MLM) that integrates geographical information, enabling AirBERT to capture the evolving patterns of sensor measurements across diverse regions utilizing extensive, unlabeled city-scale datasets. Following this, we transition into the prompt learning phase, wherein we utilize minimal labeled data in conjunction with representations derived from the pretraining phase. This approach allows AirBERT to effectively establish the mapping relationship between sensor measurements and actual pollutant concentrations in a supervised manner. Consequently, our training strategy markedly alleviates the necessity for large volumes of labeled data.

To enhance the practical applicability of CatUA, we have dedicated significant resources to deploying the software stack on our meticulously engineered Sensing Front-end. CatUA has successfully operated for over 1,200 hours, covering an experimental trajectory of more than 1,067 km^2 in a major international city, with a spatial resolution of less than 50 meters. We conducted extensive experiments across varying sampling and labeling rates for different air pollutants. Our results demonstrate that CatUA achieves an average reduction in measurement error of 96.9% with a latency of $44.9ms$, outperforming the state-of-the-art(SOTA) baseline by 42.6%.

In summary, the main contributions are as follows:

- We propose CatUA, as far as we are aware, the first accurate, city-scale, fine-grained air quality monitoring system based on mobile crowd sensing (MCS).
- We propose a novel estimation module that integrates AirBERT with a Prompt-informed Training Strategy to effectively disaggregate spatially mixed gas measurements while minimizing the reliance on labeled data at the city scale. This approach extracts non-stationary and non-Markovian features, capturing the mutual influences among sensor readings. Notably, the plug-and-play AutoPrompt enhances the pre-trained network's adaptability to specific downstream tasks.
- We design a modular Sensing Front-end with separate chambers and active airflow design, which aims to collect highly mobile air quality data under high sampling rates.
- We implement and evaluate CatUA with massive data in real-world environment. Extensive city-wide evaluation results show the effectiveness of CatUA in estimating air pollutant concentrations.

The rest of the paper is organized as follows. Section II details our motivation. Section III presents the system overview of CatUA, followed by novel designs and system implementation in Sections IV to VI. Sections VII and VIII-A evaluate our system and present practical applications. Sections IX and X review related works and conclude the paper.

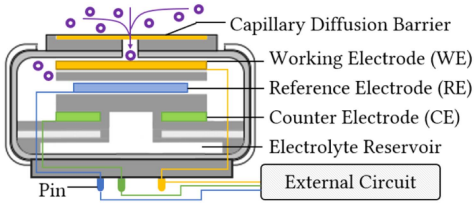


Fig. 3. The structure of an EC sensor.

II. Background of Air Quality Measurement

In the field of air quality assessment, electrochemical (EC) sensors are favored for their cost-effectiveness and portability, accounting for over 23% of global revenue [22]. Despite their widespread use, commercial EC sensors often exhibit significant measurement errors during mobile sensing, particularly at elevated sampling rates. This issue primarily arises from the differences in the chemical reactions of the sensors and the delay in their response times.

Specifically, the chemical reaction refers to the process in which gas molecules diffuse into the sensor through a porous membrane due to the concentration gradient when the external gas concentration increases. These molecules then undergo reactions on the surface of the working electrode (WE), transferring electrons from the WE to generate a current that flows through the external circuit to the counter electrode (CE), as shown in Fig. 3. However, this process may lead to issues such as reading drift, cross-sensitivity, and the need for long-term maintenance, all of which result in discrepancies between the sensor readings and the actual gas concentration. Response time refers to the time delay required for the sensor reading to stabilize after a change in gas concentration. This parameter is inherently determined by the working principle of the sensor [23] and consists of three main components: diffusion time, chemical reaction time, and electrical signal generation time.

Diffusion time: According to Fick's law, it takes time for gas molecules to diffuse from the external environment to the electrode surface, and this diffusion rate depends on the sensor structure, electrode spacing, and the diffusion coefficient of the electrolyte.

Chemical reaction time: Governed by reaction kinetics, the gas reactions on the electrode surface occur continuously over time. The chemical reaction time is influenced by factors such as gas concentration, temperature, and Gibbs free energy [24].

Electrical signal generation time: This refers to the time required to generate a current in the sensor, which can usually be neglected since the establishment of an electric field occurs at the speed of light.

III. Overview

A. Problem Formulation

1) Gas Concentration Mixing Mode: To achieve a high sampling rate, the reading from a sensor would represent the sample of the quality of the air over a distance as a vehicle moves. Therefore, stable and accurate concentration measurements for the sampled gas are unattainable due to the incomplete reaction. Moreover, the residual gas from the prior sampling lingers in the sensor, influencing the measurement of the currently sampled gas.

As a result, the gas in the sensor comprises both the presently sampled external gas and the residual gas from the preceding sampling. To streamline the complex processes of gas diffusion and EC reactions, the gas concentration I_t within the sensor at each sampling time is expressed as the following linear combination, accounting for both "current" and "previous" components.

$$I_t = \begin{cases} w_t R_t + (1 - w_t) I_{t-1} & for\ t \geq 1, \\ w_0 R_0 & for\ t = 0, \end{cases} \tag{1}$$

where R_t denotes the target "current" gas concentration to measure, and w_t serves as the coefficient to balance the 2 components.

Additionally, due to issues such as sensor drift, cross-sensitivity, and the lack of long-term maintenance, the actual sensor reading is typically C_t instead of I_t. Specifically, sensor drift causes the measurement accuracy to change over time, cross-sensitivity leads to interference from other gases, affecting the measurement of the target gas concentration, and insufficient long-term maintenance may result in sensor performance degradation and the accumulation of measurement errors. The combined effect of these factors results in the sensor reading C_t not accurately reflecting the true gas concentration I_t, thereby introducing errors and instability. In addition, C_t is affected by the previous sensor measurement due to the continuous nature of the EC reaction. Consequently, C_t can be expressed as a similar linear combination as follows,

$$C_t = \begin{cases} w'_t I_t + (1 - w'_t) C_{t-1} & for\ t \geq 1, \\ w'_0 I_0 & for\ t = 0, \end{cases} \tag{2}$$

where C_{t-1} refers to previous sensor measurement. w'_t balances the mixed gas concentrations and previous sensor measurement.

2) Concentration Estimation Problem: The inference of R_t can be formulated as a *hidden state estimation problem*. During mobile sensing, we can only collect a biased sensor measurement C_t instead of the real value R_t. Thus, the true gas concentration can be treated as a *hidden state*, while the corresponding sensor measurement can be viewed as *system observation*. Our objective is to infer the hidden states accurately with system observations.

Consequently, the problem can be formalized as follows,

$$argmin_f \sum_{t=1}^{T} \|f(C_t) - R_t\|_2^2, \tag{3}$$

where T is the length of the measurement sequence. Eq(3) illustrates we need to find an estimation function f to map observations to state values. Our goal is to minimize the L2-norm error between estimated results and hidden states.

B. System Design Overview

CatUA has been developed to minimize measurement errors associated with mobile EC sensors. The system architecture, as depicted in Fig. 4, comprises two main components: an

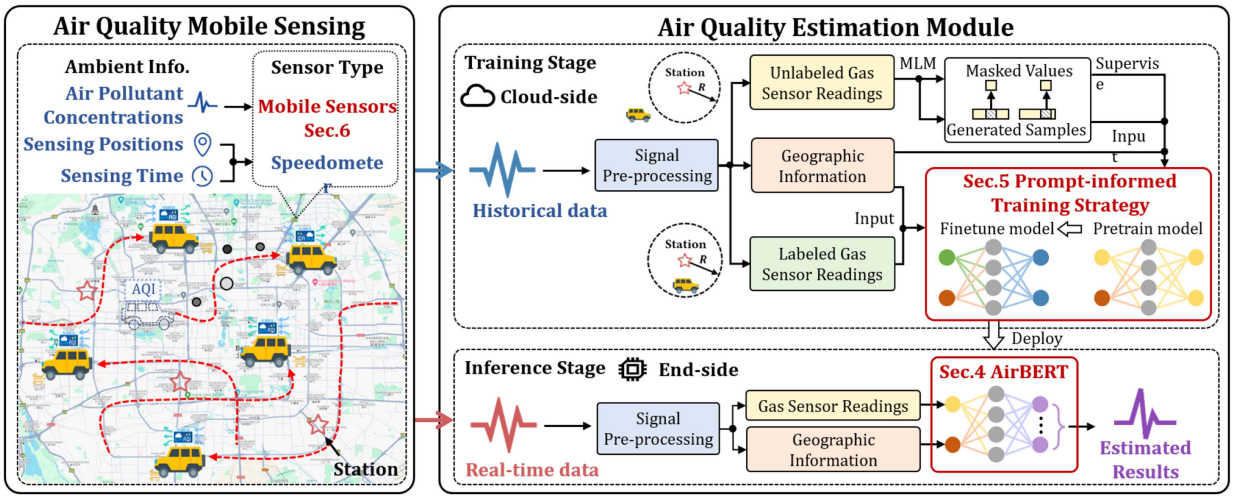


Fig. 4. CatUA consists of the air quality mobile sensing and an estimation module. The estimation module is composed of AirBERT to separate the mixed measurements of gas concentrations, an pre-train phase utilizing masked samples generated from massive unlabeled data, and a fine-tune mechanism using limited labeled data.

air quality mobile sensing module and an estimation module. Initially, CatUA collects air quality data alongside vehicular information using mobile sensors and a speedometer. Subsequently, historical sensor measurements are categorized into labeled and unlabeled datasets based on the proximity to the nearest monitoring station. These datasets are then utilized for the sequential pretraining and prompt learning of AirBERT. During the inference phase, the prompt-tuned AirBERT is deployed to accurately translate sensor readings into precise concentration values, thus enabling real-time estimation of air quality measurements across metropolitan areas. To achieve these objectives, CatUA integrates three essential components:

- *AirBERT* (Section IV): We design a tailored gas concentration separation model called AirBERT to process spatially mixed gas measurements. It models and eliminates the mixing influence of other sensor readings on a target measurement by extracting the correlation among adjacent readings.
- *Prompt-informed Training Strategy* (Section V): To ensure the effectiveness of AirBERT across urban environments, we propose a pretraining task utilizing masked language modeling (MLM). This task generates masked sample-label pairs from extensive unlabeled sensor readings, facilitating AirBERT's ability to learn the spatial and temporal patterns of sensor measurements throughout the city. Following this, a minimal set of labeled measurements is employed for prompt learning, enhancing AirBERT's predictive accuracy by aligning its outputs more closely with the ground truth data obtained from monitoring stations through structured prompts.
- *Sensing Front-end* (Section VI): We meticulously design the CatUA sensing front-end to facilitate the acquisition of mobile air quality data at high sampling rates. Our sensing front-end incorporates various innovative features, including the miniaturization of reaction chambers and an active airflow design. These enhancements aim to create a stable reaction environment and minimize the diffusing time of EC sensors.

IV. AirBERT for Concentration Separation

A. Observation

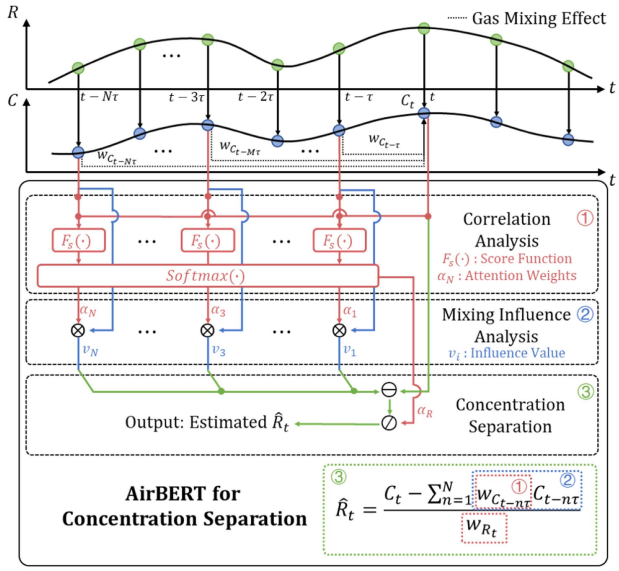
To find an effective estimation function, we first model the correlation between C_t and R_t. Combining (1) and (2), the correlation can be expressed as follows,

$$C_t = w_{R_t} R_t + \sum_{n=1}^{N} w_{C_{t-n\tau}} C_{t-n\tau}, \tag{4}$$

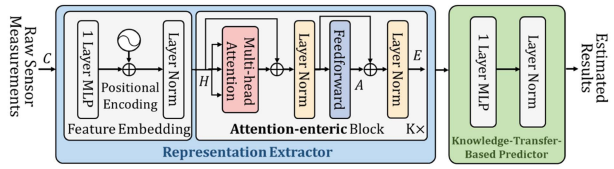
where w_{R_t} and $w_{C_{t-n\tau}}$ represent the influence coefficients of R_t and prior sensor measurements on C_t, respectively, both determined by w_t and w'_t. Eq (4) suggests the necessity of mitigating the effects of previous sensor measurements on C_t in order to accurately ascertain the true value of R_t.

According to (4), C_t displays both non-stationary and non-Markovian characteristics. First, C_t is non-stationary, as its statistical properties, including mean and variance, demonstrate temporal fluctuation and spatial variability. This behavior stems from the dynamic nature of air pollutant concentrations, which are affected by chronological and geographic factors within the urban environment. Second, C_t is non-Markovian, indicating that its values are not solely dependent on the current gas concentration but also on a sequence of prior measurements. Consequently, future values are influenced by the entire history of observations rather than merely the immediately preceding state.

Originally designed for NLP applications, Bidirectional Encoder Representations from Transformers (BERT) is an effective representation learning technique, which aims to extract semantic features and contextual representations from massive corpus data [25]. Specifically, the multi-head attention mechanism in transformer blocks can gather information from tokens in multiple positions, which allows BERT to consider



(a) AirBERT consists of 3 parts, namely correlation analysis, mixing influence analysis, and concentration separation.



(b) AirBERT network adopts an encoder (representation extractor) - decoder (Knowledge-Transfer-Based Predictor) - based structure.

Fig. 5. AirBERT architecture and design.

long-distance sequential dependencies. In addition, multiple bidirectional-connected transformer encoders also encode input sequences from both directions, which allows BERT to consider the contextual correlation.

Thanks to the extraordinary ability to learn representations from sequences, BERT has been extended to process complex time series tasks, such as pandemic prediction [26] and video representations [27], [28]. Thus, we design a novel AirBERT model to separate the mixed gas concentrations using contextual correlation among sensor measurements.

B. AirBERT Design

Fig. 5(a) illustrates the AirBERT framework, which incorporates correlation analysis, mixing influence analysis, and concentration separation. The process begins with the calculation of self-attention weights to evaluate the influence coefficients of R_t and previous sensor measurements on C_t, utilizing the dot-product as the scoring function. Subsequently, these coefficients, along with the prior sensor measurements, are multiplied to derive the influence values. In the final stage of concentration separation, AirBERT mitigates the influences from C_t through a linear combination, ultimately producing the estimated $\hat{R}_t$.

1) Why Do We Design Correlation Analysis to Assess Influences Among Measurements: Modeling the influence of $w_{C_{t-n\tau}}$ is inherently complex due to its dynamic nature, which is affected by various factors, including sampling interval, airflow, vehicular speed, and fluctuations in true gas concentration. This complexity results in correlations among measurements, exhibiting non-stationary and non-Markovian characteristics. A prevalent methodology for temporal feature extraction in time series analysis is correlation analysis, which can be employed to derive coefficients that elucidate the unknown influences within a measurement sequence. These correlation coefficients can be iteratively trained under labeled supervision to progressively approximate the actual w_{R_t} and $w_{C_{t-n\tau}}$.

2) Why Do We Adopt Self-Attention Weights for Correlation Analysis: Inspired by the self-attention mechanism in BERT, the self-attention weights α_n in AirBERT gather bidirectional information from multiple previous sensor measurements to model contextual correlation. The self-attention mechanism provides three advantages. First, it models long-term dependencies in sensor measurements beyond adjacent time steps. Second, it considers the contextual correlation throughout the entire sequence without length limitations. Third. it calculates the correlation between every two sensor measurements, enabling the capture of extensive temporal features.

3) Why do We Use Dot-Product to Calculate Attention Weights: In the linear signal space formed by all possible sensor measurements, the dot-product serves to measure the similarity between two vectors, effectively assessing the correlation between two sensor measurements. Additionally, the dot-product exhibits linear time complexity, suggesting the model's potential for on-board implementation.

C. AirBERT Structure

AirBERT, a BERT-based representation learning model, effectively captures the contextual correlations among sensor measurements. As depicted in Fig. 5(b), the model architecture commences with a multi-layer perceptron (MLP) serving as a trainable embedding module, which maps the original measurements C within a time window δ to a hidden representation $E \in \mathrm{R}^{\delta \times d}$, where d denotes the dimension of the hidden embedding. A trainable positional encoding module subsequently processes this hidden embedding, yielding H, which retains the positional information of C. Following this, K attention-centric blocks take H as input and produce the output E. Ultimately, another MLP, augmented with layer normalization, transforms the representations into the estimated results.

- *Attention-enteric Block:* Each attention-enteric block encompasses a multi-head self-attention module and a feed-forward module consisting of multiple MLP layers. In this work, K is set to 2. It is worth noticing that 2 residual blocks are adopted in each attention-enteric block, which can improve the depth and the expressiveness of the network effectively.
- *Cross-layer Parameter-sharing Mechanism:* The mechanism aims to improve parameter utilization efficiency, thereby alleviating computational overhead. Specifically, only the parameters of the first attention-centric block need to be trained, while the other attention-centric blocks share the same parameters from the first block. This means that, despite the model consisting of

multiple attention-centric blocks, the number of parameters is significantly reduced, which in turn substantially lowers computational complexity. Through this approach, the Cross-layer Parameter-sharing Mechanism effectively reduces the computational resources required during both training and inference, while preserving the model's expressiveness and performance.

V. Prompt-Informed Training Strategy

A. Observation

City-scale air quality mobile sensing requires sensors to cover the entire city. However, there exists very little ground truth data due to the sporadically distributed air quality monitoring stations. Coincidentally, in NLP tasks, many corpus data are also unlabeled due to privacy concerns, and the time and financial costs of manual labeling.

Fortunately, the self-supervised learning (SSL) method is proposed to solve the challenge of limited ground truth by learning features from unlabeled data in advance, which has been verified to bring significant performance gains on many challenging downstream tasks [29], [30], [31], [32]. Specifically, based on the characteristics of the final task, SSL designs a pretraining task and generates data-label pairs from massive unlabeled data to pretrain a model. The model is then fine-tuned to adapt to the downstream task with limited labels, combining the learned knowledge during pretraining.

Official air quality monitoring stations generally assess pollutant concentrations at specific sites. According to the United States Environmental Protection Agency, these stations are strategically located near representative urban areas, including high-traffic roads, city centers, and sites of particular concern, such as schools and hospitals, to support human health objectives. Our objective is to estimate fine-grained air quality across the city, leveraging the benefits of SSL. To this end, we propose a Prompt-informed Training Strategy based on representative areas to enhance the estimation performance of AirBERT.

B. Training Strategy Design

In Fig. 6, we provide a detailed explanation of the Prompt-informed Training Strategy for AirBERT. During the self-supervised pretraining phase, we design a masked language model (MLM) task, in which sensor measurements at the city scale are randomly masked, and these masked values are treated as generated labels. By incorporating geographic information, AirBERT effectively learns the spatio-temporal dependencies between the unmasked sensor measurements, which typically exhibit significant temporal stability. This allows AirBERT to transfer geographic knowledge to accurately recover the masked values, thereby capturing the patterns of urban pollutant distribution and adapting to different urban environments. In the subsequent supervised prompt learning phase, we fine-tune AirBERT based on the pre-trained model and a small amount of label information from monitoring stations, transferring the learned mappings from areas with existing monitoring stations to regions lacking monitoring stations. During this phase, we introduce the

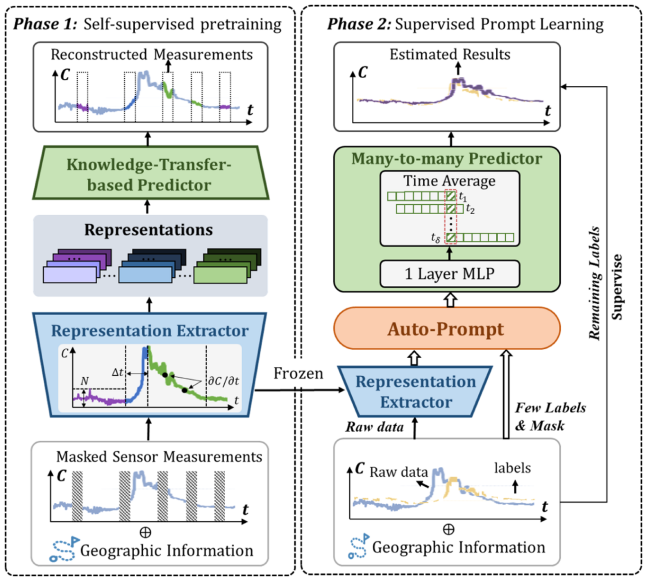


Fig. 6. Prompt-informed training strategy.

Auto-Prompt model, which further enhances AirBERT's ability to distinguish mixed gas concentrations and estimate true values. The tokens generated by Auto-Prompt serve as prompts to fine-tune the pre-trained decoder, thereby facilitating AirBERT's learning of the mapping between sensor measurements and ground truth values. This approach improves predictive accuracy while maintaining operational flexibility.

1) Why do We Design the MLM Task as the Pretrain Task: Our downstream regression task aims to estimate the real value R_t. As a more versatile pretraining task in NLP, MLM helps a language model predict masked tokens by understanding the word meaning based on contextual information, which allows the model to learn more fundamental but crucial language representations and better adapt to multiple downstream tasks. Considering the similarity between R_t estimation and tokens prediction, we design the MLM task to compel AirBERT to extract more underlying spatial and temporal correlation among sensor measurements and transfer the learned knowledge to recover masked values.

2) What Features Can AirBERT Learn in the MLM Task: From massive sensor measurements collected over a long period, AirBERT learns the general time- and spatial-invariant knowledge among measurements. Specifically, AirBERT can obtain statistical characteristics such as variance and range from a stationary sequence. It indicates the sensor noise level and can evaluate the fluctuation of measurements. Combining geographic information, AirBERT can learn the spatial variation pattern of air pollutant concentrations. Notably, these features often exhibit significant temporal stability.

3) Why Can the Prompt-Tuned Model Be Generalized to City-Scale Deployment: City-scale sensor measurements demonstrate a higher spatial density compared to ground truth data obtained from scattered monitoring stations. During the self-supervised pretraining phase, AirBERT effectively captures intricate spatial correlations among sensor measurements across

diverse urban areas. In the subsequent supervised prompt learning phase, AirBERT acquires the mapping relationship between sensor measurements sampled in proximity to monitoring stations and their corresponding ground truth values. Consequently, when deploying the prompt-tuned AirBERT in regions lacking monitoring stations, it can transfer the learned mapping relationship from nearby stations to these areas, facilitating accurate true value estimations.

4) Why do We Design Auto-Prompt During the Prompt-Learning Phase: During the pretraining phase, AirBERT elucidates the spatial variation patterns of air pollutant concentrations derived from sensor measurements. However, our downstream task necessitates the estimation of true values R_t. Consequently, the primary objective of the prompt learning phase is to guide the pre-trained model in generating the desired outputs. Given the scarcity and underutilization of labeled data in practical applications, we have developed the Auto-Prompt model. This innovative approach leverages minimal labeled data and pre-trained feature embeddings to generate efficient prompts, thereby enhancing AirBERT's precision and performance for the downstream task. Notably, as the prompt learning phase introduces new input features that modify dimensionality, Auto-Prompt's design circumvents the need for adjustments to the pre-trained AirBERT architecture, ensuring enhanced adaptability.

C. Phase1: Self-Supervised Pretraining

In this phase, we design an MLM mission to help AirBERT learn spatial and temporal correlation among sensor measurements. The MLM task randomly masks multiple pieces in measurement sequences. Then, the pretrain model learns to recover the masked values with adjacent unmasked readings.

1) Pretraining Model: The pretraining model consists of a representation extractor as an encoder and a knowledge-transfer-based predictor as a decoder. The decoder reconstructs the masked readings with representations generated by the encoder. Additionally, the MLM mission is treated as a regression task. Thus, the loss function is defined as the mean squared error (MSE) between the original readings and reconstructed values at masked positions.

2) Mask Policy: To enable AirBERT to recover the masked readings considering longer temporal dependency, we randomly mask more sensor measurements in a sequence, which provides a more challenging condition for AirBERT to learn the spatial variation trend of air pollution. The original MLM task in BERT masks only one token in a text sequence since many words have independent meanings [33]. However, in our question, AirBERT may easily degrade to copy neighbor readings as the output if masking only one position, since it may overly focus on specific positions and neglect other readings. An input with high-proportion masked pieces can provide a more challenging task to train an effective model.

D. Phase2: Supervised Prompt Learning

In this phase, Auto-Prompt initially derives more refined prompt tokens by leveraging representations produced by the pre-trained decoder and minimal labeled labels. These tokens are subsequently employed to train a novel decoder. The process is supervised by the ground truth. Prompt-informed Training Strategy is designed with minimal trainable parameters, effectively adapting to downstream tasks.

1) Auto-Prompt: Auto-Prompt is a high-level architecture that takes feature sequences extracted by the encoder and a minimal set of labeled tags to generate more informative tokens. Its primary advantage lies in its plug-and-play nature, allowing the network to learn new knowledge without altering the existing encoder and decoder structures. In our experiments, we employed a combination of random interpolation and a multilayer perceptron (MLP) structure for Auto-Prompt.

2) Many-to-Many Mapping Scheme: The input and output in this phase are sequences. Both *recent past* and *close future* measurements are included in the input sequence. Compared to one-to-one mapping [34], this many-to-many mapping scheme employs features of temporal correlation and dependency provided by the historical and future measurements, which helps achieve better results. [35].

3) Prompt Learning Model: As illustrated in Fig. 6, the fine-tuning model exhibits 3 key distinctions from the pretraining model. First, freeze the encoder of AirBERT and train only the decoder using limited ground truth data from the stations. Second, we develope an Auto-Prompt mechanism, conceptualizing prompt tokens as time-variant latent variables and task labels as observed variables influenced by these latent variables. This approach enables the derivation of context-specific prompt tokens tailored to diverse input sequences by leveraging minimal labels and embeddings from extracted representations. Third, we introduce a time-average operation within the decoder to enhance data utilization, allowing overlap between two data samples, which results in multiple updates for each sensor measurement. To estimate gas concentration at the sampling time t_1, the time-average module computes the average of all estimated results updated for t_1 at each time step, as depicted in the red dashed box in Fig. 6. The loss function is defined as MSE between the estimated results and the ground truth. Upon completion of training, the prompt-tuned model can be deployed onboard for real-time air quality estimation.

VI. Sensing Front-End

To evaluate the effectiveness of the CatUA, we delicately design a *Sensing Front-end* to collect mobile air quality data under a high sampling rate in the real world.

A. Why Do We Design Our Sensing Front-End

Most mobile monitors lack specific structural optimization, which deteriorates their sensing performance under high sampling rates. For example, Libelium [36], MSB [37], and GotchaII [38] expose sensors in the air directly or just encapsulate sensors in a box. They lack solutions to solve the gas-mixing challenge in high-mobility sensing. Consequently, we design our own CatUA Sensing Front-end. The workflow and hardware implementation are shown in Figs. 7 and 8(a).

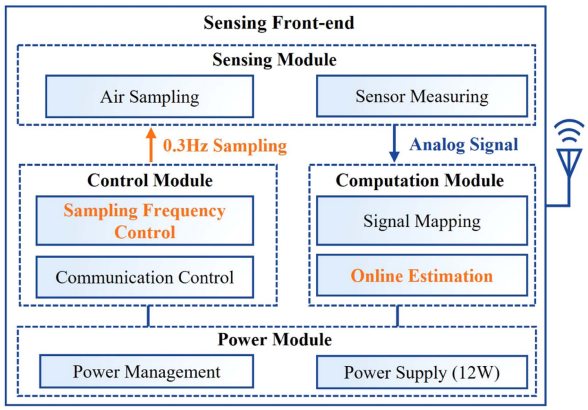


Fig. 7. Components and workflow of the CatUA sensing front-end.

TABLE I
SYSTEM PARAMETERS OF CATUA SENSING FRONT-END

Parameter	Value	Parameter	Value
Volume	195.0×134.0 ×82.0mm^3	Humidity	5% ∼ 90%
Mass	2500g	Voltage	DC 9 ∼ 36V
Temperature	-20°C∼60°C	Sampling Frequency	0.3 ∼ 1Hz
Power Consumption	10W	Air Pump Flow Rate	600ml/min
PM Chamber Specification	77.2×67.2 ×34.0mm^3	CO Chamber Specification	radius=10.3mm height=23.8mm
O_3/SO_2/NO_2 Chamber Size	radius=16.3mm height=23.8mm	VOCs Chamber Specification	radius=11.5mm height=28.6mm

During operation, the Sensing Front-end is installed on a mobile device, enabling continuous sampling along the trajectory at a frequency of 0.3 Hz. Operating at speeds between 45 and 60 km/h, it achieves a spatial resolution of under 50 meters. This high-resolution data collection ensures that the sensor captures feature patterns across a relatively complete spatial range, thereby enhancing the overall accuracy.

B. How Does the Sensing Front-End Work

The measurement process encompasses four stages: air pump inhaling gas into reaction chambers, gas diffusing into the EC sensor, gas reacting on the electrode surface, and electric signals mapped into concentration values. Initially, AR9331 sets the air pump flow rate to 600 ml/min and the sampling rate to 1 Hz. The air pump sequentially draws external gas into each reaction chamber at a constant airflow. Subsequently, gas molecules permeate the sensor due to the gas concentration gradient. Then, gas molecules react on the electrode surface, generating analog electrical signals transmitted to the central processor. The processor further de-noises and amplifies the electrical signal. After mapping the signal into gas concentrations, the real value is estimated online. All system parameters with specific settings of the Sensing Front-end are detailed in Table I. In addition, the mechanical structure plays an important role in sensing quality enhancement under high sampling rates. Specifically, compared to other devices, we consider both cavity size and airflow path in structure design, as shown in Fig. 8(b).

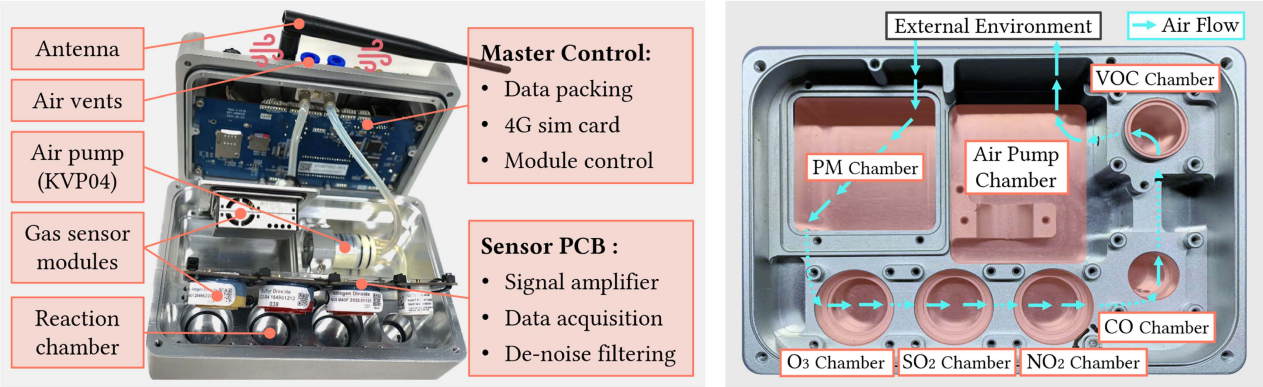
In the context of city-scale sensor deployments (e.g., thousands of sensors), the independent chambers and active airflow design of the Sensing Front-end effectively mitigate interference from mixed gases (see Fig. 8(b)). This design ensures that, even in complex environments, the sensor data remains highly accurate and reliable. Moreover, to address the computational load and signal interference caused by the simultaneous operation of thousands of sensors, the Sensing Model in the Sensing Front-end adopts a hierarchical transmission architecture (see Fig. 7). Specifically, each sensor first uploads its measurement data to a local gateway via WiFi, which then transmits the data to the cloud via a 4 G network. This hierarchical design effectively reduces the burden on the cellular network, while minimizing signal interference through WiFi. To further optimize communication performance and resource utilization, the Sensing model dynamically allocates bandwidth based on data priority (e.g., labeled vs. unlabeled data), ensuring efficient data transmission.

C. Separate Chamber Design

We assign separate reaction chambers for each gas and design the cavity size and shape for each sensor encapsulation in the front-end, as the red regions shown in Fig. 8(b). Compared to devices putting all sensors in a common space, separate chambers significantly diminish the gas-exchanging interval which is a key component of diffusing time. This design helps alleviate the gas-mixing challenge at its source. Additionally, sealing rings are also used to avoid sensing errors due to gas leakage.

D. Air Pump for Active Airflow

Airflow determines how fast gas molecules traverse reaction chambers. Unstable airflow may change the pressure in the device, causing errors in sensor measurements. Passive airflow is susceptible to external environmental factors like wind [39]. Thus, we integrate an air pump to maintain a constant airflow and stable pressure. As the blue arrows shown in Fig. 8(b), after inhaled from the external environment by the pump, air samples flow sequentially through the seven cascaded reaction chambers until being discharged. Our Sensing Front-end comprises a modular and detachable array of gas sensors designed to monitor air quality. This system quantifies six principal air pollutants, following emission source guidelines established by the U.S. Environmental Protection Agency (EPA) [40]. The Sensing Front-end assesses concentrations of O_3, SO_2, and NO_2 via Alphasense EC sensors (OX-B431, SO2-B4, NO2-B43F) selected for their proven pre-calibration accuracy and stability under diverse environmental conditions and gas concentration variances. For $PM_{2.5}$ and PM_{10} measurements, we employ the Plantower PMS5003T sensor, which additionally provides ambient temperature and humidity data. The VOCs levels are monitored using the ION MiniPID2PPB sensor. The technical details of these sensors are comprehensively listed in Table II.



(a) The Sensing Front-end samples and measures gas concentrations, and transmits data to the cloud through an antenna.

(b) The internal chamber structure, function zoning and air flow design of CatUA Sensing Front-end.

Fig. 8. Implementation of Sensing Front-end for Mobile Air Quality Data Collection.

TABLE II
SENSOR SPECIFICATIONS OF THE CATUA SENSING FRONT-END

Sensor	Gas Type	Noise	Range	Sensitivity	Tested T_{90}
PMS5003T	$PM_{2.5}/PM_{10}$	$\pm 10\mu g/m^3$@$<100\mu g/m^3$ $\pm 10\%$@$>100\mu g/m^3$	0-1000$\mu g/m^3$	50%-0.3μm 98%-0.5μm and Larger	>60s
OX-B431	O_3	±15ppb	0-20ppm	-225∼-550nA/ppm at 1ppm O_3	>60s
SO2-B4	SO_2	±5ppb	0-100ppm	275∼475nA/ppm at 2ppm SO_2	>80s
NO2-B43F	NO_2	±15ppb	0-20ppm	-175∼-450nA/ppm at 2ppm NO_2	>40s
CiTiceL@4CM	CO	±4ppb	0-2000ppm	70±15nA/ppm	>30s
MiniPID2PPB	VOCs	-	0-40ppm	>30mV/ppm	>30s

E. How is the Sensing Front-End Used

The Sensing Front-end is typically mounted inside mobile vehicles (e.g.in the trunk or on rear seats). To collect air samples with a steady airflow, two thin flexible tubes are used to connect the device with the external environment. Specifically, the two tubes are used for suction and exhaust, respectively. One end of each tube is connected to the air vent and the other end is secured to the outer surface of the vehicle body through the window. In addition, the GPS locator can be attached to the top of the vehicle by means of a magnetic device to ensure the signal quality.

Due to hardware aging or other factors, sensors may experience signal drift after prolonged use. Therefore, we perform annual calibration of the sensors in the laboratory, update hardware components, and provide reference measurement bias values. The calibrated sensors continue to be installed on mobile vehicles for data collection. Based on this, CatUA further incorporates calibration from theoretical perspectives using limited monitoring station labels to enhance the accuracy of the measurement values.

VII. EVALUATION & RESULTS

A. Data Collection

We collected the fine-grained data in a representative international city for over 1200 hours. Our Sensing Front-ends were deployed on a fleet of 15 vehicles traversing over 90% of diverse urban areas, covering an area of over 1000 km^2. The collected data includes sampling time, positions, and observations on volatile organic compounds (VOCs) concentrations, a prevalent air pollutant associated with health risks.

It is worth noting that CatUA demonstrates strong cross-city adaptability, which is attributed to the Prompt-informed Training Strategy. Specifically, during the deployment of CatUA, Sensing Front-ends are first installed on mobile devices to collect geographic information and pollutant concentration measurements across most urban areas along the vehicle's route. These data are then uploaded to the cloud for training, enabling AirBERT to learn and extract spatial features associated with different geographic locations. Next, the pre-trained model is fine-tuned with a small amount of monitoring station label data, allowing for the estimation of true pollutant concentrations. Finally, the trained model is deployed on edge devices for inference. As a result, even in cities with different pollution patterns and sensor distributions, CatUA can effectively mine the spatial features from mobile measurements and transfer these features into the mapping relationship between the measurements and monitoring station labels, thereby significantly improving prediction accuracy in most urban areas.

Observations of gas concentrations comprise *Sensor Measurements* and corresponding *Reference Data* (ground truth). Sensor measurements were collected by the front-end. Besides, a mass spectrometer is also deployed on the same vehicle, as

Fig. 9. Deployment of Sensing Front-ends and mass spectrometer on a mobile vehicle.

shown in Fig. 9. The mass spectrometer can analyze the composition and structure of substances by measuring the mass-to-charge ratio of charged ions using electromagnetic and electric field effects. Thus, this device aims to quantitatively provide fine-grained reference data. Both our front-end and the mass spectrometer operated at a sampling rate of 1 Hz.

Fig. 10 illustrates the spatial distribution of our collected data across 4 months. The air quality data always have high coverage and resolution during various collection periods. Statistically, the overall data resolution can be up to $1729.08/(\mathrm{km}^2 \cdot h)$. The spatial resolution could achieve up to 24.05m.

B. Experimental Methodology

1) Pre-Processing: First, we calculate the vehicular speed with the sampling interval and the distance change between adjacent sampling positions. Next, all input features are linearly scaled to [0,1]. Then, we use a sliding window to slice the data into sequence samples. In particular, the moving step of the window equals 1, which means there exists a 19-measurement overlapping between two adjacent samples. Each position will have a higher probability of being masked, which aims to improve data utilization and let the model learn how gas concentrations vary at any time.

We partition the dataset into three distinct subsets: the training set (60%), validation set (20%), and test set (20%). Within the training set, we further divide the data based on the labeling rate, resulting in a labeled subset comprising 1% and an unlabeled subset constituting 99%. The extensive unlabeled subset is utilized during the pretraining phase, enabling AirBERT to acquire time- and spatial-invariant knowledge. Subsequently, the labeled subset is used to train the prompt learning model through supervised learning. Considering that in practical scenarios we may have access to only a small amount of labeled data, 10% of the labeled data is directly fed into the network, while 90% is used for supervision. The validation set is integrated into the model selection process, and the performance of the trained model is ultimately assessed using the test set.

2) Training Details: The entire estimation module is implemented using Python and PyTorch, and training is conducted on a server equipped with four NVIDIA RTX A6000 GPUs (48 GB memory) and an Intel(R) Core(TM) i7-11700 2.50 GHz CPU. We configure AirBERT with two self-attention heads and use a single MLP layer with layer normalization as the knowledge-transfer-based predictor. We also apply random interpolation combined with a single MLP layer as the Auto-Prompt. The Adam optimizer is employed to update the parameters of the pre-trained model. In the multi-spot masking policy, the mask ratio is set to 0.15. Additionally, the sequence length is set to 20. During both pretraining and fine-tuning, the learning rate and batch size are set to 1×10^{-5} and 128, respectively.

3) Models for Comparison: We compare the performance of our CatUA with the following SOTA methods.

- *Naïve*: No estimation is performed. Raw sensor measurements are reported as the estimated results.
- *MLR* [41]: MLR maps over two variables to a reference result, which is widely used for sensor calibration.
- *RF* [34]: As a non-parametric estimation method, RF learns nonlinear functions for air pollution state.
- *SensorFormer (SF)* [35]: As a sequence-to-sequence model, SF uses both recent past and close future sensor data. which has been validated to surpass other methods, such as AirNet.
- *AirNet* [42]: AirNet introduces historical data sequences from both mobile devices and reference static stations, modeling the calibration of mobile sensors as a sequence-to-point mapping problem.

4) Evaluation Metric: Our final goal is to minimize the deviation between the results of model estimation and the real concentrations of air pollutants. Therefore, we adopt mean absolute error (MAE) for performance comparison.

$$MAE = \frac{1}{\tau}\sum_{i=1}^{\tau}|\tilde{x} - x|, \tag{5}$$

where τ is the length of a test sequence. $\tilde{x}$ and x represent estimated results and reference measurements for the ith element, respectively. As a common metric to assess sensing precision, MAE can evaluate the effectiveness of a model.

C. Sensing Front-End Validation

We first validate the performance of CatUA Sensing Front-end in air quality data collection. We deployed 2 front-ends with the same configurations in the lab and on a moving vehicle sequentially for in-lab static sensing and outdoor mobile sensing, respectively. In particular, we set 3 VOCs concentration levels for the in-lab test to simulate the possible pollution levels in real environments. Fig. 11 depicts the great consistency between 2 front-ends. Moreover, the average relative error of the in-lab test is 0.04 which is rather low. The detailed CDF is shown in Fig. 2.

D. Overall Performance

Fig. 12 illustrates the performance comparison between the proposed CatUA and three baseline models for VOC concentration estimation. The cumulative distribution function (CDF) underscores the superior effectiveness of CatUA, which consistently outperforms all baselines. CatUA achieves an average MAE of 5.74 $\mu g/m^3$, representing a 96.9% reduction in error

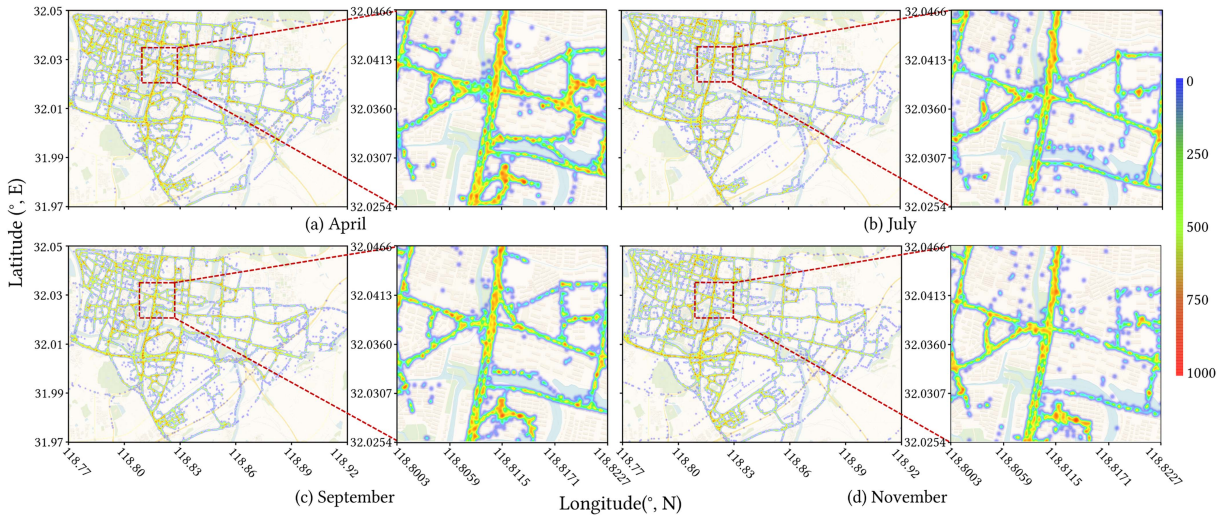


Fig. 10. Spatial distribution of data density collected by CatUA sensing front-end over 4 months.

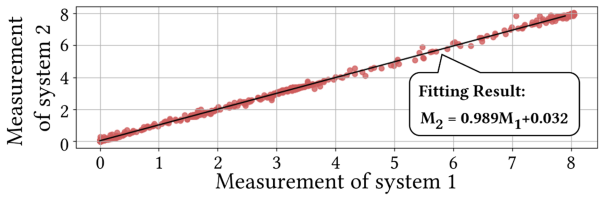


Fig. 11. Consistency validation between 2 devices with the same configuration.

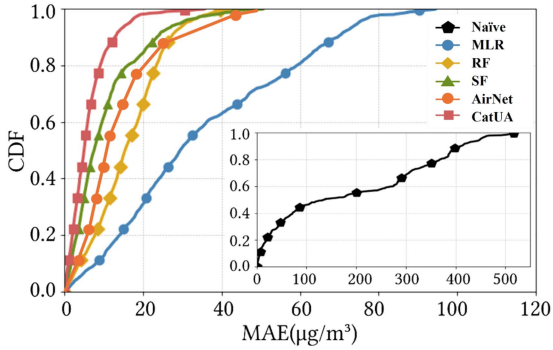


Fig. 12. Estimation performance comparison of CatUA and baselines with 1 Hz sampling rate, which is shown in CDF. The labeling rate is 1%, which means only 1% of the labeled data is used to train models.

compared to the Naïve approach (185.22 $\mu g/m^3$). Moreover, CatUA surpasses the state-of-the-art (SOTA) method SF by 42.6%, outperforms AirNet by 47.8%, exceeds RF by 63.5%, and outperforms MLR by 83.6%. Furthermore, we assess the model's effectiveness across diverse scenarios, including different air pollutant types as well as varying sampling and labeling rates.

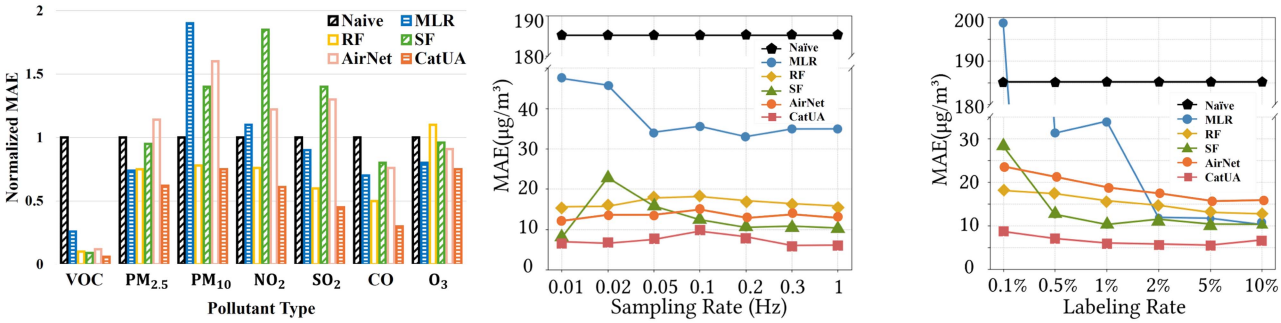
1) Effectiveness for Various Pollutants: We conducted extensive experiments using data from six common air pollutants collected by our Sensing Front-end to evaluate CatUA's performance in complex environments. These pollutants include $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3, as shown in Fig. 13(a). We normalized all Naïve MAE values to 1 and used the normalized MAE as the evaluation metric. As demonstrated, CatUA consistently outperforms other baselines across all pollutant types. Specifically, CatUA outperforms the SOTA SF by 37.6% for $PM_{2.5}$, 45.3% for PM_{10}, 69.0% for NO_2, 70.4% for SO_2, 59.2% for CO, 16.6% for O_3.

2) Impact of Sampling Rate: We then evaluate the effectiveness of CatUA by sampling VOCs concentrations at varying data rates. The results are presented in Fig. 13(b). As shown, CatUA consistently outperforms the other three baselines across all sampling rates. Specifically, as the sampling rate increases from 0.01 Hz $\sim$ 1Hz, CatUA's MAE rises from 6.63 $\mu g/m^3$ to 9.09 $\mu g/m^3$, and then drops to 5.74 $\mu g/m^3$. This behavior can be explained as follows: At lower sampling rates, the gas sampled has sufficient time to fully react inside the EC sensor before the next sample enters, leading to a lower chance of gas mixing and, consequently, lower measurement error. At higher sampling rates, the effect of gas mixing inside the sensor becomes more pronounced. However, the increased data availability allows CatUA to learn more temporal and spatial-invariant features from a large amount of unlabeled data, and more labeled data helps further fine-tune the model. Similar trends are also observed for RF and SF.

3) Impact of Labeling Rate: We further investigate the effect of the labeling rate, which defines the ratio of labeled to unlabeled data used to train CatUA. As shown in Fig. 13(c), when the labeling rate decreases from 10%$\sim$0.1%, the MAE of CatUA increases from 5.62 $\mu g/m^3$ to 8.41 $\mu g/m^3$. A similar trend is observed for the other three baseline models. Moreover, CatUA consistently outperforms the baseline models across all labeling rates, with particularly notable performance gains at lower labeling rates.

E. Ablation Study

We then experimentally analyze some core components of CatUA, and particularly, the performance gains that each of them brings into the overall system.



(a) The average and normalized MAE comparison when monitoring various air pollutants under the sampling rate of 1Hz and the labeling rates varying from 0.1%∼10%.

(b) Impact of the sampling rate. The sampling rates vary among 0.01Hz∼1Hz. All models are trained with 1% labeled data. The result is shown in MAE($\mu g/m^3$).

(c) Impact of the labeling rate. The labeling rates vary among 0.1%∼10%. All data are collected under the sampling rate of 1Hz. The result is shown in MAE($\mu g/m^3$).

Fig. 13. Estimation performance comparison of CatUA and baselines under different sampling rates and labeling rates. Our model can also be expanded to scenarios with multiple pollutants.

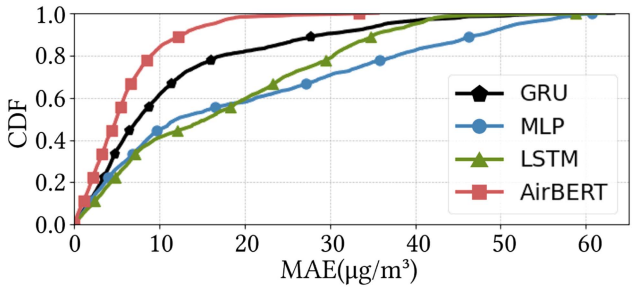


Fig. 14. Influence of pretrain encoders shown in CDF. The labeling and sampling rates are 1% and 1 Hz.

1) Impact of Pre-Train Encoder: In this section, we demonstrate that the proposed AirBERT model outperforms other pre-trained encoders in the estimation task. Specifically, we compare the AirBERT framework with three other encoder architectures: Long Short-Term Memory (LSTM) [43], Gated Recurrent Unit (GRU) [44], and Multilayer Perceptron (MLP). As shown by the CDF in Fig. 14, CatUA with AirBERT achieves a lower estimation error compared to the baseline models. AirBERT improves estimation accuracy by 18.0% over GRU, 68.3% over LSTM, and 74.2% over MLP on average.

2) Impact of Prompt-Tune Decoder: In this part, we compare the estimation performance using different prompt-tuning decoders during the supervised learning phase. Specifically, we replace the MLP layers in the CatUA Predictor with three alternative decoder structures: Gated Recurrent Unit (GRU), multi-head attention (ATTN) [45], and Long Short-Term Memory (LSTM). As shown in Table III, CatUA with MLP achieves an average improvement of more than 8.3%, 24.0%, and 10.0% over GRU, ATTN, and LSTM, respectively, across all labeling rates. However, compared to the improvements seen with pre-trained encoders, the performance gains for the fine-tuned decoders are less significant.

3) Impact of Auto-Prompt: In this section, we demonstrate the performance gains achieved during the prompt learning phase by incorporating the Auto-Prompt module. Specifically, we compare CatUA with its variant AirBERT-noAP, which omits the Auto-Prompt module. As shown in Table III, CatUA consistently outperforms AirBERT-noAP by an average of 5.4% across all labeling rates. This result demonstrates that Auto-Prompt effectively enhances the network's adaptability to downstream tasks.

TABLE III
PERFORMANCE COMPARISON OF CATUA WITH DIFFERENT PROMPT LEARNING STRATEGIES, MEASURED IN MAE ($\mu g/m^3$)

Model \ LR	0.1%	0.5%	1%	2%	5%	10%
CatUA	7.28	**5.78**	**5.74**	**5.52**	**5.16**	6.46
AirBERT-GRU	**6.46**	6.32	6.87	6.48	6.23	6.88
AirBERT-ATTN	7.38	7.53	6.27	7.93	9.88	8.341
AirBERT-LSTM	8.11	6.32	6.34	6.39	6.38	**6.42**
AirBERT-noAP	7.70	6.08	6.06	5.83	5.59	6.75

Sampling rate: 1Hz. LR indicates labeling rate.

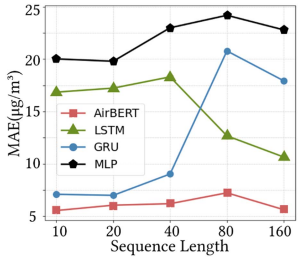


Fig. 15. Impact of sequence length with 4 prertain encoders. The labeling and sampling rates are 1% and 1 Hz.

F. Micro-Benchmark

To further explore the effectiveness of AirBERT in representation extraction, we conduct micro-benchmark experiments to inspect CatUA and evaluate its sensitivity to various system settings in the self-supervised training phase.

1) Impact of Sequence Length: As shown in Fig. 15, the performances of all methods tested are not positively related to the sequence length. The reasons are as follows. First, longer

RD \ MR	0.1	0.2	0.3	0.4	Average
18	11.36	11.52	9.98	7.47	10.08
36	6.88	7.41	5.74	7.22	**6.81**
72	13.89	14.23	9.58	11.98	12.42
144	8.11	6.29	11.75	10.95	9.27
Average	10.06	9.86	**9.26**	9.41	-

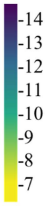


Fig. 16. CatUA performance shown in MAE($\mu g/m^3$) under different representation dimensions and mask ratios. The labeling and sampling rates are 1% and 1Hz.

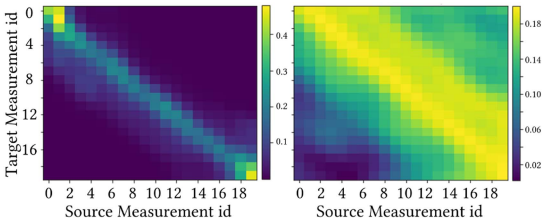


Fig. 17. Attention maps of head 0 (left) and head 1 (right) of CatUA. Head 0 concentrates more on current readings and near-future readings, while head 1 pays more attention to recent-past readings.

measurement sequences may increase the model complexity, probably causing over-fitting and yielding high estimation error. Second, when ground truth is limited, the increase in sequence length corresponds to fewer labeled samples, which may cause the severe issue of label scarcity in the supervised training process. Thus, longer sequences do not always yield lower estimation errors for tested models.

2) Impact of Masking Policy: As shown in Fig. 16, span masking with a higher ratio results in lower estimation error compared to lower masking ratios. This indicates that masking more positions can be beneficial for effective representation learning from sensor data. The setting of $r_m = 0.3$ achieves the best overall performance, and thus, we adopt it in CatUA.

3) Impact of Representation Dimension: We further investigate the impact of the representation dimension (RD) on CatUA's performance. After self-supervised training, AirBERT generates representations across various dimensions, ranging from 18∼144, with the corresponding estimation errors shown in Fig. 16. Overall, MAE decreases from 10.08 $\mu g/m^3$ to 6.81 $\mu g/m^3$ as the dimension increases from 18 to 36, but rises again to 9.27 $\mu g/m^3$ at a dimension of 144. This behavior can be explained by the relationship between representation dimension and the computational complexity of CatUA. While higher dimensions can improve the model's fit to the data, they also increase the risk of overfitting, creating a classic trade-off. Additionally, larger dimensions negatively affect the efficiency of CatUA on edge devices. Based on our experimental results, we set the representation dimension to 36, which offers the best performance.

4) Attention Map Visualization: To better understand how CatUA works, we examine the learned attention maps of 2 heads in AirBERT, to get an idea of the area the model is focusing on when processing a measurement sequence. The result is shown in Fig. 17. At each moment, both 2 heads focus on the sensor measurement generated at this moment. Moreover, head 0 also pays attention to the close future readings, which is shown in the brighter area below the diagonal. In contrast, head 1 assigns larger weights to the recent past measurements. The result shows that both recent past and close future sensor measurements can provide extra information for reference during SSL to help train effective CatUA which can achieve accurate results in the downstream estimation task.

TABLE IV
COMPUTATION EFFICIENCY COMPARISON OF CATUA AND BASELINES IN 4 ASPECTS

Model	Parameters	Size	Train Time	Infer. Time
CatUA	10.9K	56KB	13.7ms	44.9ms
SF	5.6K	35KB	7.5ms	110.4ms
RF	34K	388KB	32ms	1.7ms
MLR	0.003K	94KB	3.9ms	0.36ms

5) Computation Overhead: As a reminder, compared to existing model- and learning-based methods, CatUA can be deployed on commercial edge devices without significant resource overhead. Table IV compares CatUA with baseline methods in terms of computational efficiency. As shown, the AirBERT model in CatUA has 10.9 K parameters and a model size of 56 KB. With the support of the Prompt-informed training strategy, the training time is 13.7 ms, and the inference time is 44.9 ms. To further reduce the computational burden on edge devices and improve real-time inference speed, CatUA uploads measurement data to the cloud for continuous training. The trained and optimized model is then deployed on the edge device for inference, thus ensuring real-time performance while maintaining computational efficiency. Specifically, the inference time is the execution time to infer one sample (20 sensor readings) on a Raspberry Pi (4 Model B with quad-core Cortex-A72 processor and 4 GB RAM). All results are the average of 1000 repeated experiments. As seen, the inference time of CatUA is less than half of that of SOTA SF, although CatUA has a longer training time due to extra pretraining. Moreover, compared to the model-based methods, the size of CatUA is reduced as well thanks to the cross-layer parameter-sharing mechanism. These results indicate the lightweight CatUA can achieve real-time estimation efficiently when deployed on edge devices with high sampling rates. Although CatUA incurs slightly more parameters than some other methods, the overhead is affordable for most mobile air quality sensing scenarios.

When deploying CatUA on resource-constrained edge devices, the following three optimization strategies can be considered to further reduce the model size. *Pruning:* By removing less important neural network parameters, pruning effectively reduces both the model size and computational load. *Quantization:* Converting the model's floating-point weights to lower precision (e.g., 8-bit integers) not only reduces storage requirements but also accelerates computation. *Knowledge Distillation:* Transferring the knowledge from a large model to a smaller student model helps to reduce computational and storage demands while maintaining high prediction accuracy.

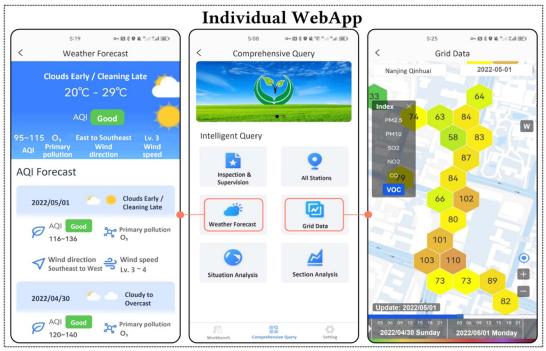


Fig. 18. Application1: Individual WebApp for surrounding AQI acquisition.

Fig. 19. Application2: Power BI platform for city-scale pollution sources monitoring.

These optimization strategies can effectively lighten the load on edge devices while enhancing the overall performance of the system. *FPGA Acceleration:* Leveraging FPGAs to design custom parallel processing architectures tailored to the specific computational patterns of AirBERT, which can substantially reduce latency, improve energy efficiency, and further optimize real-time inference performance on edge devices.

VIII. Applications

Based on the CatUA, we developed an application to collect air quality data at an urban scale. Furthermore, we explore pattern mining of the temporal and spatial distribution of pollutants.

A. Application Development

Thanks to the outstanding performance of CatUA, we develop applications with the collected city-scale data. We introduce two applications facing different users, as shown in Figs. 18 and 19.

1) Individual WebApp: Residents can access daily weather and AQIs of the entire city from the Individual WebApp. These data help users decide daily dress and whether it is suitable for travel. Additionally, the air pollution distribution at the hundred-meter scale can also be visualized. These data help citizens arrange travel routes effectively to avoid high-pollution areas and improve their travel experience. As of April 2023, Individual WebApp has accumulated 850 users.

2) Power BI Platform: Regulators can obtain a city-scale heatmap of air pollution through a BI visualization platform. The real-time spatial distribution of major pollution sources is available to help regulators infer the cause of pollution. There are mainly three procedures for source and causal prediction. First, we use the CatUA Sensing Front-ends deployed on vehicles to collect air pollution concentrations at a certain sampling rate. Then, with the collected data, we employ gaussian process regression (GPR) model to infer the concentrations of uncovered positions and reconstruct the continuous air pollution field. Finally, we take meteorological data into consideration to infer and capture the air pollution source and position. To ensure the precision of the inference process, we also combine video streams from urban cameras. Fig. 19 provides an example. The red parts in the northern area refer to areas with high O_3 concentrations started from 11:00 on May 1st. Local camera footage shows that non-standard construction is the main reason causing air pollution.

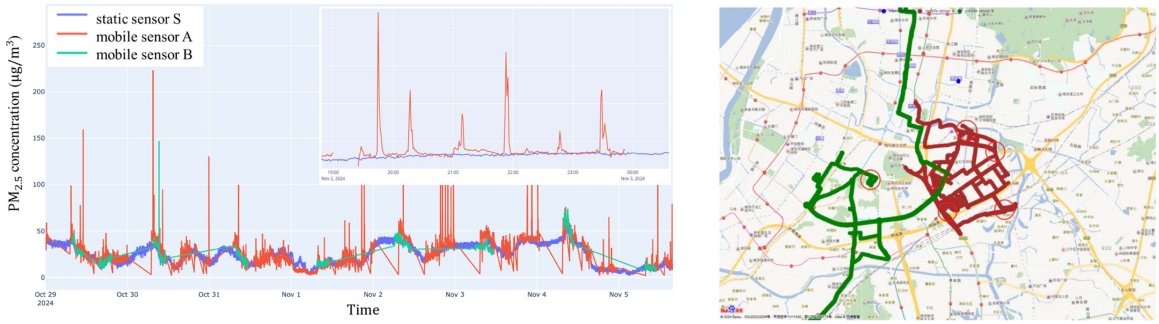
B. Pattern Mining

The developed applications have been deployed in multiple cities. Next, we explore pattern mining of the temporal and spatial distribution of pollutants.

1) Small-Scale Pollutant Activity Captured by Mobile Sensing (Pattern 1): In the same region, both mobile and stationary sensors exhibit similar general trends, including slow-changing and low-frequency components. However, during movement, the mobile sensors equipped with the CatUA model are able to detect distinct pollutant peaks more accurately, which correspond to small-scale pollutant activities that are typically difficult for stationary sensors to capture, as shown in Fig. 20(a). These pollutant peaks generally last for less than 3 minutes. Considering the common movement speed of sensors (10 to 20 km/h), the spatial scale of these pollutant activities typically affects an area of several hundred meters. By extracting these pollutant peaks and categorizing them as pollution events, their spatial aggregation can be analyzed to further identify pollution hotspot areas, as shown in Fig. 20(b). These characteristics highlight the robust performance of CatUA in handling sudden pollutant spikes and extreme environmental conditions, demonstrating its efficiency and accuracy in dealing with rapidly changing, small-scale pollutant events. This provides solid experimental validation for evaluating CatUA's robustness in complex real-world scenarios.

2) Significant Seasonal Characteristics of Single Pollutant Activities (Pattern 2): By aggregating and analyzing pollution events, the intensity of a specific pollutant's activity over a period of time can be assessed. Based on the bus-mounted sensor system deployed in Changshu City, Jiangsu Province, China, the spatial distribution of $PM_{2.5}$ pollution events during the morning and evening rush hours was examined in the first and third quarters of 2023. The key finding is that pollutant activity in the first quarter (winter and spring) was significantly more active than in the third quarter (summer and autumn), as shown in Fig. 21(a) and (b).

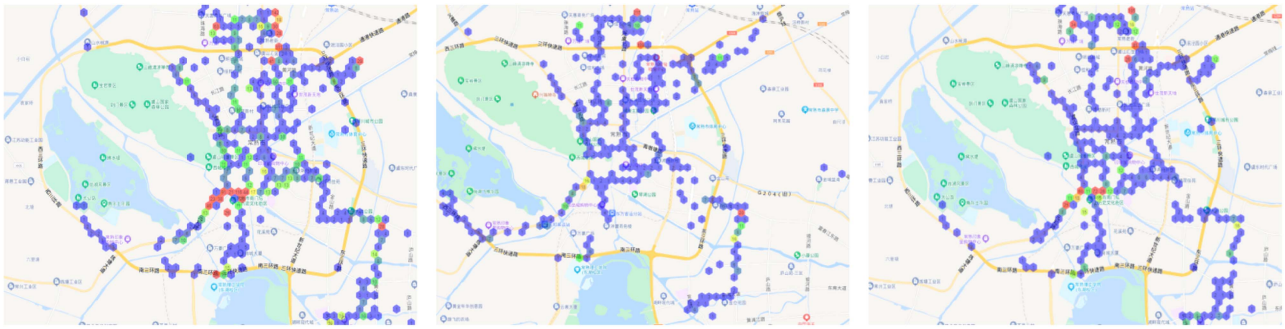
3) Spatial Distribution Correlation of Associated Pollutant Activities (Pattern 3): In the bus-mounted sensor system deployed in Changshu City, Jiangsu Province, China, both $PM_{2.5}$ and PM_{10} sensors were installed. A comparison of the spatial distribution of hotspot areas for both pollutants during the morning rush hour in the first quarter of 2023 was conducted, as shown



(a) Long time sequence diagram of two mobile sensors and one static sensor.

(b) The trajectory of the two moving sensors, with the red circle indicating the location where the peaks gather more frequently.

Fig. 20. Small scale pollutant activity.



(a) The morning peak of $PM_{2.5}$ in the first quarter of 2023.

(b) The morning peak of $PM_{2.5}$ in the third quarter of 2023.

(c) The morning peak of PM_{10} in the first quarter of 2023.

Fig. 21. Spatial distribution of PM pollutants. The number indicates the frequency of events, and the color indicates the frequency.

in Fig. 20(a) and (c). The key findings and explanations are as follows:

a) Correlation: There is a significant spatial correlation between the two pollutants. $PM_{2.5}$ and PM_{10} typically appear together, and their concentrations show a positive correlation. When one pollutant is generated, the other is often produced simultaneously.

b) Differences: Although the concentration of PM_{10} is generally higher than that of $PM_{2.5}$, when measuring pollution activity through pollution events, PM_{10} exhibits lower frequency and smaller hotspot area coverage compared to $PM_{2.5}$. A key reason for this is that dust suppression measures, such as watering, are commonly applied in the area to reduce localized pollution activities. These measures have a more significant impact on PM_{10} than on $PM_{2.5}$, suppressing the activity of PM_{10} while having little effect on the activity of $PM_{2.5}$.

IX. Related Work

A. Air Quality Sensing Systems

Most air quality sensing systems are designed for low-mobility carriers or static scenarios [46], [47], [48], [49]. The deployment cost is high for the hundred-meter-level data granularity. Although many MCS systems with high-mobility sensors have been proposed [38], [50], [51], [52], [53], significant deviations of sensor measurements still exist. Taking Sniffer4D and GotchaII [38] as an example, the hardware structure is not carefully designed and they neglect the gas mixing effect under high sampling rates. Thus, more elaborate cross-calibration is required.

B. Air Quality Estimation Methods

Existing estimation technologies can be categorized as model-based and learning-based methods. However, a precise model for the defined problem is impossible to obtain due to the stochasticity of individual drivers. Therefore, this vitiates the Gaussian mixture model-based methods [54], [55], [56], non-parametric methods [57], [58], and particle filter methods [53], [59]. Many promising learning-based methods were proposed recently for air pollution estimation. These methods rely heavily on reference data from official stations [35], [42], [60]. However, the official measurements are not representative due to the spatial sparsity, leaving mobile sensors to perform without calibration. Thus, these methods lack enough labeled data for supervised training.

X. Conclusion

In this paper, we introduced CatUA, the first fine-grained, city-scale air quality estimation system utilizing mobile sensors with high sampling rates. We developed AirBERT to handle spatially mixed gas measurements through advanced feature extraction and mutual influence modeling. Additionally, we designed

a Prompt-informed Training Strategy using Auto-Prompt, enabling CatUA to scale across cities with minimal labeled data. Deployed in an international city, our system demonstrates a 96.9% reduction in sensing errors with a latency of only 44.9 ms, surpassing the SOTA baseline by 42.6%. In future work, we plan to integrate CatUA technology into flexible mobile platforms, such as drones and robots [61], [62], [63]. By developing an air-ground collaborative communication system [64], we aim to enhance air quality monitoring capabilities in complex environments, such as high-altitude regions and vegetated terrain. The integration of this system will facilitate more precise data acquisition, thereby further expanding monitoring coverage and improving accuracy.

REFERENCES

[1] Y. Liu et al., "Mobiair: Unleashing sensor mobility for city-scale and fine-grained air-quality monitoring with airBERT," in *Proc. 22nd Annu. Int. Conf. Mobile Syst., Appli. Serv.*, New York, NY, USA, 2024, pp. 223–236, doi: 10.1145/3643832.3661872.

[2] X. Li, L. Jin, and H. Kan, "Air pollution: A global problem needs local fixes," *Nature*, vol. 570, no. 7762, pp. 437–439, 2019.

[3] Y. Liu, X. Liu, F. Man, C. Wu, and X. Chen, "Fine-grained air pollution data enables smart living and efficient management," in *Proc. 20th ACM Conf. Embedded Netw. Sensor Syst.*, 2022, pp. 768–769.

[4] "AQMesh Official Website, AQMesh," 2024. Accessed: Apr. 22, 2025. [Online]. Available: https://www.aqmesh.com

[5] C. Daepp et al., "Eclipse: An end-to-end platform for low-cost, hyperlocal environmental sensing in cities," in *Proc. 21st ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2022, pp. 28–40.

[6] Ecomsmart–Connected Air Quality Monitoring System, Ecomesure. Accessed: Apr. 22, 2025. [Online]. Available: https://ecomesure.com/en/connected-systems/ecomsmart

[7] Ecomtrek – Connected Environmental Monitoring System, Ecomesure. Accessed: Apr. 22, 2025. [Online]. Available: https://ecomesure.com/en/connectedsystems/ecomtrek

[8] H. Wang, X. Chen, Y. Cheng, C. Wu, F. Dang, and X. Chen, "H-SwarmLoc: Efficient scheduling for localization of heterogeneous MAV swarm with deep reinforcement learning," in *Proc. 20th ACM Conf. Embedded Netw. Sensor Syst.*, 2022, pp. 1148–1154.

[9] H. Wang et al., "Transformloc: Transforming MAVs into mobile localization infrastructures in heterogeneous swarms," in *Proc. IEEE Conf. Comput. Commun.*, 2024, pp. 1101–1110.

[10] X. Chen, A. Purohit, C. R. Dominguez, S. Carpin, and P. Zhang, "DrunkWalk: Collaborative and adaptive planning for navigation of micro-aerial sensor swarms," in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst.*, 2015, pp. 295–308.

[11] Z. Li et al., "Quest: Quality-informed multi-agent dispatching system for optimal mobile crowdsensing," in *Proc. IEEE Conf. Comput. Commun.*, 2024, pp. 1811–1820.

[12] Industrial gas sensors, Honeywell. Accessed: Apr. 22, 2025. [Online]. Available: https://sps.honeywell.com/gb/en/products/advanced-sensing-technologies/industrialsensing/industrial-sensors/industrial-gas-sensors

[13] Sense market intelligence and search platform, AlphaSense. Accessed: Apr. 22, 2025. [Online]. Available: https://www.alpha-sense.com/

[14] Gas sensors, FIGARO Engineering Inc., Accessed: Apr. 22, 2025. [Online]. Available: https://www.figarosensor.com/

[15] Electrochemical gas sensors, Membrapor AG, 2024. Accessed: Apr. 22, 2025. [Online]. Available: https://www.membrapor.ch/

[16] Gas sensor & detector technology leaders SGX sensortech. [Online]. Available: https://www.sgxsensortech.com/

[17] R. Shi et al., "Phy-APMR: A physics-informed air pollution map reconstruction approach with mobile crowd-sensing for fine-grained measurement," *Building Environ.*, vol. 272, 2025, Art. no. 112634.

[18] H. Wang, Y. Liu, C. Zhao, J. He, W. Ding, and X. Chen, "Califormer: Leveraging unlabeled measurements to calibrate sensors with self-supervised learning," in *Proc. Adjunct: Adjunct 2023 ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, 2023, pp. 743–748.

[19] Z. Li, F. Man, X. Chen, B. Zhao, C. Wu, and X. Chen, "TRACT: Towards large-scale crowdsensing with high-efficiency swarm path planning," in *Proc. Adjunct: Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, 2022, pp. 409–414.

[20] X. Chen et al., "Deliversense: Efficient delivery drone scheduling for crowdsensing with deep reinforcement learning," in *Proc. Adjunct: Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, 2022, pp. 403–408.

[21] X. Chen et al., "PAS: Prediction-based actuation system for city-scale ridesharing vehicular mobile crowdsensing," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3719–3734, May 2020.

[22] Gas Sensor Market Size, Share & Trends Analysis Report 2023, Grand View Research, 2023. Accessed: Apr. 22, 2025. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/gas-sensorsmarket

[23] R. Li et al., "A flexible and physically transient electrochemical sensor for real-time wireless nitric oxide monitoring," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 3207.

[24] Y. Su, K.-I. Otake, J.-J. Zheng, S. Horike, S. Kitagawa, and C. Gu, "Separating water isotopologues using diffusion-regulatory porous materials," *Nature*, vol. 611, no. 7935, pp. 289–294, 2022.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[26] J. Wang et al., "Global evidence of expressed sentiment alterations during the Covid-19 pandemic," *Nature Hum. Behav.*, vol. 6, no. 3, pp. 349–358, 2022.

[27] X. Chen, D. Liu, C. Lei, R. Li, Z.-J. Zha, and Z. Xiong, "BERT4SessRec: Content-based video relevance prediction with bidirectional encoder representations from transformer," in *Proc. 27th ACM Int. Conf. Multimedia*2019, pp. 2597–2601.

[28] R. Wang et al., "BEVT: BERT pretraining of video transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 14733–14743.

[29] B. Fernando, H. Bilen, E. Gavves, and S. Gould, "Self-supervised video representation learning with odd-one-out networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3636–3645.

[30] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," 2019, *arXiv: 1909.11942*.

[31] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer, "S4L: Self-supervised semi-supervised learning," in *Proc. Int. Conf. Comput. Vis.*, 2019, pp. 1476–1485.

[32] R. Krishnan, P. Rajpurkar, and E. J. Topol, "Self-supervised learning in medicine and healthcare," *Nature Biomed. Eng.*, vol. 6, pp. 1346–1352, 2022.

[33] H. Xu, P. Zhou, R. Tan, M. Li, and G. Shen, "LIMU-BERT: Unleashing the potential of unlabeled data for imu sensing applications," in *Proc. 19th ACM Conf. Embedded Netw. Sensor Syst.*, 2021, pp. 220–233.

[34] B. Maag, Z. Zhou, and L. Thiele, "A survey on sensor calibration in air pollution monitoring deployments," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4857–4870, Dec. 2018.

[35] Y. Cheng, O. Saukh, and L. Thiele, "SensorFormer: Efficient many-to-many sensor calibration with learnable input subsampling," *IEEE Internet Things J.*, vol. 9, no. 20, pp. 20577–20589, Oct. 2022.

[36] S. Kaivonen and E. C.-H. Ngai, "Real-time air pollution monitoring with sensors on city bus," *Digit. Commun. Netw.*, vol. 6, no. 1, pp. 23–30, 2020.

[37] S. Devarakonda, P. Sevusu, H. Liu, R. Liu, L. Iftode, and B. Nath, "Real-time air quality monitoring through mobile sensing in metropolitan areas," in *Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput.*, 2013, pp. 1–8.

[38] X. Xu, X. Chen, X. Liu, H. Y. Noh, P. Zhang, and L. Zhang, "Gotcha II: Deployment of a vehicle-based environmental sensing system," in *Proc. 14th ACM Conf. Embedded Netw. Sensor Syst.*, 2016, pp. 376–377.

[39] Y. Cheng et al., "SniffySquad: Patchiness-aware gas source localization with multi-robot collaboration," 2024, *arXiv:2411.06121*.

[40] U.S. Environmental Protection Agency, Air emissions sources, 2017. Accessed: Apr. 22, 2025. [Online]. Available: https://19january2017snapshot.epa.gov/air-emissions-inventories/air-emissions-sources

[41] B. Maag, O. Saukh, D. Hasenfratz, and L. Thiele, "Pre-deployment testing, augmentation and calibration of cross-sensitive sensors," in *Proc. Int. Conf. Embedded Wireless Syst. Netw.*, 2016, pp. 169–180.

[42] H. Yu, Q. Li, Y.-A. Geng, Y. Zhang, and Z. Wei, "AirNet: A calibration model for low-cost air monitoring sensors using dual sequence encoder networks," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1129–1136.

[43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[44] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv: 1412.3555*.

[45] A. Vaswani et al., "Attention is all you need," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Curran Associates, Inc., 2017, pp. 1–11. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[46] E. Lagerspetz, S. Varjonen, F. Concas, J. Mineraud, and S. Tarkoma, "MegaSense: Megacity-scale accurate air quality sensing with the edge," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 843–845.

[47] F. Kizel et al., "Node-to-node field calibration of wireless distributed air pollution sensor network," *Environ. Pollut.*, vol. 233, pp. 900–909, 2018.

[48] I. Godfrey, J. P. S. Brenes, M. M. Cruz, and K. Meghraoui, "Using UAS with sniffer4D payload to document volcanic gas emissions for volcanic surveillance," *Adv. UAV*, vol. 2, no. 2, pp. 86–99, 2022.

[49] Y. Miyagawa, N. Segawa, M. Yazawa, and M.-Y. Yamamoto, "Development of a low-cost gas sensor unit for wide area air pollution monitoring system (poster)," in *Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, 2019, pp. 574–575.

[50] Y. Gao et al., "Mosaic: A low-cost mobile sensing system for urban air quality monitoring," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1–9.

[51] S. Pal, A. Ghosh, and V. Sethi, "Vehicle air pollution monitoring using IoTs," in *Proc. 16th ACM Conf. Embedded Netw. Sensor Syst.*, 2018, pp. 400–401.

[52] J. Huang et al., "A crowdsource-based sensing system for monitoring fine-grained air quality in urban environments," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3240–3247, Apr. 2019.

[53] X. Chen et al., "PGA: Physics guided and adaptive approach for mobile fine-grained air pollution estimation," in *Proc. 2018 ACM Int. Joint Conf. Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput.*, 2018, pp. 1321–1330.

[54] G. Li, X. Liu, Z. Wu, Y. Wang, and L. Zhang, "Robust calibration for low-cost air quality sensors using historical data," in *Proc. 19th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2020, pp. 349–350.

[55] J. Luo, Y. Hu, C. Yu, C. Hong, X.-P. Zhang, and X. Chen, "Field reconstruction-based non-rendezvous calibration for low cost mobile sensors," in *Proc. Adjunct: Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, 2023, pp. 688–693.

[56] X. Chen et al., "Adaptive hybrid model-enabled sensing system (HMSS) for mobile fine-grained air pollution estimation," *IEEE Trans. Mobile Comput.*, vol. 21, no. 6, pp. 1927–1944, Jun. 2022.

[57] W. Hernandez, A. Mendez, A. M. Diaz-Marquez, and R. Zalakeviciute, "PM2.5 concentration measurement analysis by using non-parametric statistical inference," *IEEE Sensors J.*, vol. 20, no. 2, pp. 1084–1094, Jan. 2020.

[58] Y. Sun, Y. Liu, Z. Wang, X. Qu, D. Zheng, and X. Chen, "C-RIDGE: Indoor CO_2 data collection system for large venues based on prior knowledge," in *Proc. 20th ACM SenSys*, 2022, pp. 1077–1082.

[59] X. Chen, X. Xu, X. Liu, H. Y. Noh, L. Zhang, and P. Zhang, "HAP: Fine-grained dynamic air pollution map reconstruction by hybrid adaptive particle filter," in *Proc. 14th ACM Conf. Embedded Netw. Sensor Syst.*, 2016, pp. 336–337.

[60] F. Concas et al., "Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis," *ACM Trans. Sensor Netw.*, vol. 17, no. 2, pp. 1–44, 2021.

[61] X. Chen et al., "DDL: Empowering delivery drones with large-scale urban sensing capability," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 3, pp. 502–515, Mar. 2024.

[62] X. Chen et al., "Soscheduler: Toward proactive and adaptive wildfire suppression via multi-UAV collaborative scheduling," *IEEE Internet Things J.*, vol. 11, no. 14, pp. 24858–24871, Jul. 2024.

[63] Z. Jian, Z. Liu, H. Shao, X. Wang, X. Chen, and B. Liang, "Path generation for wheeled robots autonomous navigation on vegetated terrain," *IEEE Robot. Automat. Lett.*, vol. 9, no. 2, pp. 1764–1771, Feb. 2024.

[64] J. Ren, Y. Xu, Z. Li, C. Hong, X.-P. Zhang, and X. Chen, "Scheduling uav swarm with attention-based graph reinforcement learning for ground-to-air heterogeneous data communication," in *Proc. Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, New York, NY, USA, 2023, pp. 670–675. doi: 10.1145/3594739.3612905.

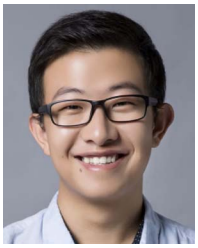
Nan Zhou received the BE degree from the School of Telecommunications Engineering, Xidian University, China, in 2024. She is currently working toward the master's degree with the Tsinghua Shenzhen International Graduate School, Tsinghua University, China. Her research interests include air pollution, spatio-temporal feature mining, and mobile computing.

Yuxuan Liu received received the BEng degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China, in 2021 and the MSc degree from Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China, in 2024. His current research interests include Artificial Intelligence of Things, environmental senisng, mobile computing, and systems.

Haoyang Wang received the BE degree from the School of Computer Science and Engineering, Central South University, China, in 2022. He is currently working toward the PhD degree with the Tsinghua Shenzhen International Graduate School, Tsinghua University, China. His research interests include mobile computing, AIoT, and distributed & embedded AI.

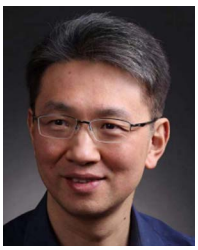
Fanhang Man received the BS degree from the University of California San Diego, La Jolla, CA, USA, in 2021. He is currently working toward the PhD degree in data science and information technology with Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. His research interests involves machine learning, optimization, large language models, etc.

Jingao Xu (Member, IEEE) received the BE and PhD degrees from Tsinghua University, in 2017 and 2022, respectively. He is now a postdoctoral researcher in School of Software, Tsinghua University. His research interests include Internet of Things, mobile computing, and edge computing.



Fan Dang (Senior Member, IEEE) received the BE degree from the School of Software, Tsinghua University, in 2013. He is currently working toward the PhD degree in the School of Software, Tsinghua University. His research interests include mobile computing and security.

Chaopeng Hong received the BE and PhD degrees from Tsinghua University, in 2012 and 2017, respectively. He is currently an assistant professor with Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. His research interests lie in the coupled human-environment systems and sustainable systems analysis, including especially climate change, air quality, agriculture, energy and water systems



Yunhao Liu (Fellow, IEEE) received the BE degree from the Department of Automation, Tsinghua University, Beijing, in 1995, the MA degree from Beijing Foreign Studies University, Beijing, in 1997, and the MS and PhD degrees in computer science and engineering from Michigan State University, East Lansing, in 2003 and 2004, respectively. He is a professor in the Department of Automation and the dean of the Global Innovation Exchange, Tsinghua University. He is a Fellow of CCF and ACM. His research interests include Internet of Things, wireless sensor networks, indoor localization, the Industrial Internet, and cloud computing.

Xiao-Ping Zhang (Fellow, IEEE) received the BS and PhD degrees in electronic engineering from Tsinghua University, Beijing, China, in 1992 and 1996, respectively. He is the chair professor with Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. Since Fall 2000, he has been with the Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Toronto, ON, Canada, where he is currently a professor and the director of the Communication and Signal Processing Applications Laboratory. In 2015 and 2017, he was a visiting scientist with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include sensor networks and the Internet of Things (IoT), machine learning, statistical signal processing, image and multimedia content analysis, and applications in Big Data, finance, and marketing. Dr. Zhang is a fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada. He was a recipient of the 2020 Sarwan Sahota Ryerson Distinguished Scholar Award and the Ryerson University Highest Honor for scholarly, research, and creative achievements. He was selected as the IEEE Distinguished Lecturer by the IEEE Signal Processing Society from 2020 to 2021 and the IEEE Circuits and Systems Society from 2021 to 2022.

Yali Song received the PhD degree in ecology from Beijing Forestry University, Beijing, China, in 2015, and as a joint PhD Student in the State University of New Jersey in the United States from 2012 to 2014. She is currently an associate professor with the College of Soil and Water Conservation, Southwest Forestry University. She is mainly engaged in research on planned burning and greenhouse gas emissions, rules and control of water and soil loss in river basins, ecological restoration in fragile mountain areas, soil respiration, and soil biological ecological processes.

Qiuhua Wang received the PhD degree from the China Academy of Forestry Sciences, in 2010. From August 2015 to August 2016, he served as a state-appointed visiting scholar with the University of South Australia. In 2017, he was appointed as a professor. Currently, his primary focus is on teaching and researching forest fire prevention.

Xinlei Chen (Member, IEEE) received the BE and MS degrees in electronic engineering from Tsinghua University, China, in 2009 and 2012, respectively, and the PhD degree in electrical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2018. He was a postdoctoral research associate in Electrical Engineering Department, Carnegie Mellon University. He is currently an associate professor with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong, China. His research interests include AIoT, artificial intelligence, pervasive computing, cyber physical system, robotics, urban sensing, brain computer interface and human computer interface.