IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 15, 2022, accepted May 3, 2022, date of publication May 23, 2022, date of current version May 31, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174853

Deep-AIR: A Hybrid CNN-LSTM Framework for Fine-Grained Air Pollution Estimation and Forecast in Metropolitan Cities

QI ZHANG, YANG HAN, VICTOR O. K. LI, (Life Fellow, IEEE), AND JACQUELINE C. K. LAM
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Corresponding authors: Victor O. K. Li (vli@eee.hku.hk) and Jacqueline C. K. Lam (jcklam@eee.hku.hk)

This work was supported in part by the Theme-Based Research Scheme of the Research Grants Council of Hong Kong, under Grant T41-709/17-N.

ABSTRACT Air pollution presents a serious health challenge in urban metropolises. While accurately monitoring and forecasting air pollution are highly crucial, existing data-driven models have yet fully captured the complex interactions between the temporal characteristics of air pollution and the spatial characteristics of urban dynamics. Our proposed Deep-AIR fills this gap to provide fine-grained city-wide air pollution estimation and station-wide forecast, by exploiting domain-specific features (including Air Pollution, Weather, Urban Morphology, Transport, and Time-sensitive features), with a hybrid CNN-LSTM structure to capture the spatio-temporal features, and 1×1 convolution layers to enhance the learning of temporal and spatial interaction. Deep-AIR outperforms compatible baselines by a higher accuracy of 1.5%, 2.7%, and 2.3% for Hong Kong and 1.4%, 1.4% and 3.3% for Beijing in fine-grained 1-hr pollution estimation, and 1-hr and 24-hr forecasts, respectively. Saliency analysis reveals that for Hong Kong, spatial features, including street canyon and road density, are the best predictors for NO_2, while temporal features, including historical air pollutants and weather, are the best predictors for $PM_{2.5}$. For Beijing, historical air pollutant data, traffic congestion, wind direction and seasonal indicator are the best predictors for all pollutants. PM_{10} in Hong Kong is achieving the best estimation and forecast accuracy, whilst CO in Beijing is achieving the best results.

INDEX TERMS Fine-grained air pollution estimation and forecast, spatial-temporal data, deep learning, CNN-LSTM, street canyon effect, traffic speed, traffic congestion, domain-specific knowledge, saliency analysis, city-wide, station-wide, Hong Kong, Beijing.

I. INTRODUCTION

Over the past few decades, rapid socio-economic development and urbanization have resulted in severe air pollution in many parts of the world, especially in prominent cities in China and India, such as Beijing and New Delhi. Many adverse health outcomes, such as respiratory and cardiovascular diseases [1], mental health problems [2], or even Covid-19 infection and mortality [3] were associated with the increase in the amount of pollutant exposures. Providing city-wide air pollutant information has significant implications on healthy living and the quality of life of the citizens. On the one hand, accurate air pollutant information can inform citizens especially the vulnerable children and elderly, and the asthmatics, across different parts of a city to improve their health and quality of life by avoiding travelling to highly polluted areas or reducing their personal activities; on the other hand, fine-grained air pollutant estimation and forecast can facilitate evidence-based environmental and public health policy-making, such as setting location-specific traffic control plans across the highly polluted areas. However, air pollution monitoring stations are often geographically sparse across a city (for instance, there are only 18 monitoring stations in Hong Kong, covering an area of 1,106 square kilometers), making it extremely challenging to provide accurate and timely air pollution report covering all of the city.

Over the last few decades, many intra-city-based air pollution models have been proposed [4]. These urban air pollution models can be subdivided into two categories, including

The associate editor coordinating the review of this manuscript and approving it for publication was Zhiwei Gao.

the physical-based and the data-driven approach. Physical-based modelling utilizes numerical models to represent the air pollution process [5], whereas data-driven modelling learns data patterns such as historical air pollution trends using statistical and machine learning methods [6]. More recently, big data and deep learning approach has been put forward, which pushes beyond the boundaries of traditional data-driven models. Generally, such approach has achieved comparably better performance as compared to the traditional physical modelling [7]–[9]. Urban air pollution modelling carries two main objectives. The first objective is to estimate air pollution in areas without monitoring stations (referred to as the fine-grained air pollution estimation) [7]. The second objective is to predict air pollution (also referred to as the air pollution forecast) [8]. Although previous studies have focused on fine-grained air pollution estimation at the city-wide level for the current hour, or air pollution forecast at the station-level for the subsequent hours, few data-driven models have aimed at fulfilling both objectives simultaneously [9], [10].

Up till now, it remains a challenge to obtain fine-grained air pollution estimation at the city-wide level and air pollution forecast at the station-level with high accuracy and low computational complexity simultaneously. As compared with other applications, such as natural language processing or facial recognition, where deep learning models have seen to achieve fundamental breakthroughs, deep learning-based urban air pollution modelling is constrained by sparse, incomplete or missing historical data. There are many missing values in air pollutant datasets in both temporal and spatial dimensions. The resultant lack of training data and noise due to missing data have severely undermined the performance of deep learning-based air pollution models. A large amount of readily available urban proxy data (also referred to as urban dynamics) can be utilized by the deep learning models to address the missing data and the data sparsity challenges. These urban dynamics and their interactions can directly or indirectly influence the spatial-temporal variation of air pollution levels in a city.

A. FACTORS AFFECTING AIR POLLUTION IN AN URBAN ENVIRONMENT

Deep learning studies have identified factors that affect air pollution in an urban setting [8], [9], [11]. These factors can be categorized into those that relate to: (a) the reaction, diffusion, or transport of air pollutants, such as weather (with forecast) [8] and urban morphology, e.g., points of interests (POIs) such as buildings and parks [11]; (b) primary emission sources, such as vehicles and factories [9]; (c) secondary sources due to chemical reactions between multiple pollutants, e.g. NO_2 and O_3 [8]; and (d) fixed effects of the unobserved factors, such as human activities, which contribute to the seasonal variation of air pollution [8].

The air pollutant and the urban dynamics data tend to be closely correlated to each other. The two interact in a complicated manner across different urban environments. For instance, $PM_{2.5}$ concentrations in Beijing, China, are highly affected by meteorological conditions [12]. A study that covered the city of Madrid, Spain, showed that meteorological factors, including wind speed and cloud type, have a strong impact on CO, NO, NO_2, and O_3 concentrations, whilst local traffic conditions have a minimal impact on PM_{10} concentrations [13]. Another study revealed that in Suzhou, China, motor vehicular emission is the most influential factor contributing to NO_2 emissions [14]. In addition to weather and traffic conditions, street canyon effect, a consequence of the complex interactions between air pollutants, weather, traffic, and urban morphology (road networks and building geometries), can often be observed in urban environments. A street canyon has been initially defined as a relatively narrow street with buildings lined up continuously along both sides. The term has also been used to represent bigger streets not necessarily flanked by buildings continuously on both sides [15]. High air pollution levels have been identified within the urban street canyons in Hong Kong [16] and Beijing [17]. Meanwhile, time-sensitive features such as weekends/weekdays, working days/public holidays, peak hours/non-peak hours, and seasonal variation, will also affect air pollution levels [18].

B. MOTIVATION AND RESEARCH SIGNIFICANCE

Air pollution and other urban dynamics data are temporally and spatially correlated. The complex temporal-spatial interactions between air pollutants and urban dynamics must be addressed to capture the variation in pollution levels across different urban environments in a fine-grained scale. Figure 1 shows the temporal and spatial correlation between air pollutants and urban dynamics in two urban environments (between a street canyon and an open area). First, the air pollution concentration in one location can be correlated with the historical air pollution concentration in the same location, depending on local conditions such as traffic and meteorological conditions. For example, the air pollution concentration within a street canyon can be significantly increased during the peak hours due to the complex interactions between traffic emissions and meteorological conditions (e.g., temperature and solar radiation), and photochemical reactions. Second, the air pollution concentration in one location is often dependent on the air pollution concentration in the surrounding areas due to air pollutant transport and diffusion. Although the air pollution of an open area in the absence of vehicular emissions tend to be lower, as compared to a street canyon, it may experience a rapid deterioration due to the transport of air pollutants from the nearby areas, given the right wind speed and direction.

However, few deep learning studies have investigated the complex spatial interactions between air pollutants and urban dynamics in high-density urban settings, where pollutant concentrations can be varied by traffic conditions and street canyons (see Section II for a detailed review of related works). Existing deep learning-based studies have only taken into account the temporal correlations between air pollutants

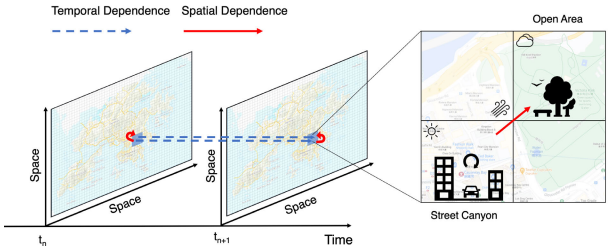


FIGURE 1. An illustration of the temporal and spatial correlation of urban dynamics data.

and urban dynamics via sequential models such as recurrent neural networks (RNNs) and their hybrid long short-term memory (LSTM) networks [19]. More advanced models have addressed the spatial dependence between air pollution and other urban dynamics, utilizing spatial models such as convolutional neural networks (CNNs) [20] and graph convolutional networks (GCNs) [11]. However, these urban air pollution models are yet to incorporate the important features indicative of the street canyon effect in urban areas, such as building density and height, and street canyon. Given that street canyon effect may serve an important estimator or predictor of air pollution in an urban setting, it would be fruitful to explore what advance spatial model structure can better capture the characteristics of spatial urban dynamics data and their interactions with temporal air pollutants data.

This study aims to fill this gap by developing Deep-AIR, a hybrid deep learning framework that enables fine-grained air pollution estimation at the city-level for the current hour and air pollution forecast at the station-level for the subsequent hours. Our proposed framework, taking the city-wide urban dynamics as image-like data, incorporates a CNN component with 1×1 convolution layers to extract the spatial feature representations, and an RNN component, executed via an LSTM model to learn the temporal correlations of the extracted features. The 1×1 convolution layers are employed to strengthen the learning of cross-feature spatial representations between air pollutants and key urban dynamic features, including weather, traffic (traffic speed and congestion), and urban morphology (road density, building density/height, and street canyon). The performance of Deep-AIR will be compared with compatible baseline models developed for urban air pollution estimation and forecast, based on the air pollutant and urban dynamics data collected from Hong Kong and Beijing (see Table 1).

Although Deep-AIR models are based in Hong Kong and Beijing, our deep learning urban air pollution estimation and forecast framework can be transferred to other highly populated and polluted areas/countries whenever urban proxy data are readily available, such as New Delhi, India, or Los Angeles, USA. Our novel methodology which integrates a CNN component with a LSTM component (see Table 1 and Section II-E) to capture the complex interactions of domain-specific spatial-temporal features, can potentially be extended to a wide range of interdisciplinary research topics covering urban computing and social sciences, such as city-wide crowd/traffic flow prediction and fine-grained wealth estimation.

The preliminary results of our former related work was archived [21]. This study extends the preliminary work by (1) providing a comprehensive literature review to include the most recent studies, (2) utilizing additional domain-specific features, including road networks and building geometries to examine the street canyon effect, and time-sensitive features, such as weekends/weekdays, peak hours/non-peak hours, seasonal indicators etc. to account for their spatial or temporal effects on air pollution variation (3) providing a saliency analysis to reveal the most salient domain-specific features in predicting urban air pollution. The rest of this paper is organized as follows. Section II reviews the related works in urban air pollution modelling and highlights the research gaps and the significance of our contribution. Section III illustrates our proposed novel methodology for fine-grained air pollution estimation at the city-level and air pollution forecast at the station-level. Section IV describes the experimental setting and results in details. Section V discusses the implications of the experimental results and charts future research directions. Section VI concludes the study.

II. RELATED WORKS

Our literature review highlights the strengths and limitations of existing air pollution modelling studies and the challenges of urban air pollution modelling using deep learning.

A. PHYSICAL-BASED URBAN AIR POLLUTION MODELLING

Physical-based models have been proposed to simulate the air pollution process in urban areas characterized by complex building geometries and road networks [5]. Utilizing pollutant emissions from different sectors such as industry, household, and transportation, fine-grained air pollution estimation can be achieved by solving computational fluid dynamics (CFD) equations that describe the physical and chemical processes in urban environments [22]. With simplified assumptions on the pollutant distribution, semi-empirical models such as the Gaussian dispersion model can reduce the computational complexity of CFD [23]. Although physical-based models have capitalized on the scientific understanding of the pollution diffusion process, they have drawbacks, including the high computational cost [24] and the inaccuracies and uncertainties in time-dependent inputs such as traffic emission estimates [25]. Such limitations have made it difficult for physical-based models to provide fine-grained air pollution estimation in a large geographical scale such as the entire city in real-time.

Apart from CFD-based models which have examined the dispersion of air pollution, chemical transport models (CTM) have focussed on the source and transport of chemical species, including CMAQ [26], MOZART [27], WRF-chem [28], NAQPMS [29], etc. They have forecasted air

pollution by simulating the physical and chemical processes and providing the pollutant plus its chemical components. However, the forecasting accuracy of these models are highly constrained as compared to the empirical data-driven models [30]–[32]. This may be caused by the uncertainty of emission inventories and the structural uncertainty associated with complex atmospheric chemical and physical processes [33], [34]. Research showed that emission rates estimated by these emissions models may result in ±50% uncertainty in most situations [35]. Besides, most CTM simulations work offline because they consume significant computational resource and time due to high computational complexity [36]. Thus, the applicability of CTM is constrained though they perform better in interpretability as compared to the empirical data-driven counterparts.

B. TRADITIONAL DATA-DRIVEN URBAN AIR POLLUTION MODELLING

Data-driven approaches to urban air pollution modelling, departing from physical-based models, are based on patterns learned from historical data. With some assumptions on the air pollution process, such models have achieved a lower computational cost and a comparable or better performance. Early attempts adopted statistical models for urban air pollution modelling, including inverse distance weighting (IDW), Kriging, and land-use regression for fine-grained air pollution estimation [37], [38], and autoregressive integrated moving average (ARIMA) for air pollution forecast at monitoring stations [39]. More advanced data-driven air pollution modelling studies were undertaken based on machine learning models such as support vector regression (SVR), random forest (RF), and artificial neural network (ANN) [6]. These data-driven models capitalized on the strengths of machine learning in capturing the non-linear relationship between air pollution and urban dynamics data. However, the number of air pollution monitoring stations is often limited, making it challenging to train machine learning models given that ground-truth air pollution measurements are sparse.

Two major approaches have been adopted to tackle the data sparsity issue. On the one hand, to improve the coverage of real-time air pollution measurements, portable sensors can be deployed to different urban environments via participatory sensing [40], vehicular sensing [41], or unmanned aerial vehicle (UAV) sensing [42]. However, a large-scale sensor deployment throughout the city tends to be highly costly and requires significant effort for sensor calibration [43]. On the other hand, some advanced data-driven models have sought to better capture the spatial correlation of air pollution and proxy data for fine-grained air pollution estimation at the city-wide level. These studies often divided a city into disjoint grids (e.g., 1km x 1km) and assumed that air pollutant values in the same grid remain constant. Zheng *et al.* [44] proposed a semi-supervised machine learning method to estimate air pollution in grids not covered by monitoring stations by jointly training a spatial classifier (ANN) utilizing spatial features including POIs and road networks, and a temporal classifier (conditional random field (CRF)) using temporal features including meteorology, traffic, and human mobility. Along this line, Chen *et al.* [45] proposed a semi-supervised ensemble learning model for air pollution estimation at a target grid, highlighting the importance of selecting spatial features from the nearest grids having monitoring data and sharing similar characteristics. Further, Zhu *et al.* [46] proposed a Granger-causality-based data-driven model to estimate air pollution levels in a target grid, based on Granger-causal urban dynamics obtained from the most influential grids (which could be geographically far away). By selecting the most relevant urban dynamics data, the Granger-causality-based model achieved higher accuracy than baseline models using all urban dynamics data from nearby grids. Moreover, several advanced data-driven models have attempted to forecast air pollution at monitoring stations, while utilizing the spatial correlation of air pollution and proxy data. By modelling the spatial dependence and temporal dependence between air pollution and urban proxy data separately, Zheng *et al.* [47] developed a hybrid machine learning framework consisting of a spatial predictor (ANN) and a temporal predictor (CRF) to forecast hourly air pollution levels at monitoring stations in the next two days. Zhao *et al.* [10] proposed a multi-task learning framework to jointly estimate city-wide fine-grained air pollution the current hour and forecast hourly air pollution at monitoring stations the next three hours. Nevertheless, given that traditional machine learning models have not learned complex non-linearities from deep representations of spatial-temporal data [48], it remains difficult for traditional data-driven models to capture urban air pollution accurately.

C. DEEP LEARNING-BASED DATA-DRIVEN URBAN AIR POLLUTION MODELLING

Deep learning or deep neural network models have advanced the state-of-the-art in data-driven urban air pollution modelling. By learning deep representations and complex non-linear relationships from a large amount of heterogeneous spatial-temporal data in urban environments, deep learning models have achieved higher accuracy in air pollution modelling tasks, including fine-grained air pollution estimation at the city-wide level and air pollution forecast at the monitoring station level.

1) FINE-GRAINED AIR POLLUTION ESTIMATION

A number of deep learning models have been proposed for fine-grained air pollution estimation at the city-wide level. Given that air pollution measurements are usually geographically sparse, deep learning models, like other machine learning approaches that are often data-intensive, have faced significant challenges due to the lack of ground truths. Vehicular sensing platforms have been developed to tackle the data sparsity issue. Ma *et al.* [49] proposed an autoencoder framework to recover a real-time high-resolution air pollution map covering a district in a city, based on a ConvLSTM model. Similarly, Do *et al.* [50] proposed an

autoencoder framework to recover real-time high-resolution air pollution in discrete locations in a city, based on a GCN model. However, the need of highly costly auxiliary sensors for experimentation often limits the generalizability of this approach. Yan *et al.* [51] proposed to utilize satellite images for fine-grained air pollution forecast, but the limited time granularity and long time-lag render the utility of open satellite images in real-time applications ineffective. Some deep learning models have been proposed to better capture the spatial-temporal variation of air pollution across the city, utilizing urban proxy data that are readily available. Cheng *et al.* [7] proposed an attention-based hybrid deep learning framework based on the intuition that not all monitoring data contributed equally to air pollution levels at a specific location. The attention model integrated an LSTM model for sequential dynamic data (air pollution and meteorology) modelling and a feedforward neural network for spatial static data (POI and road network) modelling to automatically learn the weights of air pollution monitoring stations for estimation in new locations the current hour. Along this line, Han *et al.* [52] further proposed a multi-channel attention-based GCN model to fuse the static and the dynamic aspect of spatial correlation. Along another line of study, Ma *et al.* [53] utilized a multi-task learning framework based on the observed air pollutant data and fine-grained air pollution estimations generated by a dispersion model, highlighting the use of physical-based models to guide the machine learning process. However, these methods have mainly focused on the temporal features and paid limited attention to the spatial features. The graph data structure that is deployed to model the spatial relationships have only taken into account the distance between different stations, ignoring the directional information and other more complex spatial morphological features, such as building height/density.

2) AIR POLLUTION FORECAST

By exploiting the strengths of deep learning in capturing the non-linear temporal correlation of time series data, many studies have demonstrated better performance of deep learning models for air pollution forecast at monitoring stations. Earlier studies utilized the recurrent neural network (RNN) model and its variants, including the LSTM [19] and gated recurrent unit (GRU) model [54], to capture the temporal dependence of air pollution. More recent studies have extended the RNN-based air pollution modelling using a variety of techniques, such as bi-directional LSTM [55], variational mode decomposition (VMD) to decompose time series data according to the frequency domain [56], filling in missing data via iterative training [57], incorporating weather forecast data [58], focusing on the most relevant information using attention mechanisms [59], accounting for forecast uncertainties using Bayesian methods [60], and using transfer learning to forecast air pollution at a newly built station [61]. Other deep learning models covering denoising autoencoder [62] have also been adopted for air pollution forecasting. However, these extensions have yet addressed the spatial dependence of air pollutant data. By incorporating a CNN or GCN component into the RNN-based model, hybrid deep learning models such as CNN-LSTM [63], [64], GCN-LSTM [65], and GCN-GRU [66] were proposed for air pollution forecast, to take into account the spatial dependence of nearby observations including air pollution and auxiliary data such as meteorology and urban morphology. In addition to the RNN-based modelling, the one-dimensional CNN (1D-CNN) model was used to extract the temporal dependence of urban dynamics observed at a station [67] or nearby stations [68]. A self-attention variational autoencoder was developed to capture the time-series nature of air pollutants through latent space modelling [69]. A deep fusion network consisting of multiple deep feedforward neural networks for air pollution forecast was proposed by capturing the complex interactions between different influential factors such as air pollutants and meteorological conditions [8]. Moreover, a few deep learning models have been proposed to simultaneously provide fine-grained air pollution estimations for the entire city and air pollution forecasts for monitoring stations. Chen *et al.* [9] developed a multi-task CNN-LSTM framework to estimate fine-grained air pollution for the current hour and forecast air pollution at monitoring stations in the next 48 hours through shared spatial-temporal representations across different grids in a city. Similar to the aforementioned, these studies have relied on relatively simple spatial modelling strategies such as graph or one-dimensional CNN, taking into account the nearest monitoring stations and ignoring directional information. In addition, earlier studies have yet to fully exploit important street canyon-related features, such as road density, building density/height, and street canyon. Hence, it remains unclear how these advanced spatial modellings, such as CNNs, can be better utilized to capture the street-level spatial characteristics, and their interactions with air pollutant dispersion processes in an urban environment. Recently, Wang *et al.* [70] explored the application of convolutional LSTM (ConvLSTM), which fully capitalizes on the geographical information to forecast $PM_{2.5}$ concentration in the next 24 hours. However, such a deeply coupled model (by coupling the convolutional layers with the recurrent layers) is complex and computationally expensive. A hybrid CNN-LSTM model that achieves improved performance by capturing both the spatial and temporal dimensions of air pollution characteristics, while exploiting just a few parameters, is desirable for fine-grained air pollution estimation and forecast.

D. DOMAIN-SPECIFIC DEEP LEARNING FOR URBAN AIR POLLUTION MODELLING

Although deep learning models have achieved state-of-the-art performance in air pollution modelling, they tend to suffer from model overfitting due to limited and biased data. Moreover, the interpretability of deep learning models is often low, given their "black box" nature. Existing studies have highlighted the importance of domain-specific modelling to improve the generalizability and interpretability

TABLE 1. A summary of deep learning studies in urban air pollution modelling and our new contributions to this study.

Category	Deep Learning Models	Domain-specific Modelling		Reference
(a) Fine-grained air pollution estimation at the city-wide level				
Autoencoder	ConvLSTM	Features	Air pollutant and meteorology (collected from portable sensors)	[49]
	GCN	Features	Air pollutant (collected from portable sensors)	[50]
Hybrid S-T Model	Att-LSTM-FNN	Features	AQI, meteorology, station location, road network, POI	[7]
	Att-LSTM-GCN	Features	Air pollutant, AQI, meteorology, station location, road network, POI	[52]
		Model Structure	Using attention to learn the air pollutant dispersion across stations	
	ConvLSTM	Features	Air pollutant, meteorology, nighttime light	[70]
Physical-inspired Model	FNN	Features	Air pollutant (observed and simulated data)	[53]
		Model Training	Model training guided by estimations from a dispersion model	
	ConvLSTM	Features	Air pollutant and meteorology (collected from portable sensors)	[49]
		Model Training	Model training guided by the connection between training a convolutional LSTM model and estimating parameters of a dispersion model	
(b) Air pollution forecast at the monitoring station level				
RNN-based Model	RNN (LSTM)	Features	Air pollutant, meteorology, traffic speed	[57]
	RNN (GRU)	Features	Air pollution, meteorology	[54]
	Seq2Seq (GRU)	Features	Air pollutant, meteorology, weather forecast	[58]
	Att-LSTM	Features	Air pollutant, meteorology, weather forecast	[59]
		Model Structure	Using attention to learn the impact of wind speed and wind direction	
	Bayesian GRU	Features	Air pollutant, meteorology, weather forecast	[60]
		Model Training	Model training guided by the high correlation between $PM_{2.5}$ and PM_{10}	
	Transfer Learning (LSTM)	Features	Air pollutant (existing stations and a new station)	[61]
	LSTM-Mode Decomposition	Model Structure	Decompose air pollutant data according to the frequency domain	[56]
Hybrid S-T Model	CNN-LSTM	Features	Air pollutant, meteorology, PBL height, AOD	[20]
	1D-CNN-LSTM	Features	Air pollutant, meteorology	[68]
	GCN-GRU	Features	Air pollutant, meteorology, traffic volume	[66]
	GCN-GRU	Features	AQI, meteorology, geographic features (e.g., building type)	[11]
		Model Structure	Using a diffusion convolution method to extract the spatial relationship based on the important geographic features that can affect air quality	
	GCN-GRU	Features	Air pollutant, meteorology, weather forecast, PBL height	[71]
		Model Structure	Utilizing a knowledge graph of air pollution and meteorology	
Autoencoder	Att-FNN	Features	Air pollutant	[69]
	CNN-DAE	Features	Air pollutant, meteorology	[62]
Ensemble Model	FNN	Features	Air pollutant, AQI, meteorology, weather forecast, time features, station ID	[8]
		Model Structure	Designing a set of FNN models based on the interactions of direct and indirect factors that can affect air quality	
	FNN-Linear Regression	Features	Air pollutant, meteorology, radiosonde, satellite	[51]
(c) Fine-grained air pollution estimation at the city-wide level and air pollution forecast at the monitoring station level				
Hybrid S-T Model	Graph Embedding plus CNN-LSTM	Features	AQI, meteorology, weather forecast, traffic speed, traffic congestion, factory pollutant emission, road network, POI (e.g., factory)	[9]
		Model Structure	Using a graph embedding layer to preserve the spatial relationship in the POI and road network graphs	
	CNN-LSTM with 1x1 convolution layers (proposed model)	Features	Air pollution, meteorology, traffic speed, traffic congestion, road density, **building density/height**, **street canyon**, **background pollution**, time features	N.A.
		Structure	**Utilizing 1x1 convolution layers to capture the cross-feature interactions between air pollutants, meteorology, traffic speed and congestion, urban morphology (road density, building density/height, and street canyon), thereby addressing the street canyon effects**	

Abbreviations
AOD: Aerosol optical depth; **AQI**: Air quality index; **Att**: Attention; **CNN**: Convolutional neural network; **ConvLSTM**: Convolutional long short-term memory; **FNN**: Feedforward neural network; **GCN**: Graph convolutional neural network; **GRU**: Gated recurrent unit; **LSTM**: Long short-term memory; **PBL**: Planetary boundary layer; **POI**: Point of interest; **RNN**: Recurrent neural network; **S-T**: Spatial-temporal.

of deep learning-based air pollution modelling. On the one hand, domain-specific auxiliary features that are highly relevant to the air pollution process have been taken into account, such as meteorology [59], weather forecast [72], POIs such as buildings [11] and factories [73], traffic [57], road networks [7], factory emissions [9], and time features such as month and day of the week [72]. On the other hand, the incorporation of domain-specific knowledge to guide the model learning process has been investigated. Han *et al.* [60] incorporated a specific regularization term into the model

training procedure to penalize PM forecast inconsistent with domain knowledge, particularly, the high correlation between $PM_{2.5}$ and PM_{10} pollutants observed in empirical studies. Ma *et al.* [53] utilized domain knowledge adopted in an air pollution dispersion model to calculate a neural network model's loss function based on the observed air pollutant data and the simulated pollution data generated by the dispersion model. Moreover, Ma *et al.* [49] interpreted the connection between a ConvLSTM model and a simplified dispersion model, demonstrating that the dispersion model's coefficients were automatically learned from data during the model training. Nevertheless, without proper changes in the model structure, domain-specific model training might still fail to address the complex interactions between various factors contributing to air pollution changes over space and time.

Domain knowledge has been exploited in tailor-making deep learning model structures to better capture the air pollution process. Yi *et al.* [72] proposed an ensemble deep learning framework, where each component (deep feedforward network) was designed according to domain knowledge, i.e., the direct and indirect factors that can affect air pollution. Attention-based LSTM models were used to learn the impacts of wind speed and wind direction [59] and industrial processes [73] on $PM_{2.5}$ and the air pollutant dispersion process across the stations [52]. However, the deep feedforward network or LSTM model was incapable of addressing the complicated spatial relationship between those factors. Lin *et al.* [11] constructed a unidirectional graph based on the similarity between the monitoring locations and the nearby influential features such as roads and buildings, using a diffusion convolution method to extract the spatial dependence from the graph-structured data. Chen *et al.* [9] proposed a hybrid CNN-LSTM model for air pollution estimation and forecast, utilizing a graph embedding layer to generate high-level representations of spatial data as inputs to a CNN model, while preserving the spatial relationship among nodes in the POI and road network graphs. Wang *et al.* [71] proposed a knowledge-graph-based hybrid GCN-GRU model for $PM_{2.5}$ pollution forecast, where domain knowledge was explicitly encoded into a bidirectional graph as attributes of nodes (such as wind speed) and edges (such as the impact of wind speed on $PM_{2.5}$ transport from one node to another). However, until now, the street canyon effect has largely been overlooked in existing domain-specific deep learning studies. The spatial interactions between various urban dynamics have yet to be fully addressed by deep learning-based air pollution models to capture the street-level variation of air pollution in urban areas characterized by high-rise buildings and complex traffic conditions.

E. RESEARCH GAPS AND NEW CONTRIBUTIONS

Existing data-driven or deep learning air pollution estimation and prediction studies are yet to explore in details how deep learning models can better capture the characteristics of the complex spatial interactions among air pollutants and urban dynamics data, such as background pollution level and the street canyon effect. Until recently, no rigorous deep learning model has yet been developed to take fully into account the urban street canyon effect for fine-grained air pollution estimation at the city-wide level and air pollution forecast at the station-level (see Table 1 for a summary of the related works and new contributions of our study). Based on our preliminary works in urban air pollution modelling [21], our study fills this gap by proposing a hybrid CNN-LSTM model, Deep-AIR, to capture the spatial-temporal correlations between different air pollutants and other important urban dynamics (e.g., weather, traffic speed, traffic congestion, road density, building density and height, and street canyon), utilizing 1×1 convolution layers that facilitate the spatial information exchange across different urban dynamics. Specifically, our work presents the following novelties:

1) Deep-AIR presents a first hybrid CNN-LSTM deep learning model for both fine-grained air pollution estimation and forecast, with the CNN component being integrated to an LSTM model for learning the spatial variations of air pollutant characteristics; in particular, a 1×1 convolution layer is designed to capture the spatial interactions between air pollutants and important urban dynamic features, including, weather, traffic (traffic speed and congestion), and urban morphology (road density, building density/height, and street canyon).
2) Deep-AIR captures the domain-specific features, such as background pollution and street canyon in the CNN-LSTM model, and achieves the best performance as compared to other baselines (see Table 1).
3) Saliency analysis of input features are conducted to reveal the relative significance and contribution of different domain-specific features in estimating (at the city-wide level) and forecast (at the station-level) urban air pollution levels in Hong Kong and Beijing, which makes Deep-AIR more interpretable, and benefits future model improvements.

III. METHODOLOGY

This study proposes Deep-AIR, a deep learning framework to estimate fine-grained air pollution at the city-wide level for the current hour and forecast air pollution at monitoring stations for up to 24 hours, utilizing readily available urban dynamics data. Hybrid deep learning models that combine CNN and LSTM have been extensively used for spatial-temporal data [48]. Our proposed framework utilizes a spatial model to generate a high-level representation at each time step and learns the temporal correlation of these representations through a temporal model. Compared to a parallel structure where the final outputs of a spatial model and a temporal model have been combined, our current Deep-AIR framework exploits a sequential structure to capture the historical interactions between the spatial-temporal features. Specifically, the framework

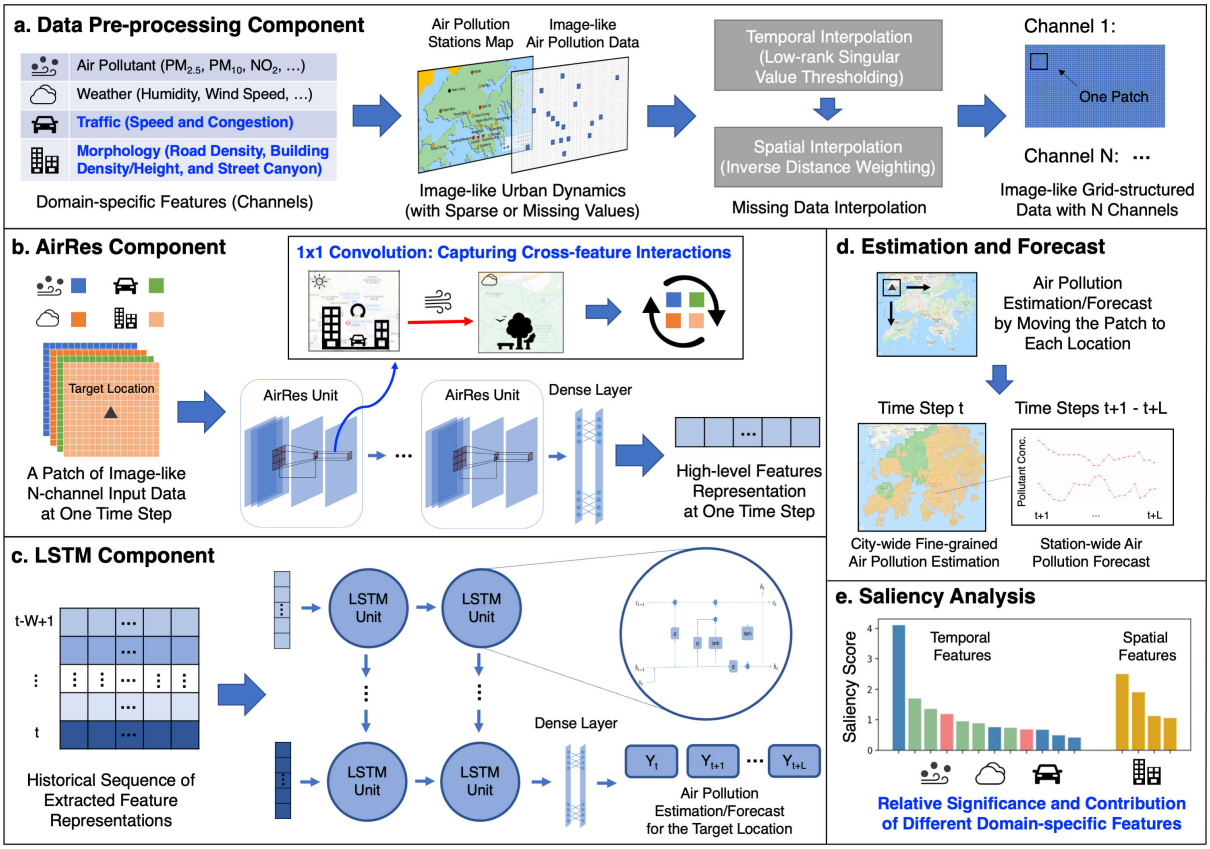


FIGURE 2. Overall structure of Deep-AIR, a hybrid CNN-LSTM deep learning framework.

consists of three sequential components, including (1) a data pre-processing component to generate an image-like grid-structured dataset, with sparse/missing values interpolated temporally and spatially, (2) a residual CNN (AirRes) component for extracting spatial features and their interactions, using 1×1 convolution layers that facilitate the information exchange across different urban dynamic features, and (3) an LSTM component for modelling the temporal dependence of the extracted spatial representations for air pollution estimation and forecast. Figure 2 shows the overall structure of Deep-AIR.

Under our proposed framework, two deep learning models were developed for fine-grained air pollution estimation and air pollution forecast separately, using the same network structure except for the final output layer (see Sections III-B and III-C for more details). The key difference between the two is whether historical ground truth values are available and how they are utilized (see Section III-A for more details). For fine-grained air pollution estimation, historical air pollutant data are only available at air pollution monitoring stations, whilst the estimation model aims to estimate the air pollutant concentrations at all locations, including those without air pollution monitoring stations. In contrast, for air pollution forecast at the monitoring stations, historical air pollutant data are available at each station.

A. DATA COLLECTION AND PRE-PROCESSING COMPONENT

A list of key urban proxy features was collected and processed for model input, with specific data types and definitions being listed in Table 2. Apart from the commonly used air pollutant concentration, meteorological and traffic condition data, our model incorporated street-canyon and background pollutant domain-specific features. To account for the street canyon effect, which is a typical feature of the urban environments [15], urban morphological data linked to the formation of street canyon, including road density, building density, and building height were collected. Besides, a binary indicator was used to indicate the presence/absence of a road, a tall building, and traffic congestion in parallel to indicate whether a street canyon is likely to exist. Our model has also addressed the effect of background pollution on pollutant concentration in the urban environments. Pollutant concentration consists of both background pollution and local pollution [74]. They are both generated from regional sources and local sources, respectively, and exhibit different spatial-temporal patterns. The 10^{th} percentile of the pollutant measurements at each time step was taken as the background pollution [75]. Both street-canyon and background pollutant data were concatenated to the regular air pollutant data to form the input vector, which enhances the capability of our

proposed Deep-AIR to model the complex physical processes for accurate inference.

The air pollution and urban proxy data were pre-processed as follows. First, a city map was divided into thousands of disjoint grids by longitude and latitude. Each grid was associated with air pollutant data and other urban dynamics data, including weather, traffic (traffic speed and congestion), and urban morphology (road density, building density and height, and street canyon), for every time step. As a result, the input data structure for the whole city was like a sequence of n-channel images. Each pixel in the image-like data corresponded to a grid on the map, and each channel corresponded to one kind of air pollutants or other urban dynamics data. Second, a two-stage interpolation method was adopted to fill in missing values in the temporal and spatial dimensions separately to reduce the noise due to missing data, address the sparsity of monitored data, and obtain a fixed input size for model training. In the first stage, an interpolation of historical data in the temporal dimension was conducted to recover missing values due to issues such as equipment failure. We used the Single Value Thresholding (SVT) algorithm proposed in [76], which formulates the missing data recovery problem as a low-rank matrix completion problem. The SVT-based interpolation performed better than other interpolation methods in the temporal dimension because it can capture the low-rank property of the air pollutant data [76]. In the second stage, a spatial interpolation for air pollution and meteorological data was performed to impute each grid's inherently missing values due to the sparsity of air pollutant and meteorological data monitoring stations. The spatial interpolation method was based on the nearest observations, using the inverse distance squared weighting function. The squared weighting function was chosen by trial and error (across the linear, squared, and cubic weighting functions). Finally, after the temporal and spatial interpolations, each input feature (except for categorical features) was normalized by subtracting its mean and dividing by its standard deviation. A complete and normalized grid-structure dataset was generated separately for our proposed fine-grained air pollution estimation model and forecast model. The key difference between the two is in the interpolation of missing data. For fine-grained air pollution estimation, historical air pollutant data are only available at locations with air pollution monitoring stations, whilst the estimation model aims to estimate the air pollutant concentrations at all locations, including those without air pollution monitoring stations. The missing air pollutant values at these locations without historical air pollutant data cannot be interpolated temporally. Instead, they are interpolated spatially, based on observations from the nearest monitoring stations. In contrast, for air pollution forecast at the monitoring stations, historical air pollutant data are available at each station, and may be used to recover missing data at that station via temporal interpolation. In addition, missing data may also be recovered via spatial interpolation of observations from the nearest monitoring stations.

Time-sensitive indicators, including seasonal-specific, calendar-specific, and peak hour indicators, were included as input features. Specifically, a seasonal-specific indicator representing four seasons was included to account for the seasonal effect [18]. A calendar-specific indicator was also included to distinguish the difference in pollutant concentration between the working days and the non-working days (weekends and public holidays) [77]. In addition, a peak hour indicator was also included to reflect the difference in the peak hour and the non-peak hour concentrations due to human activities in both the morning and the evening [78]. The peak hours are defined as the three consecutive peak hours in the morning and in the evening. Experimentally, these calendar indicators facilitated the training speed and improved inference accuracy.

B. DEEP RESIDUAL COMPONENT

After data pre-processing, a sequence of city-wide ''picture'' of urban dynamics was obtained. The picture-like data was fed into a CNN model, capable of extracting spatial information from high-dimensional data. To better capture the complex spatial relationship between various urban dynamics, the structure of the CNN model was modified to improve (1) the capability in learning spatial representations through deeper network structures and (2) the information exchange across different channels (urban dynamics).

1) DEEP RESIDUAL NETWORK

Deep CNN models have achieved outstanding performance in learning high-level representations from spatial data. However, as the neural network layers continue to deepen, it becomes challenging to train the network model due to the gradient exploding and gradient vanishing problems. Sometimes, adding more layers to a network model may even deteriorate the performance [79]. A deep residual network (ResNet) model was proposed to overcome the gradient exploding/vanishing problems when training deeper neural networks [79]. ResNet is a type of CNN that adds an identity mapping on each network block. A ResNet is made up of a series of blocks (residual units), and a residual unit consists of a few convolutional layers and an identity mapping, as shown in Equation 1.

$$X^{(l+1)} = X^{(l)} + \mathcal{F}(X^{(l)}) \tag{1}$$

where

$X^{(l)}$ denotes the input matrix of the l^{th} unit,
$X^{(l+1)}$ denotes the output matrix of the l^{th} unit,
$\mathcal{F}$ represents the identity mapping function applying to the input.

The residual units create a shortcut for the information flow, thus benefiting the training process of very deep networks. The capability of ResNet to capture spatial features through deeper networks has been demonstrated in other urban computing scenarios using spatial-temporal data, such as traffic flow prediction [80].

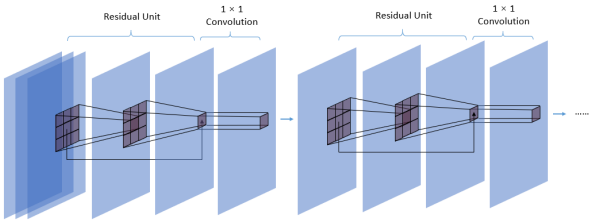


FIGURE 3. The structure of the proposed AirRes model.

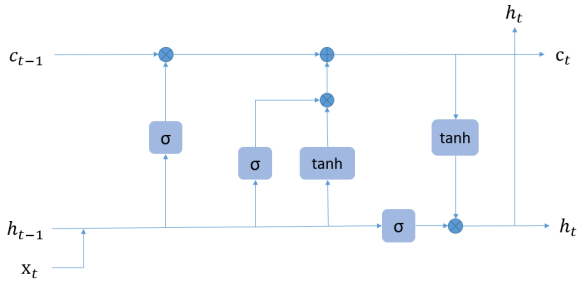


FIGURE 4. The structure of an LSTM block.

A CNN model with a deep structure is needed to extract high-level representations from the spatial correlations between different urban dynamics for urban air pollution modelling. Therefore, a ResNet model was utilized in our framework, but with modifications to better address the spatial interactions between different urban dynamics (see Section III-B2 for more details). The deep residual component was constructed using a series of modified residual units, processing the n-channel grid-structure input data at each time step, and mapping them into a sequence of feature vectors, representing the extracted spatial information from urban dynamics data.

2) 1×1 CONVOLUTION

Although the image-like urban dynamics input data can be readily utilized by the ResNet model adopted in Deep-AIR, the unique characteristics of the air pollution process, particularly the cross-feature spatial interactions across air pollutants and important urban dynamics, are yet to be fully taken into account. As mentioned in Section I and Figure 1, the dispersion of air pollutants in urban environments is strongly dependent on influential factors such as weather (e.g., wind speed and direction can affect the transport of $PM_{2.5}$ pollutants) and street canyons (e.g., the levels of traffic-related pollution tend to be higher in high-density areas than open areas). Therefore, the ResNet model needs to be modified to strengthen the information exchange of different urban dynamic input channels. We developed a tailored ResNet model for air pollution modelling, named AirRes, to address this challenge (see Figure 3). In the modified ResNet model, a 1×1 convolutional layer was inserted between each two adjacent residual units. 1×1 convolution is widely known for reducing the number of channels in GoogLeNet architecture [81]. However, it can also facilitate information interflow across channels [82] because the output of a 1×1 convolutional layer is equivalent to a linear combination of different feature maps.

C. LSTM COMPONENT

An LSTM model is a special kind of RNN model characterized by advanced memory blocks rather than simple neurons at each time step. An LSTM's memory block consists of three gates to control the information flow within the memory block, namely, an input gate, a forget gate, and an output gate. Figure 4 shows the structure of an LSTM block. Due to the carefully designed gates, LSTM networks can avoid the gradient exploding/vanishing problem in RNN while remembering the long-term temporal correlation of sequential features, making it better in modelling time series data.

After extracting high-level spatial features through the deep residual component for each time step, the extracted feature matrix (a sequence of feature vectors for all time steps) was fed into the LSTM component. Whilst certain spatial feature, such as building density, remains unchanged over time, its high-level representation was incorporated into the LSTM component to account for the interactions between the static features and the time-varying features throughout the air pollution estimation and forecast process. For example, despite building density being a static feature, the interactions between building density and traffic congestion can still vary across different hours, especially across the rush and the non-rush hours.

An LSTM's memory block is defined in Equation 2. The final hidden state of the LSTM component was used for air pollution estimation and forecast, using a fully connected layer. For fine-grained air pollution estimation at the city-wide level, one output was generated for an input matrix, representing the air pollution in a grid for the current hour. For air pollution forecast at monitoring stations, multiple outputs were generated for an input matrix, representing the hourly air pollution levels at a monitoring station for the subsequent hours.

$$\begin{aligned}
\mathbf{i}_t &= \sigma(\mathbf{W}_i\mathbf{x}_t + \mathbf{U}_i\mathbf{h}_{t-1}) \\
\mathbf{f}_t &= \sigma(\mathbf{W}_f\mathbf{x}_t + \mathbf{U}_f\mathbf{h}_{t-1}) \\
\mathbf{o}_t &= \sigma(\mathbf{W}_o\mathbf{x}_t + \mathbf{U}_o\mathbf{h}_{t-1}) \\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_t + \mathbf{i}_t \odot \tanh(\mathbf{W}_c\mathbf{x}_t + \mathbf{b}_c) \\
\mathbf{h}_t &= \tanh(\mathbf{c}_t) \odot \mathbf{o}_t
\end{aligned} \tag{2}$$

where

t denotes a time step,
$\mathbf{W}$ and $\mathbf{U}$ are matrices of the network parameters,
$\mathbf{x}_t$ is the extracted feature vector at the time step t,
$\mathbf{h}_t$ represents the hidden state at the time step t,
$\mathbf{i}_t$ represents the input gate at the time step t,
$\mathbf{f}_t$ represents the forget gate at the time step t,
$\mathbf{o}_t$ represents the output gate at the time step t,

$\mathbf{c}_t$ corresponds to the cell unit at the time step t,
$\odot$ denotes elementwise product.

D. MODEL TRAINING, ESTIMATION, AND FORECAST

The details of training the proposed Deep-AIR framework are shown as follows (see Algorithm III-D). After data pre-processing, a patch training algorithm was used to train two models for fine-grained air pollution estimation at the city-wide level and air pollution forecast at the monitoring station-level, separately. During model training, for each pair of the input and output data, the input was a patch of the grid-structured map, with one monitoring station located at the center of the patch, and the output was the air pollutant value measured at the center.

Algorithm 1a Patch Training for City-Wide Fine-Grained Air Pollution Estimation

Require:
the period of historical data T, stations S and grids G, patch size N, observed air pollution measurements $\mathbb{P} = \{p_s^t\}_{t\in T, s\in S}$, historical grid-structured air pollution dynamics $\hat{\mathbb{Q}} = \{\hat{q}_g^t\}_{t\in T, g\in G}$, (where air pollution dynamics at each monitoring station are interpolated by other stations), historical grid-structured urban proxy dynamics $\mathbb{R} = \{r_g^t\}_{t\in T, g\in G}$, the length of historical urban dynamics (as the model input) W, network structure f^1, using randomly initiated network parameters θ, learning rate λ

repeat
  for t from 1 to T **do**
    Sample a station s from 1 to S
    Obtain a set of $N \times N$ grids $\hat{G}$, with $\hat{G}$ centered on s
    for $t' \in [t - W + 1, t]$ **do**
      Crop a patch $\hat{d}_s^{t'} = \{\hat{q}_k^{t'} \oplus r_k^{t'}\}_{k\in\hat{G}}$ from $\hat{\mathbb{Q}}$ and $\mathbb{R}$ ($\oplus$ denotes matrix concatenation)
    end for
    Estimate the current air pollutant for station s: $y_s^t = f_\theta^1(\hat{d}_s^{t-W+1}, \ldots, \hat{d}_s^t)$
    Calculate the loss $\mathcal{L} = \|p_s^t - y_s^t\|^2$
    Perform backpropagation to update θ: $\theta \leftarrow \theta - \lambda\partial\mathcal{L}/\partial\theta$
  end for
until stopping criteria is met

Output: A network model with optimized parameters f_θ^1

The details of estimating fine-grained air pollution and forecasting air pollution using our proposed Deep-AIR framework are as follows (see Algorithm 2). For fine-grained air pollution estimation, a patch of historical urban dynamics data was generated for each grid on the map, including areas not covered by air pollution monitoring stations. For air pollution forecast, a patch of historical urban dynamics data was generated for each air pollution monitoring station on the map. Using the generated patches as the model inputs,

Algorithm 1b Patch Training for Air Pollution Forecast at Monitoring Stations

Require:
the period of historical data T, stations S and grids G, patch size N, observed air pollution measurements $\mathbb{P} = \{p_s^t\}_{t\in T, s\in S}$, historical grid-structured air pollution dynamics $\mathbb{Q} = \{q_g^t\}_{t\in T, g\in G}$, ((where air pollution dynamics at each monitoring station are the observed ground truths)), historical grid-structured urban proxy dynamics $\mathbb{R} = \{r_g^t\}_{t\in T, g\in G}$, the length of historical urban dynamics (as the model input) W, the length of forecast hours L ($L \geq 1$), network structure f^2, using randomly initiated network parameters ϕ, learning rate λ

repeat
  for t from 1 to T **do**
    Sample a station s from 1 to S
    Obtain a set of $N \times N$ grids $\hat{G}$, with $\hat{G}$ centered on s
    for $t' \in [t - W + 1, t]$ **do**
      Crop a patch $d_s^{t'} = \{q_k^{t'} \oplus r_k^{t'}\}_{k\in\hat{G}}$ from $\mathbb{Q}$ and $\mathbb{R}$ ($\oplus$ denotes matrix concatenation)
    end for
    Forecast air pollutant at station s: $[y_s^{t+1}, \ldots, y_s^{t+L}] = f_\phi^2(d_s^{t-W+1}, \ldots, d_s^t)$
    Calculate the loss $\mathcal{L} = \sum_{t+1}^{t+L} \|p_s^t - y_s^t\|^2/L$
    Perform backpropagation to update ϕ: $\phi \leftarrow \phi - \lambda\partial\mathcal{L}/\partial\phi$
  end for
until stopping criteria is met

Output: A network model with optimized parameters f_ϕ^2

the fine-grained estimation model predicted a fine-grained air pollution estimation map for the entire city the current hour, and the air pollution forecast model predicted air pollution for each air pollution monitoring station the subsequent hours.

E. SALIENCY ANALYSIS

In addition to model training and forecast, a post-hoc explainable module to the proposed deep learning framework was added. Specifically, a saliency analysis to better understand the influential features affecting air pollution forecast was conducted. As illustrated by the previous work [83], the gradients with respect to the input values can reflect how much each input feature contributes to the output value. The output (i.e., air pollution forecast) near a single point can be approximately expressed by Equation 3a. The magnitude of each dimension of the gradient indicates the sensitivity of the output values to the particular input feature. By taking the average of the absolute gradient for each model input over the whole training set, the saliency score is defined by Equation 3b.

$$y \approx w(\mathbf{x})^T\mathbf{x} + b \tag{3a}$$

$$\mathbf{s} = \sum_{\mathbf{x}, y\in D} \frac{|w(\mathbf{x})|}{|D|} \tag{3b}$$

Algorithm 2 Fine-Grained Air Pollution Estimation and Air Pollution Forecast

Require:
the current time $t^\star$, fitted network models f_θ^1 and f_ϕ^2, stations S and grids G, patch size N, the period of historical data T, historical grid-structured air pollution dynamics $\mathbb{Q}$, (where air pollution dynamics at each monitoring station are the observed ground truths), historical grid-structured air pollution dynamics $\hat{\mathbb{Q}}$ (where air pollution dynamics at each monitoring station are interpolated by other stations), historical grid-structured urban proxy dynamics $\mathbb{R}$, the length of historical urban dynamics (as the model input) W, the length of forecast hours L ($L \geq 1$)

for g from 1 to G **do**
  Obtain a set of $N \times N$ grids $\hat{G}$, with $\hat{G}$ centered on g
  for $t' \in [t^\star - W + 1, t^\star]$ **do**
    Crop a patch $\hat{d}_g^{t'} = \{\hat{q}_k^{t'} \oplus r_k^{t'}\}_{k \in \hat{G}}$ from $\hat{\mathbb{Q}}$ and $\mathbb{R}$
    if $g \in S$ **then**
      Crop a patch $d_g^{t'} = \{q_k^{t'} \oplus r_k^{t'}\}_{k \in \hat{G}}$ from $\mathbb{Q}$ and $\mathbb{R}$ ($\oplus$ denotes matrix concatenation)
    end if
  end for
  Estimate the current air pollutant for grid g: $y_g^{t^\star} = f_\theta^1(\hat{d}_g^{t^\star - W + 1}, \ldots, \hat{d}_g^{t^\star})$
  if $g \in S$ **then**
    Forecast air pollutant for grid g: $[y_g^{t^\star+1}, \ldots, y_g^{t^\star+L}] = f_\phi^2(d_g^{t^\star - W + 1}, \ldots, d_g^{t^\star})$
  end if
end for

Output: a fine-grained air pollutant map at $t^\star$, air pollution forecasts of the air pollution monitoring stations from $t^\star+1$ to $t^\star + L$

where

$\mathbf{x}$ is the input data consisting of different features (urban dynamics),

y is the forecasted air pollution value,

w and b are, respectively, the weight parameters and the bias parameter to approximate the relationship between $\mathbf{x}$ and y (w is the derivative of y with respect to $\mathbf{x}$),

D is the whole training set,

$\mathbf{s}$ is the saliency score vector that corresponds to the input features

IV. EXPERIMENTAL SETTING AND RESULTS

A. EXPERIMENTAL SETTING

We collected two datasets, namely, Hong Kong for 16 months (Dec 2018-Mar 2020) and Beijing for 19 months (Jan 2017-Jul 2018). Table 2 details the data types and how the data are derived or labelled. The data source and the number of data points covering the data collected for the two cities have been listed in Table 2. After data collection, for the data updated more frequently than once per hour, they were averaged for each hour so that the frequency of every kind of data was aligned to one hour.

We pre-processed the Hong Kong and Beijing datasets as follows. First, we constructed grid-structured datasets. We divided Hong Kong into 1km×1km grids, so the grid structure was a 26-channel 44×60 map for each time step. Similarly, we divided Beijing into 3km×3km grids, so the grid structure was a 24-channel 50×55 map for each time step. Second, we used a random 80/10/10 split of the grid-structured dataset for each city as the training set, the validation set, and the test set. Finally, we interpolated the missing values in the grid-structured input data, and normalized each numerical feature using its mean and standard deviation derived from the training set (see Section III-A for more details about the interpolation procedure).

After data pre-processing, we conducted an experiment to train and evaluate our proposed model for fine-grained estimation (see Algorithm III-D). We also visualized the fine-grained $PM_{2.5}$ pollution estimates in Hong Kong and Beijing across four seasons to better understand the geographical and seasonal variation of the estimated levels of air pollution. Moreover, we conducted another experiment to train and evaluate our proposed model for forecasting 1-hr and 24-hr air pollution at monitoring stations (see Algorithm III-D). In addition to model training and evaluation, we conducted a saliency score analysis to better understand the key features that predict air pollution levels in Hong Kong and Beijing.

We trained our proposed models, namely, the fine-grained air pollution estimation model and the air pollution forecast model, using the following settings. The patch size was set to 15. The AirRes component consisted of four residual units, and each residual unit had two 3 × 3 convolution layers with batch normalization and ReLU activation function. A 1 × 1 convolution layer was added between each two of the residual units. For the LSTM component, the length of the model input (past hourly observations) was set to 48 (hours), the number of LSTM layers was set to one or two, and the hidden unit size was set to 128, 256, or 512. A stochastic gradient descent optimizer was used to train the model, and the learning rate of the optimizer was set to 10^{-4}. The training process was stopped when the validation error was not improved in the latest five epochs. The best hyper-parameters, including the number of LSTM layers and the hidden unit size, were selected based on the performance evaluated on the validation set.

Five baseline models were included for model evaluation and comparison. First, an autoregressive integrated moving average (ARIMA) model, a statistical method for time series analysis, was selected. Second, a standard LSTM model, a widely used method for deep learning-based air pollution modelling, was developed. Third, a standard ConvLSTM model, a typical structure for spatial-temporal data processing that integrates convolutional structures into

TABLE 2. Urban big data collected in Hong Kong from December 2018 to March 2020 and Beijing from January 2017 to July 2018.

(a) Features

Feature Category	Feature Type	Feature Name	Unit	Definition
Air Pollutant (A)	S-T	$PM_{2.5}$	$\mu g/m^3$	$PM_{2.5}$ concentration
	S-T	PM_{10}	$\mu g/m^3$	PM_{10} concentration
	S-T	NO_2	$\mu g/m^3$	NO_2 concentration
	S-T	SO_2	$\mu g/m^3$	SO_2 concentration
	S-T	O_3	$\mu g/m^3$	O_3 concentration
	S-T	CO	$\mu g/m^3$	CO concentration
	T	Background concentration	$\mu g/m^3$	The 10^{th} percentile of all measurements of one pollutant type at one timestep
Weather (W)	S-T	Pressure	hPa	Atmospheric pressure
	S-T	Humidity	%	Relative humidity
	S-T	Temperature	°C	Ambient temperature
	S-T	Wind speed	km/h	Wind speed
	S-T	Wind direction	N/A	Wind direction (eight possible values: E, W, N, S, NE, NW, SE, SW)
	S-T	Precipitation	mm	Precipitation
	S-T	Cloudiness	%	Cloud cover percentage
Traffic (T)	S-T	Traffic congestion	N/A	The average congestion level of the road segmentation within a grid
	S-T	Traffic speed	km/h	The average speed of the vehicles on the road segmentation within a grid
Morphology (M)	S	Road density	N/A	Number of road segments within a grid area
	S	Building density	N/A	Number of buildings within a grid area
	S	Building height	m	The average of the building heights within a grid area
	S	Street canyon effect	N/A	Binary indicator (1: there is a street canyon effect in this grid; 0: there is no street canyon effect in this grid)
Time-sensitive (S)	T	Seasonal indicator	N/A	Quaternary indicator (-1.5, -0.5, 0.5, 1.5 representing spring, summer, autumn and winter respectively)
	T	Calendar indicator	N/A	Binary indicator (1 representing a weekend or a public holiday; 0 representing a working day)
	T	Peak hour indicator	N/A	Binary indicator (1 representing a peak hour; 0 representing a non-peak hour)

Notes
1. S-T refers to Spatial-temporal feature; S refers to Spatial feature; T refers to Temporal feature
2. CO data in Hong Kong are not included because only some stations have reported CO concentrations.
3. Morphological data in Beijing are not available.

(b) Data Availability

Feature Category	Number of Data Points	Update Frequency	Open Data	Data Source
Hong Kong				
Air Pollution (A)	16 stations	60 minutes	Y	HKEPD [84]
Weather (W)	44 stations	10 minutes	Y	HKO [85], WWO [86]
Traffic (T)	618 roads	5 minutes	Y	HKTD [87]
Morphology (M)	44 * 60 grids	N/A	N	HKLD [88]
Time-sensitive Indicator (S)	44 * 60 grids	60 minutes	Y	N/A
Beijing				
Air Pollution (A)	35 stations	60 minutes	Y	BJMEMC [89]
Weather (W)	18 stations	60 minutes	Y	CMDSC [90], WWO [86]
Traffic (T)	227 roads	60 minutes	N	Gaode [91]
Time-sensitive Indicator (S)	50 * 55 grids	60 minutes	Y	N/A

Notes
1. Air pollution data from two newly built air pollution monitoring stations in Hong Kong are not included.
2. The official air pollution and meteorological data in Beijing were collected in real-time. A third-party data archive is available to download the historical data [92].

an LSTM model, was constructed. Fourth, an attention-based LSTM model for air pollution was implemented based on [59]. Fifth, a hybrid model consisting of a GCN and an LSTM structure (GCN-LSTM) was selected [66]. The fourth and fifth baseline models are considered as state-of-the-art models for air pollution forecast. The mean absolute percentage error (MAPE) was used as the performance metric for fine-grained estimation (see Equation 4a) and forecast (see Equation 4b).

$$\text{MAPE}_{\text{estimation}} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i^c - y_i^{\star c}|}{y_i^c} \times 100\% \tag{4a}$$

$$\text{MAPE}_{\text{forecast}} = \frac{1}{nL}\sum_{i=1}^{n}\sum_{t=1}^{t=L}\frac{|y_i^t - y_i^{\star t}|}{y_i^t} \times 100\% \tag{4b}$$

where

n is the sample size,
i identifies one test sample,
L is the length of the forecasting period (ranging from 1 to 24),
$y_i^{\star c}$ is the ground truth air pollution value at the current time,
y_i^c is the estimated air pollution value at the current time,
$y_i^{\star t}$ is the ground truth air pollution value at time step t,
y_i^t is the forecasted air pollution value at time step t.

B. RESULTS

The proposed models and the baseline models were tested based on the Hong Kong and Beijing datasets. Results show that our proposed Deep-AIR framework has achieved the best performance in fine-grained air pollution estimation and air pollution forecast, while providing interpretations on which features are most important in predicting urban air pollution. We present the detailed results in the remaining parts of this section. We will discuss why our proposed model works better than the baseline models and which parts can be improved in Section V.

Our proposed model has achieved better performance than compatible baseline models when performing city-wide fine-grained air pollution estimations for the current hour. Table 3 shows the average error rate of different models for fine-grained air pollution estimation at the station-level when the local air pollution information was removed. The results show that our proposed model has achieved the lowest error as compared to the baseline models. On average, the error rate of our proposed fine-grained estimation model is 31.1% and 34.2% in Hong Kong and Beijing, respectively.

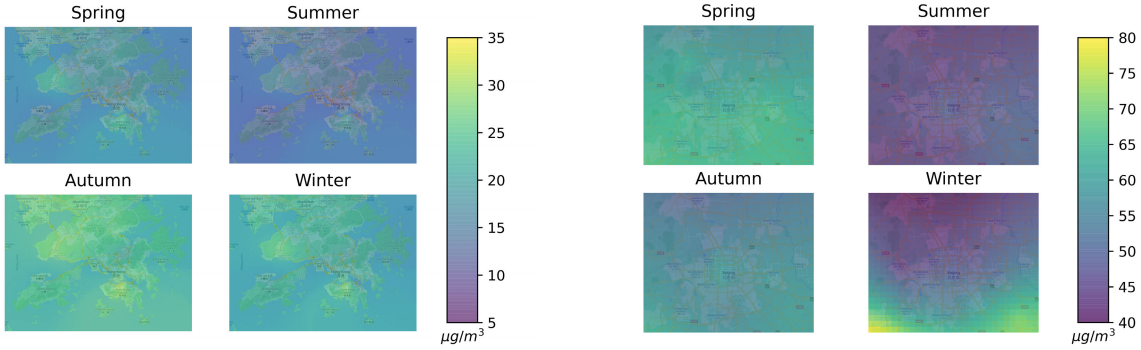
Our proposed fine-grained estimation model has made possible the evaluation of air pollution levels at a fine-grained scale throughout the city. The city-wide fine-grained estimation results are used to visualize the seasonal and geographical patterns of $PM_{2.5}$ pollution in Hong Kong and Beijing for one year. Figure 5 shows the average $PM_{2.5}$ estimated values in Hong Kong (March 2019 to February 2020) and Beijing (March 2017 to February 2018) across four seasons, including spring (from March to May), summer (from June to August), autumn (from September to November), and winter (from December to February). The average $PM_{2.5}$ level was higher in Beijing than in Hong Kong during the study period. The two air pollution visualization maps demonstrate seasonal variations in $PM_{2.5}$ estimated values. For example, the $PM_{2.5}$ estimates were slightly higher in winter and spring but lower in summer, especially in Beijing, which can be attributable to winter heating. Moreover, the two maps illustrate the geographical variations of $PM_{2.5}$ pollution estimates. For example, the $PM_{2.5}$ estimates were higher in the southern part of Beijing than in other parts of the city during the winter, which can be attributable to the regional transport of air pollutants across the Beijing-Tianjin-Hebei region. Nevertheless, the geographical variation of air pollution levels was small, especially in Hong Kong, probably when air pollution episodes became insignificant after averaging the hourly $PM_{2.5}$ estimates of individual seasons. Based on the hourly fine-grained air pollution information, pollution episodes and hotspots can be further identified for future analyses.

Moreover, our proposed model has outperformed the baselines for forecasting air pollution at monitoring stations in the next hour and over the next 24 hours. Tables 4 and 5 show the average forecast error rate of our proposed model and the baselines. Different rows in these two tables correspond to different experiments based on a different set of urban features. Time-sensitive features (S) are included for all Deep-AIR experiments. Although the forecast air pollution levels for the next 24 hours is less accurate for all models, our proposed model achieves the lowest error compared to the baseline models, utilizing all available urban dynamics data. In Hong Kong, the lowest forecast error is 21.1% in the next hour and 32.3% over the next 24 hours. In Beijing, the lowest forecast error is 23.9% in the next hour and 35.2% over the next 24 hours. Moreover, when evaluating different sets of input features, all models have achieved better results when more types of features are inputted and poorer results when less types of features are inputted. When two or more types of urban dynamics are included, our proposed model has outperformed the baseline models. Table 6 details the size of parameters across different models. As shown in the Table, our Deep-AIR model carries a larger parameter size than that of Attention-LSTM and GCN-LSTM, though it has a smaller parameter size than that of ConvLSTM. Given that our model achieves improvement in performance, a higher complexity is considered acceptable.

Although our proposed model has achieved the best performance for forecasting 1-hr air pollution, the forecast MAPE values has varied among our observed pollutants. Figure 6 shows the correlation between the 1-hr forecast values and the ground truth values of $PM_{2.5}$ in Hong Kong and Beijing. As shown on the two scatter plots, the points are distributed closely around the identity line. The R-squared value between the forecasted values and the ground truth values is 94% in Hong Kong and 90% in Beijing, suggesting that the 1-hr $PM_{2.5}$ forecast values are highly consistent with the ground truths. Table 7 shows the MAPE of 1-hr to 24-hr air pollution forecast values by Deep-AIR for each pollutant type in Hong Kong and Beijing. In addition to $PM_{2.5}$ values, Table 7 shows the MAPE of our proposed Deep-AIR when forecasting the 1-hr air pollution level of each pollutant type. Results show that our proposed model performs best when forecasting $PM_{2.5}$ and PM_{10} values in Hong Kong and CO values in Beijing. The 1-hr forecast MAPE for $PM_{2.5}$ and PM_{10} is 14.4% and 13.8% for Hong Kong, respectively whilst the 1-hr forecast MAPE for CO values for 17.9% in Beijing. However, for both cities, Deep-AIR may not be the best to predict O_3 in reasonable accuracy, likely due to the

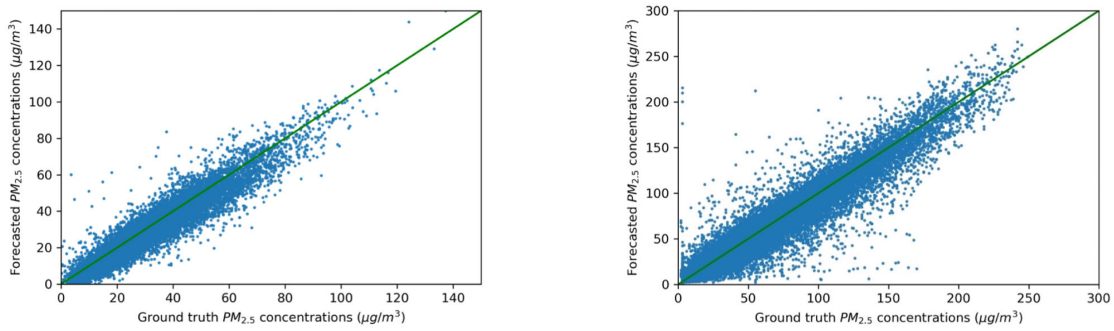
TABLE 3. MAPE of the fine-grained air pollution estimation the current hour for Deep-AIR and other compatible models (After removing the local air pollution information).

Model	ARIMA	LSTM	ConvLSTM	Attention-LSTM	GCN-LSTM	Deep-AIR
MAPE (Hong Kong)	35.0	33.0	32.6	32.8	32.6	**31.1**
MAPE (Beijing)	38.7	36.2	35.9	36.0	35.6	**34.2**



(a) Hong Kong (March 2019 to February 2020) (b) Beijing (March 2017 to February 2018)

FIGURE 5. Average $PM_{2.5}$ concentrations across four seasons in Hong Kong and Beijing.



(a) Hong Kong) (b) Beijing

FIGURE 6. 1-hr $PM_{2.5}$ forecast values by Deep-AIR vs. the ground-truth $PM_{2.5}$ values in Hong Kong and Beijing.

TABLE 4. MAPE of the 1-hr and the 24-hr air pollution forecasts for Deep-AIR and other compatible models in Hong Kong.

Feature(s)	MAPE											
	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr
	ARIMA		LSTM		ConvLSTM		Attention-LSTM		GCN-LSTM		Deep-AIR	
S+A	29.5	45.5	27.8	41.0	27.9	41.3	27.6	40.5	27.8	41.2	27.0	39.9
S+A+W	27.3	43.7	25.9	39.0	26.3	39.2	25.6	38.5	26.0	38.9	24.5	36.5
S+A+W+T	26.0	42.3	24.8	37.9	24.6	37.7	24.4	37.4	25.1	37.8	22.5	33.6
S+A+W+T+M	25.7	42.0	24.2	37.1	24.0	37.0	23.8	36.1	24.3	37.0	**21.1**	**32.3**

Note: A, W, M, T, and S refer to Air pollution, Weather, Urban Morphology, Traffic, and Time-sensitive data collected from Hong Kong for Deep-AIR analysis. Their corresponding definitions can be found in Table 2

irregularity in O_3 patterns as compared to that of the other pollutants. Given such irregularity in O_3 patterns as time evolves, it might be less desirable to learn the O_3 patterns entirely on historical data. As expected, the forecast MAPE increases across a longer time frame. The average forecast MAPE of the first 6 hours lie within 26% for Hong Kong and lie within 29% in Beijing. However, for Hong Kong, the respective forecast values of MAPE of the first 6 hours for $PM_{2.5}$ and PM_{10} are 18.7% and 18.2% only. Table 8 shows the MAPE of our proposed Deep-AIR for different pollutants in different seasons. In Beijing, the model performs better in Summer and Autumn, and relatively worse in

TABLE 5. MAPE of the 1-hr and the 24-hr air pollution forecasts for Deep-AIR and other compatible models in Beijing.

Feature(s)	MAPE											
	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr	1-hr	24-hr
	ARIMA		LSTM		ConvLSTM		Attention-LSTM		GCN-LSTM		Deep-AIR	
S+A	32.6	54.2	30.8	44.4	30.4	43.6	28.5	42.8	30.6	44.1	27.8	41.5
S+A+W	31.4	52.8	28.8	41.9	28.3	40.8	27.2	40.1	28.8	41.5	25.0	38.1
S+A+W+T	29.3	52.0	27.6	39.8	26.9	38.9	26.2	38.5	27.7	39.5	**23.9**	**35.2**

Note: A, W, T and S refer to Air pollution, Weather, Traffic, and Time-sensitive data collected from Beijing for Deep-AIR analysis. Their corresponding definitions can be found in Table 2

TABLE 6. Size of parameters of Deep-AIR and other compatible models for the fine-grained air pollution estimation.

Model	LSTM	ConvLSTM	Attention-LSTM	GCN-LSTM	Deep-AIR
Parameter size (M)	1058	7208	1511	1189	3503

TABLE 7. MAPE of 1-hr to 24-hr air pollution forecast values by Deep-AIR for each air pollutant type in Hong Kong and Beijing.

(a) Hong Kong

Pollutant	$PM_{2.5}$	PM_{10}	NO_2	SO_2	O_3	Average
1 hour	14.4	**13.8**	24.0	19.4	34.1	**21.1**
3 hours	16.9	**16.0**	26.3	22.2	37.0	**23.6**
6 hours	18.7	**18.2**	29.0	24.3	39.3	**25.9**
12 hours	21.2	**20.9**	31.5	26.9	42.1	**28.5**
24 hours	25.2	**24.6**	35.4	30.6	45.7	**32.3**

(b) Beijing

Pollutant	$PM_{2.5}$	PM_{10}	NO_2	SO_2	O_3	CO	Average
1 hour	23.7	24.3	23.9	22.2	31.5	**17.9**	**23.9**
3 hours	26.5	27.2	27.0	26.7	34.7	**20.5**	**27.1**
6 hours	28.4	28.8	28.7	26.4	36.6	**22.6**	**28.6**
12 hours	31.4	32.1	31.2	30.4	39.0	**25.4**	**31.6**
24 hours	35.2	35.9	35.4	33.0	42.8	**29.2**	**35.2**

TABLE 8. MAPE of 1-hr forecast values by Deep-AIR for each air pollutant type in each season.

(a) Hong Kong

Pollutant	$PM_{2.5}$	PM_{10}	NO_2	SO_2	O_3	Average
Spring	14.2	16.1	23.9	21.9	34.4	22.1
Summer	18.5	14.9	27.0	19.1	35.6	23.0
Autumn	12.1	10.5	25.4	14.5	31.3	18.8
Winter	13.6	14.0	21.5	21.0	34.7	21.0

(b) Beijing

Pollutant	$PM_{2.5}$	PM_{10}	NO_2	SO_2	O_3	CO	Average
Spring	21.9	26.0	22.3	25.3	31.4	17.8	24.1
Summer	23.6	24.6	21.1	14.5	29.2	19.9	22.2
Autumn	21.5	23.5	21.2	21.9	32.4	16.0	22.8
Winter	26.1	22.5	27.3	25.7	32.1	17.1	25.1

Spring and Winter. In Hong Kong, the model performs better in Autumn, and has comparative performance in the other seasons.

Table 9 shows the ablation study for the added new features based on domain-specific knowledge and the newly proposed model structure. It shows that domain-specific features that accounted for the street canyon effects has improved Deep-AIR's performance in both fine-grained pollution estimation and forecast. The revised ResNet in our model could make use of the abundant underlying geographical information and model the related physical process. Although the background pollutant feature has not improved the accuracy of 1-hour forecast substantially, it can benefit long-term forecasting and fine-grained estimation. The ablation analysis has further confirmed that the deployment of 1×1 convolution layers can better capture the spatial interactions between different urban dynamics for both short-term and long-term air pollution forecast.

Furthermore, the saliency analysis has shed more insights on which temporal and spatial features are important predictors for urban air pollution in Hong Kong and China. Figure 7 highlights the saliency scores of the temporal and spatial features that best predict air pollution in Hong Kong. For Hong Kong, spatial urban morphological features, such as street canyon effect and road density, are relatively more salient in predicting NO_2, as compared to other temporal features such as wind speed. Similarly, spatial urban morphological features such as building density and height, and street canyon effect are relatively more salient in predicting SO_2 as compared to other temporal features such as traffic speed. In contrast, temporal features, such as historical air pollutant features, NO_2, O_3, and PM_{10}, and historical meteorological features, such as temperature, pressure, and humidity, are relatively more salient in predicting $PM_{2.5}$ and PM_{10}, as compared to spatial urban morphological features, such as road density. Meanwhile, temporal features such as humidity, temperature, wind speed, and wind direction are relatively more salient in predicting O_3 as compared to spatial urban morphological features such as building density and height, and road density. In the absence of key spatial urban morphological features from the Beijing dataset, temporal features such as historical air pollutants, traffic congestion, wind direction, and seasonal indicator are the most salient temporal predictors of air pollutants level in Beijing (see Figure 8).

V. DISCUSSION AND FUTURE WORKS

This study aims to estimate fine-grained air pollution at the city-wide level for the current hour, while forecasting air pollution at monitoring stations in the short term

TABLE 9. Ablation study for domain-specific features and model structures.

	Model	1-hr forecast	24-hr forecast	Fine-grained estimation
MAPE in Hong Kong	w/o the street canyon features	22.5	33.6	32.2
	w/o the background pollutant feature	21.3	33.2	32.5
	w/o the 1x1 conv layers	22.6	34.2	31.3
	Deep-AIR model only	**21.1**	**32.3**	**31.1**
MAPE in Beijing	w/o the background pollutant feature	24.3	36.8	35.0
	w/o the 1x1 conv. layers	25.5	37.4	34.5
	Deep-AIR model only	**23.9**	**35.2**	**34.2**

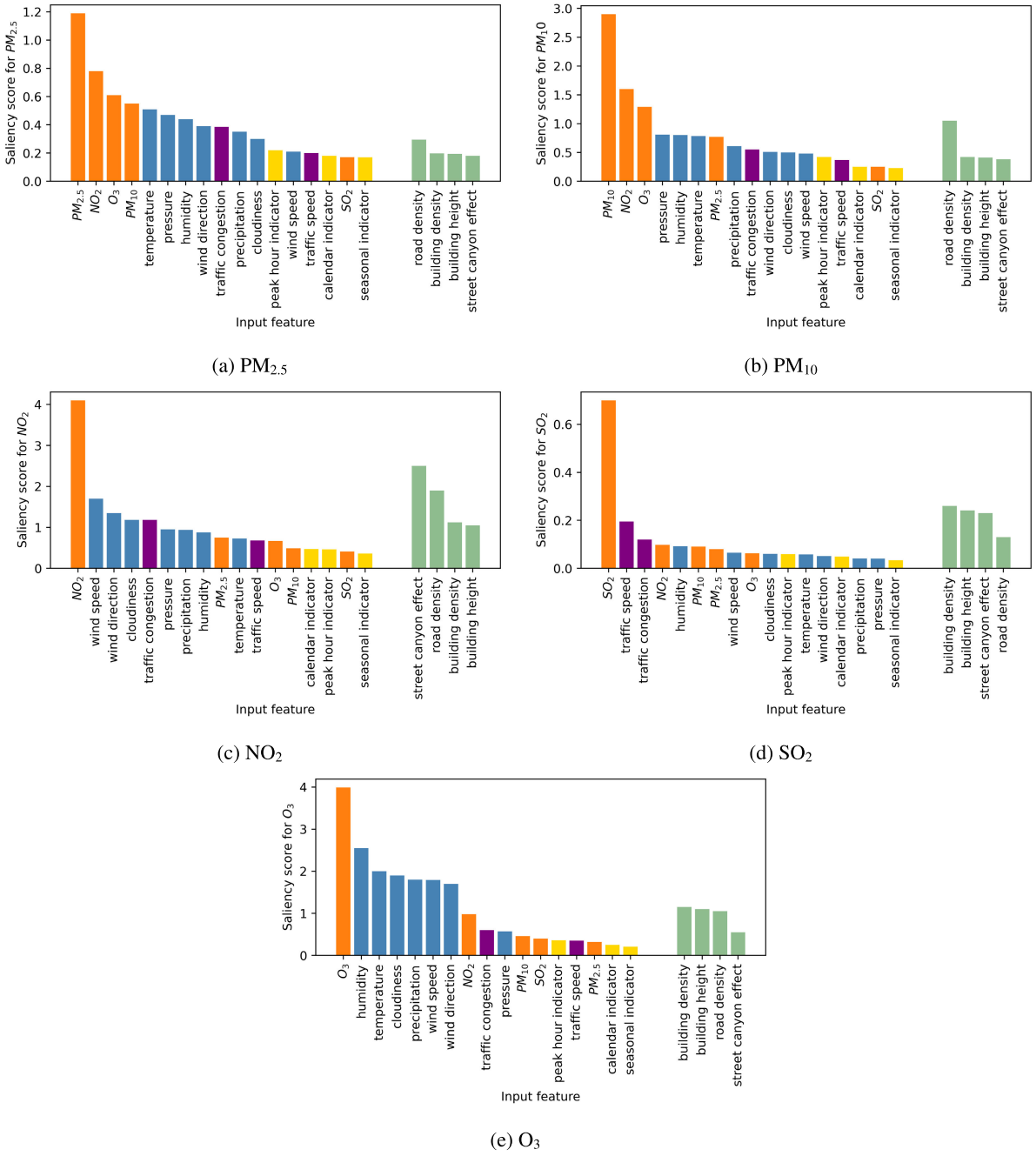


FIGURE 7. Contributions of individual temporal-spatial features to air pollutants forecast in Hong Kong via saliency analysis (Temporal and spatial features separated).

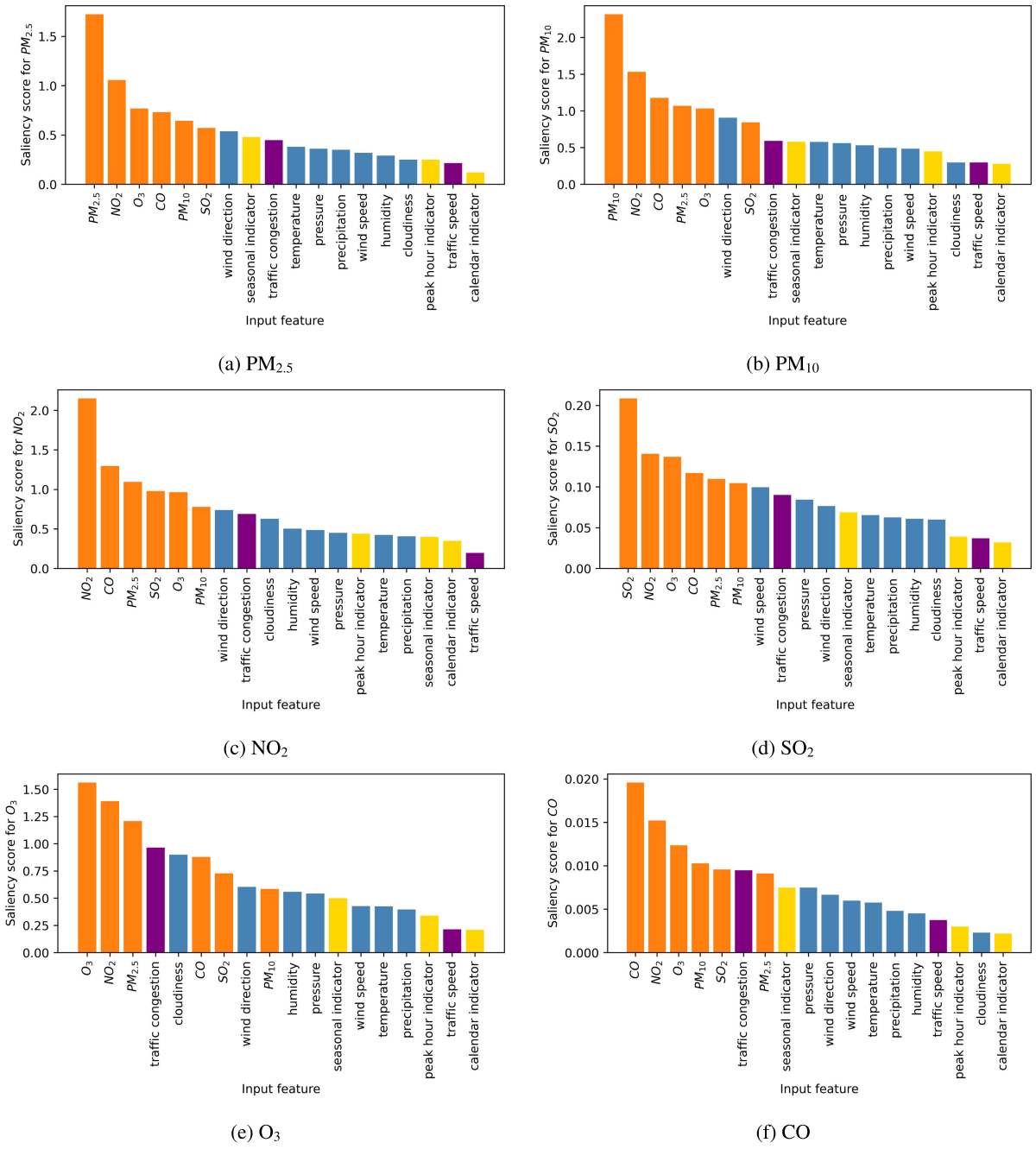


FIGURE 8. **Contributions of individual temporal-spatial features to air pollutants forecast in Beijing via saliency analysis.**

(one hour ahead) and the long term (24 hours ahead). Using Hong Kong and Beijing as the case studies, our proposed novel deep learning framework, Deep-AIR, utilizing extensively the readily available urban proxy data, while using 1×1 convolution layers to better capture cross-feature spatial interactions between air pollutants and various important urban dynamic features, including weather, traffic (traffic speed and congestion), and urban morphology (road density, building density and height, and street canyon), and time-sensitive feature (weekends/weekdays, working days/public holidays, peak hours/non-peak hours, and seasonal variation). In addition, background pollutant feature and time-sensitive feature have also been incorporated into Deep-AIR to improve model estimation and forecasting performance.

Deep-AIR has achieved better performance than the baselines, including statistical and deep learning models, in fine-grained air pollution estimation and air pollution forecast. Experimental results show that Deep-AIR has outperformed the best baselines by 1.7% and 1.4% for

fine-grained air pollution estimation in Hong Kong and Beijing, respectively (see Table 3). Deep-AIR has also outperformed the best baselines by 1.7% and 2.3% for 1-hr air pollution forecast in Hong Kong and Beijing, and by 3.8% and 3.3% for 24-hr air pollution forecast in Hong Kong and Beijing, respectively (see Tables 4 and 5). Compared to the pure temporal models such as ARIMA, LSTM, and Attention-LSTM, other temporal models that carry spatial structures, including ConvLSTM, GCN-LSTM, and Deep-AIR, have generally achieved a higher accuracy, suggesting the importance of spatial information in air pollution estimation, and forecast. For pollution forecast based on the air pollution (A) features only, or pollution forecast based on additional features, including weather (W), traffic (T), urban morphology (M), Time-sensitive features (S), have been added, our proposed model has achieved the greatest improvement among all baselines. This implies that our proposed Deep-AIR forecast model is best performed when the domain-specific urban dynamics and time-sensitive data have been incorporated. Furthermore, Deep-AIR outperforms the baselines in forecast, even when fewer input features have been used.

The reason why Deep-AIR performs better as compared to other hybrid spatial-temporal baselines may be attributable to its 1×1 convolutional layers. First, compared to the traditional stacked model (i.e., GCN-LSTM and ResNet-LSTM), Deep-AIR can better preserve the cross-feature spatial interactions between the air pollutant and other important urban dynamic features through the 1×1 convolutional layers. Second, ConvLSTM may fail to capture the long-term temporal dependence of high-dimensional urban dynamics directly. In contrast, a one-dimensional spatial representation at each time step is fed into the LSTM component in Deep-AIR. Such one-dimensional representation can better capture the spatial interactions of different features, with the incorporation of the 1×1 convolutional layers of the CNN component. Although 1×1 convolution layers have already been used for modelling cross-channel correlations [93], few have adopted this structure in deep learning-based air pollution modelling. In one recent study, a 1×1 convolution layer has been applied to the final output of a ConvLSTM model for forecasting city-wide air pollution level [94]. However, the forecast error of ConvLSTM increases when all input features, including weather and traffic, have been included [94]. This suggests that unlike our Deep-AIR model structure, the combination of ConvLSTM with the 1×1 convolution layer is less capable of capturing the spatial interactions between different types of urban dynamics.

Moreover, our experimental results show that estimating real-time fine-grained air pollution is more challenging than forecasting 1-hr air pollution at monitoring stations. The lowest error rate of our proposed model for 1-hr air pollution forecast was 21.1% and 23.9% in Hong Kong and Beijing, respectively. For fine-grained air pollution estimation, the current hour, the lowest error rates of our proposed model are 31.1% and 34.2% in Hong Kong and Beijing, respectively. Compared to 1-hr air pollution forecasts at monitoring stations where historical air pollution information is available; the error rate is significantly higher when estimating real-time air pollution levels in locations without local air pollution information. This suggests that the sparsity of ground truths has limited the accuracy of fine-grained air pollution estimation. Such finding is consistent with a previous data-driven air pollution modelling study where the results of both fine-grained air pollution estimations and 1-hr air pollution forecasts were compared [10]. Nevertheless, it remains challenging to provide accurate long-term air pollution forecasts, even though historical air pollution measurements are available. The lowest error rate of our proposed model for 24-hr air pollution forecast is 32.3% and 35.2% in Hong Kong and Beijing, respectively. Forecasting long-term air pollution is inaccurate due to various issues such as the accumulation of errors during multi-step forecast [95] and the lack of future information (proxy data) such as weather and traffic forecast [8]. In Beijing, Deep-AIR performs better in Summer and Autumn, as compared to Winter and Spring. This may be attributable to the extremely high concentration values in Winter and Spring, and the significant contribution of cross-border transmission from neighboring provinces. Such pattern is more difficult to be learned, when the model depends on historical data inputs. In Hong Kong, the model achieves the best result in Autumn, though the pollution values are high in Autumn. This may indicate that the pollutant sources and the change in concentration patterns are more regular, and can be better reflected from the available data.

Furthermore, our saliency analysis has highlighted the importance of domain-specific features for urban air pollution modelling (see Figures 7 and 8). Such findings can improve the interpretability of our proposed framework, while providing more insights for deep learning-based air pollution modelling to account for the most important domain-specific features. The saliency scores obtained from our proposed model are consistent with domain knowledge. When examining the temporal and the spatial features individually, spatial urban morphological features, including street canyon effect and road density, are more important predictors of NO_2 in Hong Kong, as compared to temporal features such as wind speed and wind direction, and traffic congestion, since Hong Kong is a city characterized by closely packed high-rise buildings and heavy road transports, thereby serious NO_2 emissions. Our finding reinforces a previous study which identified NO_2 a key road-based pollutant for urban metropolises [96]. Similarly, the spatial urban morphological feature, building density, are more salient than the temporal feature, traffic speed, in predicting SO_2 in Hong Kong. However, spatial urban morphological features are less salient in predicting $PM_{2.5}$ and O_3 in Hong Kong as compared to the temporal meteorological features such as temperature, humidity, and wind speed and direction, likely due to the fact that weather features, such as wind speed and

direction are more responsible for (1) long-range transports of PM pollutants [96] and (2) generation of O_3 as a secondary pollutant [15]. Nevertheless, temporal traffic features are less salient for all air pollutants in Hong Kong, likely due to the lack of traffic data in high geographical coverage. When traffic data are available in large-scale, further investigation regarding the importance of traffic speed and congestion can be made. In contrast, historical air pollutants, traffic congestion, and wind speed and direction are the most salient temporal predictors of air pollution in Beijing, a city facing serious air pollution due to severe local vehicular emissions and regional $PM_{2.5}$ transports from neighbouring cities. Although the spatial urban morphological data are unavailable from the Beijing dataset, previous physical modelling studies demonstrated the importance of canyon geometries for estimating the high-resolution NO_2 across the city [17]. Given that the urban morphological features have already played an important role in predicting traffic-related pollution in Hong Kong, relevant data can be collected in Beijing to improve air pollution estimation and forecast of Deep-AIR in the future.

The current study can be improved in the following aspects. First, the lack of air pollution monitoring stations and the sparsity of features relevant to the air pollution process have constrained the performance of our proposed model, especially for fine-grained air pollution estimation. The geographical coverage of the collected traffic data in Hong Kong remains low, whereas urban morphology data are unavailable in Beijing. In the future, more real-time traffic congestion data can be collected from the Google Maps Hong Kong, to complement the city's traffic data collected from the official sources, whilst street-view images from the Baidu Maps Beijing can be collected for estimating the city's urban canyon geometries, and be included as model inputs to improve the model performance of CNNs [97]. Well-calibrated portable sensors can also be deployed in schools and vehicles across different parts of Hong Kong to collect more empirical air pollutant data to supplement that provided by existing air pollution monitoring stations. Second, the importance of different domain-specific features may be better considered in our proposed model. Gradient-based saliency assessment has provided new domain-specific heuristics as regard to what an optimal subset of features should be desirable for model inputs [98]. In the future, we can investigate more advanced deep learning model structures to tailor-make model structures that utilize the most relevant features for air pollution estimation and forecast. Third, the generalizability and transferability of our proposed framework can be further examined. To improve the generalizability, one can examine more closely the similarity and the variability in saliency of the temporal and spatial features and their relative contributions to air pollution estimation and forecast across different cities. In the future, transfer learning of the temporal and spatial data [99] can be utilized in cross-city pollution estimation and forecast. Fourth, forecasting long-term air pollution can be further improved. Weather forecast data can be incorporated into new deep learning models to improve forecast accuracy.

Finally, based on our proposed hybrid deep learning framework, forecasting fine-grained air pollution at the city-wide level would be a viable research direction. As compared to fine-grained air pollution estimation in the existing hour, fine-grained air pollution forecast is more complicated because in real life air pollution measurements over both space and time are sparse or missing. Several deep learning studies have attempted to forecast the spatial-temporal variation of air pollution across different parts of the city. Ma *et al.* constructed a fine-grained $PM_{2.5}$ forecast map by interpolating station-level forecasts from a bidirectional LSTM model, using an IDW-based layer with optimized parameters [100]. Moreover, Qi *et al.* proposed a deep feedforward neural network to estimate fine-grained air pollution in the current hour and subsequent hours [101], utilizing a feature selection layer and the spatial and temporal information obtained from neighboring air pollutant and meteorological observations. Hähnel *et al.* conducted a physical-inspired study and trained an RNN model to forecast city-wide air pollution while addressing the consistency constraints of partial differential equations that govern the pollution dispersion process [102]. Nevertheless, these studies have neither fully exploited the advanced deep learning techniques such as CNN models that can better capture the spatial characteristics of the data structure, nor the complex spatial dependence of air pollution and urban proxy data. Recently, Le *et al.* proposed a ConvLSTM model to forecast fine-grained air pollution at the city-wide level, but urban morphology features has not been included in the model [94]. Up till now, how to reshape advanced spatial-temporal models to fully incorporate the high saliency spatial features and their interactions in fine-grained air pollution forecast at the city-wide level remains to be fully investigated. The air pollution forecast model in our current framework has been trained and tested at locations where historical ground-truth air pollution measurements are available. In the future, we will extend our proposed air pollution forecast model to provide fine-grained air pollution forecast in areas not covered by air pollution monitoring stations.

VI. CONCLUSION

A hybrid deep learning framework, Deep-AIR, for fine-grained air pollution estimation at the city-wide level for the current hour and air pollution forecast at the station-level for the next 24 hours has been put forward. To the best of our knowledge, this is the first deep learning-based air pollution study that integrates domain-specific features indicative of the street canyon effect, including building density and height, for air pollution estimation and forecast. Deep-AIR incorporates 1×1 convolutional layers into a CNN model to capture the complex spatial interactions between air pollutants and other domain-specific urban dynamic

features. Our experimental results show that Deep-AIR has achieved higher accuracy and outperformed the best baselines for fine-grained hourly air pollution estimation, 1-hr air pollution forecast, and 24-hr air pollution forecast by 1.4%, 2.7%, and 3.3%, respectively. For Hong Kong, Deep-AIR has achieved 68.9%, 78.9%, and 67.7% accuracy for fine-grained air pollution estimation, 1-hr air pollution forecast, and 24-hr air pollution forecast, respectively. For Beijing, Deep-AIR has achieved 65.8%, 76.1%, and 64.8% accuracy for fine-grained air pollution estimation, 1-hr air pollution forecast, and 24-hr air pollution forecast, respectively. Our saliency analysis has revealed that for Hong Kong, urban morphological features, including street canyon and road density, are the best predictors for NO2, whilst historical air pollution and meteorological features, are the best predictors for $PM_{2.5}$. For Beijing, in the absence of spatial street canyon feature, historical air pollutants, traffic congestion, wind direction and wind speed, and seasonal indicator, are the best predictors for all air pollutants. Such findings imply that future fine-grained air pollution estimation and forecast model can be improved by incorporating spatial features indicative of the street canyon effect. Further, large-scale domain-specific feature inputs, such as traffic data in Hong Kong, urban morphology data in Beijing, and weather forecast data for both Hong Kong and Beijing, can be inputted to improve the performance of Deep-AIR in the future. Further, tailor-making model structures to account for the optimal set of high-saliency temporal/spatial domain-specific features would improve model accuracy. In the future, more comprehensive experiments can be performed to evaluate the generalizability of Deep-AIR in city-wide air pollution forecast. and its transferability in fine-grained air pollution estimation across different metropolitan cities of the world.

ACKNOWLEDGMENT

The authors would like to thank Microsoft for providing cloud computing services under Microsoft Azure. They acknowledge the Environmental Protection Department, the Hong Kong Observatory, and the Transport Department of the HKSAR Government, for publicizing air pollution, weather, and traffic data of Hong Kong, respectively, and the Lands Department of the HKSAR Government for providing information on roads and buildings in Hong Kong. They would also like to thank the Beijing Municipal Environment Monitoring Center and the National Meteorological Information Center for publicizing air pollution and meteorological data in Beijing, and the map service of AutoNavi (Gaode) for providing the real-time traffic data in Beijing.

(Qi Zhang and Yang Han contributed equally to this work.)

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] T. Xue, T. Zhu, Y. Zheng, and Q. Zhang, "Declines in mental health associated with air pollution and temperature variability in China," *Nature Commun.*, vol. 10, no. 1, pp. 1–8, Dec. 2019.

[3] C. Copat, A. Cristaldi, M. Fiore, A. Grasso, P. Zuccarello, S. S. Signorelli, G. O. Conti, and M. Ferrante, "The role of air pollution (PM and NO_2) in COVID-19 spread and lethality: A systematic review," *Environ. Res.*, vol. 191, Dec. 2020, Art. no. 110129.

[4] M. Jerrett, A. Arain, P. Kanaroglou, B. Beckerman, D. Potoglou, T. Sahsuvaroglu, J. Morrison, and C. Giovis, "A review and evaluation of intraurban air pollution exposure models," *J. Exposure Anal. Environ. Epidemiol.*, vol. 15, no. 2, pp. 185–204, 2005.

[5] M. Lateb, R. N. Meroney, M. Yataghene, H. Fellouah, F. Saleh, and M. Boufadel, "On the use of numerical modelling for near-field pollutant dispersion in urban environments—A review," *Environ. Pollut.*, vol. 208, pp. 271–283, Jan. 2016.

[6] Y. Rybarczyk and R. Zalakeviciute, "Machine learning approaches for outdoor air quality modelling: A systematic review," *Appl. Sci.*, vol. 8, no. 12, p. 2570, Dec. 2018.

[7] 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. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1.

[8] X. Yi, Z. Duan, R. Li, J. Zhang, T. Li, and Y. Zheng, "Predicting fine-grained air quality based on deep neural networks," *IEEE Trans. Big Data*, early access, Dec. 24, 2021, doi: 10.1109/TBDATA.2020.3047078.

[9] L. Chen, Y. Ding, D. Lyu, X. Liu, and H. Long, "Deep multi-task learning based urban air quality index modelling," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 1, pp. 1–17, Mar. 2019.

[10] X. Zhao, T. Xu, Y. Fu, E. Chen, and H. Guo, "Incorporating spatio-temporal smoothness for air quality inference," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 1177–1182.

[11] Y. Lin, N. Mago, Y. Gao, Y. Li, Y.-Y. Chiang, C. Shahabi, and J. L. Ambite, "Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning," in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2018, pp. 359–368.

[12] X. Liang, S. Li, S. Zhang, H. Huang, and S. X. Chen, "$PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities," *J. Geophys. Res., Atmos.*, vol. 121, no. 17, pp. 10220–10236, 2016.

[13] I. Laña, J. Del Ser, A. Padró, M. Vélez, and C. Casanova-Mateo, "The role of local urban traffic and meteorological conditions in air pollution: A data-based case study in Madrid, Spain," *Atmos. Environ.*, vol. 145, pp. 424–438, Nov. 2016.

[14] F. Costabile, G. Bertoni, F. Desantis, F. Wang, H. Weimin, L. Fenglei, and I. Allegrini, "A preliminary assessment of major air pollutants in the city of Suzhou, China," *Atmos. Environ.*, vol. 40, no. 33, pp. 6380–6395, Oct. 2006.

[15] S. Vardoulakis, B. E. A. Fisher, K. Pericleous, and N. Gonzalez-Flesca, "Modelling air quality in street canyons: A review," *Atmos. Environ.*, vol. 37, no. 2, pp. 155–182, Jan. 2003.

[16] Y. Shi, K. K.-L. Lau, and E. Ng, "Developing street-level $PM_{2.5}$ and PM_{10} land use regression models in high-density Hong Kong with urban morphological factors," *Environ. Sci. Technol.*, vol. 50, no. 15, pp. 8178–8187, 2016.

[17] X. Fu, J. Liu, G. A. Ban-Weiss, J. Zhang, X. Huang, B. Ouyang, O. Popoola, and S. Tao, "Effects of canyon geometry on the distribution of traffic-related air pollution in a large urban area: Implications of a multi-canyon air pollution dispersion model," *Atmos. Environ.*, vol. 165, pp. 111–121, Sep. 2017.

[18] K. J. Maji, V. O. Li, and J. C. Lam, "Effects of China's current air pollution prevention and control action plan on air pollution patterns, health risks and mortalities in Beijing 2014–2018," *Chemosphere*, vol. 260, Dec. 2020, Art. no. 127572.

[19] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, and T. Chi, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environ. Pollut.*, vol. 231, pp. 997–1004, Dec. 2017.

[20] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, and Y. Hu, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *Sci. Total Environ.*, vol. 654, pp. 1091–1099, Mar. 2019.

[21] Q. Zhang, J. C. K. Lam, V. O. K. Li, and Y. Han, "Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution forecast," 2020, *arXiv:2001.11957*.

[22] K.-H. Kwak, J.-J. Baik, Y.-H. Ryu, and S.-H. Lee, "Urban air quality simulation in a high-rise building area using a CFD model coupled with mesoscale meteorological and chemistry-transport models," *Atmos. Environ.*, vol. 100, pp. 167–177, Jan. 2015.

[23] A. Riddle, D. Carruthers, A. Sharpe, C. McHugh, and J. Stocker, "Comparisons between FLUENT and ADMS for atmospheric dispersion modelling," *Atmos. Environ.*, vol. 38, no. 7, pp. 1029–1038, Mar. 2004.

[24] V. Mallet, A. Tilloy, D. Poulet, S. Girard, and F. Brocheton, "Meta-modeling of ADMS-urban by dimension reduction and emulation," *Atmos. Environ.*, vol. 184, pp. 37–46, Jul. 2018.

[25] Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov, "Real-time air quality forecasting, Part I: History, techniques, and current status," *Atmos. Environ.*, vol. 60, pp. 632–655, Dec. 2012.

[26] D. Byun and K. L. Schere, "Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (CMAQ) modeling system," *Appl. Mech. Rev.*, vol. 59, no. 2, pp. 51–77, Mar. 2006.

[27] X. Tie, G. P. Brasseur, C. Zhao, C. Granier, S. Massie, Y. Qin, P. Wang, G. Wang, P. Yang, and A. Richter, "Chemical characterization of air pollution in eastern China and the eastern United States," *Atmos. Environ.*, vol. 40, no. 14, pp. 2607–2625, May 2006.

[28] G. A. Grell, S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, "Fully coupled 'online' chemistry within the WRF model," *Atmos. Environ.*, vol. 39, no. 37, pp. 6957–6975, 2005.

[29] Z. Wang, T. Maeda, M. Hayashi, L.-F. Hsiao, and K.-Y. Liu, "A nested air quality prediction modeling system for urban and regional scales: Application for high-ozone episode in Taiwan," *Water, Air, Soil Pollut.*, vol. 130, nos. 1–4, pp. 391–396, Aug. 2001.

[30] W. G. Cobourn, "An enhanced $PM_{2.5}$ air quality forecast model based on nonlinear regression and back-trajectory concentrations," *Atmos. Environ.*, vol. 44, no. 25, pp. 3015–3023, 2010.

[31] T. V. Vu, Z. Shi, J. Cheng, Q. Zhang, K. He, S. Wang, and R. M. Harrison, "Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique," *Atmos. Chem. Phys.*, vol. 19, no. 17, pp. 11303–11314, Sep. 2019.

[32] W. Wang, X. An, Q. Li, Y.-A. Geng, H. Yu, and X. Zhou, "Optimization research on air quality numerical model forecasting effects based on deep learning methods," *Atmos. Res.*, vol. 271, Jun. 2022, Art. no. 106082.

[33] Z. Huang, J. Zheng, J. Ou, Z. Zhong, Y. Wu, and M. Shao, "A feasible methodological framework for uncertainty analysis and diagnosis of atmospheric chemical transport models," *Environ. Sci. Technol.*, vol. 53, no. 6, pp. 3110–3118, Mar. 2019.

[34] J. Fine, L. Vuilleumier, S. Reynolds, P. Roth, and N. Brown, "Evaluating uncertainties in regional photochemical air quality modeling," *Annu. Rev. Environ. Resour.*, vol. 28, no. 1, pp. 59–106, Nov. 2003.

[35] E. Debry and V. Mallet, "Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and PM_{10} on the Prev'Air platform," *Atmos. Environ.*, vol. 91, pp. 71–84, Jul. 2014.

[36] H. Zhang, J. C. Linford, A. Sandu, and R. Sander, "Chemical mechanism solvers in air quality models," *Atmosphere*, vol. 2, no. 3, pp. 510–532, Sep. 2011.

[37] D. J. Briggs, C. de Hoogh, J. Gulliver, J. Wills, P. Elliott, S. Kingham, and K. Smallbone, "A regression-based method for mapping traffic-related air pollution: Application and testing in four contrasting urban environments," *Sci. Total Environ.*, vol. 253, nos. 1–3, pp. 151–167, May 2000.

[38] M. Wu, J. Huang, N. Liu, R. Ma, Y. Wang, and L. Zhang, "A hybrid air pollution reconstruction by adaptive interpolation method," in *Proc. 16th ACM Conf. Embedded Netw. Sensor Syst.*, 2018, pp. 408–409.

[39] K. Kumar, A. Yadav, M. Singh, H. Hassan, and V. Jain, "Forecasting daily maximum surface ozone concentrations in Brunei Darussalam—An ARIMA modeling approach," *J. Air Waste Manage. Assoc.*, vol. 54, no. 7, pp. 809–814, 2004.

[40] N. Nikzad, N. Verma, C. Ziftci, E. Bales, N. Quick, P. Zappi, and K. Patrick, "Citisense: Improving geospatial environmental assessment of air quality using a wireless personal exposure monitoring system," in *Proc. Conf. Wireless Health*, 2012, pp. 1–8.

[41] J. Huang, N. Duan, P. Ji, C. Ma, and Y. Ding, "A crowdsource-based sensing system for monitoring fine-grained air quality in urban environments," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3240–3247, Apr. 2018.

[42] Y. Yang, Z. Zheng, K. Bian, L. Song, and Z. Han, "Real-time profiling of fine-grained air quality index distribution using UAV sensing," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 186–198, Feb. 2018.

[43] 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.

[44] 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*, 2013, pp. 1436–1444.

[45] L. Chen, Y. Cai, Y. Ding, M. Lv, C. Yuan, and G. Chen, "Spatially fine-grained urban air quality estimation using ensemble semi-supervised learning and pruning," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2016, pp. 1076–1087.

[46] J. Y. Zhu, C. Sun, and V. O. K. Li, "An extended spatio-temporal Granger causality model for air quality estimation with heterogeneous urban big data," *IEEE Trans. Big Data*, vol. 3, no. 3, pp. 307–319, Sep. 2017.

[47] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 2267–2276.

[48] S. Wang, J. Cao, and P. Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Trans. Knowl. Data Eng.*, early access, Sep. 22, 2020, doi: 10.1109/TKDE.2020.3025580.

[49] 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, 2020.

[50] 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.

[51] X. Yan, Z. Zang, Y. Jiang, W. Shi, Y. Guo, D. Li, C. Zhao, and L. Husi, "A spatial-temporal interpretable deep learning model for improving interpretability and predictive accuracy of satellite-based $PM_{2.5}$," *Environ. Pollut.*, vol. 273, Mar. 2021, Art. no. 116459.

[52] Q. Han, D. Lu, and R. Chen, "Fine-grained air quality inference via multi-channel attention model," in *Proc. 13th Int. Joint Conf. Artif. Intell. (IJCAI)*, Z.-H. Zhou, Ed. Montreal, QC, Canada, Aug. 2021, pp. 2512–2518.

[53] R. Ma, X. Xu, Y. Wang, H. Y. Noh, P. Zhang, and L. Zhang, "Guiding the data learning process with physical model in air pollution inference," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 4475–4483.

[54] V. Athira, P. Geetha, R. Vinayakumar, and K. P. Soman, "DeepAirNet: Applying recurrent networks for air quality prediction," *Proc. Comput. Sci.*, vol. 132, pp. 1394–1403, Dec. 2018.

[55] P. Nath, P. Saha, A. I. Middya, and S. Roy, "Long-term time-series pollution forecast using statistical and deep learning methods," *Neural Comput. Appl.*, vol. 33, no. 19, pp. 12551–12570, Oct. 2021.

[56] Z. Zhang, Y. Zeng, and K. Yan, "A hybrid deep learning technology for $PM_{2.5}$ air quality forecasting," *Environ. Sci. Pollut. Res.*, vol. 28, no. 29, pp. 39409–39422, 2021.

[57] V. O. K. Li, J. C. K. Lam, Y. Chen, and J. Gu, "Deep learning model to estimate air pollution using M-BP to fill in missing proxy urban data," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[58] Z. Luo, J. Huang, K. Hu, X. Li, and P. Zhang, "AccuAir: Winning solution to air quality prediction for KDD Cup 2018," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 1842–1850.

[59] D.-R. Liu, S.-J. Lee, Y. Huang, and C.-J. Chiu, "Air pollution forecasting based on attention-based LSTM neural network and ensemble learning," *Expert Syst.*, vol. 37, no. 3, 2020, Art. no. e12511.

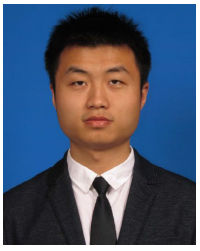
[60] Y. Han, J. C. K. Lam, V. O. Li, and Q. Zhang, "A domain-specific Bayesian deep-learning approach for air pollution forecast," *IEEE Trans. Big Data*, early access, Jun. 29, 2020, doi: 10.1109/TBDATA.2020.3005368.

[61] J. Ma, Z. Li, J. C. P. Cheng, Y. Ding, C. Lin, and Z. Xu, "Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network," *Sci. Total Environ.*, vol. 705, Feb. 2020, Art. no. 135771.

[62] K. K. R. Samal, K. S. Babu, and S. K. Das, "Temporal convolutional denoising autoencoder network for air pollution prediction with missing values," *Urban Climate*, vol. 38, Jul. 2021, Art. no. 100872.

[63] D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, and B. Zhang, "A novel combined prediction scheme based on CNN and LSTM for urban $PM_{2.5}$ concentration," *IEEE Access*, vol. 7, pp. 20050–20059, 2019.

[64] J. Yang, R. Yan, M. Nong, J. Liao, F. Li, and W. Sun, "$PM_{2.5}$ concentrations forecasting in Beijing through deep learning with different inputs, model structures and forecast time," *Atmos. Pollut. Res.*, vol. 12, no. 9, Sep. 2021, Art. no. 101168.
[65] L. Ge, K. Wu, Y. Zeng, F. Chang, Y. Wang, and S. Li, "Multi-scale spatiotemporal graph convolution network for air quality prediction," *Appl. Intell.*, vol. 51, pp. 3491–3505, Nov. 2020.
[66] V.-D. Le, T.-C. Bui, and S.-K. Cha, "Spatiotemporal graph convolutional recurrent neural network model for citywide air pollution forecasting," *TechRxiv*, Jul. 2021, doi: 10.36227/techrxiv.14958552.v1.
[67] K. K. R. Samal, K. S. Babu, and S. K. Das, "Multi-directional temporal convolutional artificial neural network for $PM_{2.5}$ forecasting with missing values: A deep learning approach," *Urban Climate*, vol. 36, Mar. 2021, Art. no. 100800.
[68] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2412–2424, Jun. 2021.
[69] A. Dairi, F. Harrou, S. Khadraoui, and Y. Sun, "Integrated multiple directed attention-based deep learning for improved air pollution forecasting," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021.
[70] W. Wang, W. Mao, X. Tong, and G. Xu, "A novel recursive model based on a convolutional long short-term memory neural network for air pollution prediction," *Remote Sens.*, vol. 13, no. 7, p. 1284, Mar. 2021.
[71] S. Wang, Y. Li, J. Zhang, Q. Meng, L. Meng, and F. Gao, "$PM_{2.5}$-GNN: A domain knowledge enhanced graph neural network for $PM_{2.5}$ forecasting," in *Proc. 28th Int. Conf. Adv. Geographic Inf. Syst.*, 2020, pp. 163–166.
[72] X. Yi, J. Zhang, Z. Wang, T. Li, and Y. Zheng, "Deep distributed fusion network for air quality prediction," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 965–973.
[73] D. Liu, Y. K. Hsu, H. Y. Chen, and H. J. Jau, "Air pollution prediction based on factory-aware attentional LSTM neural network," *Computing*, vol. 103, no. 1, pp. 75–98, 2021.
[74] A. Lau, A. Lo, J. Gray, Z. Yuan, and C. Loh. (2007). *Relative Significance of Local vs. Regional Sources: Hong Kong's Air Pollution*. Institute for the Environment of the University of Science and Technology and Civic Exchange. [Online]. Available: http://www.civicexchange.org/publications/2007/airmarch.pdf
[75] O. A. Popoola, D. Carruthers, C. Lad, V. B. Bright, M. I. Mead, M. E. Stettler, J. R. Saffell, and R. L. Jones, "Use of networks of low cost air quality sensors to quantify air quality in urban settings," *Atmos. Environ.*, vol. 194, pp. 58–70, Dec. 2018.
[76] Y. Yu, J. J. Q. Yu, V. O. K. Li, and J. C. K. Lam, "A novel interpolation-SVT approach for recovering missing low-rank air quality data," *IEEE Access*, vol. 8, pp. 74291–74305, 2020.
[77] P. Sicard, E. Paoletti, E. Agathokleous, V. Araminienė, C. Proietti, F. Coulibaly, and A. De Marco, "Ozone weekend effect in cities: Deep insights for urban air pollution control," *Environ. Res.*, vol. 191, Dec. 2020, Art. no. 110193.
[78] J. Lin and Y. E. Ge, "Impacts of traffic heterogeneity on roadside air pollution concentration," *Transp. Res. D, Transp. Environ.*, vol. 11, no. 2, pp. 166–170, Mar. 2006.
[79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[80] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1655–1661.
[81] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[82] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
[83] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," 2013, *arXiv:1312.6034*.
[84] *Inquire and Download Air Quality Monitoring Data*. Accessed: Mar. 1, 2021. [Online]. Available: https://cd.epic.epd.gov.hk/EPICDI/air/station/
[85] *Hong Kong Observatory Open Data*. Accessed: Mar. 1, 2021. [Online]. Available: https://www.hko.gov.hk/en/abouthko/opendata_intro.htm
[86] *World Weather Online*. Accessed: Mar. 1, 2021. [Online]. Available: https://www.worldweatheronline.com/
[87] *Traffic Speed Map (City Dashboard Version)*. Accessed: Mar. 1, 2021. [Online]. Available: https://data.gov.hk/en-data/dataset/hk-ogcio-da_div_02-city-dashboard-traffic-speed
[88] *Geo-Reference Database (iG1000)*. Accessed: Mar. 1, 2021. [Online]. Available: https://www.landsd.gov.hk/en/survey-mapping/mapping/other-products/geo-reference-database.html
[89] *Beijing Air Quality [in Chinese]*. Accessed: Mar. 1, 2021. [Online]. Available: http://zx.bjmemc.com.cn/
[90] *Hourly Data From Surface Meteorological Stations in China*. Accessed: Mar. 1, 2021. [Online]. Available: https://data.cma.cn/en/?r=data/detail&dataCode=A.0012.0001
[91] *Gaode Map API—Traffic Status [in Chinese]*. Accessed: Mar. 1, 2021. [Online]. Available: https://lbs.amap.com/api/webservice/guide/api/trafficstatus
[92] *Historical Air Quality and Meteorology Data in China [in Chinese]*. Accessed: Mar. 1, 2021. [Online]. Available: https://beijingair.sinaapp.com/
[93] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[94] V.-D. Le, T.-C. Bui, and S.-K. Cha, "Spatiotemporal deep learning model for citywide air pollution interpolation and prediction," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2020, pp. 55–62.
[95] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, and Y.-S. Wang, "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts," *J. Clean Prod.*, vol. 209, pp. 134–145, Feb. 2019.
[96] R. Tang, M. Blangiardo, and J. Gulliver, "Using building heights and street configuration to enhance intraurban PM_{10}, NO_X, and NO_2 land use regression models," *Environ. Sci. Technol.*, vol. 47, no. 20, pp. 11643–11650, 2013.
[97] M. Qi and S. Hankey, "Using street view imagery to predict street-level particulate air pollution," *Environ. Sci. Technol.*, vol. 55, no. 4, pp. 2695–2704, Feb. 2021.
[98] M. Wojtas and K. Chen, "Feature importance ranking for deep learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 5105–5114.
[99] L. Wang, X. Geng, X. Ma, F. Liu, and Q. Yang, "Cross-city transfer learning for deep spatio-temporal prediction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1893–1899.
[100] J. Ma, Y. Ding, V. J. Gan, C. Lin, and Z. Wan, "Spatiotemporal prediction of $PM_{2.5}$ concentrations at different time granularities using IDW-BLSTM," *IEEE Access*, vol. 7, pp. 107897–107907, 2019.
[101] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. Zhang, "Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2285–2297, Dec. 2018.
[102] P. Hähnel, J. Mareček, J. Monteil, and F. O'Donncha, "Using deep learning to extend the range of air pollution monitoring and forecasting," *J. Comput. Phys.*, vol. 408, May 2020, Art. no. 109278.



QI ZHANG received the B.E. degree in electronic engineering and the B.Ec. degree in economics from Tsinghua University, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research interests include deep-learning and its applications on air pollution, spatio-temporal data analysis, and urban computing. He holds the Hong Kong Ph.D. Fellowship.

YANG HAN received the bachelor's degree from Beihang University, the M.Sc. degree (Hons.) in computer science from The University of Hong Kong (HKU), and the M.Phil. degree in technology policy from the University of Cambridge. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, HKU. He has published widely in environmental science and policy, data and policy, and several IEEE journals. His research interests include AI and machine learning with applications on air pollution and health management.

VICTOR O. K. LI (Life Fellow, IEEE) received the S.B., S.M., E.E., and Sc.D. degrees in electrical engineering and computer science from MIT. He was a Visiting Professor with the Department of Computer Science and Technology, University of Cambridge, from April 2019 to August 2019. He was the Head of the Department of Electrical and Electronics Engineering, the Associate Dean (Research) of engineering, and the Managing Director of Versitech Ltd. He is currently the Chair of information engineering and the Cheng Yu-Tung Professor in sustainable development at the Department of Electrical and Electronic Engineering (EEE), The University of Hong Kong. He is also the Director of the HKU-Cambridge Clean Energy and Environment Research Platform and the HKU-Cambridge AI to Advance Well-Being and Society Research Platform, which are interdisciplinary collaborations with Cambridge University. He serves on the board of Sunevision Holdings Ltd., listed on the Hong Kong Stock Exchange and co-founded Fano Labs Ltd., an artificial intelligence (AI) company with his Ph.D. student. Previously, he was a Professor of electrical engineering at the University of Southern California (USC), Los Angeles, CA, USA, and the Director of the USC Communication Sciences Institute. In January 2018, he was awarded a U.S. $6.3 million RGC Theme-Based Research Project to develop deep learning techniques for personalized and smart air pollution monitoring and health management. Sought by government, industry, and academic organizations, he has lectured and consulted extensively internationally. His research interests include big data, AI, optimization techniques, and interdisciplinary clean energy and environment studies. He is a fellow of the Hong Kong Academy of Engineering Sciences, the IAE, and the HKIE. He has received numerous awards, including the PRC Ministry of Education Changjiang Chair Professorship from Tsinghua University, the U.K. Royal Academy of Engineering Senior Visiting Fellowship in communications, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of the HKSAR.

JACQUELINE C. K. LAM was the Hughes Hall Visiting Fellow at the University of Cambridge. She has been a Visiting Senior Research Fellow and an Associate Researcher at the Energy Policy Research Group, Judge Business School, University of Cambridge, since 2013. In 2019, she visited MIT, where she was a Visiting Scholar at CEEPR. She is currently an Associate Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong, and the Co-Director of HKU-AI WiSe, which embeds the HKU-Cambridge Clean Energy and Environment Research Platform and the HKU-Cambridge AI to Advance Well-Being and Society Research Platform. Her work has been published in IEEE TRANSACTIONS, AAAI, ACM, *Nature Scientific Reports*, *Environment International*, *Applied Energy*, *Environmental Science and Policy*, and *Energy Policy*. She has received four research grants funded by the Research Grants Council, HKSAR Government, from 2011 to 2020. The funded amount total U.S. $8million in PI or Co-PI capacity. Her environmental research study, in joint collaboration with Yang Han and Victor O. K. Li on $PM_{2.5}$ pollution and environmental inequality in Hong Kong. She has been published in environmental science and policy, and widely covered by more than 30 local and overseas newspapers and TVs. Her current research interests include the interface between AI, air pollution, and public health, also between AI and neuroscience, particularly the Alzheimer's Disease. Recently, she has won a Prestigious U.S. National Academy of Medicine-Healthy Longevity Catalyst Award, in 2021, on AI-driven drug repurposing, in collaboration with PI Prof. Victor O. K. Li, together with four other members from HKU, HK, and Tel Aviv University, Israel.