IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 17, 2022, accepted March 30, 2022, date of publication April 4, 2022, date of current version April 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3164506

Enhanced Air Quality Inference via Multi-View Learning With Mobile Sensing Memory

NING LIU[1], XINYU LIU[2], PO-TING LIN[1], YUE WANG[1], AND LIN ZHANG[2]
[1]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2]Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

Corresponding author: Lin Zhang (linzhang@tsinghua.edu.cn)

This work was supported in part by the Shenzhen Science and Technology Program under Grant KQTD20170810150821146, and in part by the Tsinghua-Toyota Joint Research Institute Cross-discipline Program.

ABSTRACT Fine-grained air quality can provide essential urban environmental information for administrators and residents. With advances in communication and sensing technologies, low-cost portable sensors installed on vehicles enable high-coverage air quality monitoring. However, data collected by low-cost mobile sensors may be inaccurate and inconsistent in complex operation environments, which brings the issue of data uncertainty. Moreover, due to uncontrolled vehicles and human activities, the coverage of mobile nodes is dynamic over time, leading to uneven or sparse spatial distribution. To address these challenges, we propose AQI-M^3, a novel framework for fine-grained air quality inference via multi-view learning with mobile sensing memory. Specifically, an encoder-decoder structure is applied in the region view for modeling the spatial dependencies in pollution maps. More importantly, sensing gradients are extracted in the trajectory view to enable the utilization of uncertain mobile sensing data. In addition, a memory network is designed to capture the spatial patterns from the historical sensing data and provide the global patterns as a complemental guide to overcome dynamic coverage sampling. Extensive experiments are conducted on three real-world deployments of hybrid sensing systems with both static and mobile sensors. Experimental results show that our proposed approach outperforms competitive baselines with 17%∼29% reduction in mean absolute error. Furthermore, detailed evaluations demonstrate the effectiveness and robustness of the proposed framework under dynamic coverage.

INDEX TERMS Air pollution, mobile computing, multi-view learning, spatiotemporal memory.

I. INTRODUCTION

Air pollution brings a severe threat to human health. Prolonged exposure to polluted air causes respiratory diseases [1] and further exacerbates infection rates and mortality of other diseases, such as diabetes [2], coronavirus disease [3], etc. Therefore, air quality monitoring has become a worldwide concern. In particular, fine-grained air quality information in cities can support regulatory policies and health protection. For city administrators, detailed spatial variations of air quality can enable accurate emission discovery and traffic control. For residents, fine-grained air quality information can help travel arrangements and route planning to reduce exposure risks.

However, governmental stations for routine air quality monitoring are rarely adopted for large-scale deployment due to the high costs and bulk installations. Recently, IoT platforms [4]–[8] with low-cost sensors have been widely deployed as complements to collect high-coverage monitoring data. Nevertheless, some nodes in these systems, e.g., vehicles, are unavailable at certain times due to environmental impacts and human activities [9], [10]. In addition, the long-term operation of numerous devices remains costly. For example, periodic calibration must be performed for low-cost sensors to maintain accurate measurements [7], [11], [12], resulting in remarkably high maintenance costs for large-scale systems. Therefore, an inference algorithm is still necessary to reconstruct fine-grained air quality information from existing coarse-grained data, thereby supporting an accurate understanding of the city environment and reducing long-term operational costs.

As machine learning algorithms develop and the collected data grows, data-driven methods [13], [14] have been proposed to model the mapping of available sensing samples to

The associate editor coordinating the review of this manuscript and approving it for publication was Ilaria De Munari.

target locations. However, these methods suffer from computational complexity when the number of sensors explodes. Inspired by the image super-resolution studies [15], [16], reconstruction algorithms for pollution maps are recently proposed in [6], [9], [17], in which the interest area is treated as a grid and the sensing concentrations are aggregated as pixel values. Nevertheless, these methods lack the effective utilization of fine-grained information from mobile sensing data due to the gridding operation on the original samples. Two major challenges still remain in the fine-grained air quality inference with mobile sensing as follows:

i) *Data Uncertainty*. On the one hand, mobile sensors are deployed in complicated environments, e.g., outdoor rooftops and car trunks, which may degrade the measurement accuracy of low-cost sensors. On the other hand, the response parameters, such as sensitivity, can be inconsistent due to unsynchronized calibration between multiple sensors in long-term operation.

ii) *Dynamic Coverage*, including uneven and sparse sampling coverage. Many mobile sensors are deployed on vehicles, e.g., taxis, with uncontrollable routes, causing uneven distribution issue. For the sparse coverage problem, a common phenomenon is that most mobile nodes congregate in busy urban areas during the day, and stop operations at night.

To address the above challenges, we propose a novel framework for fine-grained **a**ir **q**uality **i**nference via **m**ulti-view learning with **m**obile sensing **m**emory (**AQI-M^3**) in this paper. Our proposed framework includes region-view and trajectory-view learning for modeling data with different granularity. Under region view, an encoder-decoder structure is adopted to capture the spatial correlation in pollution maps. To utilize the mobile sensing data with uncertainty, sensing gradients are extracted on each mobile device under the trajectory view as current fine-grained spatial patterns. Furthermore, a memory network is designed to exploit the historical mobile sensing information and provide global spatial patterns related to the context. Finally, the two aspect spatial patterns are adaptively integrated to guide air quality inference by reinforcing more informative regions. Therefore, the dynamic coverage challenge is addressed by introducing global spatial information. Evaluations on three real-world datasets show that our proposed framework achieves better performance and robustness in the air quality inference with mobile sensing.

The main contributions of this paper can be summarized as follows:

- We propose a framework with multi-view learning for air quality inference, in which sensing gradients are extracted under trajectory view to enable the utilization of fine-grained mobile sensing data with uncertainty.
- We design a novel mobile sensing memory network to capture global spatial patterns, in which historical information is effectively utilized to tackle the dynamic coverage issue of mobile sensing.
- Extensive experiments are conducted on real-world systems deployed in three cities, demonstrating that our proposed AQI-M^3 obtains better performance and robustness compared with state-of-the-art methods.

The rest of this paper is organized as follows. Section II introduces related work on air quality inference methods and memory networks for spatiotemporal inference. In Section III, we formulate the air quality inference problem, present the AQI-M^3 framework with multi-view learning, and further elaborate the mobile sensing memory in the trajectory view. Section IV evaluates our algorithm and compares it with baselines. Finally, we conclude the paper and discuss some future work in Section V.

II. RELATED WORKS

In this section, we introduce some existing air quality inference algorithms and memory networks employed in spatiotemporal inference. Besides, we discuss the comparison between the existing works and our method.

A. AIR QUALITY INFERENCE IN SENSOR NETWORKS

With the development of Internet-of-Things (IoT) technologies, wireless sensing systems with static and mobile sensors are deployed for air quality monitoring, including *OpenSense* [18], *AirCloud* [13], *Gotcha* [19], *City of Things* [20], and so on. Due to the increasing number of observed samples collected in these systems, data-driven air quality inference algorithms have been widely investigated. OpenSense [18] applies land-use regression (LUR) to model the correlation between pollution concentrations and external factors, which enables inference over uncovered areas. Similarly, AirCloud [13] employs Gaussian Process Regression (GPR), considering location information and weather conditions, to achieve inference on unknown concentrations. Recently, researchers focus on deep learning techniques due to their powerful modeling capabilities. The neural attention model [21] is presented to estimate measurements of target locations with dynamic weights among multiple monitoring nodes. However, these approaches are applied to static sensing networks and lack consideration of the information from mobile sensing. To incorporate mobile sensing samples into a unified learning framework, some works, e.g., convolutional networks [9], [17], ConvLSTM [6], and variational autoencoder [20], are deployed to perform air quality inference based on gridded pollution maps. Nevertheless, the fine-grained information of mobile sensing is still not efficiently utilized in these algorithms due to the gridding process.

Therefore, the effective utilization of fine-grained mobile sensing data for air quality inference remains an understudied problem in existing works. Different from previous methods, we adopt a multi-view learning framework in which the gridded pollution map and mobile sensing data are respectively exploited under the region view and trajectory view. Furthermore, we extract sensing gradients to guide the inference of pollution maps instead of directly employing the original

mobile sensing samples, thus addressing the challenge of data uncertainty.

B. SPATIO-TEMPORAL MEMORY NETWORKS

Memory networks [22] can preserve long-term patterns from historical data and are therefore widely used in spatio-temporal inference application, e.g., visual understanding [23], [24], traffic prediction [25]–[27], etc. Inspired by the periodicity and trendiness of urban activities, temporal memory [26], [28] directly introduces multi-scale spatio-temporal data to provide historical information for inference. However, the assumptions of periodicity and trendiness can not be met in the fine-grained air quality inference problem. This is because urban air quality depends not only on the activity within the city, but also on other external factors, including pollution transport, meteorological conditions, etc. Besides, spatial memory [23]–[25], [27] is applied in spatio-temporal prediction, where information from similar regions is aggregated into location features, and then provided as supplementary information for inference at each location. Nevertheless, clustered location features are insufficient to capture comprehensive and complex urban spatial patterns for air quality inference.

Unlike previous memory networks, our memory network can more accurately capture spatial patterns in different contexts by introducing external context knowledge related to air pollution. Moreover, the historical sensing gradient information from fine-grained mobile sensing is stored and read according to external factors, representing the corresponding comprehensive spatial pattern of pollution.

III. AQI-M^3: MULTI-VIEW LEARNING FRAMEWORK WITH MOBILE SENSING MEMORY

In this section, we first formulate the fine-grained air quality inference problem mathematically and introduce the proposed AQI-M^3 framework in detail. Then, we elaborate on the mobile sensing memory network in our proposed approach. Finally, we present the details of model training.

A. PROBLEM FORMULATION

To formulate the problem of fine-grained air quality inference clearly, we first give some relevant definitions as follows:

Sensing Region. Given an air quality monitoring area, we can partition it into $H \times W$ grids of equal size according to the longitude and latitude, where each grid cell denotes a *sensing region*. Partitioning with larger H and W, which means smaller *sensing regions*, we can obtain more fine-grained sensing of air quality.

Pollution Map. The air quality monitoring system consists of multiple sensors, which are deployed at fixed locations or on mobile nodes to collect measurements. Thus, for a particular time slot t, we can reformulate the samples to construct a *pollution map* denoted as $\boldsymbol{Q}_t \in \mathbb{R}_+^{H\times W}$, where each entry $q_{h,w,t}$ is the average value of all pollution concentration measurements in the *sensing region* (h, w) at that time slot.

Pollution Trajectory. The mobile sensors in the system report the pollution concentration with a time stamp and location coordinates at a specific frequency. Therefore, the *pollution trajectory* of the mobile sensor S_n in the time slot t can be denoted as $\Theta_t^{(n)} = \{(x_1, y_1, v_1), (x_2, y_2, v_2), \cdots\}$, where x and y are the longitude and latitude coordinates, and v is the concentration value of the sample.

Then we define the problem of fine-grained air quality inference as follows: for a certain time slot t, given the coarse-grained pollution map $\boldsymbol{Q}_t^c \in \mathbb{R}_+^{H\times W}$, the corresponding external features $\boldsymbol{e}_t$, the pollution trajectory set $\Theta_t = \{\Theta_t^{(1)}, \cdots, \Theta_t^{(N)}\}$ from N mobile sensors and an upscale factor m, the objective is estimating the fine-grained pollution map $\boldsymbol{Q}_t \in \mathbb{R}_+^{Hm\times Wm}$, that is,

$$\hat{\boldsymbol{Q}}_t = \arg\max_{\boldsymbol{Q}_t} p(\boldsymbol{Q}_t | \boldsymbol{Q}_t^c, \boldsymbol{e}_t, \Theta_t). \tag{1}$$

B. FRAMEWORK OF AQI-M^3

1) OVERVIEW

We first present the proposed AQI-M^3 framework, as shown in Fig. 1. The framework adopts multi-view learning scheme, which consists of two views: ***region view*** and ***trajectory view***. More specifically, under the *region view*, an *Encoder-Decoder* structure is applied to learn spatial dependencies in pollution map for fine-grained air quality reconstruction. Besides, there are three main components in the *trajectory view*: a *Gradient-based Attention Branch* to extract the sensing gradient and generate spatial attention from the pollution trajectory, a *Mobile Sensing Memory Net* to preserve global spatial patterns based on auxiliary knowledge, and a *Neural Gating Net* to fuse the global information and the current gradient-based attention. Finally, the spatial attention obtained under the trajectory view can guide the region-view learning, by allowing the *Encoder-Decoder* network to focus on the high-frequency information in the pollution map. We then introduce the details of multi-view learning in the framework individually as follows.

2) REGION-VIEW LEARNING WITH ENCODER-DECODER STRUCTURE

To capture the spatial dependencies in the pollution map, we adopt an encoder-decoder structure for region-view learning in the proposed framework. The encoder-decoder network is widely used in machine translation [29], speech recognition [30], image restoration [31], image super-resolution [32], etc. Following some spatio-temporal inference models [9], [15], we further introduce the residual block into the encoder-decoder network to infer the fine-grained pollution maps.

Specifically, we first apply a convolutional layer in *Encoder* to extract low-level features:

$$\boldsymbol{F}_0^c = E_0(\boldsymbol{Q}_t^c), \tag{2}$$

where E_0 denotes the convolutional layer and Q_t^c is the coarse-grained pollution map. Then, the *Encoder* gradually

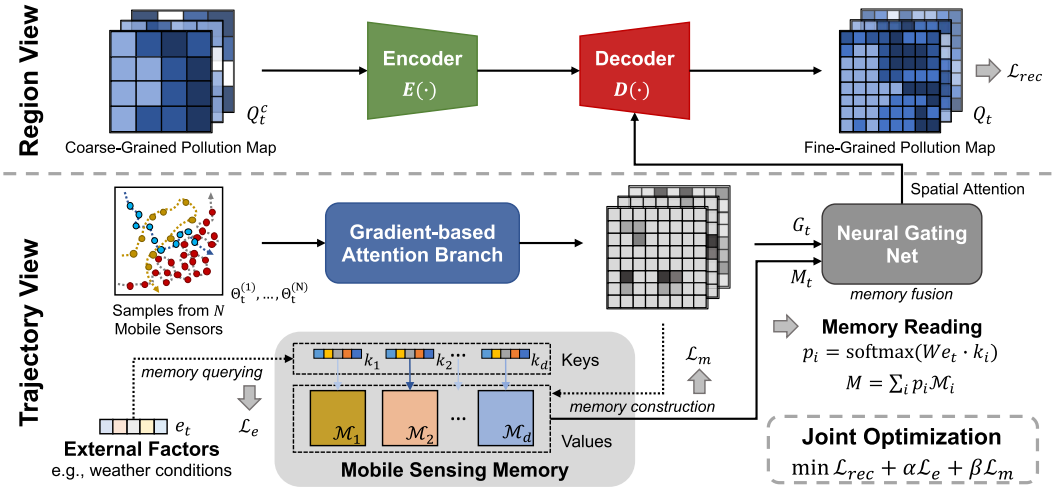


FIGURE 1. AQI-M^3 framework for air pollution inference.

extracts high-level spatial features by R stacked residual blocks, where a skip connection is applied to provide lower-level feature information and avoid the gradient vanishing problem. The output feature map of the r-th residual block can be represented by:

$$\boldsymbol{F}_r^c = E_r(\boldsymbol{F}_{r-1}^c) + \boldsymbol{F}_{r-1}^c, \tag{3}$$

where E_r denotes the convolutional layers in the r-th residual block and more details of its structure can be found in the Appendix V. Eventually, the coarse-grained residual feature is extracted by the *Encoder*, and then fed into the *Decoder* to reconstruct the fine-grained information further.

In the *Decoder*, we first employ a convolutional layer D_0 with a long skip connection to combine the low-level and high-level features extracted by the *Encoder*:

$$\boldsymbol{F}_0 = D_0(\boldsymbol{F}_R^c) + \boldsymbol{F}_0^c. \tag{4}$$

Subsequently, the sub-pixel blocks [33] are introduced in the *Decoder* to reconstruct the fine-grained information from the output features of *Encoder*, in which each block can convert the input feature map into a $2\times$ upscaling map. Afterward, a convolutional layer with tanh activation function maps the feature map to an upscaling residual pollution map, which is added to the nearest neighbor upsampling of the coarse-grained map to generate the fine-grained pollution map.

However, all the samples are aggregated into a grid map according to their timestamp and coordinates under this region view, thereby sacrificing the fine-grained sensing information from large-scale sensors, especially from the mobile sensors carried by vehicles.

3) TRAJECTORY-VIEW LEARNING WITH GRADIENT INSIGHTS

In order to exploit the abundant information in mobile sensing, we propose trajectory-view learning with gradient insights. On the one hand, the mobile sensors can provide massive samples with high temporal and spatial resolution, but they also bring more noise due to the complicated working conditions. Hence, due to their uncertainty, these original samples cannot be used directly as high-dimensional features. On the other hand, the spatial variation of sensing concentration in a region contains information about pollution characteristics of this region and its neighbor, e.g., sudden fluctuation often means the existence of a nearby pollution source. Therefore, we introduce the sensing gradient in trajectory view to represent the intensity of this variation, which can be used to indicate where the more informative regions are.

Specifically, as shown in Fig. 2, our proposed *Gradient-based Attention Branch* includes three phases: pollution trajectory denoising, sensing gradient extraction, and spatial attention generation. First, we adopt the Daubechies 4 wavelet method on each pollution trajectory to reduce the high-frequency sampling noise from the low-cost sensor. Then, for a pollution trajectory $\Theta^{(n)}$ through the region (h, w), we select the first-K and last-K samples in the descending order of concentration values as the set $\mathbb{F}^{(n)} = \{f_1, f_2, \cdots, f_K\}$ and $\mathbb{L}^{(n)} = \{l_1, l_2, \cdots, l_K\}$ respectively for the further extraction, where sample pairs $\{(f_1, l_1), \cdots, (f_K, l_K)\}$ can be organized from the set $\mathbb{F}$ and $\mathbb{L}$. And the sensing gradient of each sample pair can be calculated by:

$$g_k^{(n)} = \frac{c_{f_k} - c_{l_k}}{d_k}, \tag{5}$$

where c_{f_k} and c_{l_k} are the pollution concentration of the k-th sample pair, and d_k represents the geospatial distance between the k-th sample pair. Later, the sensing gradient in the region (h, w) from trajectory $\Theta^{(n)}$ can be written as:

$$\mathcal{G}_{h,w}^{(n)} = \frac{\sum_{k=1}^{K^{(n)}} g_k^{(n)}}{K^{(n)}}. \tag{6}$$

The gridded sensing gradient $\mathcal{G}$ can be obtained by applying an average operation on pollution trajectories in each

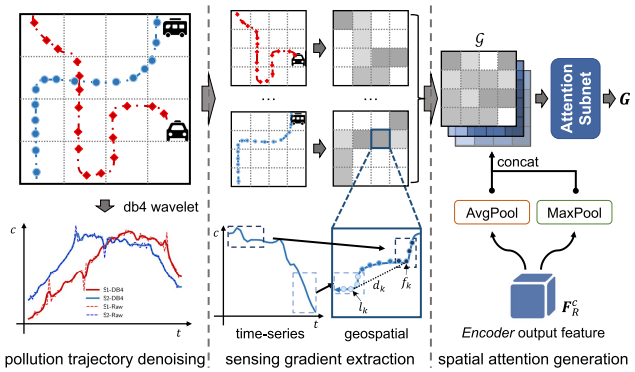


FIGURE 2. Diagram of *Gradient-based Attention Branch*, which includes three phases: pollution trajectory denoising, sensing gradient extraction, and spatial attention generation.

region. Subsequently, the gradient $\mathcal{G}$ is concatenated with the max-pooling output and average-pooling output of the encoded feature maps $\boldsymbol{F}_R^c$, and then fed into an attention subnet to generate current gradient-based attention matrix $\boldsymbol{G}$ as:

$$\boldsymbol{G} = \text{AttSubnet}(\mathcal{G}, MaxPool(\boldsymbol{F}_R^c), AvgPool(\boldsymbol{F}_R^c)), \quad (7)$$

where $\text{AttSubnet}(\cdot,\cdot,\cdot)$ is the attention subnet and more details of its structure can be found in the Appendix V.

However, the samples from mobile sensors are usually dynamic and uneven across regions, so gradient information of a location may be unavailable at certain times. To avoid this limitation, we introduce a memory network to utilize the long-term historical gradient information. The historical gradient information is queried by the current external factors, and fused with the current sensing gradient by the *Neural Gating Net*. Then *Neural Gating Net* generates spatial attention map for the *Decoder* in region view. Guided with gradient-based attention map, the backbone network in the region view can focus on learning the informative regions in the pollution map. We elaborate on the mobile sensing memory network design in the following subsection.

C. SPATIAL ATTENTION WITH MOBILE SENSING MEMORY

1) MOBILE SENSING MEMORY NETWORK IN TRAJECTORY VIEW

To address the dynamic sampling problem mentioned in the above subsection, we propose a novel *Mobile Sensing Memory Net* (MSMN) illustrated in Fig. 3. The spatial pattern of air pollution are dynamic and related to external factors, such as weather conditions [34] and human activities [35]. Therefore, different from the previous work [25], [27], we redesign a key-value pair scheme in MSMN, including the context knowledge refined keys $\boldsymbol{k}_i \in \mathbb{R}^P$ and the mobile sensing related values $\boldsymbol{\mathcal{M}}_i \in \mathbb{R}^{H\times W}$. Under the design of the key-value pair, MSMN consists of four main operations: memory querying, key refining, value construction, and memory reading. The details of the operations are introduced as follows:

Memory Querying by External Factors. We adopt the external factors related to the spatial patterns of air pollution as the original querying input. Each categorical feature in the external factors is converted into a low-dimensional vector through the embedding layer. Then, we concatenate them all together with the other continuous features as the external feature vector $\boldsymbol{e}_t$. The vector $\boldsymbol{e}_t$ is projected into the same feature space as key $\boldsymbol{k}_i$ by a fully connected layer, and then a relevant probability can be calculated by:

$$p_{t,i} = \sigma(\boldsymbol{W}\boldsymbol{e}_t \cdot \boldsymbol{k}_i), \quad (8)$$

where $\cdot$ denotes inner product, $\boldsymbol{W}$ is the parameter matrix of the fully connected layer, and $\sigma(x_i) = e^{x_i}/\sum_j e^{x_j}$.

Key Refining with Context Knowledge. To ensure that the queried values have higher correlation with the corresponding external factors, we introduce a key refining with context knowledge in MSMN. Specifically, for a memory with d key-value pairs, we first cluster the external factors into d clusters. Based on the clustering result, the key refining loss can be formulated as:

$$\mathcal{L}_e = -\sum_t \sum_i o_{t,i} \log(p_{t,i}), \quad (9)$$

where $o_{t,i}$ is the i-th element in the one-hot encoding $\boldsymbol{o}_t$ of the clustering result at time t.

Value Construction with Sensing Gradient. We adopt current gradient-based attention matrix $\boldsymbol{G}$ to construct the values in the memory, and define a memory construction loss as follows:

$$\begin{aligned} u_t &= \arg\max_i \ p_{t,i}, \\ \mathcal{L}_m &= \sum_t \left\| \boldsymbol{G}_t - \boldsymbol{\phi}_t \odot \boldsymbol{\mathcal{M}}_{u_t} \right\|_{1,1}, \end{aligned} \quad (10)$$

where $\odot$ is the element-wise multiplication, and $\boldsymbol{\phi}_t$ denotes the mask matrix padded with 1 in the available sensing region. In other words, we expect that in the available sensing region, the memory value matrix under a specific context has a substantial similarity with the corresponding gradient-based attention matrix.

Memory Reading with Attention Mechanism. Inspired by the attention mechanism, we weight each memory value matrix according to the relevant probability obtained in memory querying, and the output matrix of memory reading is defined as follows:

$$\boldsymbol{M}_t = \sum_i^d p_{t,i} \boldsymbol{\mathcal{M}}_i. \quad (11)$$

Thus, an attention matrix that reflects the global spatial pattern in the current context can be extracted.

After extracting the attention matrixes from current sensing and global memory, we need to integrate them for comprehensive modeling. Considering the dynamic spatial coverage of mobile sensing, we propose a *Neural Gating Net* to adaptively adjust the importance of different information for

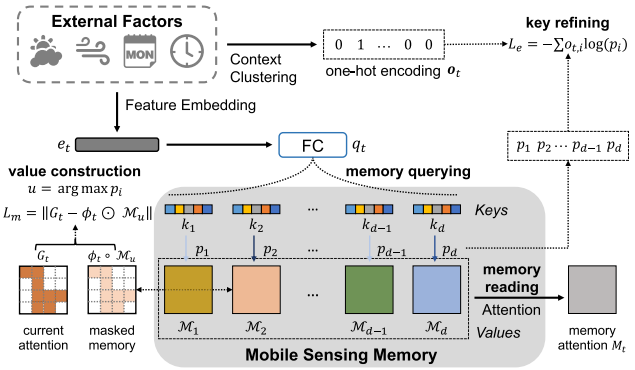


FIGURE 3. Illustration of *Mobile Sensing Memory Net* (MSMN), which consists of four main operations: memory querying, key refining, value construction, and memory reading.

attention matrixes fusion. For example, the global information should be more relied upon when vehicles equipped with mobile sensors stop operating at night. Specifically, we first learn a gate matrix by a convolutional layer based on current gradient-based attention $\boldsymbol{G}_t$ and extracted global memory $\boldsymbol{M}_t$:

$$\boldsymbol{V}_t = \sigma(H_c(\boldsymbol{G}_t) + H_g(\boldsymbol{M}_t)), \tag{12}$$

where $\sigma(\cdot)$ indicates the sigmoid activation function, H_c and H_g are the convolution functions performed on the current and global attention matrixes, respectively. Afterward, the final spatial attention map can be obtained by a nonlinear convolutional transformation with a weighted concatenation of the two types of information. Formally, the transformation operation is defined as follows:

$$\widetilde{\boldsymbol{G}}_t = H_\sigma(\boldsymbol{V}_t \odot \boldsymbol{G}_t \parallel (1 - \boldsymbol{V}_t) \odot \boldsymbol{M}_t), \tag{13}$$

where H_σ denotes the nonlinear convolutional layer with sigmoid activation function, and $(\cdot \parallel \cdot)$ denotes the concatenation operation on two tensors.

2) SPATIAL ATTENTION-GUIDED INFERENCE IN REGION VIEW

With the above fusion of current sensing and global memory, the spatial attention map can help the *Decoder* in region view to comprehensively understand the spatial pattern of air pollution, thereby guiding the *Decoder* to focus on the more informative parts of the pollution map. For example, the information from regions with drastic changes in the pollution concentrations should be emphasized, while some areas with strong stability should be suppressed. This process can be summarized as follows:

$$\hat{\boldsymbol{Q}}_t = U_m(\boldsymbol{Q}_t^c) + D(\boldsymbol{F}_0) \odot \widetilde{\boldsymbol{G}}_t, \tag{14}$$

where the U_m denotes the nearest neighbor upsampling with scaling factor m, D denotes the sub-pixel blocks and subsequent convolutional output layers in *Decoder*, and $\boldsymbol{F}_0$ is the integration of the low-level and high-level features extracted by the *Encoder* in (4).

D. MODEL TRAINING

Finally, we introduce the end-to-end training approach of the AQI-M^3 framework. Since there are no trainable parameters in pollution trajectory denoising and sensing gradient extraction, we treat them as preprocessing procedures. Considering that some regions in the ground-truth pollution map may not be covered at certain times, we propose the masked mean absolute error (MAE) as the reconstruction loss function, which can be defined as follows:

$$\mathcal{L}_r = \sum_t \left\| (\hat{\boldsymbol{Q}}_t - \boldsymbol{Q}_t) \odot \phi_t \right\|_{1,1}, \tag{15}$$

where ϕ_t denotes a binary mask matrix padded with 1 in available regions of ground truth $\boldsymbol{Q}_t$. Finally, the Adam optimizer [36] is adopted to learn the parameters of the AQI-M^3 by minimizing the following loss function:

$$\mathcal{L} = \mathcal{L}_r + \alpha \mathcal{L}_e + \beta \mathcal{L}_m, \tag{16}$$

where $\mathcal{L}_e$ is the key refining loss defined in (9), $\mathcal{L}_m$ is the memory construction loss defined in (10), and α, β are hyper-parameters to balance the importance of their corresponding losses.

IV. EVALUATION

To evaluate the proposed AQI-M^3 algorithm, we design a fine-grained air pollution sensing system with both fixed and mobile sensors. We conduct extensive experiments over three datasets collected from real-world deployments, and further analyze the performance of our proposed method for air quality inference.

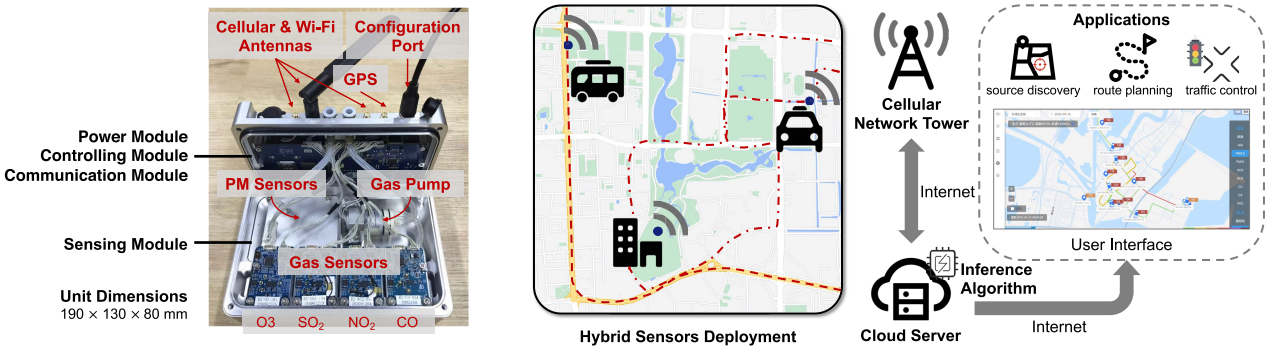
Specifically, we aim to answer the following research questions through the experiment results:

- RQ1: Whether AQI-M^3 can outperform the state-of-the-art methods for air quality inference with mobile sensing system?
- RQ2: How do the designed components, e.g., multi-view learning and mobile sensing memory, in AQI-M^3 contribute to the inference performance?
- RQ3: How do the key hyper-parameters, i.e., the number of memory slots d, loss weights α, and β, in AQI-M^3 affect its performance?
- RQ4: Whether the proposed memory mechanism can better utilize historical information and alleviate the impact of dynamics coverage in mobile sensory data?

A. SYSTEM DESIGN AND DEPLOYMENT

1) HARDWARE DESIGN AND SYSTEM ARCHITECTURE

We design an integrated air sensing unit to collect air pollution data, which can be installed on both buildings and vehicles. Fig. 4a shows the internal structure of the sensing unit, that mainly contains four modules: *i) sensing module*, including four types of gas sensors, a particulate matter sensor, a temperature sensor and a humidity sensor, *ii) controlling module*, that pumps outside air for the sensors in *sensing module*, *iii) communication module*, that



(a) Air Sensing Unit Design.

(b) Overview of System Framework.

FIGURE 4. Hardware design and system architecture of our fine-grained air quality sensing system. We design a sensing unit that can be deployed in fixed and mobile mode for data collection. The deployed units transmit the measurements to the cloud server through the 4G network. And the inference algorithms run on the cloud server based on collected data, further supporting personal and management applications.

(a) (fixed) rooftop

(b) (mobile) bus

FIGURE 5. Typical deployment locations for the air sensing units.

transmits data packet including measurement results, timestamps, GPS coordinates and instructions, *iv) power module*, that supplies power for all procedures of the sensing unit.

As shown in Fig. 4b, distributed air sensing units are deployed for air quality data collection, and then they transmit the data packets to the *cloud server* for storage and further process. In addition, the system provides a *user interface* to display air quality levels from real-time monitoring and fine-grained pollution maps produced by inference algorithms. Furthermore, the system supports various personal and management applications, including route planning without pollution exposure, pollution source location, and traffic control assistance.

2) REAL-WORLD DEPLOYMENTS

Fig. 5 presents the fixed and mobile deployment modes of our sensing unit. For the fixed configuration, the rooftop is a typical deployment location, where the operation condition is stable and maintenance is convenient. Thus this fixed configuration can enable long-term and high-precision monitoring. Besides, buses, taxis, and environmental cruisers are used as the carrier of the sensing unit in the mobile configuration. Specifically, buses follow the preset routes, while the trajectories of taxis and cruisers can be considered random. All sensing units report monitoring data to the cloud server every three seconds.

B. EXPERIMENTAL SETUP

1) DATASETS DESCRIPTION

The sensing units are deployed in specific areas of three cities in China for data collection, including Tianjin, Nanjing, and Foshan. Table 1 shows the corresponding detailed deployment time span and spatial coverage, where the number of sensors is further divided into the number of static sensors (S) and the number of mobile sensors (M). It can be noticed that these three areas are located in different provinces, and the distance between them is more than 700 kilometers. Therefore, there are different climatic conditions, geographical features and regional functions in these areas, ensuring the diversity of experimental data.

The dataset of each city contains two parts: air pollution data from our sensing units and meteorological data from official weather stations.

a: AIR POLLUTION DATA

Benefiting from our distributed sensing units with fixed and mobile deployment, we can obtain fine-grained concentration values of seven types of pollutants as shown in Table 1. We focus on one type of these pollutants, $PM_{2.5}$, for performance evaluation, which is an important factor in determining air quality levels and affecting human health according to the World Health Organization (WHO) [37].

All sensors have been calibrated in the laboratory before system deployment, and we also conduct in-field recalibration during the data collection. Therefore, we can assume that the accuracy of all sensors is consistent, so that samples from different sensors can be treated equally in the subsequent processing. Then, the raw air pollution data can be further preprocessed in the following two steps:

- *denoising*: In order to reduce the high-frequency noise in actual operation, especially the impact of physical

TABLE 1. Details of the real-world datasets collected from three cities.

Dataset	Tianjin	Foshan	Nanjing
Time Span	Jan. 1, 2018 - Dec. 31, 2018	Oct. 8, 2018 - Nov. 25, 2018	Apr. 1, 2020 - Oct. 20, 2020
Operation Hours	06:00 - 20:00	07:00 - 22:00	07:00 - 21:00
Latitude Range	39.095°N - 39.140°N	23.026°N - 23.074°N	32.010°N - 32.045°N
Longitude Range	117.725°E - 117.780°E	113.116°E - 113.170°E	118.760°E - 118.830°E
Cover Area	23.78km^2	29.43km^2	25.67km^2
# of Sensors	72 (35S+37M)	61 (39S+22M)	156 (68S+88M)
Pollution Type	$PM_{1.0}$, $PM_{2.5}$, PM_{10}, O_3, SO_2, NO_2, CO		
Auxiliary Factors	timestamp, temperature, humidity, visibility, wind speed, wind direction		

environment fluctuations during the node movement, we apply Daubechies 4 wavelet over the raw measurement of each sensor.

- *gridding*: The measurements from different sensors are aggregated to a grid according to their sampling time and location, and then the concentration of each region in the grid is assigned with the aggregated mean concentration. The time interval of aggregation is set to 60 minutes. The grid parameter we set here is 20×20, and then we perform downsampling on grid maps to 10×10 as the coarse-grained inputs to evaluate the inference performance.

With these settings, the inferred pollution map results have a granularity of about 200 meters to 300 meters, which can support street-level applications.

b: AUXILIARY FACTORS

Meteorological data is obtained from an open API of Dark Sky.[1] As shown in Table 1, we select five factors related to pollution patterns, including temperature, humidity, visibility, wind speed and wind direction. Besides, the factors related to human activity, i.e., *day (of the week)* and *hour (of the day)*, can be extracted from the timestamp of sensor recordings.

2) COMPARED BASELINES

We consider two groups of competitive baselines for performance comparison as below.

a: BASELINES WITHOUT ST MEMORY

- *Inverse Distance Weighting (IDW)*: IDW [38] is a popular spatial interpolation method for air quality inference. This method considers the distance between the observed and the predicted sample for calculating interpolation weights.
- *Gaussian Process Regression (GPR)*: GPR is a Bayesian machine learning approach that is widely used in air quality inference [13], and we adapt it for the spatiotemporal dependencies modeling as a competitive baseline.
- *ConvLSTM*: ConvLSTM [39] is an effective deep learning model for spatial-temporal inference. ConvLSTM can capture the complex nonlinearity in both spatial and temporal dependencies.
- *SRResNet*: SRResNet [15] is a deep learning model based on convolutional network integrating the residual structure, which is considered as an effective spatial-temporal inference method in some works [9], [40].

b: BASELINES WITH ST MEMORY

- *ST-ResNet*: ST-ResNet [28] utilizes both short-term and long-term information to model the spatial correlation. This model is also a deep learning model based on convolutional network and residual structure.
- *STMN*: Spatial-Temporal Memory Network [26] improves ST-ResNet by introducing ConvLSTM to capture spatial and temporal dependencies together.
- *MetaTP*: MetaTP [27] proposes an external memory network to leverage the global temporal and spatial pattern, and we adapt it into a multipoint version for gridded spatial data inference in our evaluation.

3) EVALUATION METRICS

We adopt the following three metrics to evaluate the performance of our proposed algorithm:

- Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left\|\hat{\boldsymbol{Q}}_t^{\phi} - \boldsymbol{Q}_t^{\phi}\right\|_F^2},$$

- Mean Absolute Error

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left\|\hat{\boldsymbol{Q}}_t^{\phi} - \boldsymbol{Q}_t^{\phi}\right\|_{1,1},$$

- Mean Absolute Percentage Error

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left\|(\hat{\boldsymbol{Q}}_t^{\phi} - \boldsymbol{Q}_t^{\phi}) \oslash \boldsymbol{Q}_t^{\phi}\right\|_{1,1},$$

where N is the total number of testing samples, ϕ represents the binary mask operation based on ground truth, $\hat{\boldsymbol{Q}}_t^{\phi}$ denotes the inferred fine-grained air quality map with non-zero mask at t-th time slot, and $\boldsymbol{Q}_t^{\phi}$ denotes corresponding ground truth.

4) IMPLEMENTATION DETAILS

Our implementation is based on TensorFlow 2.3.0, and all neural network models are trained using an Nvidia GTX Titan X GPU with 12 GB memory. The codes of baseline

[1] https://darksky.net/dev

TABLE 2. Performance comparison under the task of inferring pollution map from 10 × 10 to 20 × 20 (upscale factor $m = 2$) with 5 repeated trials.

Method	Tianjin			Foshan			Nanjing		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
IDW [38]	4.756	2.335	7.535	3.184	1.942	5.069	3.857	2.091	5.751
GPR [13]	4.470	2.143	7.071	3.201	1.921	4.720	4.058	2.095	5.807
ConvLSTM [39]	2.847	1.421	5.468	2.949	1.961	5.301	3.585	1.757	4.883
SRResNet [9]	2.978	1.574	6.866	2.861	1.866	4.910	3.521	1.792	5.051
ST-ResNet [28]	2.960	1.511	6.017	2.910	1.902	5.032	3.523	1.786	5.037
STMN [26]	2.838	1.447	5.597	2.918	1.937	5.141	3.562	1.708	4.861
MetaTP [27]	2.663	1.356	5.602	2.831	1.870	4.909	3.234	1.625	4.881
AQI-M^3	**2.392**	**1.122**	**4.036**	**2.151**	**1.317**	**3.047**	**2.987**	**1.271**	**3.520**

methods are from the original papers or public resources. In our experiment, Adam [36] optimizer with the learning rate of 0.001 is used for model training, and the batch size is fixed on 32 for all methods. To be fair, we set the number of stacked residual blocks $R = 16$ for all deep convolutional models and train all ANN-based models for 200 epochs with early stopping. Moreover, the mean absolute error (MAE) is adopted as the loss function for the optimization. The size of the convolution kernel is set to 3×3 in all residual blocks of the region-view module. Besides, convolution layers with 1×1 kernels are used in the spatial attention subnet of the trajectory-view module. Furthermore, the channel size of hidden convolution layers in region view and trajectory view is set to 64 and 32, respectively. For the categorical auxiliary features, including *day*, *hour*, and *visibility*, we embed them to $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^3$ respectively. Additionally, the dimension of the key in the memory network is set to 16 based on the cross-validation results.

C. EXPERIMENT RESULTS

1) OVERALL PERFORMANCE (RQ1)

We conduct experiments for performance comparison of different methods under the task of inferring the pollution map from 10×10 to 20×20. Each experiment is repeated five times independently, and the average results are shown in Table 2 as overall performance. Moreover, we summarize several important observations from the results as below.

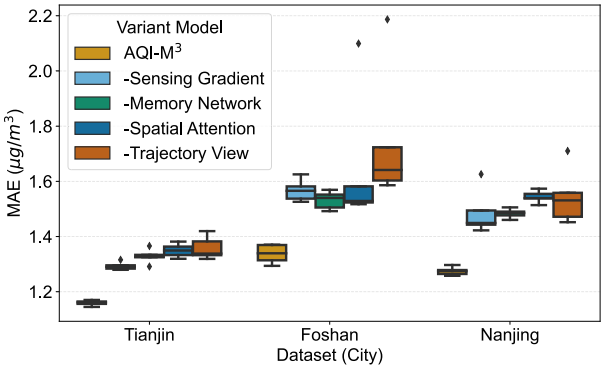
First, AQI-M^3 outperforms competitive baseline methods in three cities, demonstrating the effectiveness of our AQI-M^3 in different geographical and meteorological areas. The results show the generalization capability of our proposed inference method for different deployment scenarios. Second, compared with some deep network methods without spatio-temporal memory, i.e., ConvLSTM, SRResNet, the baselines with spatio-temporal memory (ST-ResNet, STMN, MetaTP) do not yield significant performance improvements. The result indicates that neither conventional temporal nor spatial memory can effectively extract and preserve reusable patterns for air pollution inference. This could be attributed to the fact that pollution patterns are strongly relevant to external factors in addition to inherent spatio-temporal characteristics. Our proposed memory network combines spatio-temporal information from historical mobile sensing and external contextual knowledge, which can provide more accurate and comprehensive information on pollution patterns. Third, the performance improvement of AQI-M^3 under the Foshan dataset (24%∼35%) is more significant than the datasets from other cities (8%∼28%). The possible reason is that, as shown in Table 1, there are fewer mobile sensors deployed in Foshan compared to other cities, and thus the uneven distribution of sensing data is more severe. Under such circumstances, our proposed mobile sensing memory network can provide the pollution patterns in the corresponding context as a global information supplement, bringing a more significant performance improvement.

2) ABLATION STUDY (RQ2)

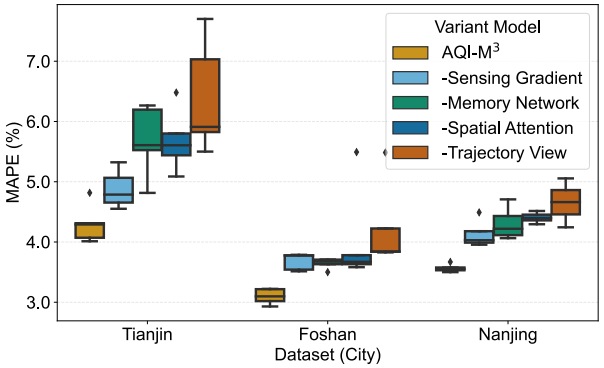
To fully investigate the effectiveness of core components in AQI-M^3, we conduct several ablation studies on its four variants with one of these components removed:

- *Sensing Gradient*: We remove the sensing gradient extraction from the trajectory-view learning, in other words, the gradient $\mathcal{G}$ in equation (7) is removed.
- *Memory Network*: We remove the mobile sensing memory from trajectory view, which means that global spatio-temporal pattern information is eliminated.
- *Spatial Attention*: We remove the spatial attention guide mechanism, and instead directly treat the sensing gradient as part of the encoder input.
- *Trajectory View*: We remove the entire trajectory-view learning, thus only the encoder-decoder network in the region view is retained for inference.

We compare the inference performance of AQI-M^3 and the above variants, and the results are presented as box plots in Fig. 6. We have several findings from these ablation experiments. First, removing trajectory-view learning causes significant performance degradation. This is because the trajectory view introduces fine-grained mobile sensing data that can greatly enhance the ability of air pollution patterns modeling, which demonstrates the effectiveness of multi-view learning. Second, applying the memory network to provide global information for inference is essential. The reason is that the memory network can store different pollution patterns from historical data, thus alleviating the spatial



(a) MAE of AQI-M³ and its Variants.



(b) MAPE of AQI-M³ and its Variants.

FIGURE 6. Performance boxplot of ablation studies on different datasets with 5 repeated trials.

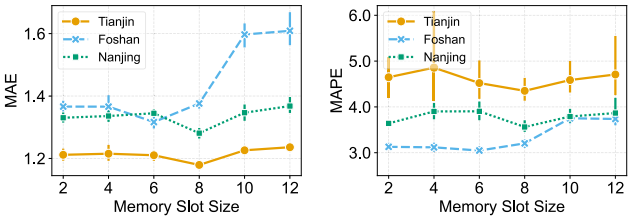
distribution unevenness of mobile sensing data. Third, the sensing gradient extraction under the trajectory view is necessary for improving the inference performance. This is because extracting sensing gradients on trajectories can help to tackle the uncertainty of the original samples while exploiting the fine-grained information. Fourth, the spatial attention guiding is also helpful for air quality inference. It is because the attention weights generated from sensing gradients encourage the model to focus on learning informative regions.

3) PARAMETER SENSITIVITY ANALYSIS (RQ3)

To analyze the parameter sensitivity of the proposed model, we explore the influence of several critical hyper-parameters, including the number of memory slots d, the loss weight α and β in equation (16).

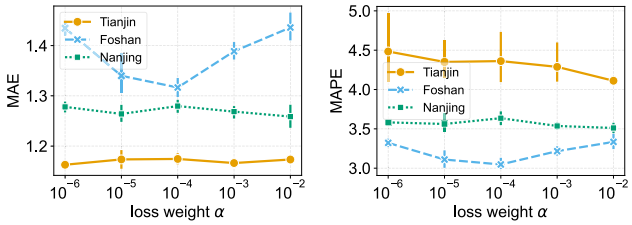
a: NUMBER OF MEMORY SLOTS

Fig. 7 shows the performance variations with different d for three datasets. From the figure, we can see that the number of memory slots d, which determines the memory size and represents the number of global pollution patterns, affects the performance of the model on all datasets. On the one hand, the performance deteriorates when the memory size is too small, because an undersized memory may introduce blended information of different patterns. On the other hand,



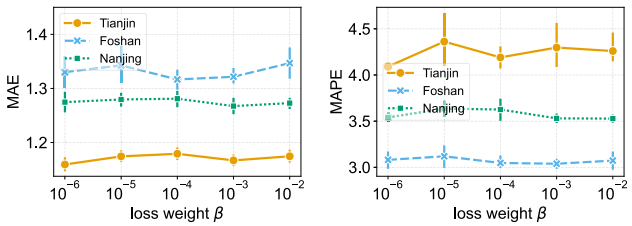
(a) MAE (b) MAPE

FIGURE 7. Performance w.r.t. memory slot size d on different datasets.



(a) MAE (b) MAPE

FIGURE 8. Performance w.r.t. loss weight α on different datasets.



(a) MAE (b) MAPE

FIGURE 9. Performance w.r.t. loss weight β on different datasets.

the model performance also degrades when the memory size is too large. The possible reason is that an excessive memory may try to store much redundant information thereby leading to overfitting. Another finding is that the optimal memory size for Foshan is smaller than that of other cities, which may be because there are fewer pollution patterns in the Foshan dataset due to its less data volume.

b: LOSS WEIGHTS FOR JOINT OPTIMIZATION

Fig. 8 presents the performance change with different α under $\beta = 10^{-4}$. It can be noticed that α can affect the inference performance only in the Foshan dataset. Since the external context changes much more slowly than fine-grained air quality, e.g., the same weather condition may last more than one day, the external factor data is sparse in the Foshan dataset due to its shorter time span. Therefore, a moderate loss weight α may be preferable to mitigate the underfitting and overfitting of the memory network. For the loss weight β, Fig. 9 shows a slight fluctuation of the model performance with different β, which indicates that the model is not sensitive to β in all datasets, as this term in loss function only aims to store historical information into memory.

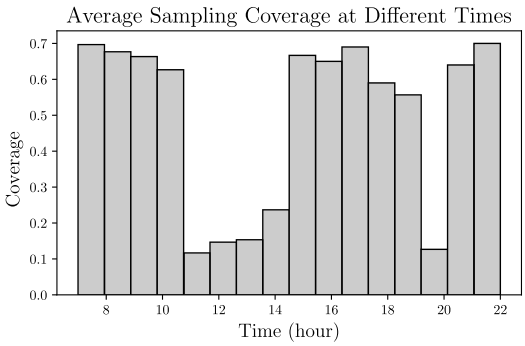
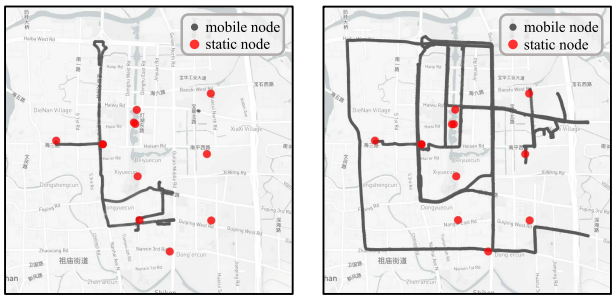


FIGURE 10. Average sampling coverage over the day in the Foshan dataset.



(a) Sparse Coverage (2 pm) (b) Dense Coverage (3 pm)

FIGURE 11. Uneven sampling distribution at different times in the Foshan dataset (Nov.22, 2018).

4) CASE STUDY (RQ4)

To further verify the effectiveness of the memory network under sampling with dynamic coverage, we conducted a case study on the Foshan dataset. Fig. 10 shows the average sampling coverage of the Foshan dataset at different times over the day. It can be noteworthy that the valley of coverage emerges near lunch and dinner time due to the break of staff. Fig. 11 shows the uneven spatial distribution of sampling from mobile sensors under real-world deployment at different times on November 22, 2018, both under sparse coverage (Fig. 11a) and dense coverage (Fig. 11b).

We compare the inference MAE of AQI-M^3 and four baselines for each operation hour of this day, including the interpolation algorithm (*IDW*), the deep-learning model without global memory (*ConvLSTM*), the deep model with global memory (*MetaTP*) and the variant of our proposed model without memory network (*Non-Mem*). First, our proposed AQI-M^3 obtains the best performance and the least variation at all times under the uneven sampling distribution. Moreover, the performance of both the conventional deep model *ConvLSTM* and the model *MetaTP* with spatio-temporal memory degrades severely during the time period with sparse coverage (e.g., 11am to 2pm). However, our proposed model shows a significant performance improvement compared to them under sparse coverage. Meanwhile, the results of *Non-Mem vs. AQI-M^3* further demonstrate the enhancement from the memory mechanism leveraging historical information during sparse sampling. In summary, the case study shows that our AQI-M^3 has better inference accuracy and robustness in sensing systems with dynamic coverage.

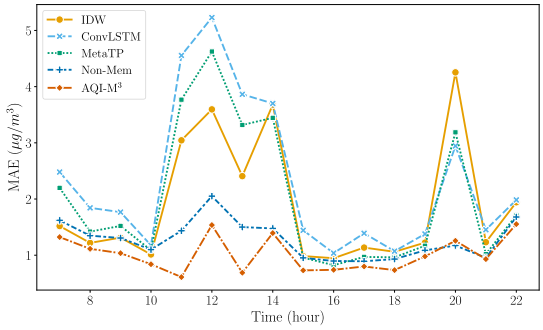


FIGURE 12. Performance comparison of five methods over time in the Foshan dataset (Nov.22, 2018).

V. CONCLUSION

In this paper, we propose a novel framework AQI-M^3 for fine-grained air quality inference, which can enhance the learning capability on mobile sensing data via the multi-view learning and global memory. Mobile sensing brings two critical challenges for inferring air pollution: the uncertainty of collected data and the dynamic coverage of sampling. To address these challenges, we employ sensing gradients extraction under trajectory view to effectively utilize mobile sensing data with uncertainty; we develop a memory network to capture global spatial patterns from historical data and overcome the issue of dynamic sampling coverage. Moreover, spatial attention is applied to further guide learning on highly informative regions, integrating current sensing information and global spatial patterns. To evaluate the performance of our proposed approach, we conduct extensive experiments on real-world sensor systems deployed in three cities, consisting of air sensing units in static and mobile configurations. The evaluation results show that the proposed method achieves about 17%∼29% MAE reduction compared to competitive baselines in different cities. Furthermore, the detailed evaluations also demonstrate the robustness of AQI-M^3 under both uneven and sparse coverage.

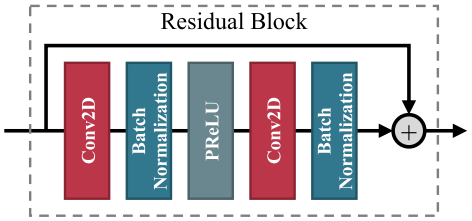
In the future, there are some topics that we can investigate further. Inspired by trajectory learning for human activities [41], [42], we can extend the sensing gradient extraction to higher-dimensional information learning on the fine-grained pollution trajectories. Another aspect is that we can try to introduce advanced graph learning on static and mobile sensors, such as graph neural networks [43], [44], to enable the collaborative inference of nodes with different uncertainty, thus further improving the inference performance.

APPENDIX A. DETAILED LAYOUT OF RESIDUAL BLOCK AND ATTENTION SUBNET

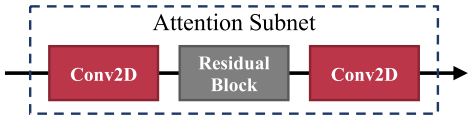
A. DETAILED STRUCTURE OF RESIDUAL BLOCK

We utilize residual blocks with learnable parameters to represent equation (3) for feature extraction in AQI-M^3. To further

illustrate the main component, i.e., residual block, used in the *Encoder* and attention subnet, we present its detailed layout in Fig. 13a. Following the previous work in [15] and [9], we introduce two convolutional layers with Batch Normalization [45] in the block. Moreover, the Parametric Rectified Linear Unit (PReLU) [46] is employed in the intermediate to serve as a nonlinear activation function. We set the kernel size of the residual blocks to 3×3 in *Encoder*. Therefore, a larger receptive field can be obtained in the deeper stacked residual blocks, which can capture the citywide spatial dependencies. In addition, skip connections in these blocks provide features at different spatial scales and avoid gradient vanishing. For fairness, we adopt the same residual block structure in all the convolution-based baseline methods.



(a) Detailed structure of residual block.



(b) Attention subnet in *Gradient-based Attention Branch*.

FIGURE 13. Detailed layout of two main components in AQI-M^3, i.e., residual block and attention subnet.

B. DETAILED STRUCTURE OF ATTENTION SUBNET

The attention subnet is designed to map the sensing gradients into an attention matrix, which is formulated in equation (7). We introduce here the detailed structure of the attention subnet as shown in Fig. 13b, which consists of an input convolutional layer, a residual block mentioned above, and an output convolutional layer sequentially, where all kernel sizes are set to 1×1. The sensing gradients are concatenated together with the max pooling and average pooling results of the pollution map as the subnet input. By this a nonlinear transformation of the subnet, the information from the sensing gradients and the pollution maps are combined in each local sensing region to generate the attention matrix. Meanwhile, the residual block in the subnet can provide information at different levels.

ACKNOWLEDGMENT

The authors would like to thank Shenzhen Environmental Thinking Science and Technology (ETST) Company Ltd. for their assistance in system deployment and data collection.

REFERENCES

[1] W.-J. Guan, X.-Y. Zheng, K. F. Chung, and N.-S. Zhong, "Impact of air pollution on the burden of chronic respiratory diseases in China: Time for urgent action," *Lancet*, vol. 388, no. 10054, pp. 1939–1951, 2016.

[2] S. Rajagopalan and R. D. Brook, "Air pollution and type 2 diabetes: Mechanistic insights," *Diabetes*, vol. 61, no. 12, pp. 3037–3045, 2012.

[3] X. Wu, R. C. Nethery, M. B. Sabath, D. Braun, and F. Dominici, "Air pollution and COVID-19 mortality in the united states: Strengths and limitations of an ecological regression analysis," *Sci. Adv.*, vol. 6, no. 45, Nov. 2020.

[4] S. Dhingra, R. B. Madda, A. H. Gandomi, R. Patan, and M. Daneshmand, "Internet of Things mobile-air pollution monitoring system (IoT-Mobair)," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5577–5584, Jun. 2019.

[5] K. R. Mallires, D. Wang, V. V. Tipparaju, and N. Tao, "Developing a low-cost wearable personal exposure monitor for studying respiratory diseases using metal–oxide sensors," *IEEE Sensors J.*, vol. 19, no. 18, pp. 8252–8261, Sep. 2019.

[6] R. Ma, N. Liu, X. Xu, Y. Wang, H. Y. Noh, P. Zhang, and L. Zhang, "Fine-grained air pollution inference with mobile sensing systems: A weather-related deep autoencoder model," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–21, Jun. 2020.

[7] S. Ali, T. Glass, B. Parr, J. Potgieter, and F. Alam, "Low cost sensor with IoT LoRaWAN connectivity and machine learning-based calibration for air pollution monitoring," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 5500511.

[8] X. Li, M. Sun, Y. Ma, L. Zhang, Y. Zhang, R. Yang, and Q. Liu, "Using sensor network for tracing and locating air pollution sources," *IEEE Sensors J.*, vol. 21, no. 10, pp. 12162–12170, May 2021.

[9] N. Liu, R. Ma, Y. Wang, and L. Zhang, "Inferring fine-grained air pollution map via a spatiotemporal super-resolution scheme," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2019, pp. 498–504.

[10] Y. Yu, V. O. K. Li, and J. C. K. Lam, "Missing air pollution data recovery based on long-short term context encoder," *IEEE Trans. Big Data*, early access, Mar. 9, 2021, doi: 10.1109/TBDATA.2020.2979443.

[11] 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.

[12] P. Ferrer-Cid, J. M. Barcelo-Ordinas, J. Garcia-Vidal, A. Ripoll, and M. Viana, "A comparative study of calibration methods for low-cost ozone sensors in IoT platforms," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9563–9571, Dec. 2019.

[13] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "AirCloud: A cloud-based air-quality monitoring system for everyone," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, New York, NY, USA, 2014, pp. 251–265.

[14] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 1436–1444.

[15] C. Ledig, L. Theis, F. Huszar, J. Caballero, and A. Cunningham, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2017, pp. 4681–4690.

[16] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[17] N. Liu, Y. Wang, J. Huang, R. Ma, and L. Zhang, "Enhanced air quality inference with mobile sensing attention mechanism," in *Proc. 17th Conf. Embedded Networked Sensor Syst.*, 2019, pp. 418–419.

[18] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, and L. Thiele, "Deriving high-resolution urban air pollution maps using mobile sensor nodes," *Pervasive Mobile Comput.*, vol. 16, pp. 268–285, Jan. 2015.

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

[20] T. H. Do, E. Tsiligianni, X. Qin, J. Hofman, V. P. La Manna, W. Philips, and N. Deligiannis, "Graph-deep-learning-based inference of fine-grained air quality from mobile IoT sensors," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8943–8955, Sep. 2020.

[21] W. Cheng, Y. Shen, Y. Zhu, and L. Huang, "A neural attention model for urban air quality inference: Learning the weights of monitoring stations," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2151–2158.

[22] S. Sukhbaatar, A. Szlam, J. Weston, and R. Fergus, "End-to-end memory networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, 2015, pp. 2440–2448.

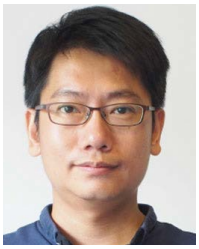
[23] Z. Liu, J. Li, G. Gao, and A. K. Qin, "Temporal memory network towards real-time video understanding," *IEEE Access*, vol. 8, pp. 223837–223847, 2020.

[24] Z. Su, C. Zhu, Y. Dong, D. Cai, Y. Chen, and J. Li, "Learning visual knowledge memory networks for visual question answering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2018, pp. 7736–7745.
[25] H. Yao, Y. Liu, Y. Wei, X. Tang, and Z. Li, "Learning from multiple cities: A meta-learning approach for spatial-temporal prediction," in *Proc. World Wide Web Conf.*, New York, NY, USA, 2019, pp. 2181–2191.
[26] X. Li, Y. Xu, Q. Chen, L. Wang, X. Zhang, and W. Shi, "Short-term forecast of bicycle usage in bike sharing systems: A spatial-temporal memory network," *IEEE Trans. Intell. Transp. Syst.*, early access, Jul. 27, 2021, doi: 10.1109/TITS.2021.3097240.
[27] W. Zhong, Q. Suo, A. Gupta, X. Jia, C. Qiao, and L. Su, "MetaTP: Traffic prediction with unevenly-distributed road sensing data via fast adaptation," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 3, pp. 1–28, Sep. 2021.
[28] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 21st AAAI Conf. Artif. Intell.*, 2017, pp. 1655–1661.
[29] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. 8th Workshop Syntax, Semantics Struct. Stat. Transl.*, 2014, pp. 103–111.
[30] S. Toshniwal, A. Kannan, C.-C. Chiu, Y. Wu, T. N. Sainath, and K. Livescu, "A comparison of techniques for language model integration in encoder-decoder speech recognition," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2018, pp. 369–375.
[31] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in Neural Information Processing Systems*, vol. 29, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2016, pp. 2802–2810.
[32] G. Cheng, A. Matsune, Q. Li, L. Zhu, H. Zang, and S. Zhan, "Encoder-decoder residual network for real super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–10.
[33] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2016, pp. 1874–1883.
[34] Y. Liu, Y. Zhou, and J. Lu, "Exploring the relationship between air pollution and meteorological conditions in China under environmental governance," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 14518.
[35] L. Li, Q. Li, L. Huang, Q. Wang, A. Zhu, J. Xu, Z. Liu, H. Li, L. Shi, R. Li, M. Azari, Y. Wang, X. Zhang, Z. Liu, Y. Zhu, K. Zhang, S. Xue, M. C. G. Ooi, D. Zhang, and A. Chan, "Air quality changes during the COVID-19 lockdown over the Yangtze river delta region: An insight into the impact of human activity pattern changes on air pollution variation," *Sci. Total Environ.*, vol. 732, Aug. 2020, Art. no. 139282.
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[37] *WHO Global Air Quality Guidelines: Particulate Matter (PM2. 5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. World Health Organization, Geneva, Switzerland, 2021.
[38] D. W. Wong, L. Yuan, and S. A. Perlin, "Comparison of spatial interpolation methods for the estimation of air quality data," *J. Exposure Sci. Environ. Epidemiol.*, vol. 14, no. 5, pp. 404–415, Sep. 2004.
[39] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
[40] Y. Liang, K. Ouyang, L. Jing, S. Ruan, Y. Liu, J. Zhang, D. S. Rosenblum, and Y. Zheng, "UrbanFM: Inferring fine-grained urban flows," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 3132–3142.
[41] X. Li, K. Zhao, G. Cong, C. S. Jensen, and W. Wei, "Deep representation learning for trajectory similarity computation," in *Proc. IEEE 34th Int. Conf. Data Eng. (ICDE)*, Apr. 2018, pp. 617–628.
[42] Q. Gao, F. Zhou, K. Zhang, G. Trajcevski, X. Luo, and F. Zhang, "Identifying human mobility via trajectory embeddings," in *Proc. IJCAI*, vol. 17, 2017, pp. 1689–1695.
[43] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.
[44] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–12.
[45] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

NING LIU received the B.Sc. degree in electronic engineering from Tsinghua University, Beijing, China, in 2015, where he is currently pursuing the Ph.D. degree with the Department of Electronic Engineering. His current research interests include urban sensing and computing, especially machine learning in sensing data analytics.

XINYU LIU received the B.Sc. degree in telecommunications engineering from Xidian University, in 2013, and the Ph.D. degree in information and communication engineering from Tsinghua University, in 2020. He is currently a Postdoctoral Researcher with Tsinghua University. His research interests include data calibration and analysis in mobile sensing systems.

PO-TING LIN received the B.Sc. degree in electronic engineering from Chiao Tung University, Hsinchu, Taiwan, in 2018. He is currently pursuing the M.Eng. degree with the Department of Electronic Engineering, Tsinghua University. His research interests include machine learning and spatio-temporal information inference.



YUE WANG received the B.Sc. and Ph.D. degrees in electronic engineering from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 1999 and 2005, respectively. He is currently an Associate Professor with Tsinghua University. His research interests include computer networks, data fusion, and complex networks.

LIN ZHANG received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1998, 2001, and 2006, respectively. He was a Visiting Professor at the University of California at Berkeley, from 2011 to 2013. He is teaching the courses in selected topics in communication networks and information theory to senior undergraduate and graduate students with Tsinghua University, where he is currently a Professor with the Tsinghua Shenzhen International Graduate School. Since 2006, he has been implementing wireless sensor networks in a wide range of application scenarios, including underground mine security, precision agriculture, industrial monitoring, 2008 Beijing Olympic Stadium (the Bird's Nest) Structural Security Surveillance Project, and a metropolitan area sensing and operating network in Shenzhen. His research interests include efficient protocols for sensor networks, statistical learning, data mining algorithms for sensory data processing, and information theory. He was a recipient of the IEEE/ACM SenSys 2010 Best Demo Awards, the IEEE/ACMIPSN 2014 Best Demo Awards, the IEEE CASE 2013 Best Paper Awards, and the Excellent Teacher Awards from Tsinghua University, in 2004 and 2010.