



MobiAir: Unleashing Sensor Mobility for City-scale and Fine-grained Air-Quality Monitoring with AirBERT

Yuxuan Liu¹, Haoyang Wang¹, Fanhang Man¹, Jingao Xu², Fan Dang³,

Yunhao Liu³, Xiao-Ping Zhang¹, Xinlei Chen^{1,4,5✉}

¹Shenzhen International Graduate school, Tsinghua University ²School of Software, Tsinghua University

³Global Innovation Exchange, Tsinghua University ⁴Pengcheng Laboratory, China

⁵RISC-V International Open Source Laboratory, China

{yuxuan-l21,haoyang-22,mfh21,dangfan,yunhao}@mails.tsinghua.edu.cn,
xujingao13@gmail.com,xpzhong@ieee.org,chen.xinlei@sz.tsinghua.edu.cn

ABSTRACT

Mobile air pollution sensing methods are developed to collect air quality data with higher spatial-temporal resolutions. However, existing methods cannot process the spatially mixed gas samples effectively due to the highly dynamic temporal and spatial fluctuations experienced by the sensor, leading to significant measurement deviations. We find an opportunity to tackle the problem by exploring the potential patterns from sensor measurements. In light of this, we propose **MobiAir**, a novel city-scale fine-grained air quality estimation system to deliver accurate mobile air quality data. First, we design *AirBERT*, a representation learning model to discern mixed gas concentrations. Second, we design a *knowledge-informed training strategy* leveraging massive unlabeled city-scale data to enhance the AirBERT performance. To ensure the practicality of MobiAir, we have invested significant efforts in implementing the software stack on our meticulously crafted *Sensing Front-end*, which has successfully gathered air quality data at a city-scale for more than 1200 hours. Experiments conducted on collected data show that MobiAir reduces sensing errors by 96.7% with only 44.9ms latency, outperforming the SOTA baseline by 39.5%.

CCS CONCEPTS

- Computer systems organization → Embedded systems; • Computing methodologies → Machine learning approaches.

KEYWORDS

AQI Monitoring, Mobile Crowd-Sensing, Self-supervised Learning

ACM Reference Format:

Yuxuan Liu, Haoyang Wang, Fanhang Man, Jingao Xu, Fan Dang, Yunhao Liu, Xiao-Ping Zhang, Xinlei Chen. 2024. MobiAir: Unleashing Sensor Mobility for City-scale and Fine-grained Air-Quality Monitoring with AirBERT. In *The 22nd Annual International Conference on Mobile Systems, Applications and Services (MOBISYS '24)*, June 3–7, 2024, Minato-ku, Tokyo, Japan. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3643832.3661872>

✉ Xinlei Chen is the Corresponding author.

Yuxuan Liu and Haoyang Wang are co-primary authors.



This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

MOBISYS '24, June 3–7, 2024, Minato-ku, Tokyo, Japan

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0581-6/24/06

<https://doi.org/10.1145/3643832.3661872>



(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[1], #2 Eclipse[2], #3 ECOMSMART[3]. Mobile AQI systems: #4 ECOMTREK[4].

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

1 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 [5–7]. Typically, air quality monitoring at the city scale 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, VOC, NO_x, CO) demonstrate significant dispersion effects over merely hundred-meter-scale distances[8]. As a result, people lack access to fine-grained air quality information in their immediate living and working environments [9].

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. In contrast, a more promising solution lies in *mobile crowd-sensing* (MCS), which utilizes crowdsourced vehicles (e.g., taxis) equipped with sensing nodes (Fig.1(b)) [10–12]. These vehicles continuously gather and report air pollutant concentration data and their sampling locations while moving, offering city-wide coverage and fine-grained measurements[13–15].

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 30~150s¹, may result in spatially mixed gas samples as vehicles move around 300m (i.e., 40km/h velocity). This issue is compounded when sensors,

¹We have analyzed the response time of air pollution sensors manufactured by 5 popular enterprises on the market, including City Tech [16], Alphasense [17], Figaro [18], Membrapor [19], and SGX Sensors [20].

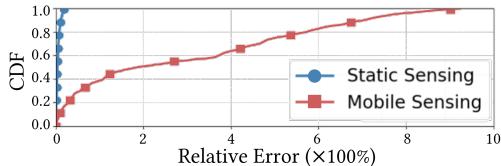


Figure 2: The average relative errors of static and mobile sensing are 4% and 174%.

retaining remnants of previously sampled gas, mix it with new samples, thus 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 estimate accurate and fine-grained air quality data from those spatially mixed gas measurement time series.

Nowadays, BERT (Bidirectional Encoder Representations from Transformers) 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 [21, 22].

(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 [23]. 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. This is particularly challenging due to the geographical variation of air pollution, exacerbating the situation [24].

To tackle these challenges, this work proposes **MobiAir**, 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 design a **knowledge-informed training strategy** to greatly reduce the demand for labeled data. First, combining geographical information, we design a pre-train phase based on mask language model (MLM), enabling AirBERT to capture changing patterns of sensor measurements in different regions

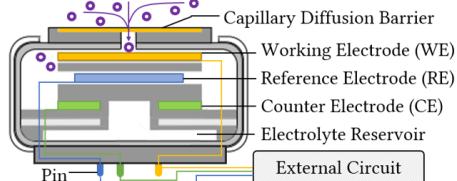


Figure 3: The structure of an EC sensor.

from extensive and unlabeled city-scale data. Subsequently, in the fine-tuning phase, the representations acquired during the pre-train phase are employed to help AirBERT effectively learn the mapping relationship between sensor measurements and true concentrations in a supervised manner. This training strategy effectively mitigates the need for a substantial amount of labeled data.

In order to render MobiAir a practical solution, we invest significant effort in deploying the software stack onto our meticulously crafted Sensing Front-end. MobiAir has operated for more than 1200 hours. The experimental trajectory has covered over 1067km² of land in a major international city. The collected data has a spatial resolution of less than 50m. We conduct extensive experiments under different sampling and labeling rates with various air pollutants. Results show that MobiAir has reduced the measurement error by 96.7% on average with 44.9ms latency, surpassing the SOTA baseline by 39.5%.

In summary, the main contributions are as follows:

- (1) We propose MobiAir, as far as we are aware, the first accurate city-scale fine-grained air quality monitoring system based on MCS.
- (2) We develop a novel estimation module fusing AirBERT with a knowledge-informed training strategy to separate the spatially mixed gas measurements with fewer demands for city-scale labeled data by extracting non-stationary and non-Markovian features and modeling the mutual influences among sensor readings.
- (3) 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.
- (4) We implement and evaluate MobiAir with massive data in real-world environment. Extensive city-wide evaluation results show the effectiveness of MobiAir in estimating air pollutant concentrations.

The rest of the paper is organized as follows. Section 2 details our motivation. Section 3 presents the system overview of MobiAir, followed by novel designs and implementation in Sections 4 to 6. Sections 7 and 8 evaluate MobiAir and present practical applications. Sections 9 and 10 review related works and conclude the paper.

2 BACKGROUND OF AIR QUALITY MEASUREMENT

In the realm of air quality assessment, electrochemical (EC) sensors are favored for their cost-effectiveness and high portability [25], accounting for over 23% of the global revenue [26]. Despite their ubiquity, commercial EC sensors often exhibit substantial measurement errors during mobile sensing under elevated sampling rates. This issue primarily stems from the prolonged response time of these sensors, ranging from 25~80s. The response time denotes the temporal delay from changes in gas concentration to the stabilization of sensor readings. This parameter is inherently dictated by the working principle of the sensor [27]. As depicted in Fig. 3, when

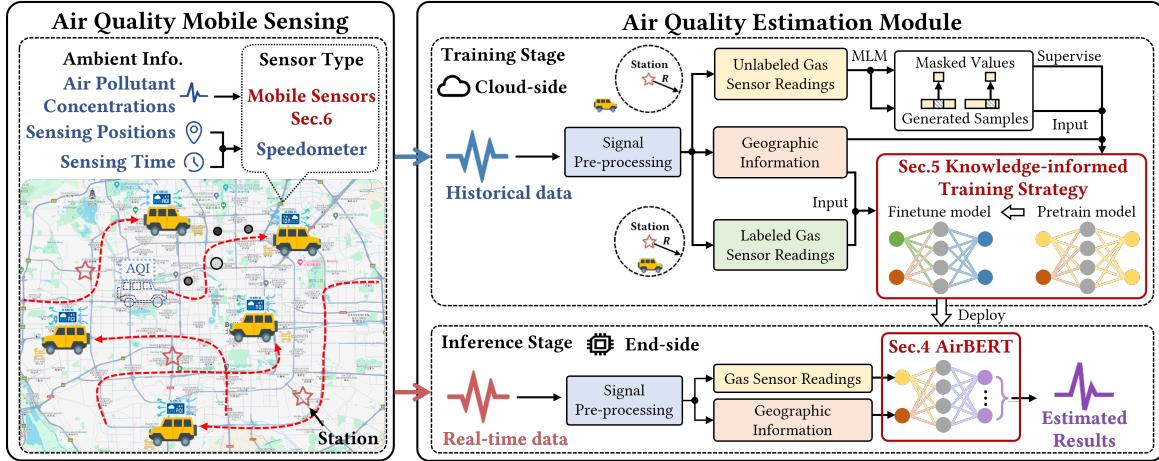


Figure 4: MobiAir consists of the air quality mobile sensing and an estimation module. The estimation module is composed of AirBERT to separate the mixed measurements of gas concentrations, a pre-train phase utilizing masked samples generated from massive unlabeled data, and a fine-tune mechanism using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

external gas concentration increases, molecules diffuse into the sensor through a porous membrane due to the concentration gradient. Subsequently, these molecules undergo reactions on the working electrode (WE) surface. Finally, electrons transfer from the WE, generating a current that flows sequentially through the counter electrode (CE) via an external circuit. It is essential to note that the response time encompasses diffusing time, chemical reaction time, and electrical signal-producing time.

- *Diffusing time*: Fick's first law tells that it takes time for gas molecules to diffuse from the external environment to the electrode surface. This depends on the sensor structure, electrode spacing, and electrolyte diffusion coefficient [28].
- *Chemical Reaction Time*: Governed by reaction kinetics, gas reactions on the electrode surface occur continuously over time. The chemical reaction time is contingent upon gas concentrations, temperature, and Gibbs free energy [29].
- *Electrical Signal-producing Time*: It refers to the time to form a current, which can be disregarded since the establishment of an electric field occurs at the speed of light.

3 OVERVIEW

3.1 Problem Formulation

3.1.1 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 electrochemical 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.

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

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

Therefore, the sensor measurement C_t represents the observation of the mixed gas concentrations I_t instead of R_t . In addition, C_t is affected by the previous sensor measurement due to the continuous nature of the electrochemical reaction. Consequently, C_t can be expressed as a similar linear combination as follows,

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

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

3.1.2 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,

$$\operatorname{argmin}_f \sum_{t=1}^T \|f(C_t) - R_t\|_2^2. \quad (3)$$

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.

3.2 System Design Overview

We developed MobiAir with the aim of minimizing measurement errors of mobile EC sensors. The system architecture of MobiAir is illustrated in Fig.4. MobiAir comprises an air quality mobile sensing module and an estimation module. MobiAir first collects air quality data and corresponding vehicular information with mobile sensors and a speedometer. Then, historical sensor measurements are categorized into labeled and unlabeled groups based on the distance between the sensing position and the nearest monitoring station. The unlabeled and unlabeled data are used to pretrain and

finetune AirBERT sequentially. In the inference stage, the finetuned AirBERT is employed to transfer the knowledge of mapping sensor readings to true concentrations for real-time estimation of city-scale sensor measurements. To achieve these goals, MobiAir integrates three core components:

- *AirBERT* (Sec. 4): 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.
- *Knowledge-informed training strategy* (Sec. 5): To guarantee the AirBERT effectiveness across the entire city, we design a pretraining task based on MLM. This task generates masked sample-label pairs with city-scale unlabeled sensor readings, which enables AirBERT to learn changing patterns of sensor measurements at different areas of a city. Then, the limited labeled measurements are used to finetune AirBERT, which optimizes estimation outcomes of AirBERT to closely align with ground truth from monitoring stations.
- *Sensing Front-end* (Sec. 6): We meticulously design the MobiAir 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.

4 AIRBERT FOR CONCENTRATION SEPARATION

4.1 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,

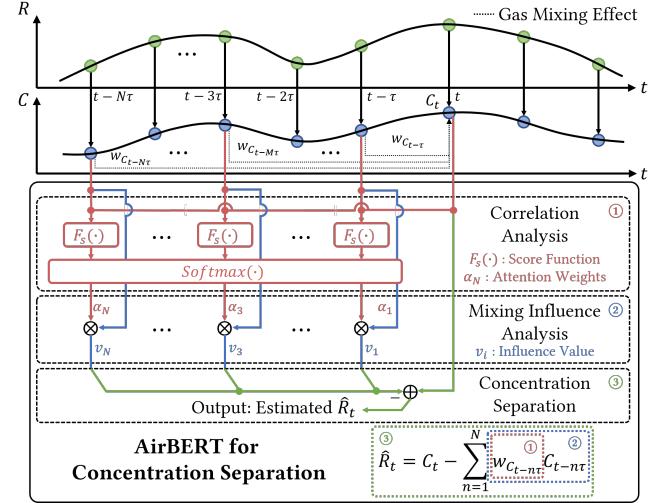
$$C_t = R_t + \sum_{n=1}^N w_{C_{t-n\tau}} C_{t-n\tau}. \quad (4)$$

$w_{C_{t-n\tau}}$ denotes the influence coefficient of previous sensor measurements on C_t , determined by w_t and w'_t . Eq.(4) indicates that we need to eliminate the influences of previous sensor measurements on C_t to get the true value R_t .

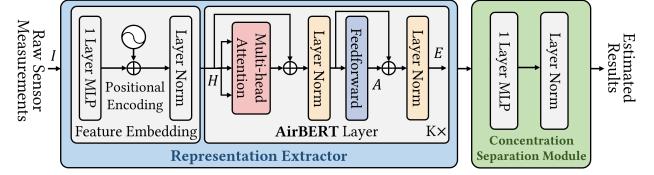
According to Eq.(4), C_t exhibits non-stationary and non-Markovian properties. First, C_t is time-dependent. Statistical characteristics like the average and standard deviation change over time and sensing positions since air pollutant concentrations vary with geographic features across a city. Second, C_t depends not only on the current true gas concentration but also on multiple previous measurements.

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 [30]. Specifically, the multi-head attention mechanism in transformer blocks can gather information from tokens in multiple positions, which allows BERT to consider 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



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



(b) AirBERT network adopts an encoder(representation extractor)-decoder(concentration separation module)-based structure.

Figure 5: AirBERT Design.

tasks, such as pandemic prediction [31] and video representations [32, 33]. Thus, we design a novel AirBERT model to separate the mixed gas concentrations using contextual correlation among sensor measurements.

4.2 AirBERT Design

Fig.5(a) delineates the AirBERT framework, encompassing correlation analysis, mixing influence analysis, and concentration separation. Initially, we calculate self-attention weights to assess the influence coefficients of previous sensor measurements on C_t , employing the dot-product as the score function. Subsequently, the coefficients and prior sensor measurements are multiplied to quantify the influence values. Finally, in the concentration separation stage, AirBERT removes the influences from C_t via linear combination and outputs estimated \hat{R}_t .

(1) *Why do we design correlation analysis to assess influences among measurements?* Modeling the influence is challenging due to the dynamic nature of $w_{C_{t-n\tau}}$ influenced by factors like sampling interval, air flow, vehicular speed, and true gas concentration variations. However, the influence manifests as correlations among measurements, reflecting non-stationary and non-Markovian properties. As a common approach for temporal feature extraction in time series, correlation analysis can be an alternative to calculate coefficients to infer the unknown influences in a measurement sequence. The correlation coefficients can be trained iteratively to approximate the real $w_{C_{t-n\tau}}$ gradually under label supervision.

(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.

4.3 AirBERT Structure

As a BERT-based representation learning model, AirBERT extracts contextual correlation among sensor measurements. The AirBERT structure is illustrated in Fig.5(b). Initially, we design a multi-layer perception (MLP) as a trainable embedding module to map the original measurements I in a time window δ , to a hidden representation $E \in \mathbb{R}^{\delta \times d}$, where d is the dimension of the hidden embedding. The trainable positional encoding module processes the hidden embedding and outputs H retaining position information of I . Subsequently, the K attention-enteric blocks take H as input and output E . Finally, an MLP with a layer norm maps representations to 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:* This mechanism aims to enhance parameter utilization efficiency. In specific, only the parameters in the first attention-enteric block need training. While other attention-enteric blocks share the same parameters in the first block. This approach reduces the computation complexity of the model effectively.

5 KNOWLEDGE-INFORMED TRAINING STRATEGY

5.1 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 [34–37]. 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

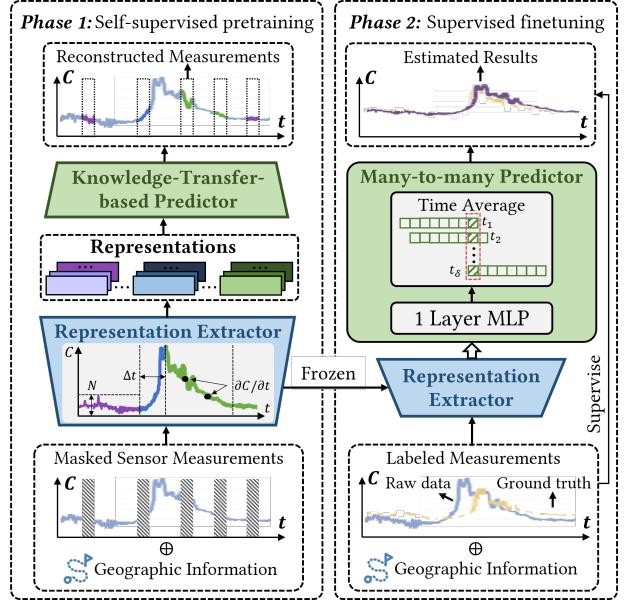


Figure 6: The MLM task assists AirBERT to learn correlation among sensor measurements.

is then fine-tuned to adapt to the downstream task with limited labels, combining the learned knowledge during pretraining.

Official monitoring stations typically measure air pollutant concentrations at specific locations. Moreover, according to the United States Environmental Protection Agency, air quality monitoring stations are designed to support human health objectives and are typically positioned near representative urban areas including busy roads, city centers, and locations of particular concern (e.g., schools, hospitals, and emissions sources). What we aim to estimate is the fine-grained air quality in a city composed of these representative areas. Inspired by the advantage of SSL, we design a representative area-based knowledge-informed training strategy to help enhance the AirBERT estimation performance.

5.2 Training Strategy Design

Fig.6 details our knowledge-informed training strategy. In the self-supervised pretraining phase, we design a masked language model (MLM) task to pretrain AirBERT. MLM randomly masks readings in city-scale sensor measurements, where the masked values are treated as generated labels. Combined with geographic information, AirBERT learns to extract spatial and temporal dependencies among sensor measurements from unmasked positions and transfer the knowledge related to geographic locations to recover the masked values. In the supervised fine-tuning phase, AirBERT is further trained with limited labeled data to separate the mixed gas concentrations and estimate true values under supervision. This enables AirBERT to learn the mapping relationship between sensor measurements and ground truth.

(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

Algorithm 1: Multi-spot Masking Policy

Input: Measurement sequence C , sequence length L
Output: Mask position set \mathcal{M} , masked value set \mathcal{V} ,
masked sequence I

```

1 Initialize:  $r_m, p, p_m$ ;
2  $l_{goal} \leftarrow \max(1, [L \times r_m]), \mathcal{M} \leftarrow \emptyset, \mathcal{V} \leftarrow \emptyset,$ 
 $l \leftarrow [1, l_{goal}], P_l \leftarrow (1-p)^{l-1}p / \sum_{k=1}^{l_{goal}} (1-p)^{k-1}p;$ 
3 while  $|\mathcal{M}| + 1 \leq l_{goal}$  do
4   sample  $l_s$  from  $l \sim P_l$ ;
5    $l_s \leftarrow \min(l_s, l_{goal} - |\mathcal{M}|)$ ,  $start \leftarrow \text{Randint}(1, L)$ ;
6   if  $start \notin \mathcal{M}$  then
7     for  $i = start$  to  $\min(start + l_s, L - 1)$  do
8        $\mathcal{M} = \mathcal{M} \cup \{i\};$ 
9       if  $\text{Rand}(0, 1) \leq p_m$  then
10          $\mathcal{V} = \mathcal{V} \cup \{C_{i \cdot}\}, C_{i \cdot} \leftarrow 0;$ 
11       end
12     end
13   end
14 end
15  $I \leftarrow C;$ 
16 return  $\mathcal{M}, \mathcal{V}, I$ 

```

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 fine-tuned model be generalized to city-scale deployment?* City-scale sensor measurements exhibit higher spatial density than ground truth from scattered monitoring stations. In the self-supervised pretraining phase, AirBERT captures detailed spatial correlation among sensor measurements across various city areas. In the supervised finetuning phase, AirBERT can learn the mapping relationship between sensor measurements sampled near monitoring stations and corresponding ground truth. Thus, when deploying the fine-tuned AirBERT in regions without monitoring stations, it can transfer the mapping relationship near the stations to these regions for true value estimation.

5.3 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.

- *pretraining Model:* The pretraining model consists of a representation extractor as an encoder and a knowledge-transfer-based

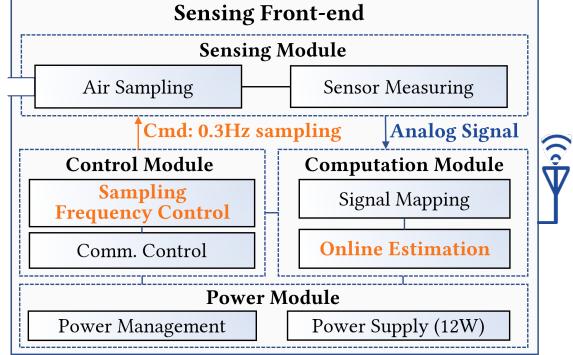


Figure 7: The specific components and work flow of the Mo-biAir sensing front-end.

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.

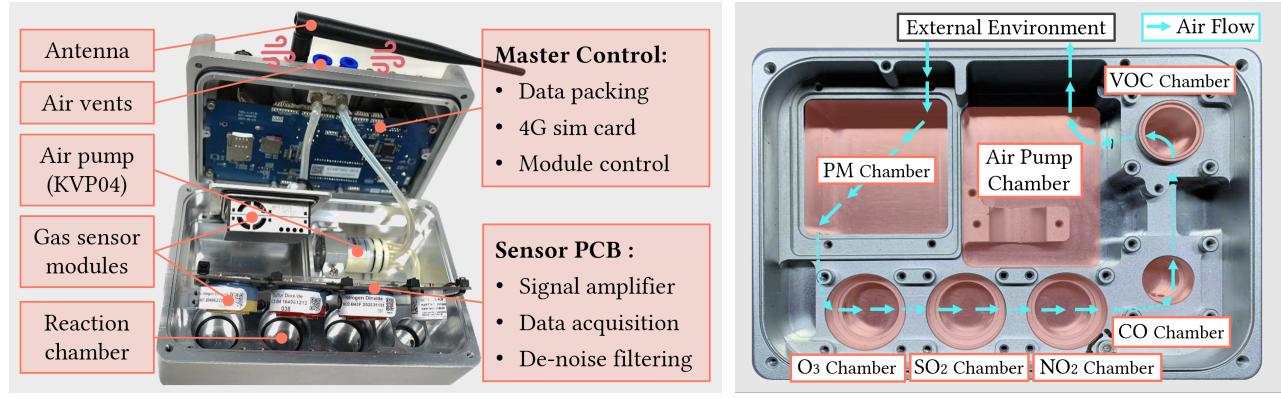
- *Mask Policy:* To enable AirBERT to recover the masked readings considering longer temporal dependency, we randomly mask more sensor measurements in a sequence, which provides a more challenging condition for AirBERT to learn the spatial variation trend of air pollution. We design the Multi-spot Masking Policy to decide the number and positions of masks and generate training data, as shown in Algorithm 1. The original MLM task in BERT masks only one token in a text sequence since many words have independent meanings [38]. 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.

5.4 Phase2: Supervised Fine-tuning

In this phase, we train a new decoder with representations generated by pretrained AirBERT to estimate real gas concentrations. The process is supervised by the ground truth.

- *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 [39], 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. [40].

- *Fine-tuning Model:* As Fig.6 illustrated, the finetuning model is similar to the pretraining model, but there are 2 main differences. First, all parameters of AirBERT are frozen and only the decoder is trained with limited ground truth from stations. Second, we introduce a time-average operation in the decoder. To improve data utilization, there is an overlap between 2 data samples. Thus, each sensor measurement is updated multiple times. To estimate the gas concentration at the sampling time t_1 , the time-average module outputs the average of all estimated results updated for t_1 at each time step, as the red block shown in Fig.6. The loss function is defined as the MSE between estimated results and ground truth. After training, the fine-tuned model can be deployed on board for online air quality estimation.



(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 MobiAir sensing front-end.

Figure 8: Sensing front-end with implementation for mobile air quality data collection.

Table 1: System parameters of MobiAir 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	VOC Chamber Specification	radius=11.5mm height=28.6mm

6 SENSING FRONT-END

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

(1) *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, Li-belium [41], MSB [42], and Gotchall [43] 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 MobiAir sensing front-end. The workflow and hardware implementation are shown in Fig.7 and 8(a). The sensing front-end has raised the sampling rate to 1Hz and declined the cost to only about \$0.5k. Additionally, the reserved reaction chambers can support measuring multiple pollutants simultaneously.

(2) *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 600ml/min and the sampling rate to 1Hz. 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 1.

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).

- *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.
- *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. 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 six 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) [44]. The sensing front-end assesses concentrations of O₃, SO₂, and NO₂ via Alphasense electrochemical 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 PM₁₀ measurements, we employ the Plantower PMS5003T sensor, which additionally

Table 2: Sensor specifications of the MobiAir sensing front-end

Sensor	Gas Type	Noise	Range	Sensitivity	Tested T ₉₀
PMS5003T	PM _{2.5} /PM ₁₀	±10µg/m ³ @<100µg/m ³ ±10%@>100µg/m ³	0-1000µg/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	VOC	-	0-40ppm	>30mV/ppm	>30s

provides ambient temperature and humidity data. The VOC levels are monitored using the ION MiniPID2PPB sensor. The technical details of these sensors are comprehensively listed in Table 2.

(3) *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. After the sensing front-end process the data collected in real time, these data will be transmitted back to the cloud server via the cellular network for storage and easy access for subsequent applications. As for regular maintenance, despite the Air Quality Estimation Module, gas sensors are not immune to signal drift after long periods of utilization. The gas sensors still need to be re-calibrated annually in the laboratory.

7 EVALUATION & RESULTS

7.1 Data Collection

We collected fine-grained data in a representative international city for over 1200 hours, covering various weather conditions across four seasons, since most existing datasets typically only have hourly measurements [40, 45–47]. Our sensing front-ends were deployed on a fleet of 15 vehicles that traversed over 90% of diverse urban areas at speeds ranging from 30 to 120 km/h, covering an area of over 1000 km². The collected data includes sampling time, positions, and observations of volatile organic compounds (VOC) concentrations, a common air pollutant associated with health risks.

Gas concentration observations consist of *Sensor Measurements* and corresponding *Reference Data* (ground truth). The front-end collected sensor measurements, while a mass spectrometer, as shown in Fig9, was also deployed on the same vehicle. The mass spectrometer analyzes the composition and structure of substances by measuring the mass-to-charge ratio of charged ions using electromagnetic and electric field effects. It provides quantitative fine-grained reference data. We selected the SPI-MS-2000 due to its ppb-level sensitivity to over 300 components, rapid response rate of 100 spectra/s, and meter-level spatial resolution. The SPI-MS-2000 exhibits high system stability, unaffected by temperature changes or vehicle vibrations. Both our front-end and the mass spectrometer operated at a sampling rate of 1Hz.



Figure 9: Sensing front-ends and a mass spectrometer are deployed on a mobile vehicle.

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/(km² · h). The spatial resolution could achieve up to 24.05m.

7.2 Experimental Methodology

7.2.1 Pre-processing. First, we calculate vehicular speed based on the sampling interval and position changes. Next, input features are scaled to [0,1]. We use a sliding window with a step size of 1 to create sequence samples, ensuring overlap between adjacent samples for better learning of gas concentration variations. The data is then divided into training (60%), validation (20%), and test (20%) subsets. Within the training set, a small labeled subset (1%) and a large unlabeled subset (99%) are created based on the labeling rate. The unlabeled data is used for pre-training, enabling the model to learn time- and spatial-invariant knowledge. The labeled data is used for fine-tuning through supervised learning. The validation set aids in model selection, and the trained model is evaluated on the test set.

7.2.2 Training Details. The entire estimation module is implemented in Python and PyTorch and trained in a server equipped with 4 NVIDIA RTX A6000 GPUs, 48GB memory, and an Intel(R) Core(TM) i7-11700 2.50 GHz CPU. We set 2 self-attention heads in AirBERT, adopt 1 MLP layer with layer norm as the knowledge-transfer-based predictor, and employ Adam optimizer to update the parameters of the pre-train model. In the multi-spot masking policy, the mask ratio, success probability, and mask probability are set to 0.15, 0.2, and 0.8. Moreover, the sequence length is set to 20. The

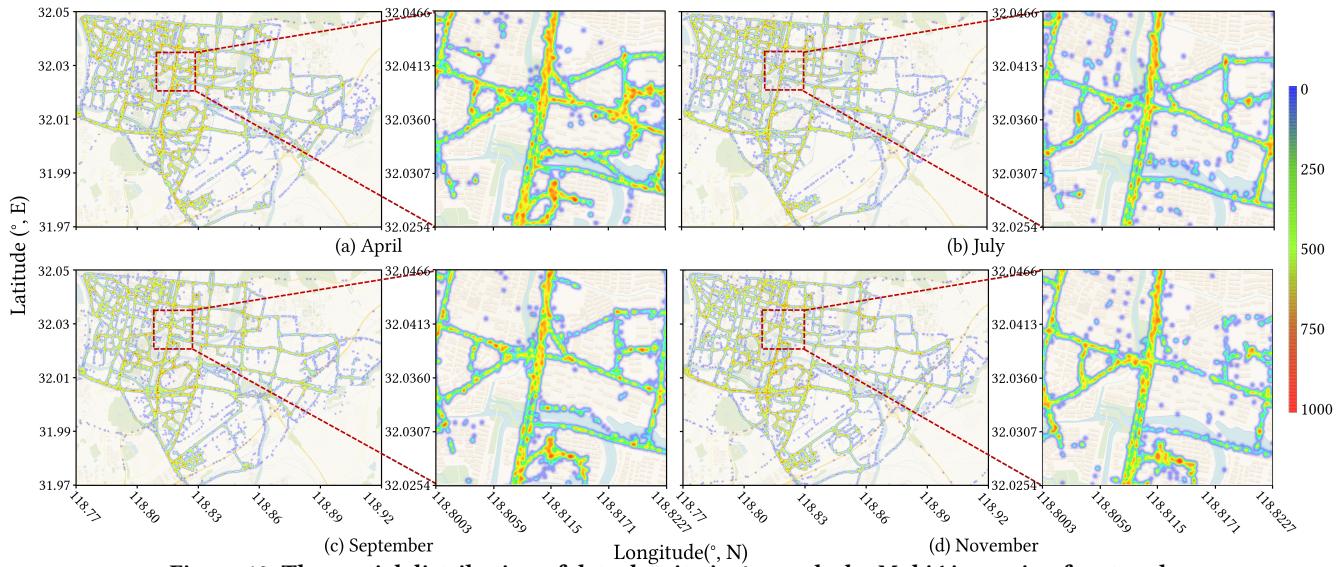


Figure 10: The spatial distribution of data density in 4 months by MobiAir sensing front-end.

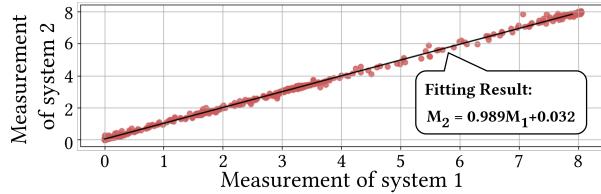


Figure 11: Consistency validation between 2 devices with the same configuration.

learning rate and batch size during pre-training and fine-tuning are set to 1×10^{-5} and 128.

7.2.3 Models for Comparison. We compare the performance of our MobiAir with the following SOTA methods.

- (1) *Naïve*: No estimation is performed. Raw sensor measurements are reported as the estimated results.
- (2) *MLR* [48]: MLR maps over two variables to a reference result, which is widely used for sensor calibration.
- (3) *RF* [39]: As a non-parametric estimation method, RF learns non-linear functions for air pollution state estimation.
- (4) *SensorFormer(SF)* [40]: 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.

7.2.4 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)**

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

for performance comparison. where τ is the length of a test sequence. \tilde{x} and x represent estimated results and reference measurements for the i -th element, respectively.

7.3 Sensing Front-end Validation

We first validate the performance of MobiAir 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

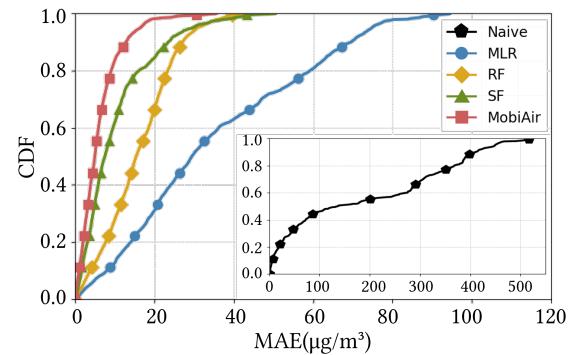
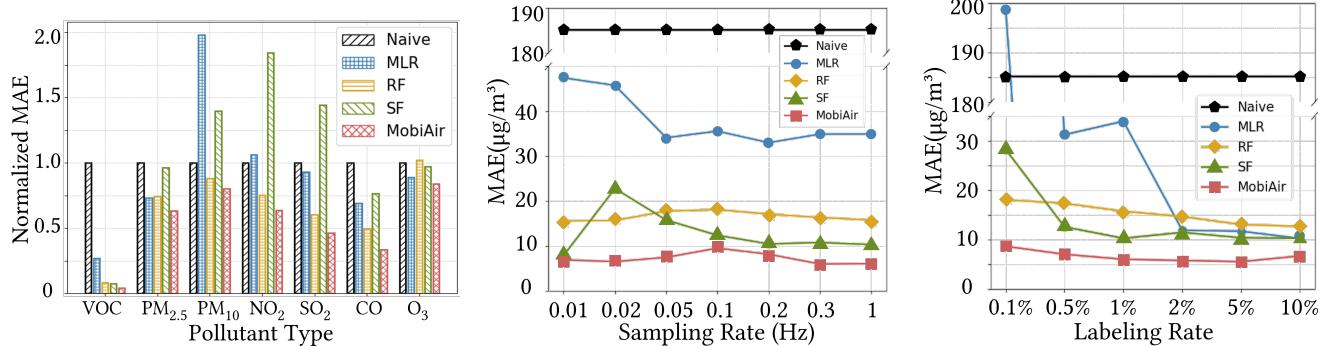


Figure 12: Estimation performance comparison of MobiAir and baselines with 1Hz sampling rate, which is shown in CDF. The labeling rate is 1%, which means only 1% of the labeled data is used to train models.

particular, we set 3 VOC 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.

7.4 Overall Performance

Fig.12 illustrates the performance of the proposed MobiAir and the other 3 baselines on VOC concentration estimation. The CDF demonstrates the effectiveness of MobiAir which outperforms all baselines. The average MAE of MobiAir is $6.06\mu\text{g}/\text{m}^3$, which reduces the error by 96.7% compared to the $185.22\mu\text{g}/\text{m}^3$ of Naïve. MobiAir outperforms the SOTA method SF by 39.5%, exceeds RF by 61.5%, and surpasses MLR by 82.7%. The reason for such outstanding results is that MobiAir can extract effective representation from unlabeled data. Furthermore, we also examine the model effectiveness under scenarios with various air pollutants under different sampling and labeling rates, both of which are important factors in the estimation performance.



(a) The average and normalized MAE comparison when monitoring various air pollutants under the sampling rate of 1Hz among 0.01Hz~1Hz. All models are trained with 1% 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($\mu\text{g}/\text{m}^3$).

(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($\mu\text{g}/\text{m}^3$).

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

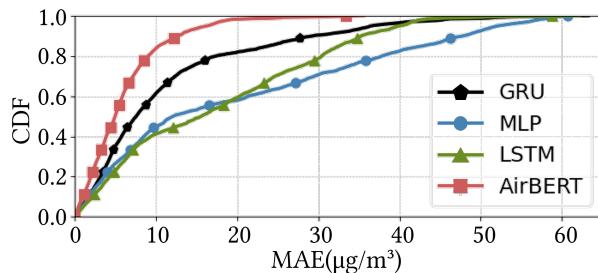


Figure 14: Influence of pretrain encoders shown in CDF. The labeling and sampling rates are 1% and 1Hz.

7.4.1 Effectiveness for various pollutants. We conduct extensive experiments with measurements of 6 common air pollutants collected by our sensing front-end to validate the performance of MobiAir in complicated environments, namely PM_{2.5}, PM₁₀, NO₂, SO₂, CO, and O₃, as shown in Fig.13(a). We scale all Naïve MAE to 1 and use the normalized MAE as the metric. As illustrated, MobiAir always surpasses other baselines for all the pollutant types. Specifically, MobiAir outperforms the SOTA SF by 34.4% for PM_{2.5}, 42.4% for PM₁₀, 65.6% for NO₂, 67.9% for SO₂, 56.1% for CO, 13.2% for O₃. However, compared to Naïve, the performance gain of MobiAir for other air pollutants is not as high as VOC. The reasons are 2-fold. First, the Naïve error depends on sensor sensitivity to pollutants which varies with EC sensor types. Second, in addition to the gas mixing error, the cross interference among pollutant types is another factor causing measurement error which cannot be ignored.

7.4.2 Impact of Sampling Rate. We then verify the effectiveness of MobiAir when the VOC concentration is sampled under different data rates. The results are shown in Fig.13(b). As seen, MobiAir always outperforms the other 3 baselines under all sampling rates. Moreover, as the data sampling rate grows from 0.01Hz~1Hz, the MAE of MobiAir increases from $6.91\mu\text{g}/\text{m}^3$ to $9.51\mu\text{g}/\text{m}^3$ and then decreases again to $6.06\mu\text{g}/\text{m}^3$. The reason is as follows. When the sampling rate is low, there is enough time for the gas sampled at the previous moment to react completely inside the EC sensor before the gas sample at the next moment enters the sensor. It indicates a lower probability for gas samples to mix, which results

Table 3: Performance comparison of MobiAir with various fine-tune decoders, which is shown in MAE($\mu\text{g}/\text{m}^3$). The sampling rate is 1Hz. LR means the labeling rate.

Model \ LR	0.1%	0.5%	1%	2%	5%	10%
MobiAir	7.70	6.08	6.06	5.83	5.59	6.75
AirBERT-GRU	6.47	6.48	7.01	6.51	6.42	7.02
AirBERT-ATTN	7.79	7.75	6.48	8.07	10.08	8.41
AirBERT-LSTM	8.60	6.47	6.48	6.47	6.48	6.49

in lower measurement error. When the sampling rate is high, the effect of mixing gas samples from previous and future moments inside the sensor is significant. However, a higher sampling rate can provide a larger amount of data. MobiAir can learn more time- and spatial-invariant knowledge from massive unlabeled data and more labeled data can better fine-tune the model. In contrast, for the sampling rate of 0.1Hz, MobiAir has no enough data to learn to extract effective representations, which makes it hard to handle the severe gas mixing error. In addition to MobiAir, RF and SF also display a similar trend as shown in Fig.13(b).

7.4.3 Impact of Labeling Rate. We further examine the impact of the labeling rate which determines the proportion of labeled and unlabeled data to train MobiAir. As illustrated in Fig.13(c), when the labeling rate decreases from 10%~0.1%, the MAE of MobiAir increases from $5.84\mu\text{g}/\text{m}^3$ to $8.69\mu\text{g}/\text{m}^3$. The other 3 baselines also show a similar tendency. The result indicates that even though pertaining with unlabeled data helps MobiAir learn and transfer knowledge to extract effective representations, labeled data is still essential for accurate estimation in a supervised manner. Moreover, MobiAir consistently achieves better performances than other baselines under each labeling rate. In particular, the performance gain is significant under low labeling rates.

7.5 Ablation Study

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

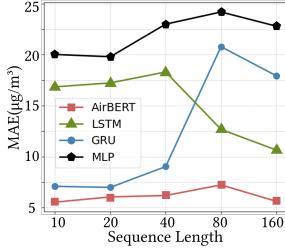


Figure 15: Impact of sequence length. The labeling and sampling rates are 1% and 1Hz.

RD \ MR	0.1	0.2	0.3	0.4	Average
	18	36	72	144	
0.1	11.87	11.67	10.12	7.76	10.35
0.2	7.01	7.72	6.06	7.63	7.10
0.3	14.15	14.52	9.87	12.23	11.94
0.4	8.36	6.62	12.06	11.24	8.97
Average	10.35	10.13	9.52	9.72	-

Figure 16: MobiAir performance shown in MAE($\mu\text{g}/\text{m}^3$) under different representation dimensions and mask ratios. The labeling and sampling rates are 1% and 1Hz.

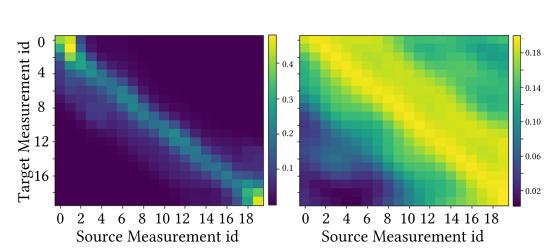


Figure 17: Attention maps of head 0 (left) and head 1 (right) of MobiAir. Head 0 concentrates more on current readings and near-future readings, while head 1 pays more attention to recent-past readings.

7.5.1 Impact of Pre-train Encoder. In this part, we demonstrate the proposed AirBERT model performs better than other pre-train encoders in the estimation task. In this part of the experiment, we compare the AirBERT framework with another 3 encoder structures, long short-term memory (LSTM) [49], Gated Recurrent Unit (GRU) [50], and multilayer perception (MLP). As CDF illustrated in Fig.14, MobiAir with AirBERT achieves lower estimation error than other baselines. Specifically, AirBERT outperforms GRU by 13.4%, surpasses LSTM by 64.8%, and exceeds MLP by 69.4% on average. The delightful results lie in our in-depth study of non-stationary and non-Markovian properties of sensor measurements which are similar to human natural language. Moreover, the BERT-encoder-based structure can better handle both properties and help MobiAir extract more effective sequential representations.

7.5.2 Impact of Fine-tune Decoder. We also compare the estimation performance with different fine-tune decoders in the supervised learning phase. We change the MLP layers in MobiAir with another 3 decoder structures, Gated Recurrent Unit (GRU), multi-head attention (ATTN) [51], and long short term memory (LSTM). As Table 3 shows, MobiAir with MLP achieves slightly enhanced estimation performance for more than 5.4%, 21.7%, and 7.3% compared to GRU, ATTN, and LSTM on average over all labeling rates, respectively. Compared to the pre-train encoders, the performance gain for the fine-tune decoder is not that significant. The result indicates that the representation extraction is more important than the representation-value mapping process. Better representations can contribute to lower estimation errors.

7.6 Micro-benchmark

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

7.6.1 Impact of Sequence Length. To investigate the impact of sequence length on MobiAir’s performance, we conducted experiments varying the length from 10 to 160. Fig.15 reveals that the performance of all tested methods is not consistently related to sequence length. This can be attributed to two reasons. Firstly, longer sequences may increase model complexity, potentially leading to overfitting and higher estimation errors. Secondly, longer sequences result in fewer labeled samples when ground truth is limited, exacerbating the issue of label scarcity during supervised training. As a result, longer sequences do not necessarily lead to lower estimation errors for the tested models.

Table 4: Computation efficiency comparison of MobiAir and baselines in 4 aspects.

Model	Parameters	Size	Train Time	Infer. Time
MobiAir	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

7.6.2 Impact of masking policy. We examine how MobiAir performs on different masking policies. We change the masking ratios (r_m) of the span masking policy from 0.1~0.4. r_m determines the number of masked positions. The span mask with larger r_m can mask more positions, which is more challenging. As shown in Fig.16, span masking with a higher masking ratio obtains a lower estimation error than the low masking ratio. This result suggests that masking more positions may be more effective for representation learning from sensor measurements. The setting of $r_m = 0.3$ achieves the best overall performances and we adopt it in MobiAir.

7.6.3 Impact of Representation Dimension. We further study the impact of the representation dimension (RD) on MobiAir performance. The AiBERT generates representations after self-supervised training. Multiple dimensions from 18~144 are applied and the estimation error is shown in Fig.16. In general, MAE decreases from $10.35\mu\text{g}/\text{m}^3$ to $7.10\mu\text{g}/\text{m}^3$ and then increases again to $8.97\mu\text{g}/\text{m}^3$ on average when the dimension grows from 18 to 144. The reason is as follows. The representation dimension is highly related to the computation complexity of MobiAir. A larger representation dimension can improve the fitting results of MobiAir but may lead to over-fitting. It is a classical trade-off. The larger dimension value also influences the efficiency of MobiAir on edge devices. We set the representation dimension to 36, which can achieve the best performance in our series of experiments. Thanks to the less complexity of sensor measurements compared to human language, the dimension here is much smaller than those in the original BERT model. It reduces the model size significantly as well.

7.6.4 Attention Map Visualization. To understand MobiAir’s operation, we analyze the attention maps of two heads in AirBERT. Fig.17 shows that both heads focus on current sensor measurements. Head 0 also considers nearby future readings, while Head 1 emphasizes recent past measurements. This indicates that leveraging both recent past and close future sensor data during self-supervised learning helps train an accurate and effective MobiAir model for downstream estimation tasks.



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

7.6.5 Computation Overhead. Compared to existing methods, MobiAir offers advantages by being deployable on commercial edge devices without significant resource requirements. Table 4 compares MobiAir with baselines in terms of parameters, model size, training time, and inference time. MobiAir can be continuously trained in the cloud, allowing the edge device to focus on inference. Despite longer training time due to pre-training, MobiAir achieves faster inference than the SOTA SF method. Additionally, MobiAir reduces model size thanks to its cross-layer parameter-sharing mechanism. These results demonstrate MobiAir’s efficiency for real-time estimation on edge devices with high sampling rates, with manageable overhead for most mobile air quality sensing scenarios.

8 APPLICATIONS

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

- *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.

- *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 MobiAir 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 O₃ concentrations started from 11:00 on May 1st. Local camera footage shows that non-standard construction is the main reason causing air pollution.



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

9 RELATED WORK

- **Air quality sensing systems.** Most air quality sensing systems are designed for low-mobility carriers or static scenarios [52–54]. The deployment cost is high for the hundred-meter-level data granularity. Although many MCS systems with high-mobility sensors have been proposed [13, 55, 56], significant deviations of sensor measurements still exist. Taking Sniffer4D and GotchaII [43] 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.

- **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 [57–59], non-parametric methods [60, 61], and particle filter methods [56, 62]. Many promising learning-based methods were proposed recently for air pollution estimation. These methods rely heavily on reference data from official stations [40, 47, 63]. 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.

10 CONCLUSION

In this paper, we propose MobiAir, the first city-scale fine-grained air quality estimation system with mobile sensors of high sampling rates. We first developed AirBERT to process the spatially mixed gas measurements by feature extraction and mutual influence modeling. We then designed a knowledge-informed training strategy based on self-supervised learning to enable MobiAir for city-scale deployment with limited labeled data. The system is deployed in an international city for operation and the experiments show that MobiAir reduces sensing errors by 96.7% with only 44.9ms latency, outperforming the SOTA baseline by 39.5%. In the future, we plan to extend MobiAir to MCS vehicles with higher mobility [64–66].

11 ACKNOWLEDGEMENT

We sincerely thank our shepherd, Xiaofan Jiang, and the anonymous reviewers for their constructive comments and feedback in improving this work. This paper was supported by the National Key R&D program of China No. 2022YFC3300703, the Natural Science Foundation of China under Grant No. 62371269, Guangdong Innovative and Entrepreneurial Research Team Program No. 2021ZT09L197, Shenzhen 2022 Stabilization Support Program No. WDZC2022081110350001, Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan No. JC20220011, and Meituan.

REFERENCES

- [1] Aqmesh. <https://www.aqmesh.com>.
- [2] Cabral Daapp, Madeleine IG et al. Eclipse: an end-to-end platform for low-cost, hyperlocal environmental sensing in cities. In *Proceedings of 21th ACM/IEEE IPSN*, pages 28–40. IEEE, 2022.
- [3] Ecomsmart.<https://ecomesure.com/en/connected-systems/ecomsmart>.
- [4] Ecomtrek.<https://ecomesure.com/en/connected-systems/comtrek>.
- [5] Xiangdong Li, Ling Jin, and Haidong Kan. Air pollution: a global problem needs local fixes. *Nature*, 570(7762):437–439, 2019.
- [6] Jos Lelieveld, John S Evans, Mohammed Fnais, Despina Giannakaki, and Andrea Pozzer. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569):367–371, 2015.
- [7] Gowri Sankar Ramachandran et al. An immersive visualization of micro-climatic data using usc air. In *Proceedings of the 17th ACM Mobicys*, pages 675–676, 2019.
- [8] Joshua S Apte, Kyle P Messier, Shahzad Gani, Michael Brauer, Thomas W Kirchstetter, Melissa M Lundin, Julian D Marshall, Christopher J Portier, Roel CH Vermeulen, and Steven P Hamburg. High-resolution air pollution mapping with google street view cars: exploiting big data. *Environmental science & technology*, 51(12):6999–7008, 2017.
- [9] Yuxuan Liu, Xinyu Liu, Fanhang Man, Chenye Wu, and Xinlei Chen. Fine-grained air pollution data enables smart living and efficient management. In *Proceedings of the 20th ACM SenSys*, pages 768–769, 2022.
- [10] Zuxin Li, Fanhang Man, Xuecheng Chen, Baining Zhao, Chenye Wu, and Xinlei Chen. Tract: Towards large-scale crowdsensing with high-efficiency swarm path planning. In *Proceedings of the ACM UbiComp*, pages 409–414, 2022.
- [11] Xuecheng Chen, Haoyang Wang, Zuxin Li, Wenbo Ding, Fan Dang, Chengye Wu, and Xinlei Chen. Deliversense: Efficient delivery drone scheduling for crowdsensing with deep reinforcement learning. In *Proceedings of the ACM UbiComp*, pages 403–408, 2022.
- [12] Xinlei Chen, Susu Xu, Jun Han, Haohao Fu, Xidong Pi, Carlee Joe-Wong, Yong Li, Lin Zhang, Hae Young Noh, and Pei Zhang. Pas: Prediction-based actuation system for city-scale ridesharing vehicular mobile crowdsensing. *IEEE Internet of Things Journal*, 7(5):3719–3734, 2020.
- [13] Yi Gao, Wei Dong, Kai Guo, Xue Liu, Yuan Chen, Xiaojin Liu, Jiajun Bu, and Chun Chen. Mosaic: A low-cost mobile sensing system for urban air quality monitoring. In *Proceedings of IEEE INFOCOM 2016*, pages 1–9. IEEE, 2016.
- [14] Susu Xu, Xinlei Chen, Xidong Pi, Carlee Joe-Wong, Pei Zhang, and Hae Young Noh. ilocus: Incentivizing vehicle mobility to optimize sensing distribution in crowd sensing. *IEEE Transactions on Mobile Computing*, 19(8):1831–1847, 2019.
- [15] Xinlei Chen, Susu Xu, Haohao Fu, Carlee Joe-Wong, Lin Zhang, Hae Young Noh, and Pei Zhang. Asc: Actuation system for city-wide crowdsensing with ride-sharing vehicular platform. In *Proceedings of the Fourth Workshop on International Science of Smart City Operations and Platforms Engineering*, pages 19–24, 2019.
- [16] Industrial gas sensors | honeywell.
- [17] AlphaSense | market intelligence and search platform.<https://www.alphasense.com/>.
- [18] Gas sensors / FIGARO engineering inc. world leader in gassensing innovation. <https://www.figarosensor.com/>.
- [19] Membrapor the global source of high quality electrochemical gas sensors. <https://www.membrapor.ch>.
- [20] Gas sensor & detector technology leaders | SGX sensor-tech.<https://www.sgxsensortech.com/>.
- [21] Tanishka Goyal, Ankita Singh, Smriti Chhaya, Aditi Vikas, Poorva Garg, Ritika Malik, and Rijurekha Sen. Low cost platform design for pollution measurement in delhi-ncr using vehicle-mounted sensors. In *Proceedings of the 24th ACM MobiCom*, pages 759–761, 2018.
- [22] Haoyang Wang, Yuxuan Liu, Chenyu Zhao, Jiayou He, Wenbo Ding, and Xinlei Chen. Califomer: Leveraging unlabeled measurements to calibrate sensors with self-supervised learning. In *Proceedings of the 2023 ACM UbiComp*, pages 743–748, 2023.
- [23] Yuzhe Yang, Zhiwen Hu, Kaigui Bian, and Lingyang Song. Imgsensingnet: Uav vision guided aerial-ground air quality sensing system. In *Proceedings of IEEE INFOCOM 2019*, pages 1207–1215. IEEE, 2019.
- [24] Ahmed Boubrima, Walid Bechkit, Hervé Rivano, and Lionel Soulhac. Poster: Toward a better monitoring of air pollution using mobile wireless sensor networks. In *Proceedings of the 23th ACM MobiCom*, pages 534–536, 2017.
- [25] Jian Ding and Ranveer Chandra. Towards low cost soil sensing using wi-fi. In *Proceedings of 25th ACM MobiCom*, pages 1–16, 2019.
- [26] Gas sensor market size & share analysis [2023 report].
- [27] Rongfeng Li, Hui Qi, Yuan Ma, Yuping Deng, Shengnan Liu, Yongsheng Jie, Jinzhu Jing, Jinlong He, Xu Zhang, Laura Wheatley, et al. A flexible and physically transient electrochemical sensor for real-time wireless nitric oxide monitoring. *Nature communications*, 11(1):3207, 2020.
- [28] A Terrence Conlisk. *Essentials of micro-and nanofluidics: with applications to the biological and chemical sciences*. Cambridge University Press, 2012.
- [29] Yan Su, Ken-ichi Otake, Jia-Jia Zheng, Satoshi Horike, Susumu Kitagawa, and Cheng Gu. Separating water isotopologues using diffusion-regulatory porous materials. *Nature*, 611(7935):289–294, 2022.
- [30] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [31] Jianghao Wang, Yichun Fan, Juan Palacios, Yuchen Chai, Nicolas Guetta-Jeanrenaud, Nick Obradovich, Chenghu Zhou, and Siqi Zheng. Global evidence of expressed sentiment alterations during the covid-19 pandemic. *Nature Human Behaviour*, 6(3):349–358, 2022.
- [32] Xusong Chen, Dong Liu, Chenyi Lei, Rui Li, Zheng-Jun Zha, and Zhiwei Xiong. Bert4sessrec: Content-based video relevance prediction with bidirectional encoder representations from transformer. In *Proceedings of the 27th ACM MM*, pages 2597–2601, 2019.
- [33] Rui Wang, Dongdong Chen, Zuxuan Wu, Yinpeng Chen, Xiyang Dai, Mengchen Liu, Yu-Gang Jiang, Luowei Zhou, and Lu Yuan. Bevt: Bert pretraining of video transformers. In *Proceedings of the IEEE CVPR*, pages 14733–14743, 2022.
- [34] Basura Fernando, Hakan Bilen, Efstratios Gavves, and Stephen Gould. Self-supervised video representation learning with odd-one-out networks. In *Proceedings of the IEEE CVPR*, pages 3636–3645, 2017.
- [35] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*, 2019.
- [36] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S4l: Self-supervised semi-supervised learning. In *Proceedings of the IEEE/CVF ICCV*, pages 1476–1485, 2019.
- [37] Rayan Krishnan, Pranav Rajpurkar, and Eric J Topol. Self-supervised learning in medicine and healthcare. *Nature Biomedical Engineering*, pages 1–7, 2022.
- [38] Huatao Xu, Pengfei Zhou, Rui Tan, Mo Li, and Guobin Shen. Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications. In *Proceedings of the 19th ACM SenSys*, pages 220–233, 2021.
- [39] Balz Maag, Zimu Zhou, and Lothar Thiele. A survey on sensor calibration in air pollution monitoring deployments. *IEEE Internet of Things Journal*, 5(6):4857–4870, 2018.
- [40] Yun Cheng, Olga Saukh, and Lothar Thiele. Sensorformer: Efficient many-to-many sensor calibration with learnable input subsampling. *IEEE Internet of Things Journal*, 9(20):20577–20589, 2022.
- [41] Sami Kaivonen and Edith C-H Ngai. Real-time air pollution monitoring with sensors on city bus. *Digital Communications and Networks*, 6(1):23–30, 2020.
- [42] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, pages 1–8, 2013.
- [43] Xiangxiang Xu, Xinlei Chen, Xinyu Liu, Hae Young Noh, Pei Zhang, and Lin Zhang. Gotcha ii: Deployment of a vehicle-based environmental sensing system. In *Proceedings of the 14th ACM SenSys*, pages 376–377, 2016.
- [44] OAR US EPA. Air emissions sources.
- [45] Balz Maag, David Hasenfratz, Olga Saukh, Zimu Zhou, Christoph Walser, Jan Beutel, and Lothar Thiele. Ozone and carbon monoxide dataset collected by the opensense zurich mobile sensor network, 2019.
- [46] Yan Tao, Shengquan Mi, Shuhong Zhou, Shigong Wang, and Xiaoyun Xie. Air pollution and hospital admissions for respiratory diseases in lanzhou, china. *Environmental pollution*, 185:196–201, 2014.
- [47] Haomin Yu, Qingyong Li, Yangli Ao Geng, Yingjun Zhang, and Zhi Wei. AirNet: A calibration model for low-cost air monitoring sensors using dual sequence encoder networks. 34(1):1129–1136.
- [48] Balz Maag, Olga Saukh, David Hasenfratz, and Lothar Thiele. Pre-deployment testing, augmentation and calibration of cross-sensitive sensors. In *EWSN*, pages 169–180, 2016.
- [49] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [50] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014.
- [51] Ashish Vaswani, Noam Shazeer, et al. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Proceedings of NIPS*, volume 30. Curran Associates, Inc., 2017.
- [52] Emiel Lagerspetz, Samu Varjonen, Francesca Concas, Julien Mineraud, and Sasu Tarkoma. Megasense: Megacity-scale accurate air quality sensing with the edge. In *Proceedings of the 24th ACM MobiCom*, pages 843–845, 2018.
- [53] Ian Godfrey, José Pablo Sibaja Brenes, María Martínez Cruz, and Khadija Meghraoui. Using uas with sniffer4d payload to document volcanic gas emissions for volcanic surveillance. *Advanced UAV*, 2(2):86–99, 2022.
- [54] Yuta Miyagawa, Norihsisa Segawa, Masato Yazawa, and Masa-yuki Yamamoto. Development of a low-cost gas sensor unit for wide area air pollution monitoring system (poster). In *Proceedings of the 17th ACM Mobicys*, pages 574–575, 2019.
- [55] Sujata Pal, Anindo Ghosh, and Vivek Sethi. Vehicle air pollution monitoring using iots. In *Proceedings of the 16th ACM SenSys*, pages 400–401, 2018.
- [56] Xinlei Chen, Xiangxiang Xu, Xinyu Liu, Shijia Pan, Jiayou He, Hae Young Noh, Lin Zhang, and Pei Zhang. Pga: Physics guided and adaptive approach for mobile

- fine-grained air pollution estimation. In *Proceedings of the ACM UbiComp*, pages 1321–1330, 2018.
- [57] Guodong Li, Xinyu Liu, Zhiyuan Wu, Yue Wang, and Lin Zhang. Robust calibration for low-cost air quality sensors using historical data. In *2020 19th ACM/IEEE IPSN*, pages 349–350. IEEE, 2020.
- [58] Ji Luo, Yiling Hu, Chengzhao Yu, Chaopeng Hong, Xiao-Ping Zhang, and Xinlei Chen. Field reconstruction-based non-rendezvous calibration for low cost mobile sensors. In *Proceedings of the 2023 ACM UbiComp*, pages 688–693, 2023.
- [59] Xinlei Chen, Susu Xu, Xinyu Liu, Xiangxiang Xu, Hae Young Noh, Lin Zhang, and Pei Zhang. Adaptive hybrid model-enabled sensing system (hmss) for mobile fine-grained air pollution estimation. *IEEE Transactions on Mobile Computing*, 21(6):1927–1944, 2020.
- [60] Wilmar Hernandez, Alfredo Mendez, Angela Maria Diaz-Marquez, and Rasa Zalakeviciute. Pm 2.5 concentration measurement analysis by using non-parametric statistical inference. *IEEE Sensors Journal*, 20(2):1084–1094, 2019.
- [61] Yifei Sun, Yuxuan Liu, Ziteng Wang, Xiaolei Qu, Dezhi Zheng, and Xinlei Chen. C-ridge: Indoor co2 data collection system for large venues based on prior knowledge. In *Proceedings of the 20th ACM SenSys*, pages 1077–1082, 2022.
- [62] Xinlei Chen, Xiangxiang Xu, Xinyu Liu, Hae Young Noh, Lin Zhang, and Pei Zhang. Hap: Fine-grained dynamic air pollution map reconstruction by hybrid adaptive particle filter. In *Proceedings of the ACM SenSys*, pages 336–337, 2016.
- [63] Yinan Liang, Ziwei Wang, Xiuwei Xu, Yansong Tang, Jie Zhou, and Jiwen Lu. Mecformer: Deploying vision transformers on microcontrollers with limited memory. *Advances in Neural Information Processing Systems*, 36, 2024.
- [64] Haoyang Wang, Xuecheng Chen, Yuhang Cheng, Chenye Wu, Fan Dang, and Xinlei Chen. H-swarmloc: Efficient scheduling for localization of heterogeneous mav swarm with deep reinforcement learning. In *Proceedings of the 20th ACM SenSys*, pages 1148–1154, 2022.
- [65] Haoyang Wang, Jingao Xu, Chenyu Zhao, Zihong Lu, Chen Cheng, Yuhang, Xuecheng Chen, Xiao-Ping Zhang, Yunhao Liu, and Xinlei Chen. Transformloc: Transforming mavs into mobile localization infrastructures in heterogeneous swarms. In *Proceedings of the IEEE INFOCOM*, 2024.
- [66] Xinlei Chen, Aaveek Purohit, Carlos Ruiz Dominguez, Stefano Carpin, and Pei Zhang. Drunkwalk: Collaborative and adaptive planning for navigation of micro-aerial sensor swarms. In *Proceedings of the ACM SenSys*, pages 295–308, 2015.