Knowledge-Based Systems 263 (2023) 110275

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A graph-attention based spatial-temporal learning framework for tourism demand forecasting

Binggui Zhou [a,b], Yunxuan Dong [a,b], Guanghua Yang [a,c,d,*], Fen Hou [b], Zheng Hu [e], Suxiu Xu [f], Shaodan Ma [b]

[a] School of Intelligent Systems Science and Engineering, Jinan University, Zhuhai 519070, China
[b] State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macau, 999078, Macao Special Administrative Region of China
[c] GBA and B&R International Joint Research Center for Smart Logistics, Jinan University, Zhuhai, 519070, China
[d] Institute of Physical Internet, Jinan University, Zhuhai 519070, China
[e] State Key Laboratory of Network and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[f] School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:
Received 16 March 2022
Received in revised form 23 October 2022
Accepted 4 January 2023
Available online 10 January 2023

Keywords:
Tourism demand forecasting
Dynamic spatial connections
Spatial-temporal learning
Graph neural network
Attention mechanism

ABSTRACT

Accurate tourism demand forecasting can improve tourism experiences and realize smart tourism. Existing spatial–temporal tourism demand forecasting models only explore pre-specified and static spatial connections across regions without considering multiple or dynamic spatial connections; however, this is not sufficient for modeling actual tourism demand. In this paper, we propose a graph-attention based spatial–temporal learning framework for tourism demand forecasting. A weight-dynamic multi-dimensional graph is organized to embed multiple explicit dynamic spatial connections and provide a node attribute sequence for learning implicit dynamic spatial connections. We further propose a heterogeneous spatial–temporal graph-attention network (called HSTGANet), which is effective in handling both explicit and implicit dynamic spatial connections, learning high-dimensional spatial–temporal features, and forecasting tourism demand. Experimental results demonstrate the effectiveness of the proposed model over baseline models in forecasting the tourism demand for six regions of Wanshan Archipelago in Zhuhai, China, and indicate that the proposed spatial–temporal learning framework may provide useful insights for developing more effective models for other spatial–temporal forecasting problems.

1. Introduction

The tourism industry has grown enormously worldwide and generated unparalleled economic and social benefits over recent decades. Accurate tourism demand forecasting can help tourists arrange their schedules properly and help tourism practitioners to appropriately allocate tourism resources; thus, this forecasting plays an increasingly important role in improving tourism experiences and realizing smart tourism.

Tourism demand forecasting problems are usually formulated as time series forecasting problems [1], spatial modeling problems [2–4], or spatial–temporal forecasting problems [5–7]. Time series forecasting and spatial modeling ignore spatial information or neglect important temporal information, respectively, which leads to limited forecasting accuracy. Spatial–temporal forecasting comprehensively attends to both the spatial and temporal information of tourism demand and is regarded as the most reasonable problem formulation of this type of forecasting. One spatial–temporal forecasting model can forecast the tourism demand for many regions under a specified spatial structure [7], since spatial information is aggregated into this model to provide accurate across-region tourism demand forecasts. However, existing spatial–temporal tourism demand forecasting models only consider pre-specified and static spatial connections across regions, rather than considering multiple and dynamic spatial connections.

To date, research has investigated various forecasting models, including statistical (e.g., autoregressive integrated moving average (ARIMA) [1] and seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) [8] models), traditional machine learning (e.g., support vector machines [9] and random forests [10]), and deep learning models (e.g., long short-term memory (LSTM) [11], bidirectional LSTM (BiLSTM) [11], Bayesian BiLSTM [12], and attention-based models [13–15]). Some studies have also reported hybrid forecasters to take advantage of different forecasting models [16,17]. Existing machine

* Corresponding author at: School of Intelligent Systems Science and Engineering, Jinan University, Zhuhai 519070, China.
E-mail address: ghyang@jnu.edu.cn (G. Yang).

https://doi.org/10.1016/j.knosys.2023.110275
0950-7051/

learning based studies, and most statistical models, formulate tourism demand forecasting purely using time series forecasting, without considering spatial interactions across regions.

Recently, spatial–temporal forecasting using deep learning models has been well studied for a variety of forecasting objectives, such as traffic flow [18–24], traffic speed [24–28], travel demand [29–32], and ambulance demand [33]. This demonstrates that deep learning models have a higher capacity to learn complex spatial–temporal features from data and have the potential to conduct accurate spatial–temporal forecasting. Among these models, graph neural networks, a type of neural network that captures the dependencies of graphs via message passing between graph nodes [34], and multi-head attention networks, a type of neural network that captures long-range dependencies within sequences via a multi-head attention mechanism [35], have shown particular effectiveness in exploring spatial–temporal features and achieved improved forecasting results compared with other neural models. To the best of our knowledge, no study has yet investigated spatial–temporal tourism demand forecasting with deep learning models, especially using graph neural networks and multi-head attention networks.

Considering the growth of accessible spatial–temporal data and the superiority of spatial–temporal forecasting in aggregating spatial information for better forecasting performance, we model a tourism demand forecasting problem as a spatial–temporal forecasting problem. Further, considering the limitations of existing spatial–temporal tourism demand forecasting models, and inspired by the recent successful works of graph neural networks and multi-head attention networks in spatial–temporal forecasting, we propose a graph-attention based spatial–temporal learning framework. The proposed framework consists of two parts. To better represent the spatial–temporal information of tourism demand, we organize the spatial–temporal data of a scenic area with several regions as a weight-dynamic multi-dimensional graph with multiple nodes and multiple dimensions of edges. To better learn useful spatial–temporal features, we propose a heterogeneous spatial–temporal graph-attention network, called HSTGANet, to extract spatial–temporal features and simultaneously forecast the tourism demand for all scenic regions.

The main contributions of this work are summarized as follows:

1. This work proposes a novel representation method to organize spatial–temporal tourism demand information, which can be utilized for spatial–temporal tourism demand forecasting.
2. The proposed HSTGANet can handle both explicit and implicit dynamic spatial connections, learn high-dimensional spatial–temporal features, and conduct accurate tourism demand forecasting. Our experimental results demonstrate the superiority of the proposed HSTGANet compared with baseline models.
3. To the best of our knowledge, this is the first work that combines graph convolutions, to model explicit dynamic spatial connections, and self-attention, to model implicit dynamic spatial connections, for general spatial–temporal forecasting, in addition to tourism demand forecasting. Hence, this work may provide useful insights for developing more effective models for other spatial–temporal forecasting problems.

The remainder of this paper is organized as follows. In Section 2, we review existing research works related to tourism demand forecasting from both the problem formulation perspective and methodology perspective. In Section 3, we introduce the proposed spatial–temporal information representation method. In Section 4, we present the heterogeneous spatial–temporal graph-attention network. In Section 5, we introduce the details of experiments, including dataset description, experimental settings, baselines, and experimental results. Finally, we conclude this work in Section 6.

2. Related works

By considering different theoretical assumptions and data availability, tourism demand forecasting can be formulated as a time series forecasting, spatial modeling (cross-sectional analysis), or spatial–temporal forecasting problem.

A time series is a type of sequential data where the data points are spaced in chronological order. In contrast, spatial data are related via distances, spatial arrangements, and other geographical information [36]. Spatial data are characterized by spatial dependence and heterogeneity [37], which reflect the spatial autocorrelation between and uniqueness of different regions, respectively. Another well-investigated spatial effect is spatial spillover, which represents the externalities generated by neighboring tourism destinations [38]. Generally, two kinds of spatial data have been studied in tourism literature: cross-sectional and panel data. Cross-sectional data are collected by observing many regions at the same time point or over the same period of time [6]. Panel data are observed by extending cross-sectional data with a time dimension [6,7,39], which incorporates both spatial and temporal information.

(1) Time Series Forecasting. Tourism demand forecasting can be regarded as a time series forecasting problem when either all available data are time series or spatial information is discarded. Specifically, if only the forecast variable is considered, the forecasting problem is regarded as univariate time series forecasting; when other explanatory variables are considered, the problem becomes a multivariate time series forecasting problem. In tourism literature, many studies have treated tourism demand forecasting as a time series forecasting problem. For example, [1] proposed two models, the seasonal autoregressive integrated moving average (SARIMA) model and autoregressive integrated moving average with exogenous factors (ARIMAX) model, to forecast airport passenger traffic for Hong Kong. The SARIMA model is a univariate time series forecasting model, while the ARIMAX model takes into account other identified endogenous or exogenous variables, e.g., originating and connecting traffic and visitors by air transport. Recently, the increase of model capability due to the development of deep learning and neural networks has enabled more studies on multivariate tourism demand forecasting. In these studies, time series forecasting models extract features from historical observations of more variables, such as weather and holiday [13], search engine data [11,40], news [41], and even social network data [42], to make forecasts for the future.

(2) Spatial Modeling. Spatial modeling, or cross-sectional analysis, used to be a widely applied method to forecast tourism demand; it estimates the spatial interaction across geographic areas and forecasts future values based on this estimated interaction. Such spatial interaction commonly covers spatial dependence, spatial heterogeneity, and other spatial effects (e.g., spatial spillover). Among these models, gravity models, which are based on the gravity law of spatial interaction, are the most widely applied models in tourism demand forecasting [2–4]. Recently, however, spatial models have been neglected in tourism demand forecasting since they do not pay enough attention to the importance of temporal information.

(3) Spatial-Temporal Forecasting. To further improve the performance of tourism demand forecasting, several studies have investigated panel data and spatial–temporal learning, considering that both spatial and temporal features of tourism demand are

important for forecasting. [5] proposed a set of panel data models for short-term domestic tourism demand forecasts for 341 cities in China, demonstrating superiority over a pooled ordinary least squares (OLS) model and naive baseline models. [6] proposed a panel model which incorporated both spatial dependence and spatial heterogeneity to improve forecasting accuracy. [7] proposed several spatial–temporal models, including dynamic spatial panel models and space–time autoregressive moving average (STARMA) models, to forecast inbound tourism demand in 29 Chinese provincial regions.

The methodology of tourism demand forecasting models can be generally divided into three categories: statistical, machine learning (including traditional machine and deep learning models), and hybrid models.

(1) Statistical Models. There are many statistical models that model tourism demand forecasting using time series forecasting; these models first model a time series using a parametric model and then forecast future values of the forecast variable. Some basic statistical models, such as autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models, assume that the time series is stationary. However, it is well recognized that tourism demand is usually seasonal and non-stationary; therefore, the basic statistical models that were used in tourism demand forecasting in the past are now almost no longer used for this problem. To tackle the non-stationary and seasonal nature of tourism demand, some studies have applied ARIMA [43], SARIMA [1] and other models to tourism demand forecasting and achieved improved performances. In addition, to incorporate variables other than the forecast variable, ARX [44] and SARIMAX models [8] have also been applied in some tourism demand forecasting studies. Some statistical models realize tourism demand forecasting using spatial or spatial–temporal modeling, such as gravity [2–4], panel data [5,7], and space–time autoregressive moving average (STARMA) models [6,7]. Such models suppose that spatial effects are indispensable to improving forecasting accuracy.

(2) Machine Learning Models. Recently, machine learning based models have drawn increasing attention in tourism demand forecasting due to the development of machine and deep learning. Compared with statistical models, machine learning models have a higher model capacity and the ability to incorporate large numbers of variables, which means that they can be applied to more practical tourism demand forecasting problems. Machine learning models can be divided into traditional machine learning models and deep learning models. Traditional machine learning models, such as support vector machines [9], random forests [10], and Bayesian models [45], have been reported to be superior to statistical models. Deep learning models have had a large impact on many research fields, such as computer vision (CV) and natural language processing (NLP), in recent years. In the tourism demand forecasting field, deep learning models have shown great efficiency in extracting complex patterns from various input data, yielding outstanding performance enhancement compared to other models. Such models include long short-term memory (LSTM) [11], Bayesian BiLSTM [12], and attention-based models [13–15].

(3) Hybrid Models. In addition to statistical and machine learning models, some hybrid models, which combine statistical and machine learning models and some advanced operations (e.g., decomposition), have been proposed. In [16], an adaptive hybrid forecaster, which combined complete ensemble empirical mode decomposition (CEEMD) with a model library consisting of an artificial neural network (ANN), SARIMAX, LSTM, and SVR, was proposed to forecast Macao's hotel demand and estimate the impact of the Coronavirus disease (COVID-19) on hotel demand. In [17], 26 stacking models of 11 single models (including multilayer perceptron (MLP), seasonal and trend decomposition using Loess (STL), SARIMA, etc.) were examined in a forecasting framework to forecast the inbound tourism demand across 20 countries. In [46], STL was combined with a duo attention deep learning model (DADLM) for the accurate decomposition of input data forecasting of monthly tourist arrivals in Hong Kong. By combining such approaches, these hybrid forecasters can provide more accurate forecasting results than single models, but may also increase the complexity of the forecasting framework.

3. Spatial-temporal information representation

Considering that both spatial and temporal domain information are critical for tourism demand forecasting, we consider a spatial–temporal forecasting problem in this work.

To the best of our knowledge, there are only limited works focusing on spatial–temporal tourism demand forecasting. Most existing works either use panel data models [5,7] or spatial–temporal autoregressive models [6,7].

A panel data model for spatial–temporal tourism demand forecasting can be formulated as

$$\hat{\mathbf{y}}_t = \lambda \mathbf{W}_s \mathbf{y}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\hat{\mathbf{y}}_t$ is a vector of tourism demand forecasts at time t for all regions, $\mathbf{y}_{t-1}$ is a vector of tourism demand at $t-1$ for all regions, λ is a coefficient, $\mathbf{W}_s \in \mathbb{R}^{N\times N}$ is a spatial weighting matrix that defines the spatial connection across N regions, $\boldsymbol{\mu}$ is a vector representing region-specific effects, and $\boldsymbol{\varepsilon}$ is a vector of the error term. In [7], three different spatial weighting matrices were proposed: a contingency-based, distance-based, and flow-based matrix, to pre-specify the spatial weighting matrix $\mathbf{W}_s$.

A spatial–temporal autoregressive model extends a traditional autoregressive model by incorporating spatial lags into the autoregressive and moving average terms of the model, which can be formulated as

$$\hat{\mathbf{y}}_t = \sum_{i=1}^{p}\sum_{j=1}^{m} \varphi_{ij} \mathbf{W}_s^{(j)} \mathbf{y}_{t-i} + \sum_{k=1}^{q}\sum_{l=1}^{n} \theta_{kl} \mathbf{W}_s^{(l)} \boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \tag{2}$$

where p and m are the maximum orders of the autoregressive term in time and space, respectively, q and n are the maximum orders of the moving average term in time and space, respectively, φ and θ are coefficients, and $\mathbf{W}_s^{(j)}$ and $\mathbf{W}_s^{(l)}$ are the jth order and lth order of the spatial weighting matrix, respectively. In [7], the three spatial weighting matrices (the contingency-based, distance-based, and flow-based matrices) were applied to build spatial–temporal autoregressive models. In [6], a k-nearest-neighbor based spatial weighting matrix was proposed.

Although these models consider the spatial connections across regions and incorporate them into the modeling, the proposed spatial weighting matrices are pre-specified and static, and do not consider multiple or dynamic spatial connections across regions. Considering this limitation, we propose spatial–temporal information representation while considering multiple and dynamic spatial connections.

We observe that although some spatial connections across regions can be explicitly inferred by their direct interactions (e.g., geographical distances, interchanging tourist flows), some spatial connections can only be implicitly inferred by the explanatory variables related to each region (e.g., local tourist volume, weather) since some spatial effects (e.g., spatial spillover) cannot be explicitly represented by direct interactions. Therefore, both the direct interactions and explanatory variables related to each region are considered here to form our spatial–temporal information representation.

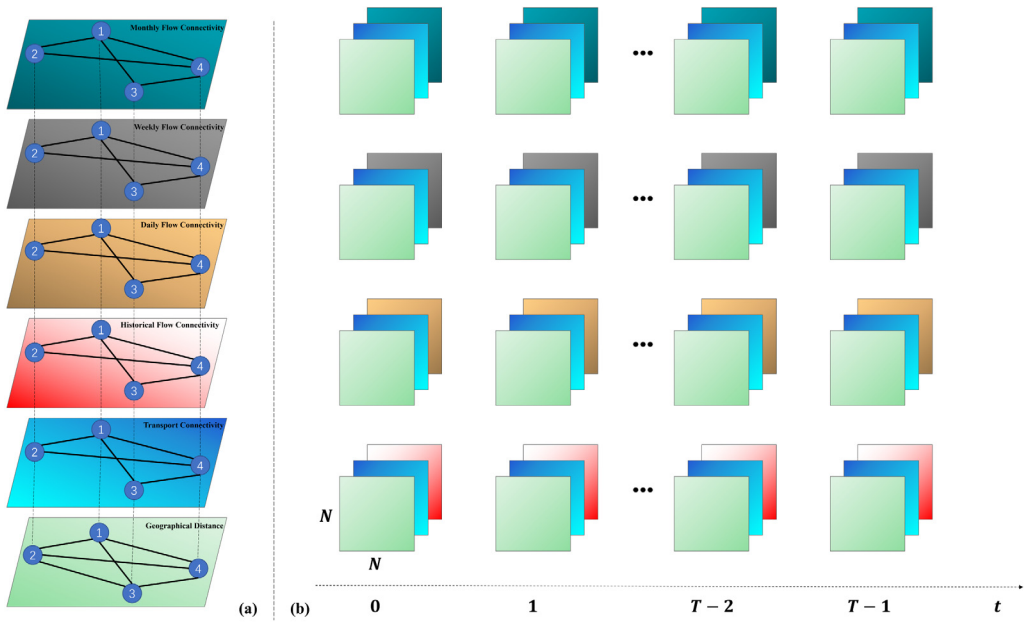


Fig. 1. Proposed weight-dynamic multi-dimensional graph and corresponding weighted adjacency matrices for each dimension: (a) graph topology; and (b) 4 sets of weighted adjacency matrices for learning historical and periodic spatial connections. N is the number of regions and T is the temporal length of the sequence data.

Consider a scenic area $\mathbb{S} = \{s_1, s_2, \ldots, s_N\}$ with N regions (as shown in Fig. 1(a)). To represent explicit dynamic spatial interactions, a weight-dynamic multi-dimensional graph, denoted by $\mathcal{G} = (\mathbb{V}, \mathbb{E})$, is formed. $\mathcal{G}$ consists of one node set $\mathbb{V} = \{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_N\}$ representing multiple regions, and d disjoint subsets, $\mathbb{E} = \{\mathbb{E}_1, \mathbb{E}_2, \ldots, \mathbb{E}_d\}$, representing multiple explicit interactions across regions.

The nodes in the multi-dimensional graph $\mathcal{G}$ are associated with historical x^{htd} and periodic (including daily x^{dtd}, weekly x^{wtd}, and monthly x_{mtd}) tourism demands[1] and a set of explanatory variables (including holiday x^{hol}, weather x^{wea}, and search index information x^{si}). Note that the usage of periodic observations (including the periodic observations of tourism demand and undermentioned tourist flow) is based on the observation that some spatial connections might be periodic due to human behaviors [47], climates [48], etc. Therefore, learning periodic spatial connections can help with tourism demand forecasting.

Considering both the spatial and temporal dimensions, a spatial–temporal sequence $\mathbf{X}$ is expressed as

$$\begin{aligned}\mathbf{X} &\in \mathbb{R}^{T\times N\times c}\\ &= \mathbf{X}^{\text{hol}} \cup \mathbf{X}^{\text{wea}} \cup \mathbf{X}^{\text{si}}\\ &\quad \cup \mathbf{X}^{\text{htd}} \cup \mathbf{X}^{\text{dtd}} \cup \mathbf{X}^{\text{wtd}} \cup \mathbf{X}^{\text{mtd}},\end{aligned} \tag{3}$$

where c is the number of channels of $\mathbf{X}$ (here $c = 7$), N is the number of regions, and T is the temporal length of the sequence data.

A weighted adjacency matrix is given for each dimension in $\mathcal{G}$ for further graph operations. Specifically, there are six explicit interactions that are considered: historical, daily, weekly, and monthly interchanging flow connectivity, transport connectivity, and geographical distance.

The four dimensions representing the historical, daily, weekly, and monthly flow connectivity across regions are measured by the historical, daily, weekly, and monthly interchanging tourist flows, respectively. The historical and periodic observations can comprehensively represent the overall connectivity between two regions. The four corresponding adjacency matrices are given as follows:

$$\mathbf{A}^{\text{his}}_{(i,j),t} = \frac{Q^{\text{his}}_{(i,j),t} + Q^{\text{his}}_{(j,i),t}}{2}, \tag{4}$$

$$\mathbf{A}^{\text{daily}}_{(i,j),t} = \frac{Q^{\text{daily}}_{(i,j),t} + Q^{\text{daily}}_{(j,i),t}}{2}, \tag{5}$$

$$\mathbf{A}^{\text{weekly}}_{(i,j),t} = \frac{Q^{\text{weekly}}_{(i,j),t} + Q^{\text{weekly}}_{(j,i),t}}{2}, \tag{6}$$

$$\mathbf{A}^{\text{monthly}}_{(i,j),t} = \frac{Q^{\text{monthly}}_{(i,j),t} + Q^{\text{monthly}}_{(j,i),t}}{2}, \tag{7}$$

where $Q_{(i,j)}$ and $Q_{(j,i)}$ indicate the flow from $\mathcal{V}_i$ to $\mathcal{V}_j$ and from $\mathcal{V}_j$ to $\mathcal{V}_i$, respectively. The subscripts of $\mathbf{A}$ and $\mathbf{Q}$ indicate that these are all dynamic terms.

The edge in the dimension representing the transport connectivity across regions indicates accessible transport between two regions. Therefore, the weighted adjacency matrix can be formally defined as

$$\mathbf{A}^{\text{trans}}_{i,j} = \begin{cases} 1 & \text{, if } \mathcal{V}_i, \mathcal{V}_j \text{ have accessible transport} \\ 0 & \text{, otherwise.} \end{cases} \tag{8}$$

An additional dimension measuring the geographical distance across regions is also considered with the weighted adjacency matrix:

$$\mathbf{A}^{\text{dis}}_{i,j} = dis(L_i, L_j), \tag{9}$$

where L_i and L_j are the locations of $\mathcal{V}_i$ and $\mathcal{V}_j$, respectively, and $dis(\cdot,\cdot)$ is a function whose output is the geodesic distance between two locations.[2] Note that $\mathbf{A}^{\text{dis}}$ and $\mathbf{A}^{\text{trans}}$ will not change over time.

[1] Historical tourism demand refers to the tourism demand within a past regular observation duration (depending on the time granularity of the collected data). Daily, weekly, and monthly tourism demands are the periodic average values of the past daily, weekly, and monthly tourism demands. The definitions of undermentioned historical and periodic observations of interchanging tourism flow are similar to that of the historical, daily, weekly, and monthly tourism demands.

[2] Although there are various definitions of geographical distance, we adopt the most frequently used definition, the geodesic distance, in this work.

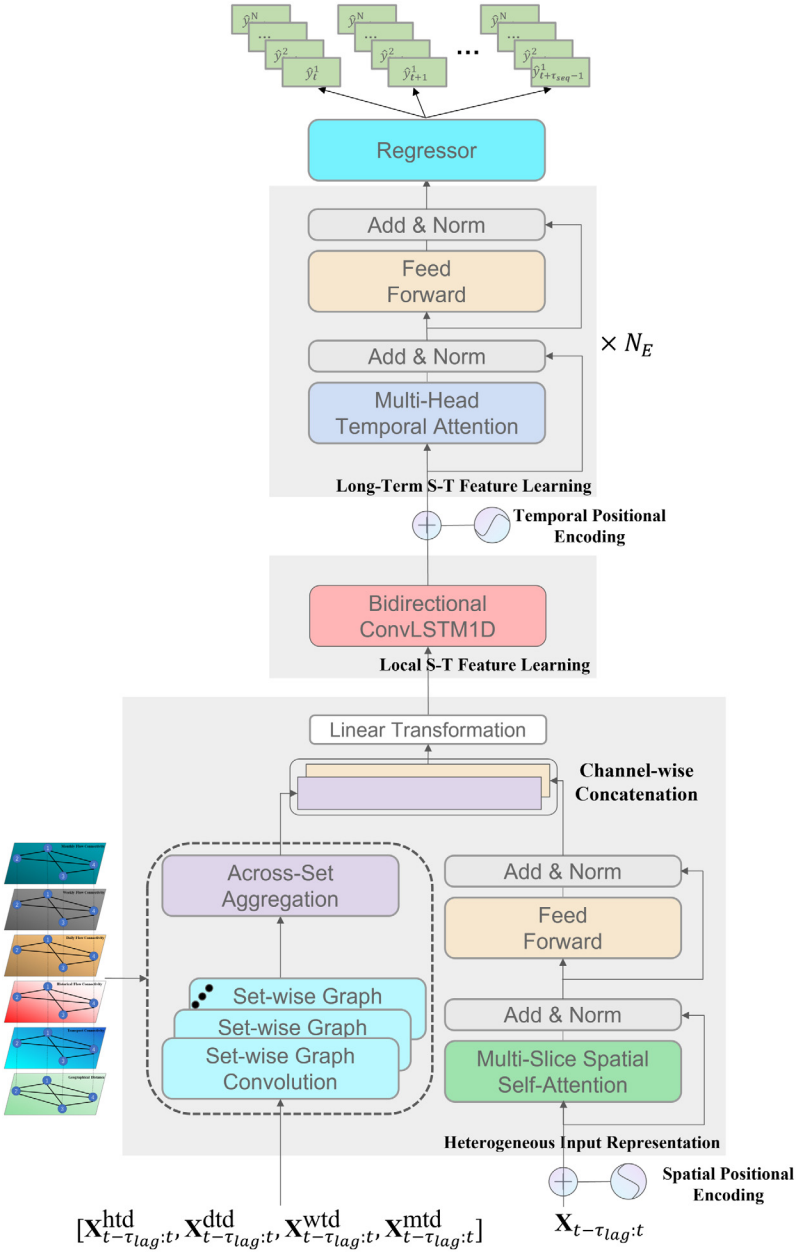


Fig. 2. Proposed heterogeneous spatial–temporal graph-attention network (HSTGANet).

Transport connectivity and geographical distance may affect the spatial connections across regions. Therefore, we combine these two graph dimensions with one of the four graph dimensions representing flow connectivity to form four graph dimension sets, which are then used to model four historical and periodic high-dimensional spatial–temporal features, respectively. The corresponding adjacency matrices of the four graph dimension sets are shown in Fig. 1(b).

With the proposed spatial–temporal information representation, high-dimensional spatial–temporal features can be learned from the multiple graph dimension sets and corresponding channels in the graph structure through graph convolutions, and from the spatial–temporal sequence with the help of self-attention. To incorporate more temporal tourism demand information, a set of slices from the spatial–temporal sequence is used as the input of the proposed heterogeneous spatial–temporal graph-attention network. Given the spatial–temporal information representation, the τ_{seq}-step tourism demand forecasting with τ_{lag}-step spatial–temporal inputs can be formulated as

$$\{\hat{\mathbf{y}}^i_{t:t+\tau_{seq}} | i = 1, 2, \ldots, N\} = F_\theta \left(\mathbf{X}_{t-\tau_{lag}:t}, \mathcal{G}\right), \tag{10}$$

where $\hat{\mathbf{y}}$ is the tourism demand forecast, F denotes the proposed HSTGANet, and θ denotes the learnable parameters.

4. Heterogeneous spatial-temporal graph-attention network

As shown in Fig. 2, the proposed HSTGANet is composed of a heterogeneous input representation layer, a local spatial–temporal feature learning layer, N_E spatial–temporal feature learning blocks, and a final forecasting layer. The heterogeneous input representation layer is used to effectively learn the hidden representation to reflect dynamic spatial connections, while the local spatial–temporal feature learning layer and long-term spatial–temporal feature learning blocks are used to learn local and long-term spatial–temporal features, respectively, that are useful for spatial–temporal tourism demand forecasting.

4.1. Heterogeneous input representation layer

As outlined in Section 3, spatial–temporal information is represented by the spatial–temporal sequence $\mathbf{X}$ and the multi-dimensional graph $\mathcal{G}$. Note that the spatial connections across regions can be explicitly inferred by the multiple graph dimension sets and corresponding channels in the graph structure, and can also be implicitly inferred by the spatial–temporal sequence. To extract the hidden representation embedded with complex spatial–temporal features from both parts of the inputs, we propose a multi-dimensional graph embedding layer and multi-slice spatial self-attention layer to form a heterogeneous input representation layer to extract both parts of the spatial–temporal features.

4.1.1. Multi-dimensional graph embedding layer

The multi-dimensional graph embedding layer extracts set-wise representations ($\mathbf{h}^{\text{his}}_{t-\tau_{lag}:t}$, $\mathbf{h}^{\text{daily}}_{t-\tau_{lag}:t}$, $\mathbf{h}^{\text{weekly}}_{t-\tau_{lag}:t}$, and $\mathbf{h}^{\text{monthly}}_{t-\tau_{lag}:t}$, which represent historical, daily, weekly, and monthly spatial connection patterns, respectively) and then aggregates these representations from all sets to form the final graph embedding (i.e., the final hidden representation learned from explicit dynamic spatial connections, $\mathbf{h}^{\text{exp}}_{t-\tau_{lag}:t}$). To simplify the description hereafter, the set-wise representations are uniformly denoted by $\mathbf{h}^{\text{set}}_{t-\tau_{lag}:t}$, where set $\in$ {his, daily, weekly, monthly}. The input for learning each set-wise representation, denoted by $\mathbf{X}^{\text{set}}_{t-\tau_{lag}:t}$, is the corresponding channel of $\mathbf{X}_{t-\tau_{lag}:t}$, which can be specified as

$$\mathbf{X}^{\text{set}}_{t-\tau_{lag}:t} = \begin{cases} \mathbf{X}^{\text{htd}}_{t-\tau_{lag}:t} & \text{, if set = his} \\ \mathbf{X}^{\text{dtd}}_{t-\tau_{lag}:t} & \text{, if set = daily} \\ \mathbf{X}^{\text{wtd}}_{t-\tau_{lag}:t} & \text{, if set = weekly} \\ \mathbf{X}^{\text{mtd}}_{t-\tau_{lag}:t} & \text{, if set = monthly.} \end{cases} \tag{11}$$

Each set-wise representation is an aggregation of dimension-wise representations obtained by graph convolution operations [49]. The aggregated representation at t, $\mathbf{h}^{\text{set}}_t$, can be specified as

$$\mathbf{h}^{\text{set}}_t = (\mathbf{h}^{\text{set,dis}}_t, \mathbf{h}^{\text{set,trans}}_t, \mathbf{h}^{\text{set,set}}_t)\mathbf{W}^{\text{set}}, \tag{12}$$

and the dimension-wise representations at t are obtained by

$$\mathbf{h}^{\text{set,dis}}_t = \tilde{\mathbf{A}}^{\text{dis}}\mathbf{X}^{\text{set}}_t\mathbf{W}^{\text{set,dis}}, \tag{13}$$

$$\mathbf{h}^{\text{set,trans}}_t = \tilde{\mathbf{A}}^{\text{trans}}\mathbf{X}^{\text{set}}_t\mathbf{W}^{\text{set,trans}}, \tag{14}$$

$$\mathbf{h}^{\text{set,set}}_t = \tilde{\mathbf{A}}^{\text{set}}_t\mathbf{X}^{\text{set}}_t\mathbf{W}^{\text{set,set}}, \tag{15}$$

where $\mathbf{W}^{\text{set}} \in \mathbb{R}^{(3\times d_{\text{exp}})\times d_{\text{exp}}}$, $\mathbf{W}^{\text{set,dis}} \in \mathbb{R}^{1\times d_{\text{exp}}}$, $\mathbf{W}^{\text{set,trans}} \in \mathbb{R}^{1\times d_{\text{exp}}}$, and $\mathbf{W}^{\text{set,set}} \in \mathbb{R}^{1\times d_{\text{exp}}}$ are learnable weights, and d_{exp} is the output dimension of the multi-dimensional graph embedding layer. Without loss of generality, $\tilde{\mathbf{A}}_t$ can be defined as

$$\tilde{\mathbf{A}}_t = \mathbf{I} + (\mathbf{D}_t)^{-\frac{1}{2}}(\mathbf{A}_t)(\mathbf{D}_t)^{-\frac{1}{2}}, \tag{16}$$

where $\mathbf{D}_t$ and $\mathbf{A}_t$ are the degree and adjacency matrices at t of the corresponding dimension in $\mathcal{G}$, and $\mathbf{I}$ is the identity matrix.

Finally, $\mathbf{h}^{\text{exp}}_t$ can be expressed as

$$\mathbf{h}^{\text{exp}}_t = ReLU((\mathbf{h}^{\text{his}}_t, \mathbf{h}^{\text{daily}}_t, \mathbf{h}^{\text{weekly}}_t, \mathbf{h}^{\text{monthly}}_t)\mathbf{W}^{\text{exp}} + \mathbf{b}^{\text{exp}}), \tag{17}$$

where $ReLU(\cdot)$ is the Rectified Linear Unit (ReLU) nonlinear activation function and $\mathbf{W}^{\text{exp}} \in \mathbb{R}^{(12\times d_{\text{exp}})\times d_{\text{exp}}}$ and $\mathbf{b}^{\text{exp}}$ are learnable weights and biases, respectively.

4.1.2. Multi-slice spatial self-attention

Multi-head attention has been widely used in various research fields, especially natural language processing and computer vision, due to its strong capability of learning long-term dependencies. The multi-head attention can be calculated as [35]

$$\text{MultiHead}(\mathbf{Q}^H, \mathbf{K}^H, \mathbf{V}^H) = \text{Concat}\left(\mathbf{O}^{\text{Head}}_1, \ldots, \mathbf{O}^{\text{Head}}_H\right)\mathbf{W}^O,$$

$$\mathbf{O}^{\text{Head}}_i = \text{Attention}\left(\mathbf{Q}^H\mathbf{W}^Q_i, \mathbf{K}^H\mathbf{W}^K_i, \mathbf{V}^H\mathbf{W}^V_i\right), \tag{18}$$

where Attention$(\cdot)$ denotes the scaled dot-product attention, which is defined as

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{19}$$

where $\mathbf{Q}^H \in \mathbb{R}^{\tau\times d_{\text{model}}}$, $\mathbf{K}^H \in \mathbb{R}^{\tau\times d_{\text{model}}}$, and $\mathbf{V}^H \in \mathbb{R}^{\tau\times d_{\text{model}}}$ are the query, key and value matrices, respectively; $\mathbf{Q} \in \mathbb{R}^{\tau\times d_k}$, $\mathbf{K} \in \mathbb{R}^{\tau\times d_k}$, $\mathbf{V} \in \mathbb{R}^{\tau\times d_v}$ are the transformed query, key and value matrices for calculating the scaled dot-product attention, respectively; τ is the input dimension of time slices; d_k is the input dimension of queries and keys; and d_v is the input dimension of the values. $\mathbf{W}^Q_i \in \mathbb{R}^{d_{\text{model}}\times d_k}$, $\mathbf{W}^K_i \in \mathbb{R}^{d_{\text{model}}\times d_k}$, $\mathbf{W}^V_i \in \mathbb{R}^{d_{\text{model}}\times d_v}$, and $\mathbf{W}^O \in \mathbb{R}^{Hd_v\times d_{\text{model}}}$, are learnable weights which project from the original representation space to different sub-spaces. H is the number of attention heads, d_{model} is the input dimension of the transformed matrices, and $d_k = d_v = d_{\text{model}}/H$. Concat denotes the concatenation operation. Multi-head attention is expected to learn various features from multiple representation sub-spaces, and thus enhance the performance of the model.

We would like to adopt self-attention to learn the implicit dynamic spatial connections hidden in the spatial–temporal sequence. However, the multi-head attention mechanism calculates spatial attention under hidden sub-spaces, which does not provide a clear visualization of the dynamic spatial connections learned from the inputs compared with the explicit spatial relations shown in Fig. 1(b). Therefore, we propose a multi-slice spatial self-attention to learn the implicit dynamic spatial connections across regions.

Fig. 3 details the proposed multi-slice spatial self-attention mechanism. To learn spatial self-attention, $\mathbf{X}^S \in \mathbb{R}^{\tau\times N\times c}$ is first split into τ slices $\{\mathbf{S}_i \in \mathbb{R}^{N\times c} | i = 1, 2, \ldots, \tau\}$, and then projected to $\mathbf{Q} \in \mathbb{R}^{N\times d_{sk}}$, $\mathbf{K} \in \mathbb{R}^{N\times d_{sk}}$ and $\mathbf{V} \in \mathbb{R}^{N\times d_{sv}}$ for the calculation of the scaled dot-product attention shown in Eq. (19).

The multi-slice spatial attention can be expressed by the following equations:

$$\text{MultiSlice}(\mathbf{X}^S) = \text{Concat}\left(\mathbf{O}^{\text{Slice}}_1, \ldots, \mathbf{O}^{\text{Slice}}_\tau\right),$$

$$\mathbf{O}^{\text{Slice}}_i = \text{Attention}\left(\mathbf{S}_i\mathbf{W}^{SQ}_i, \mathbf{S}_i\mathbf{W}^{SK}_i, \mathbf{S}_i\mathbf{W}^{SV}_i\right), \tag{20}$$

where $\mathbf{W}^{SQ}_i \in \mathbb{R}^{c\times d_{sk}}$, $\mathbf{W}^{SK}_i \in \mathbb{R}^{c\times d_{sk}}$, and $\mathbf{W}^{SV}_i \in \mathbb{R}^{c\times d_{sv}}$ are learnable weights which conduct projections from the original representation space to the slice-based temporal sub-spaces and $d_{sk} = d_{sv} = d_{\text{imp}}$, where d_{imp} is the output dimension of the multi-slice spatial self-attention layer.

With the proposed multi-slice spatial self-attention, the network can learn implicit spatial–temporal connections from the input spatial–temporal sequence, while the concatenated attention weights (denoted by $\mathbf{W}^{\text{Attn}} \in \mathbb{R}^{\tau\times N\times N}$) visualize the learned spatial–temporal connections.

4.1.3. Feed-forward layer and layer normalization

As shown in Fig. 2, there is a feed-forward layer after the multi-slice spatial self-attention layer. A residual connection is employed around each of the two layers, followed by layer normalization. The feed-forward layer is composed of two linear transformations, in between which is a ReLU activation. The output of the multi-slice spatial self-attention (i.e., MultiSlice$(\mathbf{X}^S)$) is denoted as $\mathbf{h}^{\text{at}}$ and the output of the first layer normalization is denoted as $\mathbf{h}^{\text{ln}}$. The learned implicit spatial connections $\mathbf{h}^{\text{imp}}$ can be expressed as

$$\mathbf{h}^{\text{ln}} = \text{LayerNorm}\left(\mathbf{h}^{\text{at}}\right) + \mathbf{h}^{\text{at}}, \tag{21}$$

$$\mathbf{h}^{\text{imp}} = \text{LayerNorm}\left(\left(ReLU(\mathbf{h}^{\text{ln}}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2\right)\right) + \mathbf{h}^{\text{ln}}, \tag{22}$$

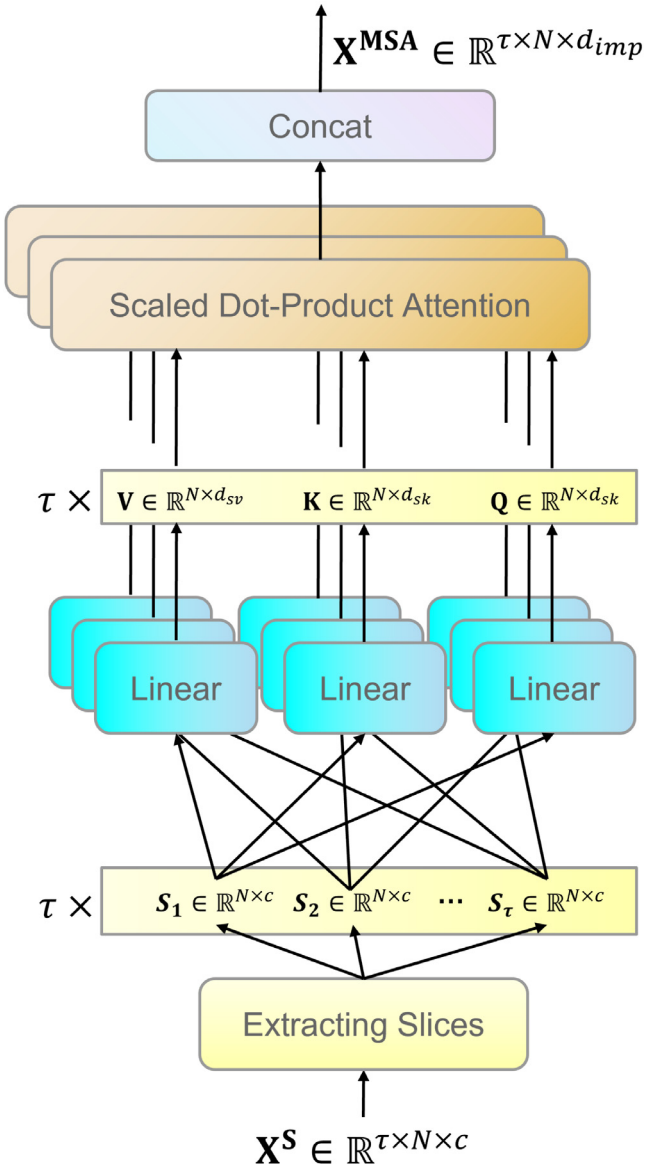


Fig. 3. Proposed multi-slice spatial self-attention.

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are the learnable weights and $\mathbf{b}_1$ and $\mathbf{b}_2$ are the learnable biases of the first and the second linear transformations, respectively.

4.1.4. Channel-wise concatenation and fully connected layer

The hidden representation learned by the multi-dimensional graph embedding layer and multi-slice spatial self-attention are concatenated with channel-wise concatenation, followed by a linear transformation, which can be expressed by

$$\mathbf{h}^{\text{emb}} = \text{Concat}\left(\mathbf{h}^{\text{exp}}, \mathbf{h}^{\text{imp}}\right)\mathbf{W}_3 + \mathbf{b}_3, \tag{23}$$

where $\mathbf{W}_3 \in \mathbb{R}^{(d_{\text{exp}}+d_{\text{imp}})\times d_{\text{emb}}}$ and $\mathbf{b}_3$ are the learnable weights and biases, and d_{emb} is the output dimension of the heterogeneous input representation layer.

4.2. Local spatial-temporal feature learning layer

To further learn spatial–temporal information from the hidden representation learned by the heterogeneous input representation layer, we adopt a one-dimensional convolutional LSTM (ConvLSTM1D) layer, modified from that in [50], immediately after the heterogeneous input representation layer. The convolution and recurrent operations in the ConvLSTM1D layer help learn local spatial–temporal features (denoted by $\mathbf{h}^{\text{local}}$). To fully exploit the temporal information, we apply the bidirectional ConvLSTM1D layer to simultaneously extract forward and backward temporal information.

4.3. Long-term spatial-temporal feature learning blocks

The stacked long-term spatial–temporal feature learning blocks are used to better learn long-term spatial–temporal features (denoted by $\mathbf{h}^{\text{long}}$). Each long-term spatial–temporal feature learning block is composed of a multi-head temporal attention layer and a feed-forward layer. The multi-head temporal attention first reshapes the input $\mathbf{h}^{\text{local}} \in \mathbb{R}^{\tau \times N \times d_{\text{local}}}$ to $\tilde{\mathbf{h}}^{\text{local}} \in \mathbb{R}^{\tau \times (N \times d_{\text{local}})}$, and then transforms it to $\mathbf{Q}^H \in \mathbb{R}^{\tau \times d_{\text{model}}}$, $\mathbf{K}^H \in \mathbb{R}^{\tau \times d_{\text{model}}}$, and $\mathbf{V}^H \in \mathbb{R}^{\tau \times d_{\text{model}}}$, and finally conducts the multi-head self-attention defined in Eq. (18). There are also two sets of residual connections and layer normalizations neighboring the multi-head temporal attention layer and a feed-forward layer, with feed-forward dimension d_{ff}, following it.

4.4. Positional encoding

Since the scaled dot-product attention treats the elements (rows) of input matrices equivalently, their spatial or temporal positions, which may also contain very useful information, are naturally ignored by the attention mechanism. Therefore, we add spatial positional encoding (SPE) to the input $\mathbf{X} \in \mathbb{R}^{\tau \times N \times c}$ before the multi-slice spatial self-attention layer, and temporal positional encoding (TPE) to $\tilde{\mathbf{h}}^{\text{local}} \in \mathbb{R}^{\tau \times (N \times d_{\text{local}})}$ before the multi-head temporal self-attention layer, to indicate the spatial and temporal positions of the input elements, respectively. We investigated two widely adopted choices of positional encodings: sinusoidal [35] and learnable positional encoding [51]. Sinusoidal positional encodings can be specified as

$$S(T)PE(pos, 2i) = sin(\frac{pos}{10000^{2i/d_{pe}}}), \tag{24}$$

$$S(T)PE(pos, 2i+1) = cos(\frac{pos}{10000^{2i/d_{pe}}}), \tag{25}$$

where pos is the spatial or temporal position of the input element, i is the dimensional index of the positional encoding, and d_{pe} is the dimension of the positional encoding. For STE, $d_{pe} = c$, and for TPE, $d_{pe} = N \times d_{\text{local}}$.

The learnable positional encoding simply adopts a set of weights to store positional information learned from training. Based on ablation studies for positional encoding, shown in Table 8, we chose sinusoidal positional encoding to form the spatial and temporal positional encodings.

4.5. Output regressor

The output regressor consists of two linear transformation layers and conducts the final regression with the given learned spatial–temporal features. Specifically, the τ_{seq}-step tourism demand forecasts can be expressed as

$$\hat{\mathbf{y}} = (\mathbf{h}^{\text{long}}\mathbf{W}_4 + \mathbf{b}_4)\mathbf{W}_5 + \mathbf{b}_5, \tag{26}$$

where $\mathbf{W}_4 \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$ and $\mathbf{W}_5 \in \mathbb{R}^{d_{\text{model}} \times \tau_{seq}}$ are learnable weights and $\mathbf{b}_4$ and $\mathbf{b}_5$ are biases.

Table 1
Statistics of collected tourism demand for the six regions of Wanshan Archipelago.

Region	mean	std	min	25%	50%	75%	max
Xiangzhou	141.83	165.68	0	0	94	213	1091
Guishan	35.90	54.75	0	0	4	63	445
Wailingding	62.12	101.49	0	0	0	97	709
Hengqin	7.42	24.87	0	0	0	0	437
Dongao	36.66	65.83	0	0	0	54	500
Wanshan	12.03	24.17	0	0	0	10	323

mean: Mean of tourism demand;
std: Standard deviation of tourism demand;
min: Minimum tourism demand;
max: Maximum tourism demand;
p%: p percentile.

4.6. Model training

The spatial–temporal tourism demand-forecasting model is optimized by minimizing the mean absolute error (MAE) between the tourism demand forecasts and the true tourism demand across all regions and forecasting horizons:

$$\underset{\theta}{\operatorname{argmin}}\ \mathcal{L}_{MAE} = \frac{1}{N * \tau_{seq}} \sum_{i=1}^{N} \sum_{t=1}^{\tau_{seq}} |y_t^i - \hat{y}_t^i|, \tag{27}$$

where θ denotes the learnable parameters of the model and y and $\hat{y}$ are the true and forecasted tourism demands, respectively.

5. Experiments

5.1. Experimental datasets

We evaluated the performance of the proposed spatial–temporal feature learning framework using a tourism demand dataset of Wanshan Archipelago, an island scenic area in Zhuhai, China. Wanshan Archipelago includes four well-developed islands, Guishan, Wailingding, Dongao, and Wanshan, and two ports located at Zhuhai city, Xiangzhou and Hengqin. In this case study, the tourism demand for the six regions was forecasted using their historical observations and an explicit multi-dimensional graph.

As public ferry is the only transportation available to tourists, the historical tourism demand and interchanging flows (including incoming and outgoing flows) can be obtained from the ferry ticketing data. The collected data ranged from Jan. 1st, 2017, to Dec. 31st, 2019. There were a total of 11 observations per day.[3] Specifically, the tourism demand of one region s at time t refers to the number of ferry tickets sold whose destination was in region s and departure time was within the last observation duration of t. Similarly, the outgoing flow from region i to j is the number of ferry tickets sold whose origin was i and destination was j; the incoming flow from j to i is the number of ferry tickets sold whose origin was j and destination was i. In addition, the daily, weekly, and monthly observations of tourism demand and interchanging flow were taken to be the average values of the past daily, weekly, and monthly tourism demands and interchanging tourist flows. The statistics of the tourism demand for the six regions of Wanshan Archipelago are shown in Table 1.

We further collected geographical locations, information about Chinese public holidays, search index data from Baidu Search Engine, and weather data from the China Meteorological Data Service Centre for all six regions to form the input sequences. In this case study, all six regions share the same holiday information.

[3] Note that there was no ferry service between 20:00–8:00 (next day), so the observations are hourly observations from 9:00–19:00 every day.

5.2. Experimental settings

The dataset is split into a training set, a validation set, and a testing set, with a ratio of around 10:1:1 (36 months of data available in total). Specifically, the tourism demand data in the validation set ranges from Oct. 1st, 2018, to Dec. 31st, 2018, while the tourism demand data in the testing set ranges from Oct. 1st, 2019, to Dec. 31st, 2019.

We conduct forward filling for several dates without ferry ticketing data. The data are normalized with min–max normalization.

The proposed HSTGANet is implemented with TensorFlow 2.7. Both τ_{lag} and τ_{seq} are set to 77, which means the model takes a 77-observation (7-day) input for 77-step (7-day) tourism demand forecasting. The HSTGANet was trained using the Adam optimizer [52] with an initial learning rate of 0.0005.

Generally, hyper-parameter optimization can be done through grid search [53,54], genetic algorithms [55,56], Bayesian optimization algorithms [57], etc. To avoid large number of model parameters which may lead to overfitting, search space for hyper-parameters of the proposed HSTGANet would not be large. Therefore, to exhaust specified subsets of hyper-parameter search space for better performance [53], we utilize the grid search algorithm to optimize hyper-parameters of the HSTGANet. Specifically, the optimization algorithm validates the model trained under each subset of search space to get the best model with the lowest validation loss. The search space and the final best hyper-parameter setting for each hyper-parameter are shown in Table 2.

Two metrics are utilized to evaluate the performance of the proposed HSTGANet and the baselines, including Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), which can be formally defined as:

$$MAE^i = \frac{1}{\tau_{seq}} \sum_{t=1}^{\tau_{seq}} |y_t^i - \hat{y}_t^i|, \tag{28}$$

$$RMSE^i = \sqrt{\frac{1}{\tau_{seq}} \sum_{t=1}^{\tau_{seq}} (y_t^i - \hat{y}_t^i)^2}, \tag{29}$$

where MAE^i and $RMSE^i$ are the evaluation metrics for region i.

5.3. Baseline models and implementations

As previously outlined, spatial models have not been applied in tourism demand forecasting for a long time. Therefore, we do not consider any spatial models in our baseline models. Instead, we use time series forecasting and spatial–temporal forecasting baseline models, which, along with their implementations, are listed as follows.

(1) Time Series Forecasting Baseline Models:

We chose the ARIMA [43] and SARIMAX models [8] as representative statistical models. When fitting the SARIMAX model, we used holiday, weather, and search index information as explanatory variables.

We chose support vector regression (SVR) [9] and random forest regression (RFR) [10] as representative traditional machine learning models. As representative deep learning models, we chose two RNN-based models, LSTM [11] and BiLSTM [11], and one attention-based model, the multi-head attention CNN (MHACNN) [15]. All deep learning models shared the same inputs as HSTGANet.

(2) Spatial-Temporal Forecasting Baseline Models:

Currently, there are only two types of models focusing on spatial–temporal tourism demand forecasting: panel data models [5,7] and spatial–temporal autoregressive models [6,7]. We chose the panel data and spatial–temporal autoregressive models

Table 2
Search space and best hyper-parameter settings for hyper-parameters of the proposed HSTGANet.

Hyper-parameter	Search space	Best setting
Hidden dimensions ($d_{exp} = d_{imp} = d_{emb} = d_{local}$)	{32, 64, 128}	64
Kernel size of ConvLSTM1D	{3, 5}	3
Number of ST learning blocks (N_E)	{1, 2, 3}	3
Number of attention heads (H)	{4, 8}	4
Multi-head attention dimension (d_{model})	{64, 128, 256}	128
Feed-forward dimension (d_{ff})	{64, 128, 256}	128
Dropout rate	{0.3, 0.5, 0.7}	0.5

Table 3
Performance comparison of 77-step tourism demand forecasting for the six regions in Wanshan Archipelago for HSTGANet and the time series forecasting baseline models.

Region	Metric	ARIMA	SARIMAX	SVR	RFR	LSTM	BiLSTM	MHACNN	HSTGANet
Xiangzhou	MAE	119.97	118.06	74.00	59.75	77.44	75.09	69.77	**50.81**
	RMSE	151.02	153.61	120.61	93.22	102.20	100.93	93.78	**79.65**
Guishan	MAE	42.42	38.34	26.46	24.77	30.75	28.64	28.12	**20.87**
	RMSE	60.43	60.18	52.62	43.19	46.81	44.48	**41.29**	42.04
Wailingding	MAE	57.34	55.79	32.61	25.91	34.70	38.39	30.89	**21.06**
	RMSE	73.57	73.61	59.18	49.87	51.66	54.73	48.42	**42.93**
Hengqin	MAE	9.06	16.14	7.88	11.49	9.67	11.91	16.28	**6.97**
	RMSE	26.67	27.04	21.26	24.27	**19.76**	20.70	22.70	22.59
Dongao	MAE	35.63	38.41	25.22	16.72	26.45	23.11	20.87	**13.34**
	RMSE	45.87	50.78	48.11	32.03	37.74	33.34	29.27	**27.1**
Wanshan	MAE	13.28	14.21	8.41	**5.86**	7.56	6.37	7.65	7.28
	RMSE	21.31	20.89	20.77	13.12	14.03	**12.52**	13.28	16.46
Overall	MAE	46.28	46.83	29.10	24.08	31.10	30.59	28.93	**20.06**
	RMSE	63.15	64.35	53.76	42.62	45.37	44.45	41.46	**38.46**

from [7] as representative baseline models. Specifically, we used a flow-based spatial weighting matrix when calculating Eqs. (1) and (2). As suggested by [7], explanatory variables were not considered in these two models.

Extending to existing spatial–temporal forecasting models, we further considered a deep convolutional neural network (DCNN) [28], spatio-temporal graph convolutional network (STGCN) [24] and spatial–temporal transformer (STTN) [21], which is one of the SOTA for long-term traffic flow forecasting, as baselines. In addition, two advanced GNN-based spatial–temporal models, the graph multi-attention network (GMAN) [58] and attention based spatial–temporal graph neural network (ASTGNN) [59], were considered for comparison.

We used the grid search algorithm to search for the best hyper-parameters for the baseline models with hyper-parameters, as this was conducted for the proposed HSTGANet. Note that for the time series forecasting models, region-specific models were trained to obtain the best forecasting performance for each region. For the spatial–temporal forecasting models, one single model was trained for all the six regions. Therefore, we acknowledge that the optimized spatial–temporal forecasting model might not outperform region-specific time series forecasting models for some regions, but instead can obtain a better overall forecasting performance.

5.4. Experimental results

A comparison of the performance of the proposed HSTGANet with the time series and spatial–temporal forecasting baseline models for 77-step tourism demand forecasting of the six regions in Wanshan Archipelago, in terms of the MAE and RMSE, are shown in Tables 3 and 4, respectively.

The proposed HSTGANet yields a better overall forecasting performance than both the time series and spatial–temporal forecasting baseline models, demonstrating the effectiveness of the proposed framework in learning spatial–temporal features by combining graph convolutions with self-attention mechanisms.

Among the time series forecasting models, the machine learning models show a better performance than that of the statistical models, since the statistical models are not suitable for fine-grained tourism demand forecasting. The RFR model is the best traditional machine learning-based tourism demand-forecasting model, while MHACNN is the best deep learning-based tourism demand forecasting model.

Among the spatial–temporal forecasting models, all of the deep learning-based models outperformed the statistical models (i.e., the panel and STARMA models). The graph-attention based spatial–temporal forecasting models (including STTN, GMAN, ASTGNN, and HSTGANet) yield a more competitive forecasting performance compared with the two convolution-based and graph convolution-based spatial–temporal forecasting models (DCNN and STGCN) in terms of the MAE and RMSE performance.

Tables 3 and 4 show that the advanced ASTGNN model outperforms the RFR and MHACNN models in terms of the MAE, and shows a very competitive performance in terms of the RMSE. The superiority of the HSTGAN and ASTGNN models shows that advanced spatial–temporal forecasting models can learn spatial–temporal features of the six regions and therefore provide more accurate tourism demand forecasts than time series forecasting models, which ignore spatial relations across regions. However, as previously discussed, the time series forecasting models were trained to obtain the best forecasting performance for each region, while one single spatial–temporal forecasting model was trained on all six regions based on the overall forecasting performance. Therefore, the optimized spatial–temporal forecasting models might not outperform region-specific time series forecasting models for some regions, especially for those with small scales in tourism demand (e.g., BiLSTM achieves the best forecasting performance in terms of RMSE for forecasting the tourism demand for Wanshan island). In addition, spatial–temporal features captured by different spatial–temporal forecasting models might differ, so the degree of attention to the six regions might also differ, resulting in inconsistent forecasting performance across regions (i.e., these models might achieve the best region-specific performances in different regions).

Table 4
Performance comparison of 77-step tourism demand forecasting for the six regions in Wanshan Archipelago for HSTGANet and the spatial–temporal forecasting baseline models.

Region	Metric	Panel	STARMA	DCNN	STGCN	STTN	GMAN	ASTGNN	HSTGANet
Xiangzhou	MAE	152.95	151.43	73.68	76.55	60.11	60.25	63.08	**50.81**
	RMSE	215.8	212.9	96.57	102.71	86.77	95.09	99.91	**79.65**
Guishan	MAE	49.14	48.25	33.36	26.55	20.33	23.82	21.1	**20.87**
	RMSE	84.54	82.57	46.38	42.63	**36.6**	49.8	44.48	42.04
Wailingding	MAE	56.72	56.57	43.99	41.56	34.4	29.61	24.82	**21.06**
	RMSE	97.57	95.99	58.11	59.4	52.18	55.25	49.58	**42.93**
Hengqin	MAE	8.89	8.85	18.03	12.25	14.31	6.82	**6.69**	6.97
	RMSE	28.09	27.12	25.14	20.2	22.78	21.63	**19.7**	22.59
Dongao	MAE	34.45	34.55	33.13	26.89	18.98	17.68	14.24	**13.34**
	RMSE	62.2	60.44	42.76	36.68	29.7	34.33	28.87	**27.1**
Wanshan	MAE	14.91	15.52	12.09	8.54	6.66	7.03	**5.64**	7.28
	RMSE	30.68	29.68	16.62	15.33	13.38	16.74	**13.26**	16.46
Overall	MAE	52.84	52.53	35.71	32.06	25.80	24.20	22.60	**20.06**
	RMSE	86.48	84.78	47.60	46.16	40.24	45.47	42.63	**38.46**

Table 5
Ablation studies for the effectiveness of the proposed multi-dimensional graph embedding layer and multi-slice spatial self-attention layer.

Model	Region	MAE	RMSE
HSTGANet	Overall	**20.06**	**38.46**
HSTGANet w/o MDGE	Overall	27.14	48.37
HSTGANet w/o MSSSA	Overall	44.23	72.92

Table 6
Ablation studies for the effectiveness of the six spatial interactions in the multi-dimensional graph.

Model	Region	MAE	RMSE
HSTGANet	Overall	**20.06**	**38.46**
HSTGANet w/o A^{dis}	Overall	21.73	40.22
HSTGANet w/o A^{trans}	Overall	23.71	43.89
HSTGANet w/o A^{his}	Overall	21.83	41.33
HSTGANet w/o A^{daily}	Overall	25.02	45.01
HSTGANet w/o A^{weekly}	Overall	28.61	52.25
HSTGANet w/o $A^{monthly}$	Overall	21.93	40.59

5.5. Ablation studies

To verify the effectiveness of the proposed graph-attention based spatial–temporal learning framework, we conducted a set of ablation studies. For simplicity, we will only show the overall performance of the studied models here.

5.5.1. Effectiveness of the proposed MDGE and MSSSA

In this section, the multi-dimensional graph embedding (MDGE) layer or the multi-slice spatial self-attention (MSSSA) layer was removed from HSTGANet to verify the effectiveness of explicit and implicit dynamic spatial connections, respectively, and the effectiveness of the two layers in extracting high-dimensional spatial–temporal features from the two inputs.

Table 5 shows that, without the multi-dimensional graph embedding layer or the multi-slice spatial self-attention layer, the forecasting performance significantly degrades, demonstrating the effectiveness of both layers.

5.5.2. Effectiveness of the multi-dimensional graph

To further verify the effectiveness of the six spatial interactions in the multi-dimensional graph, we conducted six experiments, deleting one of each of the six spatial interactions from the multi-dimensional graph in turn. The experimental results shown in Table 6 demonstrate the effectiveness of the adopted six explicit spatial interactions modeled by the multi-dimensional graph.

Table 7
Ablation studies for the effectiveness of periodic observations.

Model	Region	MAE	RMSE
HSTGANet	Overall	**20.06**	**38.46**
HSTGANet w/o X^{dtd}	Overall	21.34	38.84
HSTGANet w/o X^{wtd}	Overall	24.26	44.71
HSTGANet w/o X^{mtd}	Overall	21.80	40.00

Table 8
Ablation studies for the effectiveness of positional encodings.

Model	Region	MAE	RMSE
HSTGANet	Overall	**20.06**	**38.46**
HSTGANet w/o PE	Overall	22.88	43.74
HSTGANet w learnable SPE	Overall	23.82	44.46
HSTGANet w learnable TPE	Overall	21.97	41.92
HSTGANet w learnable PE	Overall	20.44	39.09

5.5.3. Effectiveness of periodic observations

To explore the benefits of periodic observations, we removed the daily, weekly, and monthly tourism demands from the inputs of HSTGANet.

Table 7 shows that removing any of the periodic observations has a negative impact on the prediction performance of the long-term tourism demand forecasting.

5.5.4. Effectiveness of positional encodings

Finally, to demonstrate the influence of positional information on the proposed model and investigate the effectiveness of sinusoidal positional encoding and learnable positional encoding, we conducted experiments with five positional encoding settings:

(1) HSTGANet, with sinusoidal positional encodings;

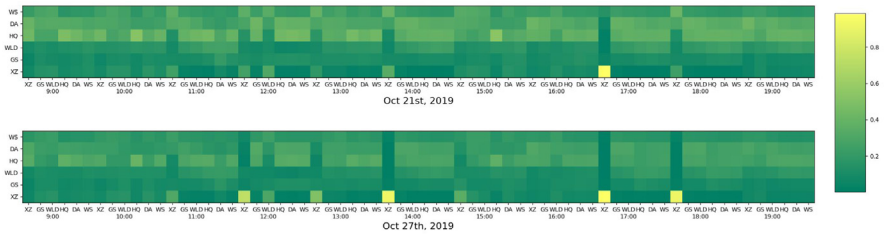
(2) HSTGANet, with positional encodings removed (HSTGANet without PE);

(3) sinusoidal SPE replaced by learnable SPE (HSTGANet with learnable SPE);

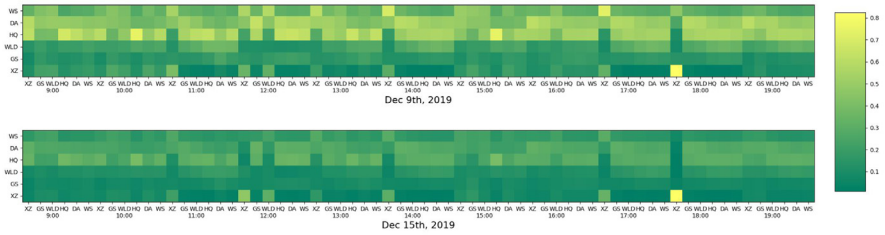
(4) sinusoidal TPE replaced by learnable TPE (HSTGANet with learnable TPE);

(5) both sinusoidal positional encodings and TPE replaced by learnable positional encodings (HSTGANet with learnable PE).

Table 8 clearly shows that positional information is important for the attention layers. Specifically, sinusoidal positional encodings are more effective than learnable positional encodings. The experimental results also suggest that using hybrid positional encodings might lead to a worse tourism demand forecasting performance.



(a) Implicit dynamic spatial connections on Oct. 21st, 2019, and Oct. 27th, 2019.



(b) Implicit dynamic spatial connections on Dec. 9th, 2019, and Dec. 15th, 2019.

Fig. 4. Implicit dynamic spatial connections learned by the proposed multi-slice spatial self-attention for October 21st, October 27th, December 9th, and December 15th, 2019.

5.6. Visualization of learned implicit dynamic spatial connections

With the proposed multi-slice spatial self-attention, an attention weighting matrix $\mathbf{W}_s^{\text{imp}} \in \mathbb{R}^{\tau_{\text{lag}} \times N \times N}$ representing the learned implicit dynamic spatial connections is obtained for a slid input. In contrast with the weighted adjacency matrices shown in Fig. 1(b), which represent the explicit dynamic spatial connections across regions, we obtain the attention weighting matrices when conducting tourism demand forecasting for two forecasting horizons: October 21st to October 27th and December 9th to December 15th, 2019. For simplicity, only the slices of the attention weighting matrices regarding the first day (Monday) and the last day (Sunday) of the two forecasting horizons are visualized (see Fig. 4).

6. Conclusion

In this paper, we proposed a graph-attention based spatial–temporal learning framework for tourism demand forecasting. The proposed spatial–temporal information representation incorporated multiple dynamic spatial connections. By proposing the multi-dimensional graph embedding and multi-slice spatial self-attention, high-dimensional dynamic spatial connections were comprehensively embedded. By leveraging the local and long-term spatial–temporal feature learning layers, the proposed HSTGANet showed remarkable ability in learning spatial–temporal features and forecasting tourism demand. The effectiveness of HSTGANet was demonstrated by experimental results in tourism demand forecasting for six regions in Wanshan Archipelago. Comparison with some outstanding baseline models for long-term traffic flow forecasting indicated that this work also provides useful insights for developing more effective models for other spatial–temporal forecasting problems.

The impact of COVID-19 on the tourism industry is significant. Due to a lack of available tourism demand data for Wanshan Archipelago from 2020 onwards, we could not investigate the impact of COVID-19 on tourism demand forecasting or the tourism industry in this paper; this is a direction we wish to explore to further improve this work in the future.

CRediT authorship contribution statement

Binggui Zhou: Data curation, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Yunxuan Dong:** Data curation, Validation, Writing – review & editing. **Guanghua Yang:** Conceptualization, Supervision, Writing – review & editing. **Fen Hou:** Writing – review & editing. **Zheng Hu:** Writing – review & editing. **Suxiu Xu:** Writing – review & editing. **Shaodan Ma:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

This work was supported in part by the Guangdong-Macau Joint Funding Project under Grant 2021A0505080015, in part by the National Key Research and Development Program of China under Grant 2017YFE0120600, in part by the Science and Technology Development Fund, Macau, under Grant SKL-IOTSC(UM)-2021-2023, and in part by the National Natural Science Foundation of China under Grant 72071093.

References

[1] W.H.K. Tsui, H. Ozer Balli, A. Gilbey, H. Gow, Forecasting of Hong Kong airport's passenger throughput, Tour. Manag. 42 (2014) 62–76, http://dx.doi.org/10.1016/j.tourman.2013.10.008.

[2] C. Morley, J. Rosselló, M. Santana-Gallego, Gravity models for tourism demand: Theory and use, Ann. Tour. Res. 48 (2014) 1–10, http://dx.doi.org/10.1016/j.annals.2014.05.008.

[3] M. Alawin, Z. Abu-Lila, Uncertainty and gravity model for international tourism demand in Jordan: Evidence from panel-GARCH model, Appl. Econom. Int. Dev. 16 (1) (2016).

[4] S.F. Witt, C.A. Witt, Forecasting tourism demand: A review of empirical research, Int. J. Forecast. 11 (3) (1995) 447–475, http://dx.doi.org/10.1016/0169-2070(95)00591-7.

[5] W. Long, C. Liu, H. Song, Pooling in tourism demand forecasting, J. Travel Res. 58 (7) (2019) 1161–1174, http://dx.doi.org/10.1177/0047287518800390.

[6] X. Jiao, G. Li, J.L. Chen, Forecasting international tourism demand: A local spatiotemporal model, Ann. Tour. Res. 83 (2020) 102937, http://dx.doi.org/10.1016/j.annals.2020.102937.

[7] Y. Yang, H. Zhang, Spatial-temporal forecasting of tourism demand, Ann. Tour. Res. 75 (2019) 106–119, http://dx.doi.org/10.1016/j.annals.2018.12.024.

[8] T. Chhorn, Tourism demand and exogenous exchange rate in Cambodia a stochastic seasonal Arimax approach, Theor. Pract. Res. Econ. Fields 9 (1) (2018) 5–16.

[9] Z.-j. Cai, S. Lu, X.-b. Zhang, Tourism demand forecasting by support vector regression and genetic algorithm, in: 2009 2nd IEEE International Conference on Computer Science and Information Technology, 2009, pp. 144–146, http://dx.doi.org/10.1109/ICCSIT.2009.5234447.

[10] S. Cankurt, Tourism demand forecasting using ensembles of regression trees, in: 2016 IEEE 8th International Conference on Intelligent Systems, IS, 2016, pp. 702–708, http://dx.doi.org/10.1109/IS.2016.7737388.

[11] S.-C. Hsieh, Tourism demand forecasting based on an LSTM network and its variants, Algorithms 14 (8) (2021) 243, http://dx.doi.org/10.3390/a14080243.

[12] A. Kulshrestha, V. Krishnaswamy, M. Sharma, Bayesian BILSTM approach for tourism demand forecasting, Ann. Tour. Res. 83 (2020) 102925, http://dx.doi.org/10.1016/j.annals.2020.102925.

[13] S. Yi, X. Chen, C. Tang, Tsformer: Time series transformer for tourism demand forecasting, 2021, arXiv:2107.10977 [Cs].

[14] M. Adil, J.-Z. Wu, R.K. Chakrabortty, A. Alahmadi, M.F. Ansari, M.J. Ryan, Attention-based STL-BiLSTM network to forecast tourist arrival, Processes 9 (10) (2021) 1759, http://dx.doi.org/10.3390/pr9101759.

[15] D.-K. Kim, S.K. Shyn, D. Kim, S. Jang, K. Kim, A daily tourism demand prediction framework based on multi-head attention CNN: The case of the foreign entrant in South Korea, 2021, arXiv:2112.00328 [Cs].

[16] L.-Y. He, H. Li, J.-W. Bi, J.-J. Yang, Q. Zhou, The impact of public health emergencies on hotel demand - estimation from a new foresight perspective on the COVID-19, Ann. Tour. Res. 94 (2022) 103402, http://dx.doi.org/10.1016/j.annals.2022.103402.

[17] R.T.R. Qiu, D.C. Wu, V. Dropsy, S. Petit, S. Pratt, Y. Ohe, Visitor arrivals forecasts amid COVID-19: A perspective from the Asia and Pacific team, Ann. Tour. Res. 88 (2021) 103155, http://dx.doi.org/10.1016/j.annals.2021.103155.

[18] X. Zhang, C. Huang, Y. Xu, L. Xia, P. Dai, L. Bo, J. Zhang, Y. Zheng, Traffic flow forecasting with spatial-temporal graph diffusion network, Proc. AAAI Conf. Artif. Intell. 35 (17) (2021) 15008–15015.

[19] S. Guo, Y. Lin, N. Feng, C. Song, H. Wan, Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, Proc. AAAI Conf. Artif. Intell. 33 (01) (2019) 922–929, http://dx.doi.org/10.1609/aaai.v33i01.3301922.

[20] C. Song, Y. Lin, S. Guo, H. Wan, Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting, Proc. AAAI Conf. Artif. Intell. 34 (01) (2020) 914–921, http://dx.doi.org/10.1609/aaai.v34i01.5438.

[21] M. Xu, W. Dai, C. Liu, X. Gao, W. Lin, G.-J. Qi, H. Xiong, Spatial-temporal transformer networks for traffic flow forecasting, 2021, arXiv:2001.02908 [Cs, Eess].

[22] L. Liu, J. Zhen, G. Li, G. Zhan, Z. He, B. Du, L. Lin, Dynamic spatial-temporal representation learning for traffic flow prediction, 2020, arXiv:1909.02902 [Cs, Stat].

[23] K. Chen, G. Chen, D. Xu, L. Zhang, Y. Huang, A. Knoll, NAST: Non-autoregressive spatial-temporal transformer for Time Series Forecasting, 2021, arXiv:2102.05624 [Cs, Stat].

[24] B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, in: IJCAI'18, AAAI Press, Stockholm, Sweden, 2018, pp. 3634–3640.

[25] Y. Xu, W. Liu, Z. Jiang, Z. Xu, T. Mao, L. Chen, M. Zhou, MAF-GNN: Multi-adaptive spatiotemporal-flow graph neural network for traffic speed forecasting, 2021, arXiv:2108.03594 [Cs].

[26] Z. Zhang, Y. Li, H. Song, H. Dong, Multiple dynamic graph based traffic speed prediction method, Neurocomputing 461 (2021) 109–117, http://dx.doi.org/10.1016/j.neucom.2021.07.052.

[27] N. Zhang, X. Guan, J. Cao, X. Wang, H. Wu, A hybrid traffic speed forecasting approach integrating wavelet transform and motif-based graph convolutional recurrent neural network, 2019, arXiv:1904.06656 [Cs, Eess].

[28] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, Y. Wang, Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction, Sensors 17 (4) (2017) 818, http://dx.doi.org/10.3390/s17040818.

[29] L. Bai, L. Yao, S.S. Kanhere, X. Wang, Q.Z. Sheng, STG2Seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting, 2019, arXiv:1905.10069 [Cs, Stat].

[30] D. Wang, Y. Yang, S. Ning, DeepSTCL: A deep spatio-temporal ConvLSTM for travel demand prediction, in: 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1–8, http://dx.doi.org/10.1109/IJCNN.2018.8489530.

[31] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, Z. Li, Deep multi-view spatial-temporal network for Taxi demand prediction, Proc. AAAI Conf. Artif. Intell. 32 (1) (2018).

[32] H. Yao, X. Tang, H. Wei, G. Zheng, Z. Li, Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction, 2018, arXiv:1803.01254 [Cs].

[33] Z. Wang, T. Xia, R. Jiang, X. Liu, K.-S. Kim, X. Song, R. Shibasaki, Forecasting ambulance demand with profiled human mobility via heterogeneous multi-graph neural networks, in: 2021 IEEE 37th International Conference on Data Engineering, ICDE, 2021, pp. 1751–1762, http://dx.doi.org/10.1109/ICDE51399.2021.00154.

[34] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, 2021, arXiv:1812.08434 [Cs, Stat].

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2017, arXiv:1706.03762 [Cs].

[36] Y. Zhang, J.-H. Xu, P.-J. Zhuang, The spatial relationship of tourist distribution in Chinese cities, Tour. Geographies 13 (1) (2011) 75–90, http://dx.doi.org/10.1080/14616688.2010.529931.

[37] L. Anselin, Spatial Econometrics: Methods and Models, in: Studies in Operational Regional Science, Springer Netherlands, Dordrecht, 1988.

[38] Y. Yang, K.K.F. Wong, A spatial econometric approach to model spillover effects in tourism flows, J. Travel Res. 51 (6) (2012) 768–778, http://dx.doi.org/10.1177/0047287512437855.

[39] H. Li, H. Song, L. Li, A dynamic panel data analysis of climate and tourism demand: Additional evidence, J. Travel Res. 56 (2) (2017) 158–171, http://dx.doi.org/10.1177/0047287515626304.

[40] H. Li, M. Hu, G. Li, Forecasting tourism demand with multisource big data, Ann. Tour. Res. 83 (2020) 102912, http://dx.doi.org/10.1016/j.annals.2020.102912.

[41] E. Park, J. Park, M. Hu, Tourism demand forecasting with online news data mining, Ann. Tour. Res. 90 (2021) 103273, http://dx.doi.org/10.1016/j.annals.2021.103273.

[42] T. Peng, J. Chen, C. Wang, Y. Cao, A forecast model of tourism demand driven by social network data, IEEE Access 9 (2021) 109488–109496, http://dx.doi.org/10.1109/ACCESS.2021.3102616.

[43] B. Petrevska, Predicting tourism demand by A.R.I.M.A. Models, Econ. Res.-Ekonomska Istraživanja 30 (1) (2017) 939–950, http://dx.doi.org/10.1080/1331677X.2017.1314822.

[44] Y. Yang, Y. Fan, L. Jiang, X. Liu, Search query and tourism forecasting during the pandemic: When and where can digital footprints be helpful as predictors? Ann. Tour. Res. 93 (2022) 103365, http://dx.doi.org/10.1016/j.annals.2022.103365.

[45] K.K.F. Wong, H. Song, K.S. Chon, Bayesian models for tourism demand forecasting, Tour. Manag. 27 (5) (2006) 773–780, http://dx.doi.org/10.1016/j.tourman.2005.05.017.

[46] Y. Zhang, G. Li, B. Muskat, R. Law, Tourism demand forecasting: A decomposed deep learning approach, J. Travel Res. 60 (5) (2021) 981–997, http://dx.doi.org/10.1177/0047287520919522.

[47] S. Suna, D. Bi, J.-e. Guo, S. Wang, Seasonal and trend forecasting of tourist arrivals: An adaptive multiscale ensemble learning approach, 2020, arXiv:2002.08021 [Cs, Econ, Stat].

[48] D.C. Wu, H. Song, S. Shen, New developments in tourism and hotel demand modeling and forecasting, Int. J. Contemp. Hosp. Manag. 29 (1) (2017) 507–529, http://dx.doi.org/10.1108/IJCHM-05-2015-0249.

[49] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2017, arXiv:1609.02907 [Cs, Stat].

[50] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, W.-c. Woo, Convolutional LSTM network: A machine learning approach for precipitation nowcasting, 2015, arXiv:1506.04214 [Cs].

[51] J. Gehring, M. Auli, D. Grangier, D. Yarats, Y.N. Dauphin, Convolutional sequence to sequence learning, 2017, arXiv:1705.03122 [Cs].

[52] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, arXiv:1412.6980 [Cs].

[53] P. Liashchynskyi, P. Liashchynskyi, Grid search, random search, genetic algorithm: A big comparison for NAS, 2019, http://dx.doi.org/10.48550/arXiv.1912.06059, arXiv:1912.06059.

[54] B. Zhou, G. Yang, Z. Shi, S. Ma, Interpretable temporal attention network for COVID-19 forecasting, Appl. Soft Comput. 120 (2022) 108691, http://dx.doi.org/10.1016/j.asoc.2022.108691.

[55] R. Chen, C.-Y. Liang, W.-C. Hong, D.-X. Gu, Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm, Appl. Soft Comput. 26 (2015) 435–443, http://dx.doi.org/10.1016/j.asoc.2014.10.022.

[56] H. Song, R.T.R. Qiu, J. Park, A review of research on tourism demand forecasting: Launching the annals of tourism research curated collection on tourism demand forecasting, Ann. Tour. Res. 75 (2019) 338–362, http://dx.doi.org/10.1016/j.annals.2018.12.001.

[57] H. Li, L.-Y. He, J.-J. Yang, Forecasting the medium-term performance of restructured tourism firms with an adaptive integrated predictor, Tour. Manag. 88 (2022) 104436, http://dx.doi.org/10.1016/j.tourman.2021.104436.

[58] C. Zheng, X. Fan, C. Wang, J. Qi, GMAN: A graph multi-attention network for traffic prediction, Proc. AAAI Conf. Artif. Intell. 34 (01) (2020) 1234–1241, http://dx.doi.org/10.1609/aaai.v34i01.5477.

[59] S. Guo, Y. Lin, H. Wan, X. Li, G. Cong, Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting, IEEE Trans. Knowl. Data Eng. (2021) 1, http://dx.doi.org/10.1109/TKDE.2021.3056502.