

Efficient inference of large-scale air quality using a lightweight ensemble predictor

Peixiao Wang, Hengcai Zhang, Jie Liu, Feng Lu & Tong Zhang

To cite this article: Peixiao Wang, Hengcai Zhang, Jie Liu, Feng Lu & Tong Zhang (10 Dec 2024): Efficient inference of large-scale air quality using a lightweight ensemble predictor, International Journal of Geographical Information Science, DOI: [10.1080/13658816.2024.2437044](https://doi.org/10.1080/13658816.2024.2437044)

To link to this article: <https://doi.org/10.1080/13658816.2024.2437044>



Published online: 10 Dec 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)








View Crossmark data [↗](#)

RESEARCH ARTICLE



Efficient inference of large-scale air quality using a lightweight ensemble predictor

Peixiao Wang^{a,b} , Hengcai Zhang^{a,b} , Jie Liu^c , Feng Lu^{a,b}  and Tong Zhang^d 

^aState Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, China; ^bCollege of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China; ^cSchool of Information Science and Engineering, University of Jinan, Jinan, China; ^dState Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

ABSTRACT

Accurate and efficient air quality prediction is crucial for public health protection and environmental sustainability. While numerous grid-based and graph-based prediction models have been developed, they encounter challenges in large-scale scenarios: (1) Grid-based models, though computationally efficient, have limited prediction accuracy in large-scale sparse scenarios; (2) Graph-based models, despite higher prediction accuracy, suffer from significant computational inefficiencies when dealing with a large number of sensors, i.e. graph nodes. To address these issues, we propose a Lightweight Ensemble Predictor (LiEnPred) for efficient air quality prediction in large-scale sparse scenarios. First, we present a data structure transformation algorithm that converts sparse monitoring sensors from graph structures to compact grid structures, preserving the connections between graph nodes. Next, we present a lightweight parameter-shared spatio-temporal dilation convolution network that efficiently captures spatio-temporal dependencies in air quality data without significantly increasing computation time or parameter scale. In our experiments, we collected air quality data from over 2000 sensors across China over the past three years and evaluated LiEnPred's prediction performance in large-scale scenarios using PM_{2.5} and NO₂ concentration data. The experimental results demonstrate that the proposed LiEnPred model matches or exceeds the predictive accuracy of eight baselines with faster time efficiency and fewer model parameters.

ARTICLE HISTORY

Received 19 August 2024
Accepted 28 November 2024

KEYWORDS

Air quality prediction; computationally efficient model; large-scale; sparse scenario

1. Introduction

With the advancement of urbanization, air pollution issues, such as PM_{2.5}, sulfur dioxide (SO₂), and nitrogen dioxide (NO₂), have become increasingly severe, posing significant challenges to the sustainable development of cities (Amato *et al.* 2020, Cheng *et al.* 2021). Studies have shown that prolonged exposure to highly polluted environments

may increase the risk of respiratory diseases and cancer (Yang *et al.* 2020, Cheng *et al.* 2023). In this context, accurate and efficient air quality prediction techniques have become essential tools for reducing human exposure to air pollution (Zhang and Zhang 2023, Guan *et al.* 2024). These technologies can forecast changes in pollutant concentrations, providing valuable insights and guidance for the public and government to implement targeted protective measure (Mengfan *et al.* 2022, Saad *et al.* 2024).

The key to air quality prediction lies in mining the complex spatio-temporal relationships within historical data to accurately and efficiently forecast future air quality. In recent years, deep learning models have become the predominant choice for air quality prediction, owing to their exceptional ability to capture complex spatio-temporal relationships (Zhang *et al.* 2021). From deep grid-based to deep graph-based prediction models, the diversity and performance of air quality prediction models have improved across various scenarios (Zhang *et al.* 2021, Guan *et al.* 2024). Although existing models have achieved satisfactory results in many scenarios, they still face challenges in large-scale scenarios with high real-time requirements (Le *et al.* 2020, Huang *et al.* 2021). As shown in Figure 1, monitoring sensors in large-scale scenarios are often sparsely distributed. The classical grid-based models for large-scale prediction are difficult to capture the spatiotemporal dependence in such sparse scenarios, significantly affecting their prediction accuracy (Asif *et al.* 2014, Guo *et al.* 2023). Moreover, grid-based models typically focus on modeling individual grid cells rather than specific sensors, making it difficult to accurately predict air quality changes at the sensor level. While graph-based models can offer high prediction accuracy for individual sensors, they often face computational inefficiencies due to the large number of sensors (graph nodes) (Li *et al.* 2023, Wang *et al.* 2024a). Overall, existing models face difficulties in balancing computational efficiency and prediction accuracy in large-scale scenarios (Cheng *et al.* 2020).

To address these challenges, we propose a Lightweight Ensemble Predictor (LiEnPred), which enables efficient air quality prediction in large-scale scenarios by combining the advantages of graph-based and grid-based models. The specific contributions of this study are as follows:

1. We present a data structure transformation (DST) algorithm that converts sparse monitoring sensors from graph structures to compact grid structures.

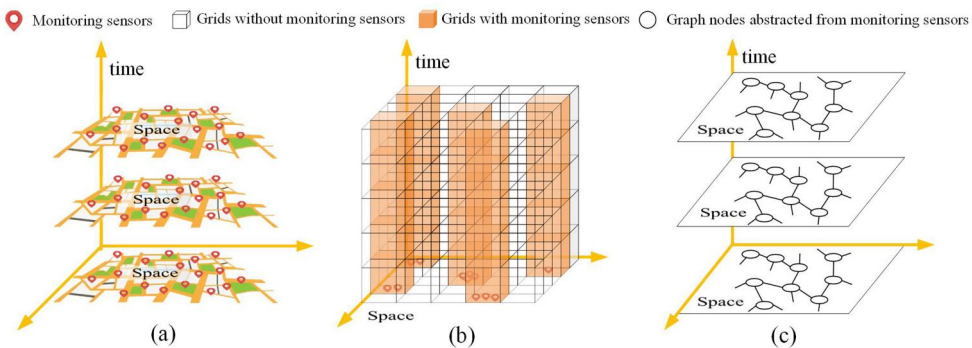


Figure 1. Large-scale air quality prediction scenarios: (a) sparse distributed sensors, (b) air quality data under grid structures, and (c) air quality data under graph structure.

The adjacency of pixels in these images implicitly reflects the connections between nodes in the original graph structure. This DST algorithm not only preserves essential graph information but also enables the LiEnPred model to develop a large-scale air quality prediction model based on grid structures.

2. We propose a lightweight parameter-shared spatio-temporal dilation convolution network (STP-DCN) that enables the LiEnPred model to swiftly capture spatio-temporal dependencies in air quality data without significantly increasing computation time or parameter scale.
3. We open-sourced an air quality dataset encompassing over 2,000 monitoring sensors across China from the past three years, and evaluated the LiEnPred's prediction performance (including prediction accuracy, inference speed, and parameter scale) using PM_{2.5} and NO₂ concentrations. Additionally, we opened the source code for the LiEnPred model to ensure the reproducibility of the experimental results.

2. Related works

Air quality prediction is fundamentally a spatiotemporal forecasting task. Consequently, we reviewed more general spatiotemporal prediction models, categorizing them into grid-based and graph-based models.

2.1. Grid-based spatio-temporal prediction models

Grid-based spatiotemporal prediction models originated from convolutional neural networks (CNNs) in computer vision (Voulodimos *et al.* 2018, Ibrahim *et al.* 2020). Numerous studies have demonstrated that CNNs can efficiently mine the nonlinear relationships between adjacent grid cells (or pixels) in images (Oprea *et al.* 2022). Given that most spatiotemporal data can be gridded, many scholars have leveraged CNNs to mine temporal relationships within images, enabling the prediction of future spatiotemporal data (Shi *et al.* 2015, 2017, Casagli *et al.* 2023). The large coverage of images makes grid-based models particularly suitable for large-scale prediction tasks, such as global-scale extreme rainfall prediction (Bi *et al.* 2023, Zhang *et al.* 2023), national-scale land use change prediction (Corner *et al.* 2014, Aburas *et al.* 2019), and city-scale travel demand prediction (Zheng *et al.* 2021, Noursalehi *et al.* 2022). Classical large-scale spatiotemporal prediction models include spatio-temporal residual networks (ST-ResNet) (Zhang *et al.* 2017), spatial-temporal 3D convolutional neural networks (ST-3DNet) (Guo *et al.* 2019b), spatio-temporal 3D multi-scale dilated dense networks (ST-3DMDDN) (He *et al.* 2024), hybrid integrated-DL model for spatio-temporal prediction (HIDLST) (Zhang *et al.* 2020), and spatio-temporal memory networks (STMN) (Li *et al.* 2022). Although many large-scale spatiotemporal prediction models have been proposed, they still face challenges, especially in scenarios that require highly accurate predictions for specific sensors. Specifically, grid-based models typically focus on modeling individual grid cells rather than specific sensors. Since each grid cell is equipped with multiple sensors, these models are unable to predict the future state of individual sensors. Furthermore, grid-based models often assume

that pixels in images are densely distributed and exhibit strong correlations with neighboring pixels (Bao *et al.* 2021, He *et al.* 2021). However, images converted from sparse monitoring stations often do not meet these assumptions in air quality prediction (Zhang and Zhang 2023).

2.2. Graph-based spatio-temporal prediction models

Graph-based spatiotemporal prediction models utilize graph structures to represent the connections or dependencies between monitoring stations, enabling the prediction of future spatiotemporal data (Kipf and Welling 2017, Schlichtkrull *et al.* 2017). Unlike grid structures, graph structures explicitly define relationships or dependencies between monitoring stations, thereby enhancing the model's predictive capability (Rossi *et al.* 2020, Wang *et al.* 2022). Classical graph-based spatiotemporal prediction models include graph neural networks and their variants, such as temporal graph convolutional networks (T-GCN) (Zhao *et al.* 2020), dynamic spatio-temporal aware graph neural network (DSTAGNN) (Lan *et al.* 2022), and Gaussian-based spatiotemporal graph convolutional network (RT-GCN) (Liu *et al.* 2024). Moreover, some scholars have established spatio-temporal prediction models based on the attention mechanism, such as Airformer (Liang *et al.* 2023), and Spatial-Temporal Transformer Networks (STTNs) (Xu *et al.* 2021). Essentially, the attention mechanism can be viewed as a special case of graph-based models, in particular the graph attention model (Veličković *et al.* 2018, Zhang *et al.* 2024b). For instance, Xu *et al.* (2023) referred to attention in the temporal dimension as temporal graph attention and in the spatial dimension as spatial graph attention. Overall, extensive studies indicate that graph-based spatiotemporal prediction models have broad application prospects in fields such as intelligent transportation, smart meteorology, urban planning, and environmental monitoring (Wang *et al.* 2023, Zhang *et al.* 2024). Despite achieving satisfactory prediction accuracy, graph-based spatiotemporal models still face challenges. Specifically, these models have primarily been tested on small-scale datasets and have not been validated on large-scale datasets with numerous graph nodes. The reason is the exponential growth in computational complexity of graph-based models with increasing node numbers, making most models unsuitable for real-time applications (Do *et al.* 2019, Dai *et al.* 2023, Li *et al.* 2023). Although some scholars have proposed lightweight spatiotemporal graph dilation neural networks (STGDN), these have only been validated on small-scale datasets (Wang *et al.* 2024a).

2.3. Strategy

In general, both grid-based and graph-based spatio-temporal prediction models exhibit mutual advantages and disadvantages. Grid-based models are particularly effective in terms of computational efficiency for large-scale spatiotemporal prediction tasks. However, they often struggle to achieve optimal accuracy when dealing with sparsely distributed datasets, owing to the constraints of their regular data structures. In contrast, graph-based models are well-suited for small-scale, high-precision

spatiotemporal predictions. However, their high complexity makes it challenging to meet real-time requirements in large-scale scenarios with numerous graph nodes.

To address these issues, we propose the LiEnPred model for efficient air quality prediction in large-scale scenarios. Specifically, we present a data structure transformation algorithm that converts sparse monitoring sensors from graph structures to compact grid structures, preserving the connections between graph nodes. Then, we present a parameter-sharing spatiotemporal dilated convolutional network to efficiently infer future spatiotemporal data using this compact grid representation.

3. Preliminaries

Before describing the LiEnPred model, Table 1 provides a brief overview of the naming conventions for mathematical notation used in this study.

Based on mathematical notation, we provide the relevant definition for LiEnPred model and the mathematical description of air quality prediction. Specifically, the sensors in the study area form a graph structure $G = \langle \{v_i\}_{i=1}^N, \mathbf{A} \rangle$, where $v_i \in V$ represents the i -th sensor, $\mathbf{A} \in R^{N \times N}$ denotes the connections or dependencies between sensors, and N is the total number of sensors. As illustrated in Figure 2, the LiEnPred model uses the graph structure of the sensors to generate a compact grid image $\mathbf{M} \in R^{P \times Q}$, where $m_{pq} > 0$ indicates that the pixel at p -th row and q -th column corresponds to a sensor, $m_{pq} = 0$ denotes that the pixel at p -th row and q -th does not correspond to any sensor. In this representation, P denotes the height and Q denotes the width of the compact grid image.

In the compact grid image, air pollutants (e.g. PM_{2.5} concentration, NO₂ concentration, CO concentration) collected over T time windows form a three-dimensional tensor $\mathcal{X} \in R^{P \times Q \times T}$, where $\mathbf{X}_t \in R^{P \times Q}$ represents the air pollutants recorded by all sensors within the t -th time window and $x_{pqt} \in R^{1 \times 1}$ denotes the air pollutant levels recorded by sensor m_{pq} within the t -th time window. For any $t \in [1, T]$, the LiEnPred model aims to efficiently infer future air quality $\hat{\mathcal{X}}_{t+1}^{t+F}$ based on the compact grid image \mathbf{M} and the historical three-dimensional tensor \mathcal{X}_{t-B+1}^t , as detailed in Equation (1).

$$\begin{cases} \mathbf{M} = \text{LiEnPred} \leftarrow G \\ \hat{\mathcal{X}}_{t+1}^{t+F} = \text{LiEnPred} \leftarrow \langle \mathcal{X}_{t-B+1}^t | \mathbf{M} \rangle \end{cases} \quad (1)$$

where $\text{LiEnPred} \leftarrow G$ denotes the generation of the compact grid image, $\langle \mathcal{X}_{t-B+1}^t | \mathbf{M} \rangle$ represents the construction of the three-dimensional tensor based on the compact grid image; $\mathcal{X}_{t-B+1}^t = \{\mathbf{X}_{t-b+1}\}_{b=1}^B \in R^{P \times Q \times B}$ indicates the historically collected air quality data with B being historical dependency step, $\hat{\mathcal{X}}_{t+1}^{t+F} = \{\hat{\mathbf{X}}_{t+f}\}_{f=1}^F \in R^{P \times Q \times F}$ refers to the prediction air quality with F being prediction step.

Table 1. Naming rules for mathematical symbols.

Symbol	Symbol type	Symbol meaning	Shape
$N, P, Q, T, B, F, E, L, D,$ and K	italic capital letter	constants with fixed values	one-dimension
$i, j, s, p, q, t, b, f, e,$ and l	italic lowercase letter	variables for iteration	one-dimension
$\mathbf{A}, \mathbf{M},$ and \mathbf{X}	bold capital letter	matrix for storing data	two-dimension
$\mathcal{X}, \mathcal{H},$ and \mathcal{W}	handwritten letter	tensor for storing data	three-dimension
$x, h, m,$ and w	lowercase letter	elements in matrix/tensor	one-dimension

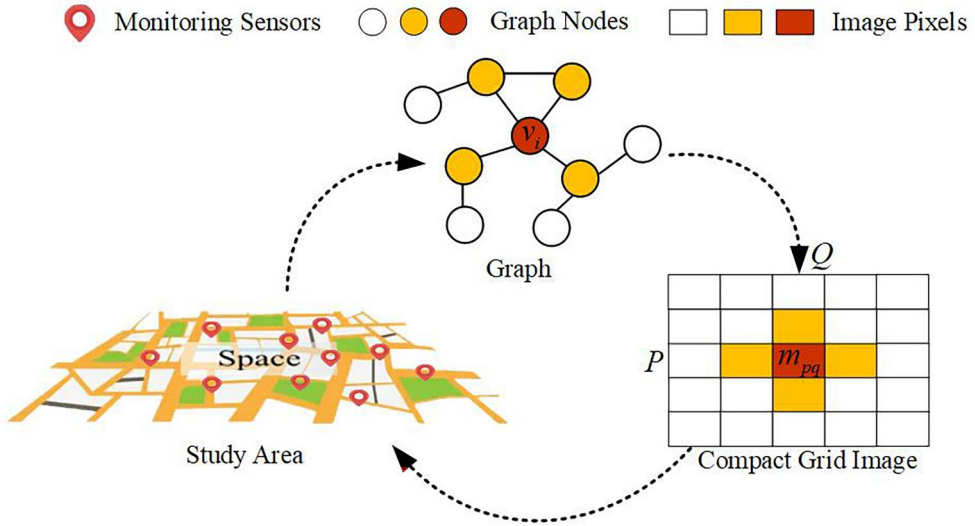


Figure 2. Data structure transformation: the closer the distance between two pixels, the stronger the dependency between the graph nodes.

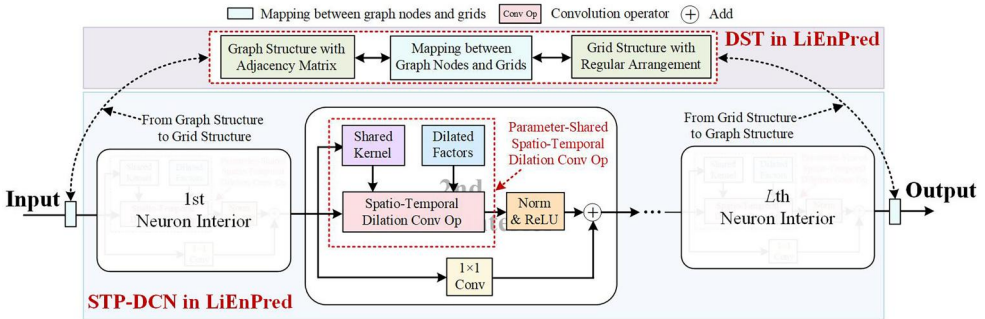


Figure 3. Workflow of the LiEnPred.

4. Methodology

In this section, we provide a detailed overview of the proposed LiEnPred model. As shown in Figure 3, the LiEnPred model consists of two main modules: the Data Structure Transformation (DST) module and the Parameter-Shared Spatio-Temporal Dilation Convolution Network (STP-DCN) module. Among them, the DST module is responsible for mapping the relationships between graph nodes and compact grid pixels, and the STP-DCN module is designed to mine the spatio-temporal dependencies between compact grid pixels. Specifically, the DST module first converts air quality data from graph structures into compact grid structures, serving as input for the STP-DCN module. Then, the STP-DCN module establishes the relationship between the input and output data within compact grid structure to complete the prediction. Finally, the DST module converts the grid structures back to graph structures for displaying the prediction results. Notably, in the STP-DCN module, L neurons share the

same set of convolutional kernels and use dilation factors to reduce both computation time and parameter scale.

4.1. Construction of the LiEnPred

4.1.1. Data structure transformation

In large-scale scenarios, achieving both accuracy and speed in air quality predictions remains challenging. For instance, although grid structures can be used for efficient predictions in large-scale scenarios, the sparse distribution of sensors can significantly impact the prediction accuracy of grid-based models. On the other hand, graph structures can accurately capture complex relationships between sensors, but the large number of graph nodes can significantly impact computational efficiency of graph-based models. To address these challenges, we propose the DST module, which creates a compact grid structure incorporating graph node connections, facilitating air quality prediction in large-scale scenarios.

To streamline the DST module, we define the compact grid image as a square structure, specifying its length and width as outlined in Equation (2).

$$\begin{cases} P = \min_p (p^2 \geq N) \\ Q = \min_q (q^2 \geq N) \end{cases} \quad (2)$$

where P and Q represent the length and width of the compact grid image, respectively; N denotes the total number of graph nodes; \min is the minimum function used to find the smallest values of $p^2 \geq N$ and $q^2 \geq N$. **Note:** The compact grid image will be fully filled with graph nodes only when $p^2 = P$ and $q^2 = Q$; otherwise, a few pixels will remain unfilled.

After defining the length and width of the compact grid image, we can establish a mapping between the graph nodes and the grid pixels based on the graph structure. In real-world scenarios, the graph structure can be represented either by an unweighted adjacency matrix, a weighted distance matrix, or even a hybrid matrix

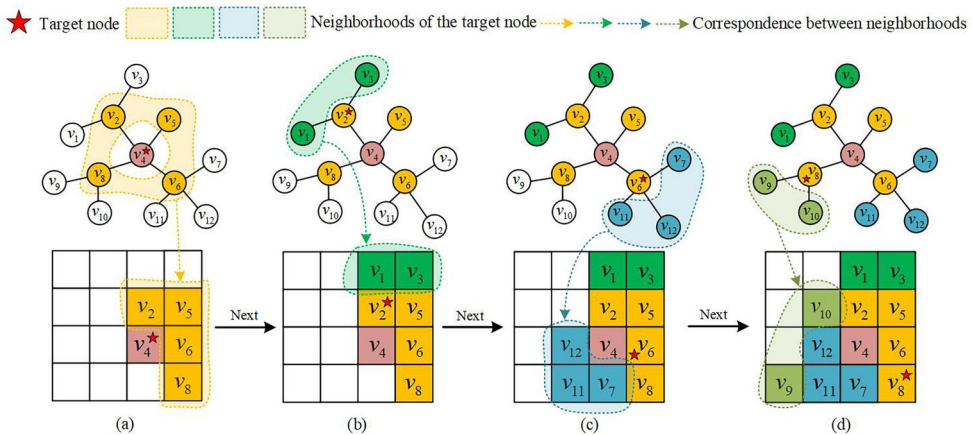


Figure 4. Generation of compact grid image: (a) neighbors of v_4 , (b) neighbors of v_2 , (c) neighbors of v_6 , and (d) neighbors of v_8 .

combining both distance and adjacency. Each matrix is used to measure the correlation or similarity between graph nodes. Therefore, we established this mapping using the graph structure matrix, as shown in Algorithm 1. In line 2, we identify two unmapped graph nodes with the highest similarity and select one of them as v_i . In line 6, we determine the graph node most similar to v_i among the already mapped nodes and position v_i within the neighborhood of this most similar graph node. In line 12, we determine the coordinates of v_j 's neighbors in a clockwise manner, such as v_4 and its neighbors $v_2, v_5, v_6,$ and v_8 in Figure 4(a). If the first-order eight neighboring pixels around v_j cannot accommodate additional graph nodes, we will search a suitable position from the second-order sixteen neighboring pixels of v_j , such as v_6 and its neighbors v_{11} and v_{12} in Figure 4(c).

Algorithm 1. Process of DST module

Input: Number of graph nodes: N
 Matrix for storing graph node relationships: A
 Length of the compact grid image: P
 Width of the compact grid image: Q
Output: mapping relations: Ω

- 1: **while** existence of unmapped graph nodes **do**
- 2: obtain one unmapped graph nodes v_i via highest similarity
- 3: **if** all graph nodes are unmapped:
- 4: obtain central position $(p_i, q_i) = (\text{int}(P/2), \text{int}(Q/2))$ of compact grid image
- 5: **else:**
- 6: obtain mapping position (p_i, q_i) by similarity of mapped nodes to v_i
- 7: add v_i into *Seeds* and add (i, p_i, q_i) into Ω
- 8: **while** existence of graph nodes in *Seeds* **do**
- 9: pop a graph node v_j from the *Seeds*
- 10: search unmapped neighbor nodes $\{v_s\}$ of v_j from the matrix A
- 11: get mapping position (p_j, q_j) of v_j from Ω
- 12: get multiple mapping positions $\{(p_s, q_s)\}$ of $\{v_s\}$ with (p_j, q_j) as the center
- 13: add collection $\{v_s\}$ into *Seeds* and add collection $\{(s, p_s, q_s)\}$ into Ω
- 14: **output** mapping relations Ω

Figure 5 presents three examples of small graphs to clearly illustrate the process of graph structure transformation. These examples emphasize the core concept of the DST module: clustering a graph node and its neighboring nodes as closely as possible within a grid structure. This arrangement enables convolutional kernels to detect all neighboring nodes of a target graph node within a limited receptive field, allowing them to effectively model the spatial relationships between the target graph node and its neighboring nodes. Even if neighboring pixels are not directly adjacent in the graph structure, the model can optimize the corresponding weights in the convolutional kernel to 0, thereby diminishing the contribution of those neighboring pixels to the target node. Furthermore, we observe that the process of the DST module closely resembles that of the breadth-first search (BFS) algorithm, as both methods involve ‘first visiting all neighboring nodes of a specific node’. This principle serves as the foundation for ‘ensuring that a target graph node and its neighbors are as close as possible’.

After applying the DST module, we obtain a compact grid image $\mathbf{M} \in R^{P \times Q}$, where $m_{pq} > 0$ indicates that the pixel at p -th row and q -th column contains a graph node, $m_{pq} = 0$ indicates that it does not. The compact grid image not only addresses the sparse issue of monitoring sensors in large-scale scenarios, but also preserves the

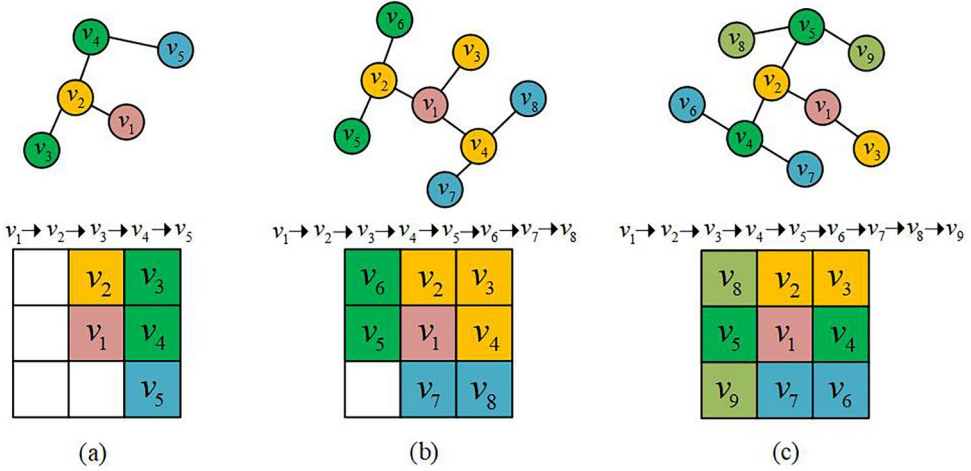


Figure 5. Example of graph structure transformation for a small graph: (a) five graph nodes, (b) eight graph nodes, and (c) nine graph nodes.

dependencies among graph nodes. Specifically, target graph nodes and their neighbors often display clustered distributions within the compact grid, making the modeling of adjacent pixels analogous to modeling adjacent graph nodes.

4.1.2. Parameter-shared spatio-temporal dilation convolution network

After applying the DST module, air quality data over T time windows will be composed into a three-dimensional tensor $\mathcal{X} \in R^{P \times Q \times T}$. In large-scale scenarios with high real-time requirements, it is crucial to quickly and accurately mine spatiotemporal dependencies from the tensor \mathcal{X} . However, most existing prediction models struggle to balance prediction accuracy with computational efficiency. High-accuracy models often suffer from overly complex designs and inefficient computations, making them inadequate for real-time applications. To address these challenges, we propose a lightweight parameter-shared spatio-temporal dilation convolution network (STP-DCN).

As illustrated in Figure 6, the STP-DCN module extends the classic 3D convolutional network. Unlike traditional 3D convolutional networks, our approach enforces a temporal dependency constraint, ensuring that the convolutional kernel only processes time windows prior to the current moment. Additionally, we incorporate dilation factors to significantly reduce the network depth while preserving prediction accuracy. Finally, by sharing convolutional kernels across different hidden layers, we ensure that the model's parameter size remains constant despite an increase in network depth.

Compared to classic 3D convolutional networks, the lightweight STP-DCN module offers three main advantages. First, it adheres to the basic constraints of temporal order in the time dimension, making its convolution operations more reasonable than those in classic 3D convolutional networks. Second, by controlling the depth of the neural network through dilation factors, the STP-DCN module achieves higher computational efficiency. Lastly, the STP-DCN module shares convolutional kernels between different layers, reducing the parameter size compared to classic 3D convolutional networks. Taking the three-dimensional tensor $\mathcal{X}_{t-B+1}^t \in R^{P \times Q \times B}$ as an example,

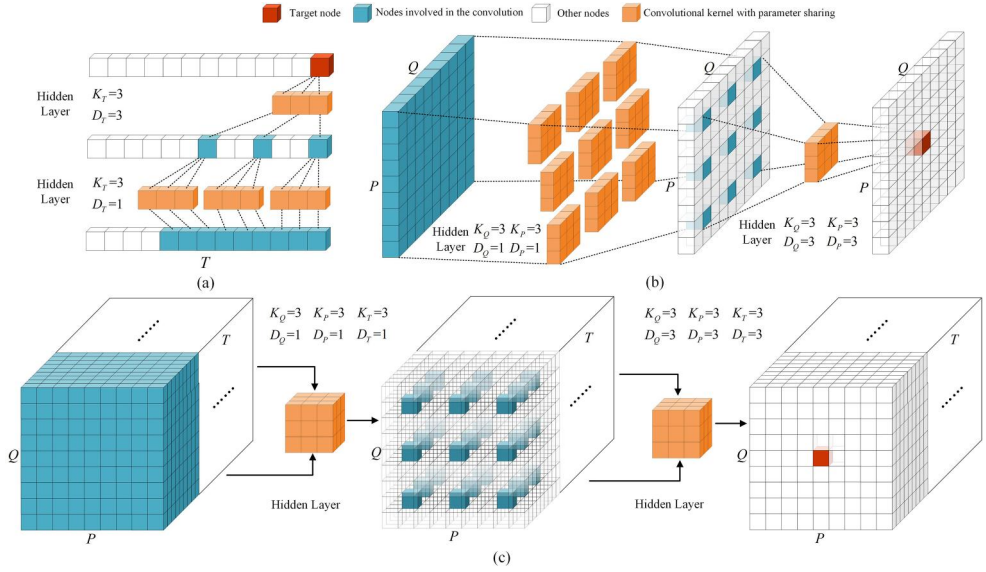


Figure 6. Forward propagation of parameter-shared spatio-temporal dilation convolution: (a) temporal dimension, (b) spatial dimension, and (c) spatio-temporal dimension.

Equations (3) and (4) illustrate the forward propagation process of the STP-DCN module.

$$STPDCN = \begin{cases} \{\mathcal{H}^{e:1}\}_{e=1}^E = STDCP\left(\text{Conv}(\mathcal{X}_{t-B+1}^t) | D_Q^1, D_P^1, D_T^1, \{\mathcal{W}^e\}_{e=1}^E\right) & l = 1 \\ \{\mathcal{H}^{e:l}\}_{e=1}^E = STDCP\left(\{\mathcal{H}^{e:l-1}\}_{e=1}^E | D_Q^l, D_P^l, D_T^l, \{\mathcal{W}^e\}_{e=1}^E\right) & 1 < l < L \\ \hat{\mathcal{X}}_{t+1}^{t+F} = \text{Conv}\left(STDCP\left(\{\mathcal{H}^{e:L-1}\}_{e=1}^E | D_Q^L, D_P^L, D_T^L, \{\mathcal{W}^e\}_{e=1}^E\right)\right) & l = L \end{cases} \quad (3)$$

$$STDCP\left(\{\mathcal{H}^{e:l-1}\}_{e=1}^E | D_Q^l, D_P^l, D_T^l, \{\mathcal{W}^e\}_{e=1}^E\right) = \begin{cases} h_{pqt}^{e:l} \in \mathcal{H}^{e:l} \\ h_{pqt}^{e:l} = \sum_{i=1}^{K_P} \sum_{j=1}^{K_Q} \sum_{s=1}^{K_T} h_{p'q't'}^{e:l} * w_{ijs}^e \\ p' = p + \frac{(2i - K_P - 1)D_P^l}{2} \\ q' = q + \frac{(2j - K_Q - 1)D_Q^l}{2} \\ t' = t - (K_T - s)D_T^l \end{cases} \quad (4)$$

where $\mathcal{X}_{t-B+1}^t \in \mathbb{R}^{P \times Q \times B}$ represents the historical air quality data; $\hat{\mathcal{X}}_{t+1}^{t+F} \in \mathbb{R}^{P \times Q \times F}$ represents the predicted air quality data; D_Q^l, D_P^l , and D_T^l denote the dilation factors in the Q -dimension, P -dimension, and T -dimension, respectively, at the l -th hidden layer; $\mathcal{W}^e \in \mathbb{R}^{K_P \times K_Q \times K_T}$ represents the shared convolutional kernels in the hidden layer, with K_Q , K_P and K_T indicating the kernel sizes in the Q -dimension, P -dimension, and T -dimension, respectively; $\{\mathcal{H}^{e:l}\}_{e=1}^E \in \mathbb{R}^{P \times Q \times T \times E}$ represents the hidden states at the l -th hidden layer, with E indicating the dimension of the hidden layer and the number of convolutional kernels; $\mathcal{H}^{e:l} \in \mathbb{R}^{P \times Q \times F}$ denotes the computed three-dimensional tensor using the e -th convolutional kernel; $h_{pqt}^{e:l} \in \mathbb{R}^{1 \times 1}$ represents an element in the three-dimensional tensor \mathcal{H}_i^e , with p , q , and t representing coordinates in the Q -dimension,

P -dimension, and T -dimension, respectively; $w_{ijs}^e \in R^{1 \times 1}$ is an element in the three-dimensional tensor, with i , j , and s as coordinates in the Q -dimension, P -dimension, and T -dimension, respectively; and $Conv$ denotes a standard convolution operation used for dimensional alignment.

4.2. Optimization of the LiEnPred

During the forward propagation process, the LiEnPred model predicts future air quality data $\{\hat{\mathbf{X}}^{t+f}\}_{f=1}^F$ using historical air quality data $\{\mathbf{X}^{t-b+1}\}_{b=1}^B$. Typically, a trained prediction model is achieved by minimizing the squared loss between the ground truth $\{\mathbf{X}^{t+f}\}_{f=1}^F$ and prediction values $\{\hat{\mathbf{X}}^{t+f}\}_{f=1}^F$. However, as the compact grid matrix \mathbf{M} may not be fully populated by graph nodes, we only incorporate the loss from specific positions into the loss function, as outlined in Equation (5).

$$\mathcal{L}(\mathcal{W}) = \min_{\mathcal{W}} \sum_{b=1}^B \sum_{p=1}^P \sum_{q=1}^Q (x_{pq(t-b)} - \hat{x}_{pq(t-b)} | m_{pq} > 0)^2 \quad (5)$$

where $x_{pq(t-b)} \in \mathcal{R}^{1 \times 1}$ represents the ground truth from sensor m_{pq} within $(t-b)$ -th time window; $\hat{x}_{pq(t-b)} \in \mathcal{R}^{1 \times 1}$ represents the prediction value from sensor m_{pq} within $(t-b)$ -th time window; B is historical dependency step; $(x_{pq(t-b)} - \hat{x}_{pq(t-b)} | m_{pq} > 0)$ indicates that the loss $x_{pq(t-b)} - \hat{x}_{pq(t-b)}$ is computed only if $m_{pq} > 0$; \mathcal{W} denotes the learnable parameters in the LiEnPred model.

5. Experiments

5.1. Air quality datasets

5.1.1. Data sources

In this study, we evaluate the prediction performance of the LiEnPred model using real large-scale air quality data. As illustrated in Figure 7, the data is sourced from 2014 monitoring sensors across China, covering seven indicators: PM_{2.5}, CO, NO₂, PM₁₀, SO₂, O₃, and AQI. All indicators are sampled at 60-minute intervals, with the data covering the period from January 1, 2021, to December 31, 2023.

5.1.2. Data characteristics and preprocessing

As shown in Figure 8, images generated with traditional grid partitioning methods are extremely sparse, with sparsity increasing as the spatial resolution becomes finer. For instance, when $Q = P = 50$, approximately 85.16% of the grid pixels lack monitoring sensors, whereas at $Q = P = 200$, 97.96% of the grid pixels are without sensors. In contrast, the compact grid image has a sparsity rate of just 0.54% (11/2025), significantly enhancing pixel utilization and improving the efficiency of subsequent modeling. **Note:** The generation of compact grid images leverages the underlying graph structure among monitoring sensors. In practical applications, users can define more complex graph structures as needed. For the purposes of this study, we simplify the graph by connecting each node to four adjacent nodes, representing the four sensors most similar to the target sensor.

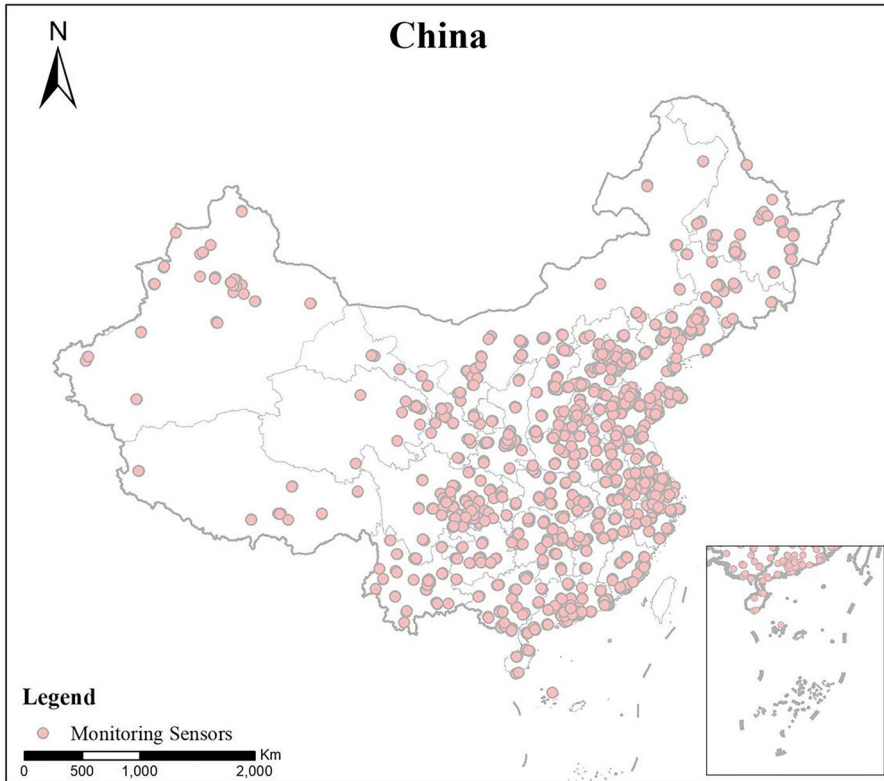


Figure 7. Spatial distribution of air quality monitoring sensors: It is IJGIS policy to remain strictly neutral with respect to jurisdictional claims on disputed territories in published maps, and the naming conventions used in maps are left to the discretion of authors.

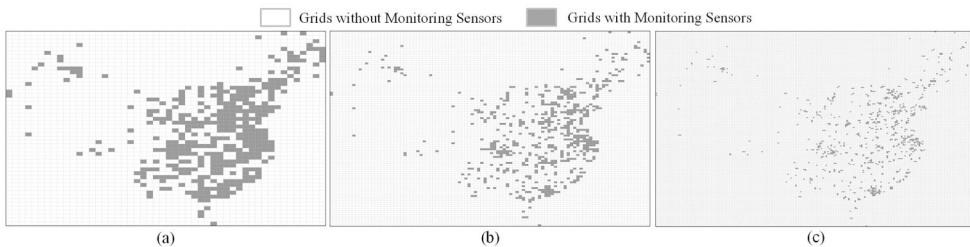


Figure 8. Proportion of grids without monitoring sensors: (a) 85.16% at $Q = P = 50$, (b) 94.27% at $Q = P = 100$, and (c) 97.96% at $Q = P = 200$.

Table 2 displays the distribution characteristics of the air quality data. The results reveal that the data includes not only outliers (such as extreme values) but also varying degrees of missing data. To support this study, we treat data between the 1st and 99th percentiles as valid and use the HA and IDW algorithms to re-estimate missing and outlier values in the original dataset. In addition, although we open-sourced seven types of air quality data, we verified the performance of the LiEnPred model exclusively using $PM_{2.5}$ and NO_2 concentrations, given its general applicability for air quality prediction.

Table 2. Spatial distribution of air quality monitoring sensors.

Data type	Min	P1	P5	P50	P95	P99	Max	Missing rate
PM _{2.5}	1	2	5	23	91	158	200000	17.96%
O ₃	1	2	6	59	145	190	1200	17.68%
CO	0.1	0.1	0.3	0.6	1.3	2.1	90	17.54%
NO ₂	1	2	5	18	63	89	1028	17.31%
PM ₁₀	1	5	11	44	156	284	100000	17.82%
SO ₂	1	1	2	7	20	40	1284	17.15%
AQI	1	12	18	48	132	234	500	19.78%

P1, P5, P50, P95, and P99 represent the 1st, 5th, 50th, 95th, and 99th percentiles of the data, respectively.

5.2. Evaluation metrics

For large-scale air quality prediction models, accuracy, computational efficiency, and parameter scale are crucial performance metrics. In this study, we use Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) to assess the prediction accuracy of the proposed model. The runtime of forward and backward propagation evaluates computational efficiency, while the number of learnable parameters measures parameter scale. As the calculations for runtime and parameter number are relatively straightforward, we provide detailed calculation methods only for the accuracy metrics, as outlined in Equations (6) and (7).

$$RMSE = \sqrt{\frac{1}{F * P * Q} \sum_{f=1}^F \sum_{p=1}^P \sum_{q=1}^Q (x_{pq(t+f)} - \hat{x}_{pq(t+f)})^2 | m_{pq} > 0} \quad (6)$$

$$MAPE = \frac{100\%}{F * P * Q} \sum_{f=1}^F \sum_{p=1}^P \sum_{q=1}^Q \left| \frac{x_{pq(t+f)} - \hat{x}_{pq(t+f)}}{\hat{x}_{pq(t+f)}} \right| | m_{pq} > 0 \quad (7)$$

where $x_{pq(t+f)}$ represents the ground truth from sensor m_{pq} in $(t + f)$ -th time window; $\hat{x}_{pq(t+f)}$ represents the prediction value from sensor m_{pq} in $(t + f)$ -th time window; F is prediction step; P denotes the height of the compact grid image; Q denotes the width of the compact grid image. Similar to Equation (5), evaluation metrics for the prediction model are calculated only when $m_{pq} > 0$.

5.3. Settings

In this subsection, we describe the experimental environment (hardware and software environment) and hyperparameter setting information.

5.3.1. Environment settings

In this study, the spatiotemporal data is processed on a PC (Intel(R) Core(TM) i7-11700 CPU @ 2.50 GHz, memory: 48.0GB). Moreover, we built our model based on PyTorch and Python3.7 on a Graphics Processing Unit (GPU) platform with 24GB of GPU memory.

5.3.2. Hyper-parameter setting

In this subsection, we present the calibration process of the LiEnPred model using $PM_{2.5}$ concentration data. The hyperparameters of the LiEnPred model include the historical dependency step B , the number of neurons L , the number of convolution kernels E , and the kernel sizes (K_p, K_Q, K_T) in the P -dimension, Q -dimension, and T -dimension, respectively, as well as the dilation factors (D_p, D_Q, D_T) for these dimensions.

In the time dimension, we apply the concept of hyper-parameter dependency from Wang *et al.* (2024b) to determine the optimal combination of hyper-parameters. First, we set the historical dependency step $B=8$ and the kernel size $K_T = 2$ in T -dimension. Then, we determined the number of neurons $L = 3$ via the formula $L = \min_l (K_T^{l-1} \geq B) - 1$. Finally, the dilation factors for the three neurons were set to $D_T^1 = 1$, $D_T^2 = 2$, and $D_T^3 = 4$.

In the spatial dimension, we determine the kernel size and dilation factor based on the coordinates (positions in the compact grid image) of the target graph nodes and their graph neighbors. As shown in Figure 9(a), most target graph nodes differ from their graph neighbors by 1–4 index in the grid structure. Therefore, we only need to set K_p and K_Q to 3 to simulate the process of graph convolution for most graph nodes in the grid structure, and the corresponding dilation factors for the three neurons are $D_p^1 = D_Q^1 = 1$, $D_p^2 = D_Q^2 = 3$, and $D_p^3 = D_Q^3 = 9$.

For the number of convolution kernels E , we used the control variable method to determine the optimal value. As shown in Figure 9(b), the model's prediction accuracy initially increases with the number of convolution kernels and then stabilizes. Considering that computational efficiency also improves with more kernels, we ultimately set E to 64.

5.4. Comparison with baselines

In this study, we compare the proposed LiEnPred model with two categories of baselines. The first category is grid-based spatiotemporal prediction models, including the

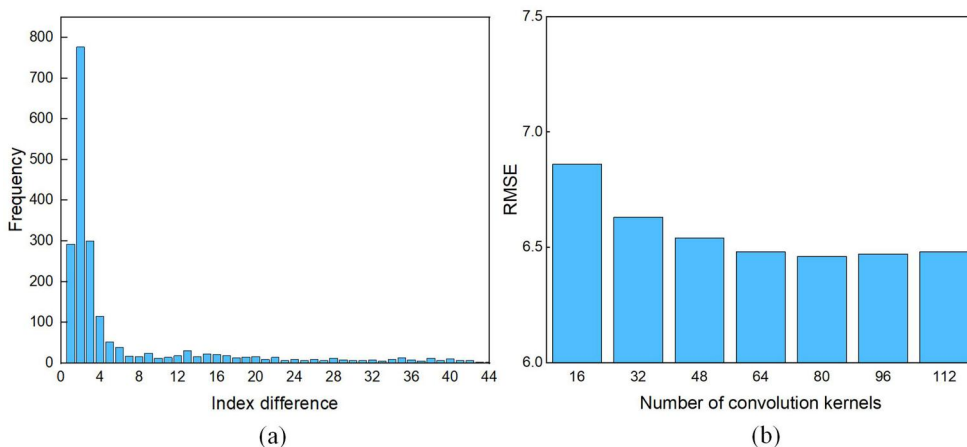


Figure 9. Determination of hyper-parameters on $PM_{2.5}$ concentration data: (a) index difference between target graph node and its adjacent graph nodes under grid structure, and (b) prediction accuracy varies with the number of convolution kernels.

ConvLSTM model (Shi *et al.* 2017), the ST-3DNet model (Guo *et al.* 2019b), and the ST-3DMDDN model (He *et al.* 2024). The second category is graph-based spatiotemporal prediction models, including the ST-GCN model (Yu *et al.* 2018), the ASTGCN model (Guo *et al.* 2019a), the DSTAGNN model (Lan *et al.* 2022), the GDGCN model (Xu *et al.* 2023), and the STGDN model (Wang *et al.* 2024a).

5.4.1. Comparison of prediction precision

To fairly compare the prediction accuracy of grid-based and graph-based models, we address the correspondence between pixels and sensors in a sparse image. Specifically, we divided the study area into a 200×200 grid, ensuring that each pixel corresponded to a single sensor (If multiple sensors corresponded to the same pixel, we reassigned them to neighboring pixels). Table 3 shows the prediction accuracies of the LiEnPred model and baselines on the $PM_{2.5}$ dataset and NO_2 dataset. The results demonstrate that graph-based models outperform grid-based models in prediction accuracy. The reason is that graph-based models explicitly define the spatial dependencies between sensors, enhancing their predictive capabilities. In contrast, grid-based models struggle to capture these dependencies in a sparse grid, resulting in reduced prediction accuracy. Moreover, the prediction accuracy of the LiEnPred model not only surpasses that of the first category of models but also matches or exceeds the performance of the second category in multi-step predictions. The reason is that although the LiEnPred model is a grid-based model, it embeds graph structure within its grid structure, enabling it to approximate the prediction accuracy of graph-based models. The results also indicate that the prediction accuracy of the LiEnPred model is slightly lower than that of the STGDN model in single-step predictions. This is because the LiEnPred model struggles to capture dynamic spatiotemporal correlations within its static grid structure, an area that requires further improvement in the future. Overall, the prediction accuracy of the LiEnPred model is superior to or comparable with the baselines, excluding considerations of operational efficiency and parameter scale. Additionally, Figure 10 illustrates the stability of the LiEnPred model, showing that it maintains relatively consistent prediction accuracy across different random seeds, further validating its ability to match the accuracy of current state-of-the-art models.

5.4.2. Comparison of computational efficiency

In this subsection, we further analyze the computational efficiency of the LiEnPred model. As there is no significant difference in its efficiency between the $PM_{2.5}$ and

Table 3. Evaluation metrics (RMSE/MAPE) of LiEnPred and baselines.

Models	$PM_{2.5}$ concentration data		NO_2 concentration data	
	1-step	7-steps	1-step	7-steps
ConvLSTM	6.75/24.17%	12.68/54.37%	6.61/26.59%	11.65/56.52%
ST-3DNet	6.63/21.30%	12.28/49.29%	6.49/24.34%	11.31/55.40%
ST-3DMDDN	6.64/21.54%	12.14/48.84%	6.53/24.76%	11.34/55.78%
ST-GCN	6.65/23.17%	12.67/52.39%	6.41/23.42%	11.57/57.64%
ASTGCN	6.51/22.68%	12.14/49.31%	6.28/23.19%	10.94/53.34%
DSTAGNN	6.48/22.04%	12.08/48.38%	6.24/22.94%	10.45/49.87%
GDGCN	6.50/21.17%	11.84/48.14%	6.18/21.79%	10.34/48.91%
STGDN	6.44/21.12%	11.80/47.97%	6.15/20.87%	10.27/46.74%
LiEnPred	6.48/20.84%	11.79/47.88%	6.18/20.62%	10.23/46.21%

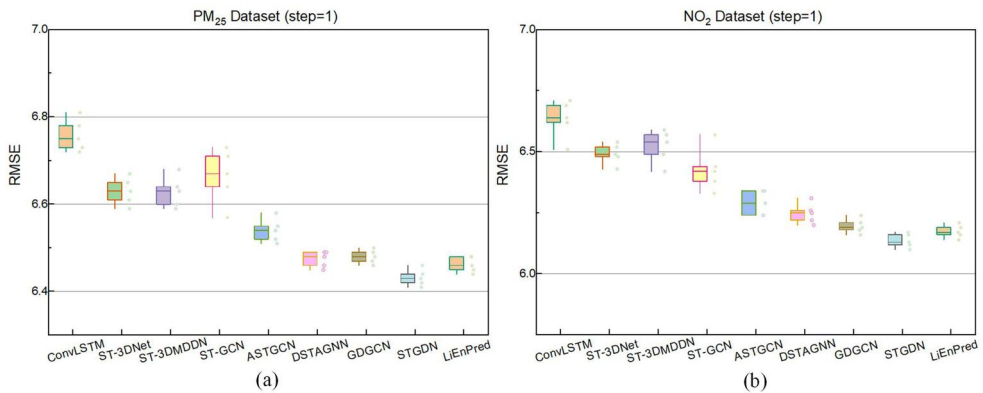


Figure 10. Stability of the LiEnPred model and baselines using different random seeds: (a) one-step prediction on $PM_{2.5}$ concentration data, and (b) one-step prediction on NO_2 concentration data.

Table 4. Running time (milliseconds) of LiEnPred and baselines with batch size being 8.

Models	Time for one inference	Time for one optimization
ConvLSTM	8.79 ± 1.24	37.91 ± 2.18
ST-3DNet	68.56 ± 7.13	99.64 ± 9.12
ST-3DMDDN	69.13 ± 8.42	108.24 ± 11.38
ST-GCN	150.86 ± 14.04	357.63 ± 41.16
ASTGCN	285.54 ± 28.34	303.43 ± 31.16
DSTAGNN	310.24 ± 32.16	340.15 ± 33.18
GDGCN	261.15 ± 27.34	618.35 ± 54.72
STGDN	183.30 ± 16.35	459.59 ± 38.29
LiEnPred	70.98 ± 8.43	102.19 ± 6.71

NO_2 datasets, we present only the results for the $PM_{2.5}$ dataset, as shown in Table 4. Among them, the time for one inference reflects the online prediction speed of the LiEnPred model, while the time for one optimization indicates the offline training speed. The results indicate that with 2,014 graph nodes, the LiEnPred model achieves a 2 to 4 times faster computation time compared to graph-based models. The reason is that the LiEnPred model is essentially grid-based model, resulting in significantly lower computational complexity compared to the graph-based models. Figure 11 presents a scatter plot of computational efficiency versus prediction accuracy. In this plot, points closer to the center indicate superior balance between computational efficiency and prediction accuracy. The results demonstrate that the LiEnPred model effectively balances prediction accuracy and computational efficiency. Additionally, we assessed the difference in computational efficiency between the LiEnPred model and the lightweight STGDN using simulated datasets, with the results illustrated in Figure 12. The findings indicate that the computational efficiency of the LiEnPred model is comparable to that of the lightweight STGDN model for small graph structures (fewer than 512 nodes). However, significant differences in their computational efficiencies emerge for larger graph structures, particularly when the number of nodes exceeds 1,024. This result demonstrates that LiEnPred is well-suited for prediction tasks in large-scale scenarios.

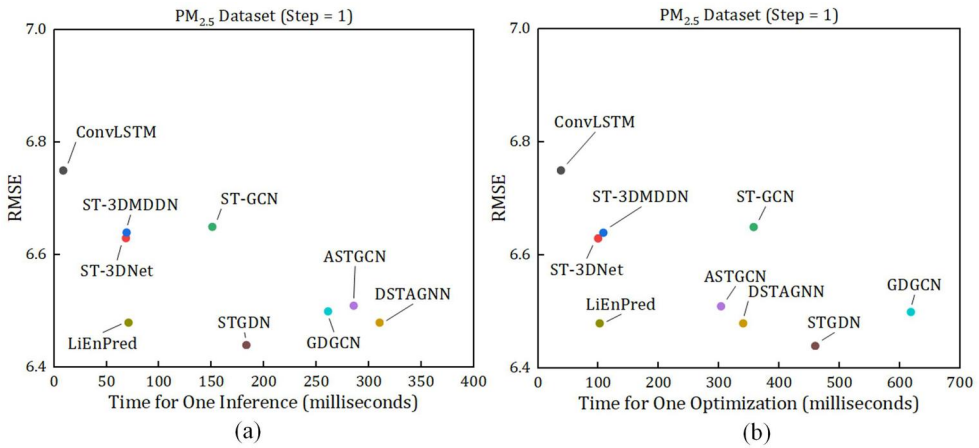


Figure 11. Prediction accuracy vs. computational efficiency on PM_{2.5} concentration data: (a) time for one inference, and (b) time for one optimization.

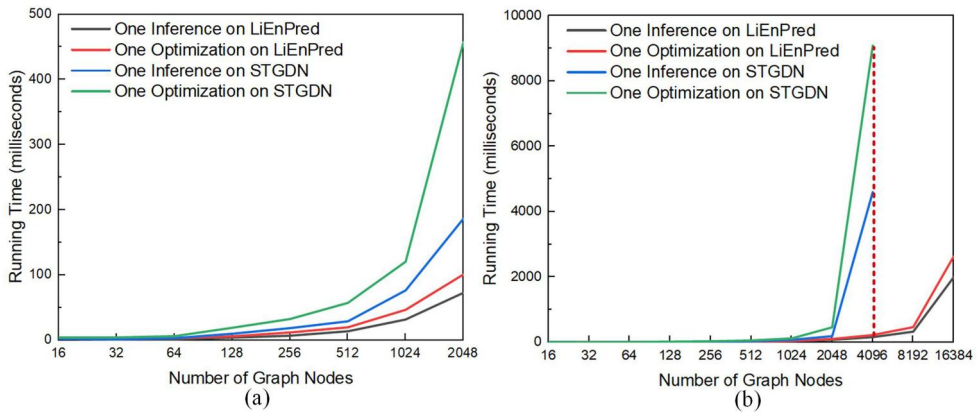


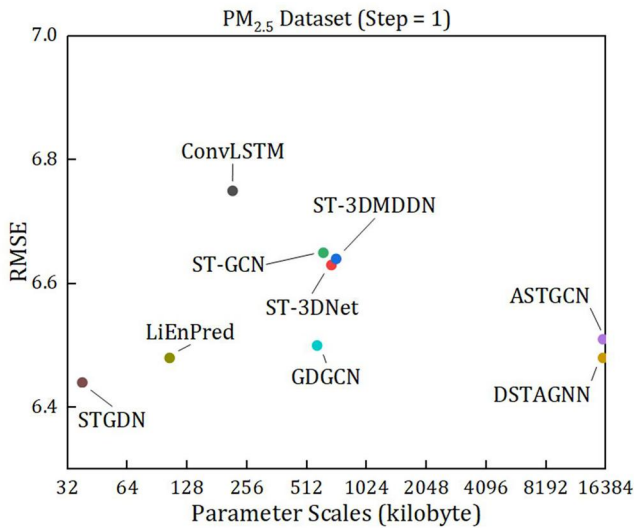
Figure 12. Computational efficiency comparison between STGDN and LiEnPred on simulated datasets: (a) when the number of graph nodes is fewer than 2048, and (b) when the number of graph nodes exceeds 2048.

5.4.3. Comparison of parameter scale

Following the analysis of computational efficiency, we assessed the parameter scales of the learnable weights in the LiEnPred model for the PM_{2.5} dataset, as shown in Table 5. The results indicate that LiEnPred significantly outperforms the ConvLSTM, ST-3DNet, ST-3DMDDN, ASTGCN, DSTAGNN and GDGCN models in terms of parameter scales. Specifically, the parameter scale of the LiEnPred model is only one-seventh to one-half that of existing grid-based models and one-fifteenth to one-fifth that of existing graph-based models. Moreover, compared to the lightweight STGDN model, LiEnPred does not introduce a significant number of additional learnable parameters. The LiEnPred model's advantage in parameter size stems from its use of parameter sharing, which effectively reduces the overall number of model parameters. Additionally, Figure 13 displays a scatter plot of parameter size versus prediction accuracy. In this plot, points closer to the center represent better prediction accuracy

Table 5. Parameter scale (kilobyte) of learnable weights for LiEnPred and baselines.

Models	Parameter scale
ConvLSTM	216.56
ST-3DNet	680.63
ST-3DMDDN	718.92
ST-GCN	620.00
ASTGCN	15849.42
DSTAGNN	15854.54
GDGCN	576.23
STGDN	37.89
LiEnPred	104.25

**Figure 13.** Prediction accuracy vs. parameter scale of learnable weights on $PM_{2.5}$ concentration data.

and lower parameter scales. The results show that LiEnPred model achieves acceptable prediction accuracy while maintaining a relatively low parameter scale.

5.4.4. Summary of comparative results

First, the LiEnPred model is fundamentally a grid-based prediction model that improves upon the prediction accuracy of existing grid-based models, achieving prediction accuracy that can be superior to or comparable with that of graph-based models. Second, the LiEnPred model demonstrates remarkable computational efficiency, operating 2 to 4 times faster than graph-based models in a scenario with 2014 graph nodes, with this efficiency advantage becoming even more pronounced when the number of graph nodes exceeds 2000. Finally, the LiEnPred model has fewer learnable parameters compared to baselines, possessing only one-seventh to one-half of the parameters found in existing grid-based models and one-fifteenth to one-fifth of those in current graph-based models. Overall, experimental results demonstrate that the proposed LiEnPred model matches or exceeds the predictive accuracy of eight baselines while offering faster computational efficiency and requiring fewer

parameters. Additionally, the findings indicate that the LiEnPred model is highly suitable for air quality prediction tasks in large-scale sparse environments.

5.5. Spatio-temporal visualization of prediction results

In this subsection, lines and maps are used to qualitatively assess the prediction performance of the LiEnPred model. Figure 14(a,c) illustrates the temporal differences between prediction values and ground truth, while Figure 14(b,d) shows the spatial differences. The results demonstrate that the LiEnPred model generally predicts the trend of air quality data accurately over time and the spatial distribution effectively. Additionally, we identified two scenarios where the model's prediction accuracy decreases. The first occurs when the trend of air quality data changes abruptly within a short period, as indicated by the blue areas in Figure 14(a,c). The second situation involves regions with severe air pollution, highlighted by the blue areas

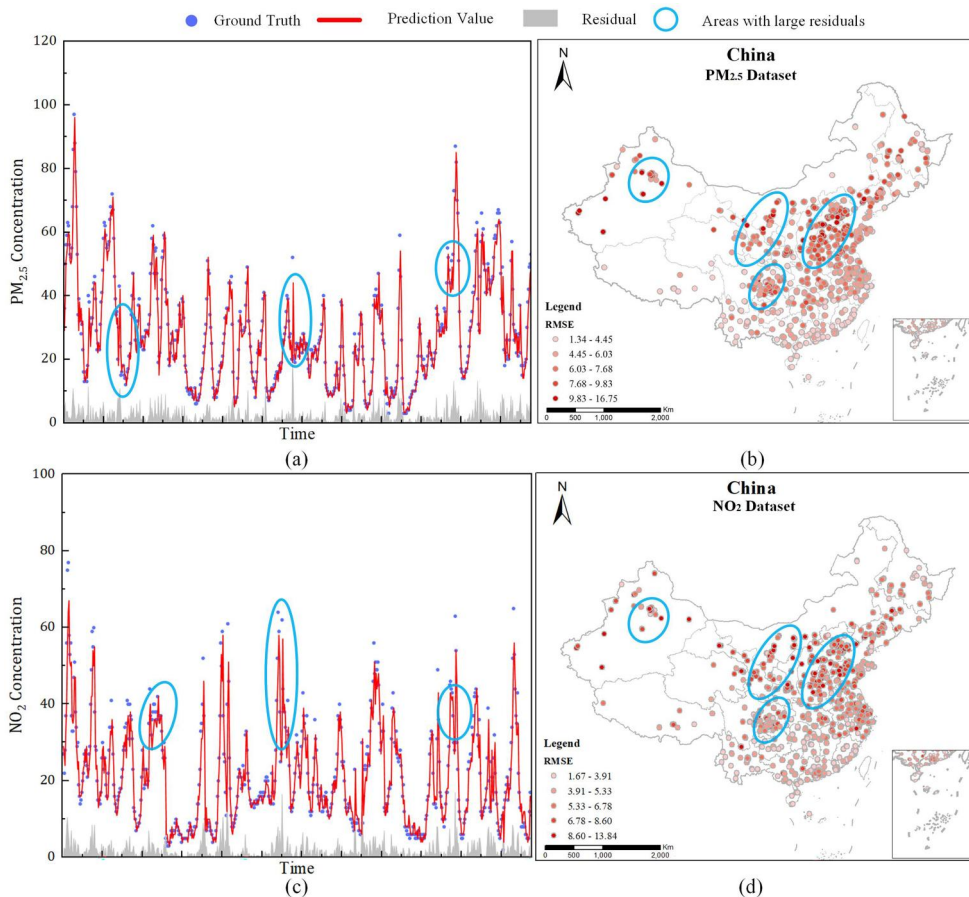


Figure 14. Prediction results of the LiEnPred model: (a) temporal dimension of the PM_{2.5} dataset, (b) spatial dimension of the PM_{2.5} dataset, (c) temporal dimension of the NO_{2.5} dataset, and (d) spatial dimension of the NO_{2.5} dataset. It is IJGIS policy to remain strictly neutral with respect to jurisdictional claims on disputed territories in published maps, and the naming conventions used in maps are left to the discretion of authors.

Table 6. Impact of design on prediction performance in LiEnPred model.

Designs		Accuracy (PM _{2.5} NO ₂ : RMSE/MAPE)	Efficiency (inference/optimization)	Scale (kilobyte)
Compact Grid Image	Y	6.48/20.84% 6.18/20.62%	171.1/572.6	–
	N	6.64/21.61% 6.32/22.18%	1124.11/3124.92	–
Dilatated Factor	Y	–	171.1/572.6	–
	N	–	508.2/1328.8	–
Parameter Sharing	Y	6.48/20.84% 6.18/20.62%	–	104.25
	N	6.41/20.47% 6.12/20.13%	–	356.44

Y represents 'Yes' and N represents 'No'.

in Figure 14(b,d). Severely polluted areas are primarily concentrated in North China, closely tied to the region's advanced industry, dry climate, and extensive transportation infrastructure. Energy-intensive sectors like steel and coal emit large quantities of pollutants. The dry climate and low wind speeds trap these emissions, limiting their dispersion. Furthermore, heavy vehicle traffic adds to the concentration of pollutants, exacerbating the issue. To predict air quality accurately in these contexts, it is essential to incorporate external variables that influence air quality changes, such as meteorological factors, seasonal variations, and industrial activities.

5.6. Ablation study

In this section, we analyze the effects of various design choices in the LiEnPred model on prediction performance, with results shown in Table 6. In the DST module, the compact grid image enhances prediction accuracy by leveraging implicit graph information, and improves computational efficiency by reducing the number of pixels. In the STP-DCN module, the dilatated factor enhances computational efficiency, while parameter sharing decreases the model's parameter scale. Although parameter sharing may slightly reduce prediction accuracy, the significant reduction in parameter size justifies this trade-off. Additionally, users can decide whether to implement parameter sharing depending on the scenario. Overall, Table 6 confirms the effectiveness of these design choices.

6. Conclusions

Accurate and efficient air quality prediction is crucial for public health protection and environmental sustainability. While numerous grid-based and graph-based prediction models have been developed, they encounter challenges in large-scale scenarios: (1) Grid-based models, though computationally efficient, have limited prediction accuracy under large-scale sparse distributions; (2) Graph-based models, despite higher prediction accuracy, suffer from significant computational inefficiencies when dealing with a large number of sensors, i.e. graph nodes

To address the above issues, we proposed a novel LiEnPred model for efficient air quality prediction in large-scale scenarios. More specifically, we presented a data structure transformation algorithm and generated a compact grid structure with an embedded graph structure, combining the high prediction accuracy of graph-based models with the high computational efficiency of grid-based models. In addition, we presented a lightweight parameter-shared spatio-temporal dilation convolution

network that efficiently captures spatio-temporal dependencies in air quality data without significantly increasing computation time or the parameter scale. In our experiments, we collected air quality data from over 2000 sensors across China over the past three years and evaluated LiEnPred's prediction performance in large-scale scenarios using PM_{2.5} and NO₂ concentration data. The experimental results demonstrated that LiEnPred offered several distinct advantages. First, the proposed LiEnPred model significantly enhanced prediction accuracy in large-scale sparse scenarios compared to classical grid-based models. Moreover, the proposed LiEnPred model achieved comparable accuracy to graph-based models while reducing parameter scales and increasing computational efficiency. Overall, we introduced new methods for spatio-temporal prediction in large-scale sparse scenarios. More specifically, the proposed LiEnPred model excels as a versatile spatio-temporal prediction model for large-scale sparse scenarios, particularly in applications that demand high training and inference speeds.

This study has limitations: First, the proposed LiEnPred model completes the prediction task using a static grid structure, limiting its ability to capture dynamic spatio-temporal relationships. In future work, we aim to integrate a lightweight attention mechanism to enhance the prediction accuracy of the LiEnPred model while maintaining its high computational efficiency. Second, the LiEnPred model was primarily validated on the PM_{2.5} dataset, not other datasets. Future works will aim to evaluate the model's performance on a broader range of datasets and assess its applicability to various applications.

Acknowledgments

The numerical calculations in this article have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This project was supported by National Key Research and Development Program of China [Grant No. 2021YFB3900803]; National Natural Science Foundation of China [Grant No. 42401524]; China National Postdoctoral Support Program for Innovative Talents [Grant No. BX20230360]; China Postdoctoral Science Foundation [Grant No. 2023M743454]; National Natural Science Foundation of Shandong Province [Grant No. ZR2024QD179]; Open Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University [Grant No. 23103]; Innovation Project of LREIS [Grant No. 08R8A092YA].

Notes on contributors

Peixiao Wang is a Postdoctoral Fellow from State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Nature Resources Research, Chinese Academy of Sciences. He received Ph.D. degree under from State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, and received the M.S. degree from The Academy of Digital China, Fuzhou University.

His research topics include spatiotemporal data mining, and spatiotemporal prediction, especially focus on spatiotemporal prediction of transportation systems.

Hengcai Zhang is an Associate Professor of State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. He received his Ph.D. degree from Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. He is the member of the Theory and Methodology Committee of the Chinese Association of Geographic Information System, and member of Chinese Branch of ACM SIGSPATIAL. His interests focus on spatial-temporal data mining and 3D-Computing.

Jie Liu received the Ph.D. degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2023. She is currently a Lecturer with University of Jinan. Her research interests include spatiotemporal prediction and physic-informed deep learning.

Feng Lu is a Professor at the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests cover trajectory data mining, computational transportation science and location-based services.

Tong Zhang is a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. He received the M.Eng. degree in cartography and geographic information system (GIS) from Wuhan University, Wuhan, China, in 2003, and the Ph.D. degree in geography from San Diego State University, and the University of California at Santa Barbara in 2007. His research topics include urban computing and machine learning.

ORCID

Peixiao Wang  <http://orcid.org/0000-0002-1209-6340>

Hengcai Zhang  <http://orcid.org/0000-0002-5004-9609>

Jie Liu  <http://orcid.org/0000-0001-7715-142X>

Feng Lu  <http://orcid.org/0000-0001-6573-2550>

Tong Zhang  <http://orcid.org/0000-0002-0683-4669>

Data and codes availability statement

The data and codes supporting this study are available at <https://doi.org/10.6084/m9.figshare.26771938>.

References

- Aburas, M.M., Ahamad, M.S.S., and Omar, N.Q., 2019. Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: A review. *Environmental Monitoring and Assessment*, 191 (4), 205.
- Amato, F., et al., 2020. A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific Reports*, 10 (1), 22243.
- Asif, M.T., et al., 2014. Spatiotemporal patterns in large-scale traffic speed prediction. *IEEE Transactions on Intelligent Transportation Systems*, 15(2), 794–804.
- Bao, H., Dong, L., and Wei, F., 2021. BEiT: BERT pre-training of image transformers. arXiv: 2106.08254 [Cs]. <http://arxiv.org/abs/2106.08254>
- Bi, K., et al., 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619 (7970), 533–538.

- Casagli, N., et al., 2023. Landslide detection, monitoring and prediction with remote-sensing techniques. *Nature Reviews Earth & Environment*, 4 (1), 51–64.
- Cheng, S., et al., 2021. Emission characteristics and control scenario analysis of VOCs from heavy-duty diesel trucks. *Journal of Environmental Management*, 293, 112915.
- Cheng, S., Peng, P., and Lu, F., 2020. A lightweight ensemble spatiotemporal interpolation model for geospatial data. *International Journal of Geographical Information Science*, 34 (9), 1849–1872.
- Cheng, S., et al., 2023. Health and economic benefits of heavy-duty diesel truck emission control policies in Beijing. *Environment International*, 179, 108152.
- Corner, R.J., Dewan, A.M., and Chakma, S., 2014. Monitoring and prediction of land-use and land-cover (LULC) change. In: A. Dewan and R. Corner, eds., *Dhaka megacity: geospatial perspectives on urbanisation, environment and health*. Springer Netherlands, 75–97. <http://www.springer.com/series/10180>
- Dai, X., Liu, G.-P., and Hu, W., 2023. An online-learning-enabled self-attention-based model for ultra-short-term wind power forecasting. *Energy*, 272, 127173.
- Do, L.N.N., et al., 2019. An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 108, 12–28.
- Guan, Q., et al., 2024. Predicting short-term PM_{2.5} concentrations at fine temporal resolutions using a multi-branch temporal graph convolutional neural network. *International Journal of Geographical Information Science*, 38 (4), 778–801.
- Guo, R., et al., 2023. A deep spatio-temporal learning network for continuous citywide air quality forecast based on dense monitoring data. *Journal of Cleaner Production*, 414, 137568.
- Guo, S., et al., 2019a. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33 (01), 922–929.
- Guo, S., et al., 2019b. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 20 (10), 3913–3926.
- He, K., et al., 2021. Masked autoencoders are scalable vision learners. arXiv:2111.06377 [Cs]. <http://arxiv.org/abs/2111.06377>
- He, R., et al., 2024. Deep spatio-temporal 3D dilated dense neural network for traffic flow prediction. *Expert Systems with Applications*, 237, 121394.
- Huang, G., et al., 2021. Large scale air pollution prediction with deep convolutional networks. *Science China Information Sciences*, 64 (9), 192107.
- Ibrahim, M.R., Haworth, J., and Cheng, T., 2020. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities*, 96, 102481.
- Kipf, T.N., and Welling, M., 2017, February 22. Semi-supervised classification with graph convolutional networks. In: *International conference on learning representations*. <http://arxiv.org/abs/1609.02907>
- Lan, S., et al., 2022. DSTAGNN: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In: *Proceedings of the 39th international conference on machine learning*, 11906–11917. <https://proceedings.mlr.press/v162/lan22a.html>
- Le, V.-D., Bui, T.-C., and Cha, S.-K., 2020. Spatiotemporal deep learning model for citywide air pollution interpolation and prediction. In: *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 55–62.
- Li, G., et al., 2023. A lightweight and accurate spatial-temporal transformer for traffic forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 35 (11), 10967–10980.
- Li, X., et al., 2022. Short-term forecast of bicycle usage in bike sharing systems: a spatial-temporal memory network. *IEEE Transactions on Intelligent Transportation Systems*, 23 (8), 10923–10934.
- Liang, Y., et al., 2023. AirFormer: predicting nationwide air quality in China with transformers. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37 (12), 14329–14337.
- Liu, Y., et al., 2024. RT-GCN: Gaussian-based spatiotemporal graph convolutional network for robust traffic prediction. *Information Fusion*, 102, 102078.
- Mengfan, T., et al., 2022. Including the feature of appropriate adjacent sites improves the PM_{2.5} concentration prediction with long short-term memory neural network model. *Sustainable Cities and Society*, 76, 103427.

- Noursalehi, P., Koutsopoulos, H.N., and Zhao, J., 2022. Dynamic origin-destination prediction in urban rail systems: a multi-resolution spatio-temporal deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 23 (6), 5106–5115.
- Oprea, S., et al., 2022. A review on deep learning techniques for video prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (6), 2806–2826.
- Rossi, E., et al., 2020. Temporal graph networks for deep learning on dynamic graphs. arXiv: 2006.10637 [Cs, Stat]. <http://arxiv.org/abs/2006.10637>
- Saad, F., et al., 2024. Scalable spatiotemporal prediction with Bayesian neural fields. *Nature Communications*, 15 (1), 7942.
- Schlichtkrull, M., et al., 2017. Modeling relational data with graph convolutional networks. arXiv: 1703.06103 [Cs, Stat]. <http://arxiv.org/abs/1703.06103>
- Shi, X., et al., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems – Volume 1*, 802–810.
- Shi, X., et al., 2017. Deep learning for precipitation nowcasting: a benchmark and a new model. In: *Advances in neural information processing systems*, vol. 30. Red Hook, NY: Curran Associates, Inc.
- Veličković, P., et al., 2018. Graph attention networks (arXiv:1710.10903). arXiv.
- Voulodimos, A., et al., 2018. Deep learning for computer vision: a brief review. *Computational Intelligence and Neuroscience*, 2018, 7068349.
- Wang, P., et al., 2024a. A lightweight spatiotemporal graph dilated convolutional network for urban sensor state prediction. *Sustainable Cities and Society*, 101, 105105.
- Wang, P., et al., 2024b. Adding attention to the neural ordinary differential equation for spatio-temporal prediction. *International Journal of Geographical Information Science*, 38 (1), 156–181.
- Wang, P., et al., 2022. A multi-view bidirectional spatiotemporal graph network for urban traffic flow imputation. *International Journal of Geographical Information Science*, 36 (6), 1231–1257.
- Wang, P., et al., 2023. Urban traffic flow prediction: a dynamic temporal graph network considering missing values. *International Journal of Geographical Information Science*, 37 (4), 885–912.
- Xu, M., et al., 2021. Spatial-temporal transformer networks for traffic flow forecasting. arXiv: 2001.02908 [Cs, Eess]. <http://arxiv.org/abs/2001.02908>
- Xu, Y., et al., 2023. Generic dynamic graph convolutional network for traffic flow forecasting. *Information Fusion*, 100, 101946.
- Yang, K., et al., 2020. Human activities and the natural environment have induced changes in the PM_{2.5} concentrations in Yunnan Province, China, over the past 19 years. *Environmental Pollution (Barking, Essex: 1987)*, 265 (Pt B), 114878.
- Yu, B., Yin, H., and Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 3634–3640.
- Zhang, J., Zheng, Y., and Qi, D., 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31 (1), 1–7.
- Zhang, L., et al., 2021. Spatiotemporal causal convolutional network for forecasting hourly PM_{2.5} concentrations in Beijing, China. *Computers & Geosciences*, 155, 104869.
- Zhang, S., et al., 2020. Network-wide traffic speed forecasting: 3D convolutional neural network with ensemble empirical mode decomposition. *Computer-Aided Civil and Infrastructure Engineering*, 35 (10), 1132–1147.
- Zhang, T., et al., 2024. Prior-guided gated convolutional networks for rainstorm forecasting. *Journal of Hydrology*, 633, 130962.
- Zhang, Y., et al., 2023. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619 (7970), 526–532.
- Zhang, Z., and Zhang, S., 2023. Modeling air quality PM_{2.5} forecasting using deep sparse attention-based transformer networks. *International Journal of Environmental Science and Technology*, 20 (12), 13535–13550.

- Zhao, L., *et al.*, 2020. T-GCN: a temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21 (9), 3848–3858.
- Zheng, H., *et al.*, 2021. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 22 (11), 6910–6920.