Annals of Tourism Research 90 (2021) 103255

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: https://www.journals.elsevier.com/annals-of-tourism-research

Tourism demand forecasting with time series imaging: A deep learning model

Jian-Wu Bi [a], Hui Li [a,*], Zhi-Ping Fan [b,c]

[a] College of Tourism and Service Management, Nankai University, Tianjin 300350, China
[b] Department of Information Management and Decision Sciences, School of Business Administration, Northeastern University, Shenyang 110167, China
[c] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Article history:
Received 30 September 2020
Received in revised form 31 May 2021
Accepted 1 June 2021
Available online xxxx

Associate editor: Haiyan Song

Keywords:
Tourism demand forecasting
Time series imaging
Deep learning
Convolutional neural networks
Long short-term memory networks

ABSTRACT

To leverage computer vision technology to improve the accuracy of tourism demand forecasting, a model based on deep learning with time series imaging is proposed. The model consists of three parts: sequence image generation, image feature extraction, and model training. In the first part, the tourism demand data are encoded into images. In the second part, the convolution and pooling layers are used to extract features from the obtained images. In the final part, the extracted features are input into long short-term memory networks. Based on historical tourism demand data, the model for forecasting future tourism demand can be obtained. The performance of the proposed model is experimentally assessed through comparing against seven benchmark models.

Introduction

Tourism demand forecasting is of great interest to scholars and tourism practitioners alike (Song et al., 2019; Wu et al., 2020). Accurate forecasts of demand can ensure that resources are efficiently allocated, and allow for rapid adjustments to the supply of tourism products and services as well as the avoidance of imbalances between supply and demand (Li et al., 2019; Li, Ge, et al., 2020; Silva et al., 2019; Zhang, Li, Muskat, & Law, 2020; Zhang, Li, Muskat, Law, & Yang, 2020). For example, if large numbers of tourist arrivals are predicted, then managers can plan accordingly, for instance in their staffing arrangements. Conversely, if small numbers of tourist arrivals are predicted, then managers might formulate pricing strategies or develop attractive tourism packages. Accurate demand forecasting is also of interest to the government at both national and local level, to inform the planning of both tourist attractions and transportation facilities (Bi, Liu, Fan, & Zhang, 2020; Li et al., 2017; Li et al., 2021). Therefore, the development of accurate tourism demand forecasting models is a critical issue.

Unlike time series data in other fields, tourism demand time series data have obvious complexity characteristics, such as nonlinearity, periodicity, and volatility, because the tourism industry itself is characterized by seasonality and uncertainty. Tourism demand time series data do, though, contain a wealth of features that are helpful for forecasting. Although many tourism demand forecasting models have been proposed (Song et al., 2019), most of them can establish only a one-dimensional mapping relationship between each observation within the lag order and the observation to be predicted. Therefore, some important features em-

* Corresponding author.
E-mail addresses: jwbi@nankai.edu.cn, (J.-W. Bi), lihuihit@gmail.com, (H. Li), zpfan@mail.neu.edu.cn. (Z.-P. Fan).

https://doi.org/10.1016/j.annals.2021.103255
0160-7383/

bedded in the tourism demand time series data cannot be fully exploited, which may lead to the loss of some important information in the process of forecasting, such as the temporal correlations and patterns between each observation within the lag order. Thus, to get more accurate forecasts, as many of the features embedded in time series as possible should be utilized in the forecasting model.

As a kind of models that can automatically extract features from data, deep learning has recently been developed for the forecasting of tourism demand (Bi, Liu, & Li, 2020; Law et al., 2019). Because long short-term memory (LSTM) networks can automatically learn the lag order of data, nearly all the deep learning models applied in tourism demand forecasting are LSTM networks (Bi, Liu, & Li, 2020; Law et al., 2019). It should be noted that convolutional neural networks (CNNs), a deep learning technology with good performance in the field of computer vision, has been ignored in this field. However, CNNs can not only extract some features from the data that are different from those extracted by LSTM networks, but also outperform LSTM networks in many fields, especially in image processing and computer vision (Radenović et al., 2018; Zhang et al., 2017). One of the main reasons why CNNs are not used in tourism demand forecasting is that they cannot fully extract data features from one-dimensional tourism demand data.

Recent studies have shown that more of the features embedded in the original time series can be retained by transforming one-dimensional time series data into two-dimensional images in some way. More importantly, CNNs can then be used to extract deeper features from the obtained images. Therefore, if the one-dimensional tourism demand data can be transformed into two-dimensional image data, and the obtained images are mined and processed by CNNs, then the advantages of CNNs in image processing will be fully realized, and the effective and diverse features embedded in the original tourism demand data can be fully extracted. Further, if these extracted features can be utilized by LSTM networks, then the advantages LSTM networks in dealing with sequence data can also be brought into full play, and thus better forecasts can be expected.

In this study, a model for tourism demand forecasting based on deep learning and time series imaging is proposed. It consists of three parts: (1) sequence image generation, (2) image feature extraction, and (3) model training. In part 1, the tourism demand data are first converted into the data type required by deep learning, and then the converted tourism demand data are encoded into images. In part 2, the convolution and pooling layers are used to extract features from the obtained images. In part 3, the extracted features are input into LSTM networks, and the whole model can be trained based on actual tourism demand data. Future tourism demand is forecast by the trained model. The present experimental study draws on real-world tourism demand data from two tourist attractions and compares the proposed new model against seven benchmark models to verify its effectiveness.

Related work

Tourism demand forecasting

The development of accurate tourism demand forecasting models is a critical issue. To this end, many such models have been proposed (Song et al., 2019). They can be broadly divided into three categories: time series models, econometric models, and artificial intelligence models (Song & Li, 2008).

Time series models have been the traditional and most widely used models for tourism demand forecasting (Athanasopoulos et al., 2011; Chan et al., 2005; Cho, 2001; Shahrabi et al., 2013). They include the autoregressive moving average model and exponential smoothing model, and their improved versions (Fildes et al., 2011; Goh & Law, 2002).

The econometric models have been used to capture the causal relationship between the number of tourists and the factors that influence that number (Shen et al., 2009; Wong et al., 2007; Wu et al., 2017). They include the vector autoregressive model, the error correction model, the autoregressive distributed lag model, and the system-of-equation model (Assaf et al., 2018; Li et al., 2006; Song et al., 2003; Turner & Witt, 2001).

Artificial intelligence models are relatively new, but they have proved to be highly accurate in forecasting tourism demand (Song & Li, 2008). They also have other advantages. For example, they do not require any assumptions to be made about the data distribution, and have dynamic adaptability and strong nonlinear fitting capabilities (Law, 2000). Several artificial intelligence models have been developed for tourism demand forecasting (Li et al., 2018). In view of their advantages, the present study focuses on this category of models.

The artificial intelligence models used for tourism demand forecasting can be divided into two categories on the basis of their "depth": shallow learning models and deep learning models (Bi, Liu, & Li, 2020). The former uses an intelligent algorithm with a simple structure and fewer hidden layers, but usually need to have data features manually constructed (Kon & Turner, 2005). Shallow learning models have been widely used in tourism demand forecasting since the 1990s, due to their good nonlinear fitting ability (Law, 2000). They include: multilayer perceptron, radial basis function neural networks, Elman neural networks, denoised neural networks, and support vector machines (Claveria et al., 2015, 2016; Kon & Turner, 2005; Sencheong & Turner, 2005).

Deep learning models use algorithms with a more complex structure and more hidden layers than shallow learning models (Kulshrestha et al., 2020; LeCun et al., 2015; Schmidhuber, 2015). They are end-to-end algorithms that can automatically extract features within a dataset. These models have been successfully applied in many fields, including image recognition and the processing of natural language, and, more recently, the forecasting of tourism demand (Bi, Liu, & Li, 2020; Law et al., 2019). For example, in view of the problem that a large number of search intensity indices are involved when using search engine data to forecast tourism demand, Law et al. (2019) proposed a deep learning approach based on LSTM networks.

It should be noted that there are two main types of deep learning models: convolutional neural networks (CNNs) and recursive neural networks (Schmidhuber, 2015). Nearly all the deep learning models applied in tourism demand forecasting are recursive neural networks, of which the LSTM networks are the most commonly used (Bi, Liu, & Li, 2020; Kulshrestha et al., 2020). Tourism demand forecasting models based on CNNs have not been reported. This may be because CNNs cannot fully extract data features from one-dimensional tourism demand data, which means that the advantages of CNNs cannot be fully realized. However, CNNs perform better than recursive neural networks in many fields, especially in image processing and computer vision (Radenović et al., 2018; Zhang et al., 2017). If the one-dimensional tourism demand data can be transformed into two-dimensional image data in some way, and the obtained images are mined and processed by CNNs, then the advantages of CNNs in image processing will be fully realized, and the effective and diverse features embedded in the original tourism demand data can be fully extracted. Further, if these extracted features can be utilized by LSTM networks, then the advantages of CNNs and LSTM networks can be brought into full play, and better forecasts can be expected.

Forecasting using deep learning architectures with time series imaging

In recent years, computer vision technologies have advanced greatly, and they are now gradually being applied in forecasting research. One of the emerging research areas is time series forecasting using deep learning architectures with imaging. Most of the studies focus on time series classification problems (Hatami et al., 2018; Martínez-Arellano et al., 2019; Wang & Oates, 2015). The main step of these studies is to first encode time series into images, then to extract features from these images through deep learning, and finally to classify the time series based on the obtained features. For example, Hatami et al. (2018) proposed a method for time series classification based on CNNs and recurrence plots. In the method, time series are first transformed into two-dimensional texture images using recurrence plots. Then, a classifier is trained through the CNNs, which learns different levels of representation from the obtained texture images. The authors verified the validity of the method with an experiment based on 20 time series classification datasets. Wang and Oates (2015) developed a novel framework for time series classification based on tiled CNNs, in which three time series encoding methods were used: Gramian angular summation fields, Gramian angular difference fields, and Markov transition fields. The effectiveness of the proposed framework was verified by comparing it with nine state-of-the-art approaches.

Inspired by the above studies, some scholars began to address the regression problems, i.e., time series regression forecasting based on deep learning architectures with time series imaging. There have, though, been few studies on this issue (Li, Kang, & Li, 2020; Zhang & Guo, 2020). Li, Kang, and Li (2020) proposed a new method for regression forecasting with time series imaging and automated features extraction. In the method, the time series are first transformed into recurrence plots. Then, computer vision technologies are used to extract local features from the obtained images. According to the obtained features, an averaging model based on nine commonly used forecast algorithms is proposed. An experimental study was carried out on a forecasting competition dataset, and the results showed that the proposed method performed about as well as other state-of-the-art methods.

Zhang and Guo (2020) proposed an ensemble method for forecasting the hourly consumption of residential electricity. In their method, related data (e.g., weather conditions and residential building data) are first decomposed into several band-limited intrinsic mode functions using variational mode decomposition. Then, the incremental kernel principal components are extracted from the band-limited intrinsic mode functions by the incremental kernel principal component analysis, which are encoded into images based on Gramian angular fields. On this basis, a novel ensemble method is proposed to forecast the hourly consumption of residential electricity, where an elitist search strategy of the multi-population genetic and an improved dragonfly algorithm are developed. The validity of the method is verified by experimental comparison with other state-of-the-art methods.

As noted above, there have been few studies of forecasting using deep learning architectures with time series imaging, as it is still an emerging research field. More particularly, there has been no research on tourism demand forecasting using deep learning with time series imaging. Furthermore, the existing studies mainly use a single method (the recurrence plots or Gramian angular fields) to convert time series into images of a single type, which essentially encodes the time series into black and white images. If the time series can be encoded into multiple images with various features by different encoding methods, and these images are regarded as three "Red-Green-Blue" channels of color images, then more information can be extracted and used in forecasting, thus better predictions can be expected.

Time series imaging

In recent years, deep learning technologies have developed rapidly and have achieved great success in the fields of computer vision, speech recognition, and the processing of natural language. However, since these technologies do not incorporate a time axis, their advantages cannot be fully realized in the forecasting of one-dimensional time series. However, if one-dimensional tourism demand data can be encoded into two-dimensional image data, then the advantages of deep learning in computer vision can be fully realized, and better forecasts can be expected. To fully extract the various features embedded in the tourism demand time series data, three methods are adopted here to transform the time series data into images: Gramian angular field (GAF), Markov transition field (MTF) and recurrence plot (RP). The principles of these three methods are briefly introduced below.

(1) GAF

The main idea of GAF is to encode the time series data as images through the Gramian matrix and polar coordinate systems (Wang & Oates, 2015). To fully capture the information embedded in the original time series data, the polar coordinates are used to represent the time series data to determine the elements in the Gramian matrix. A brief introduction to GAF is given below.

Let $X = \{x_1, x_2, \ldots, x_N\}$ denote the time series data with N observations, where x_n is the nth observation, $n = 1, 2, \ldots, N$. To avoid the result being biased by the observation with the largest value, X is first rescaled to the interval $[-1, 1]$ by Eq. (1).

$$x_n' = \frac{(x_n - \max(X)) + (x_n - \min(X))}{\max(X) - \min(X)}, \quad n = 1, 2, \ldots, N \tag{1}$$

where x_n' is the rescaled value of x_n.

Then, the rescaled data x_n' can be represented as a polar coordinate (ϕ_n, r_n), where ϕ_n and r_n are the angle and radius in polar coordinates, respectively. (ϕ_n, r_n) can be calculated using Eq. (2):

$$\begin{cases} \phi_n = \arccos(x_n') \\ r_n = \dfrac{n}{T} \end{cases} n = 1, 2, \ldots, N \tag{2}$$

where n and T respectively denote the time stamp and the constant factor for regularizing the span of the polar coordinates.

According to the obtained (ϕ_n, r_n), the element (n', n'') in the Gramian matrix can be determined by calculating the trigonometric difference between each observation:

$$GAF_{n',n''} = \sin(\phi_{n'} - \phi_{n''}), n', n'' = 1, 2, \ldots, N \tag{3}$$

Based on the obtained $GAF_{n',\ n''}$, $X = \{x_1, x_2, \ldots, x_N\}$ can be converted into an image. The obtained image can well capture the temporal correlation within different time intervals.

(2) MTF

The main idea of MFT is to convert the time series data into images through the Markov matrix of quantile bins (Wang & Oates, 2015). MTF preserves time domain information in time series by representing the Markov transition probabilities sequentially. A brief introduction to MTF is given below.

The time series data $X = \{x_1, x_2, \ldots, x_N\}$ are first classified into Q quantile bins, and each observation x_n is assigned to a corresponding bin, q_m, $m = 1, 2, \ldots, Q$. Then, a weighted adjacency matrix $W = [w_{mm'}]_{Q \times Q}$ can be constructed by counting transitions among these quantile bins along the time axis in the form of a first-order Markov chain. The element $w_{mm'}$ is determined by the frequency of a point in the quantile $q_{m'}$ followed by a point in the quantile q_m. Thus, $w_{mm'}$ can be determined by Eq. (4):

$$\begin{cases} w_{mm'} = P(x_n \in q_m | x_{n-1} \in q_{m'}) \\ \sum_{m'} w_{mm'} = 1 \end{cases} m, m' = 1, 2, \ldots, Q \tag{4}$$

By considering the temporal positions, the weighted adjacency matrix W can be expanded into a MTF matrix $D = [d_{mm'}]_{Q \times Q}$, where $d_{mm'}$ is the transition probability of $q_m \to q_{m'}$ (i.e., $P(q_m \to q_{m'})$), such that $\sum_{m'} d_{mm'} = 1$, $m, m' = 1, 2, \ldots, Q$. Finally, by averaging the pixels in each non-overlapping region, the MTF image corresponding to X can be obtained.

(3) RP

RP is an important method to analyze the periodicity, chaos, and non-stationarity of time series, which provides a way to visualize the periodicity of trajectories in a phase space (Li, Kang, & Li, 2020; Marwan et al., 2007). It can reveal the internal structure of time series, particularly in terms of similarity and stability. The RP is especially suitable for the analysis of time series data drawn from a short period. A brief introduction to RP is given below.

Given a time series $X = \{x_1, x_2, \ldots, x_N\}$, an extracted trajectory with m dimensions can be defined as:

$$\vec{x}_n = \left(x_n, x_{n+\tau}, \ldots, x_{n+(m-1)\tau}\right), n = 1, 2, \ldots, N-(m-1)\tau \tag{5}$$

where τ denotes the time delay.

Then, the RP of X can be determined by calculating the pairwise distance between the trajectories, using Eq. (6):

$$R(n, n') = \begin{cases} 1, & \|\vec{x}_n - \vec{x}_{n'}\| \geq \theta \\ 0, & \|\vec{x}_n - \vec{x}_{n'}\| < \theta \end{cases} n, n' = 1, 2, \ldots, N-(m-1)\tau \tag{6}$$

where θ denotes the threshold.

It can be seen from Eq. (6) that $R(n,n')$ is a binary function. In other words, if the difference between $\vec{x}_n$ and $\vec{x}_{n'}$ is greater than the threshold θ, then a black dot is drawn in the RP image. This binary output in standard RP may cause some information loss. To capture more information from time series, a modified RP is used in this study, i.e.,

$$R(n,n') = \begin{cases} \theta, & \|\vec{x}_n - \vec{x}_{n'}\| \geq \theta \\ \|\vec{x}_n - \vec{x}_{n'}\|, & \|\vec{x}_n - \vec{x}_{n'}\| < \theta \end{cases}, n, n' = 1, 2, \ldots, N-(m-1)\tau \quad (7)$$

According to Eq. (7), the modified RP can output more values than the standard RP.

(4) Typical examples of GAF, MTF, and RP images

To show these three kinds of images more intuitively, typical examples of GAF, MTF and RP images are given in Figs. 1–3, where "Figs. 1(b), 2(b), and 3(b)", "Figs. 1(c), 2(c), and 3(c)", and "Figs. 1(d), 2(d), and 3(d)" respectively represent the GAF, MTF and RP images corresponding to a random time series, a periodic time series, and a "periodic + trend" time series.

In Fig. 1, the distribution of the values in all three types of images corresponding to the random time series show no regularity. The main difference among these three types of images is the difference of range.

In Fig. 2, the distribution of the values in the three types of images corresponding to the periodic time series shows obvious periodicity. It can be seen that there are some differences in the distribution of the values within these three images. Specifically, the regularity of the distribution in the GAF and RP images is stronger than that of the MTF image, and the range values of the GAF and RP images is larger than that of the MTF image. The GAF and RP images are complementary to each other from the vertical axis to a certain extent, and the value ranges of the GAF and RP images are [−1, 1] and [0, 2], respectively.

In Fig. 3, the distribution of the values with respect to the three types of images corresponding to the "periodic + trend" time series shows obvious periodicity and trend. The distribution of the values of the three images is again different. Specifically, the distribution of the values within the GAF image increases periodically from the lower left corner to the upper right corner; the distribution of the values in the MTF image shows a periodically decreasing trend with the main diagonal as the symmetry axis, and a periodically increasing trend with the sub-diagonal as the symmetry axis; the distribution of the values in the RP image shows a periodically increasing trend with the main diagonal as the symmetry axis.

Methodology

Principle and framework of the model

Tourism demand is affected by a variety of factors, resulting in the obvious complexity characteristics of tourism demand time series data, such as nonlinearity, periodicity, and volatility. When the existing models are used, the features embedded in the tourism demand time series data cannot be fully exploited and utilized. The reasons are as follows: (1) The traditional models establish only a one-dimensional mapping relationship between each observation within the lag order and the observation to be predicted. For example, Fig. 4(a) is a tourism demand time series with six observations. Assuming that the lag order is 5, then the mapping relationship constructed by a traditional model is shown in Fig. 4(b). If there are relationships between observations within the lag order, such as temporal correlations and patterns between observations, these relationships may have an impact on the observation to be predicted, but they are ignored by traditional models. Furthermore, the traditional models can use only a fixed set of lagged observations for forecasting. However, the temporal dependence (lag order) may vary with circumstance, and this dynamic feature cannot be captured by traditional models. Thus, to get more accurate forecasts, more features embedded in time series should be used.

If the tourism demand time series can be transformed into the mapping relationship shown in Fig. 4(c), i.e., the tourism demand time series are encoded into images, then more features embedded in the data can be utilized in the construction of the model. Then, CNNs, which are very good at image processing, can be used to extract features from the obtained images and LSTM networks, which are well suited to processing sequence data, can be used to automatically learn the temporal dependence between the features obtained by the CNNs. Through this process, the advantages of time series imaging in feature extraction, CNNs in image processing and LSTM networks in sequence data processing can be given full play, the rich deep-seated features embedded in the tourism demand time series can be fully mined and utilized, thus better forecasts can be expected.

Based on the above analysis, the present study proposes a novel deep learning model for forecasting tourism demand. Its framework is shown in Fig. 5. The model consists of three parts, i.e., sequence image generation, image feature extraction, and model training.

Sequence image generation

Tourism demand data has obvious complexity characteristics, such as nonlinearity, periodicity, and volatility. The main purpose of this part is to extract as many as possible of the features embedded in the tourism demand data, through time series imaging. This is done in two steps: (1) converting the tourism demand data into the data type required by deep learning; and (2) encoding time series data into images.

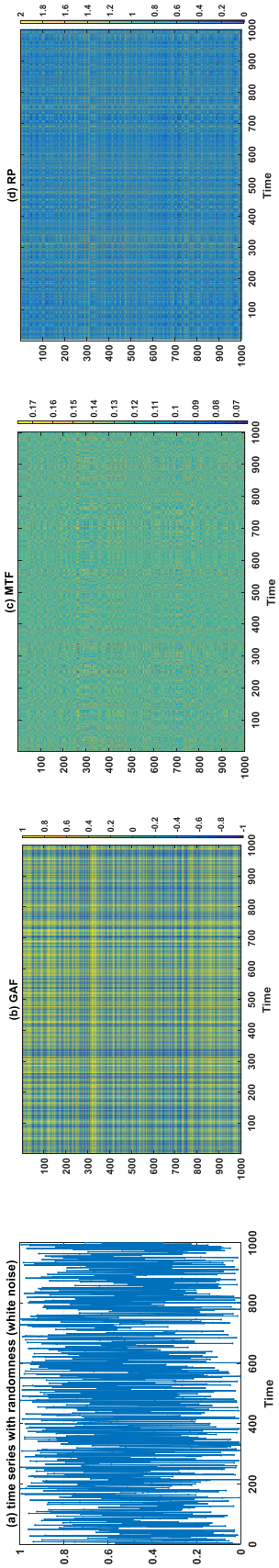

Fig. 1. Typical examples of GAF, MTF, and RP images for random time series data.

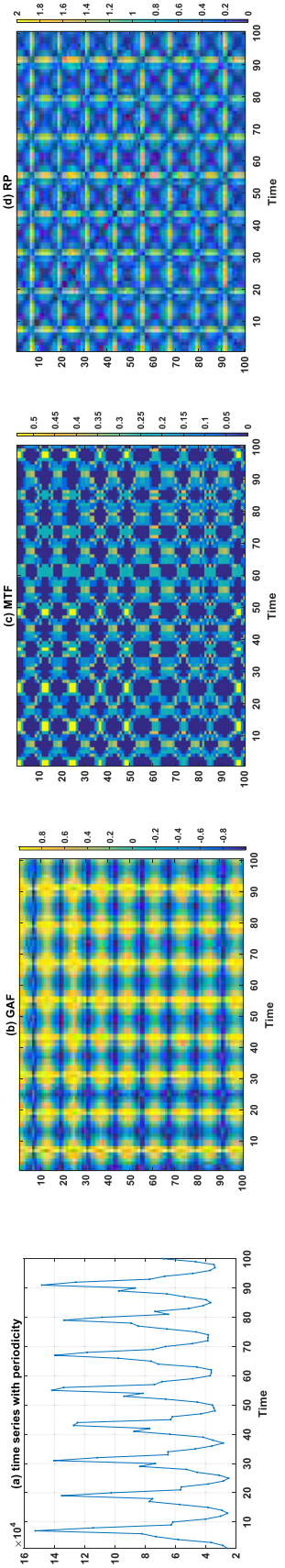

Fig. 2. Typical examples of GAF, MTF, and RP images for a periodic a time series.

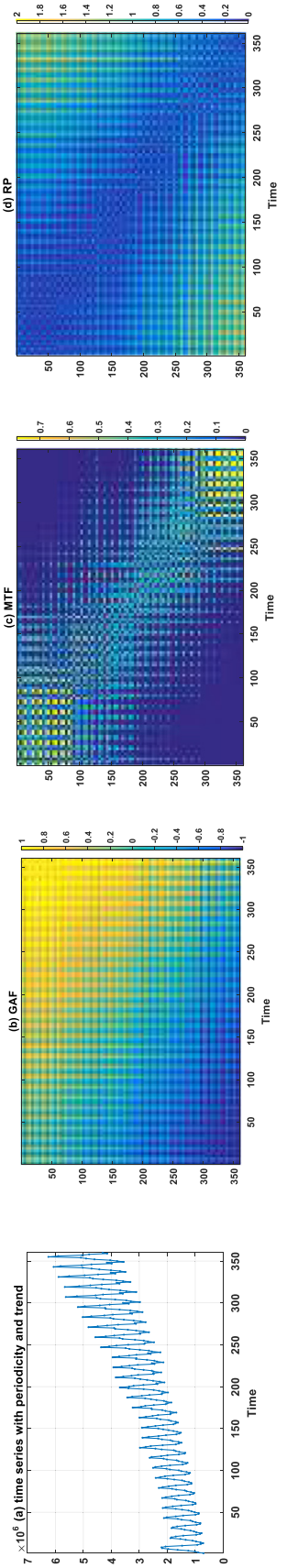

Fig. 3. Typical examples of GAF, MTF, and RP images for time series with periodicity plus a trend.

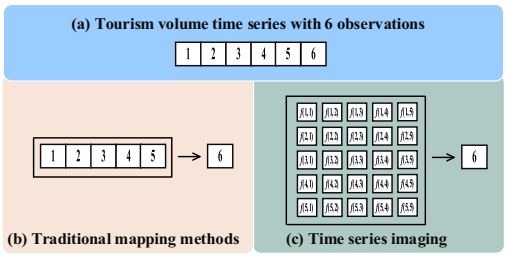

Fig. 4. An example of data conversion.

(1) Converting the tourism demand data into the data type required by deep learning

Let $X = (x_1,x_2,x_3,x_4,...,x_{T-3},x_{T-2},x_{T-1},x_T)$ represent a time series of tourism demand with T observation points, where x_t is the tourism demand at time t, $t = 1, 2, ..., T$. Tourism demand data X are in the form of a time series, but deep learning models require the data to be in the form of input-output pair sequences. Before the model can be built, it is therefore necessary to convert X into input-output pair sequences. Let the length of the input sequence be n; that is, n previous observations of tourism demand are used to forecast the tourism demand of the following (hth) observation, $n \in \{1,2,...,T-1\}$. The value of n should generally be greater than the period in the tourism demand data X. X can then be converted into sequences of input-output pairs, i.e., $X_j \to x_{n+j-1-h}$, where $X_j = (x_j,x_{j+1},...,x_{n+j-1})$, $j = 1, 2, ..., T - n + 1 - h$. $X_j \to x_{n+j-1-h}$ can be expanded into the following matrix form:

$$\begin{matrix} v_1 & v_2 & \cdots & v_n & y \end{matrix}$$
$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ \cdots & \cdots & \cdots & \cdots \\ x_j & x_{j+1} & \cdots & x_{n+j-1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{T-n+1-h} & x_{T-n+2-h} & \cdots & x_{T-h} \end{bmatrix} \to \begin{bmatrix} x_{n+h} \\ \cdots \\ x_{n+j-1-h} \\ \cdots \\ x_T \end{bmatrix}$$

where each row in the left-hand matrix represents an input sequence for the deep learning model, and the corresponding element in the right-hand matrix represents an output.

(2) Encoding time series data into images

To accurately forecast tourism demand, the forecasting model needs to be able to capture the global changes in tourism demand and to fit the local change in tourism demand. The global change in tourism demand is mainly reflected in scale variance, i.e., the global seasonality and scale (or level). The local change in tourism demand is mainly reflected in the local nonlinear trend, i.e., the local recurring patterns and temporary relations. Although deep learning models have strong nonlinear fitting capabilities, they tend to struggle with scale variance. Compared with deep learning models, although some traditional models, such as the Holt-Winter model, have no nonlinear ability, they are good at capturing the scale variance of tourism demand (Smyl et al., 2018). If the scale variance of tourism demand can be eliminated by the Holt-Winter model, then the nonlinear fitting ability of deep learning models can be fully utilized, and better forecasts can be expected. In this study, the Holt-Winter model is adopted to normalize the data:

$$\hat{x}_t = \frac{y_t}{l_t \times s_t}, j = 1,2,...,T-n+1-h \tag{8}$$

where $\hat{x}_t$ is the normalized and deseasonalized value of x_t, which can reflect the local nonlinear trends in tourism demand; y_t is the smoothed value of x_t; s_t are the seasonal correction factors (which reflect the seasonality of tourism demand); and l_t is the smoothed value of the constant part for x_t, which reflects the scale of tourism demand.

According to Eq. (8), for each X_j, a corresponding normalized vector can be obtained, i.e., $\hat{X}_j = (\hat{x}_j,\hat{x}_{j+1},...,\hat{x}_{n+j-1})$, $j = 1, 2, ..., T - n + 1 - h$. To fully extract the various features embedded in the tourism demand time series data, three methods (GAF, MTF, and RP) are adopted to encode the time series data into images. Thus, using Eqs. (1)–(3), (4), and (5)–(7), $\hat{X}_j$ is encoded into a GAF image, a MTF image, and a RP image, respectively. They are denoted as $GAF\left(\hat{X}_j\right)$, $MTF\left(\hat{X}_j\right)$ and $RP\left(\hat{X}_j\right)$, respectively, $j = 1, 2, ..., T - n + 1 - h$.

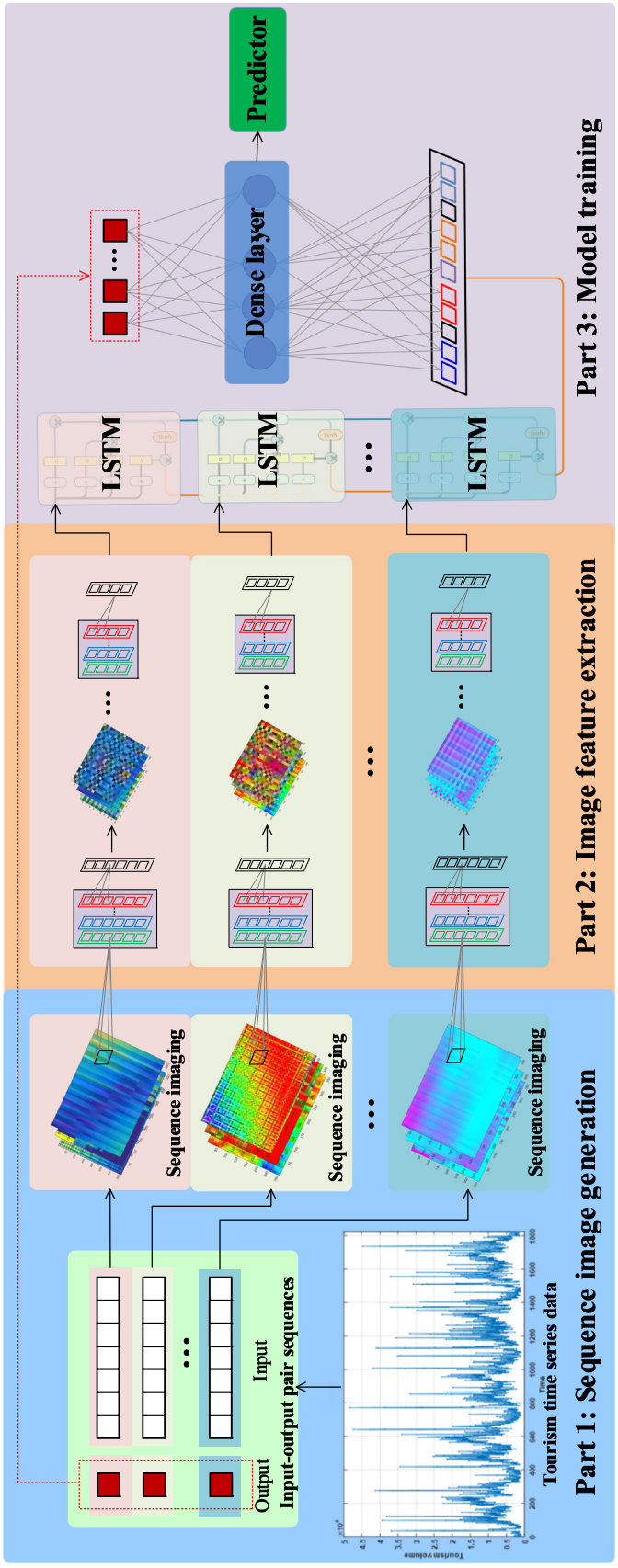

Fig. 5. The framework of the proposed model.

Image feature extraction

To extract deeper features from the obtained images, the CNNs, which are very suitable for image feature extraction, are adopted in this part. Specifically, the convolution and pooling layers are used to extract features from $GAF\left(\tilde{X}_j\right)$, $MTF\left(\tilde{X}_j\right)$ and $RP\left(\tilde{X}_j\right)$, where $GAF\left(\tilde{X}_j\right)$, $MTF\left(\tilde{X}_j\right)$ and $RP\left(\tilde{X}_j\right)$ are regarded as three channels of an image, $j = 1, 2, ..., T - n + 1 - h$.

The process of feature extraction using convolution layers is shown in Fig. 6, where the image is regarded as a two-dimensional matrix. The specific steps of image convolution are as follows: slide the convolution kernel on the image, multiply the pixel value on the image point by the value on the corresponding convolution kernel, and then add all the multiplied values. Repeat this process until the convolution kernel slides the entire image. Drawing on Krizhevsky et al. (2017) and Bi, Liu, and Fan (2020), the mathematical principles are as follows.

Let $I^{\#}\left(\tilde{X}_j\right)$ denote the image corresponding to $\tilde{X}_j$, $\# \in \{GAF, MTF, RP\}$. Without loss of generality, $I^{\#}\left(\tilde{X}_j\right)$ can be written as an $n \times n$ matrix, i.e.,

$$I^{\#}\left(\tilde{X}_j\right) = \begin{bmatrix} \delta_{11}^{j\#} & \cdots & \delta_{1k}^{j\#} & \cdots & \delta_{1n}^{j\#} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \delta_{k1}^{j\#} & \cdots & \delta_{kk}^{j\#} & \cdots & \delta_{kn}^{j\#} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \delta_{n1}^{j\#} & \cdots & \delta_{nk}^{j\#} & \cdots & \delta_{nn}^{j\#} \end{bmatrix}, j = 1, 2, ..., T-n+1-h, n \in \{1, 2, ..., T-1\}.$$

Let $I_S^{\#}\left(\tilde{X}_j\right)$ be a $(k + 1) \times (k + 1)$ submatrix of $I^{\#}\left(\tilde{X}_j\right)$, i.e.,

$$I_S^{\#}\left(\tilde{X}_j\right)_{k':(k'+k-1),k'':(k''+k-1)} = \begin{bmatrix} \delta_{k'k''}^{j\#} & \cdots & \delta_{(k')(k''+k)}^{j\#} \\ \cdots & \cdots & \cdots \\ \delta_{(k'+k)(k'')}^{j\#} & \cdots & \delta_{(k'+k)(k''+k)}^{j\#} \end{bmatrix}, k, k', k'' \in N^+, k + k', k + k'' \le n.$$

Let $c_{k'k''}^{j\#}$ denote the convolution feature generated from $I^{\#}\left(\tilde{X}_j\right)_{k':(k'+k-1),k'':(k''+k-1)}$, and $c_{k'k''}^{j\#}$ can be calculated by:

$$c_{k'k''}^{j\#} = \Theta\left(W_C * I^{\#}\left(\tilde{X}_j\right)_{k':(k'+k-1),k'':(k''+k-1)} + b\right), k, k', k'' \in N^+, k + k', k + k'' \le n \tag{9}$$

where Θ denotes the activation function, $W_C \in \mathbb{R}^{k \times k}$ is the weight matrix of the convolution kernel, $*$ denotes the convolution operation, and $b \in \mathbb{R}$ is the bias term.

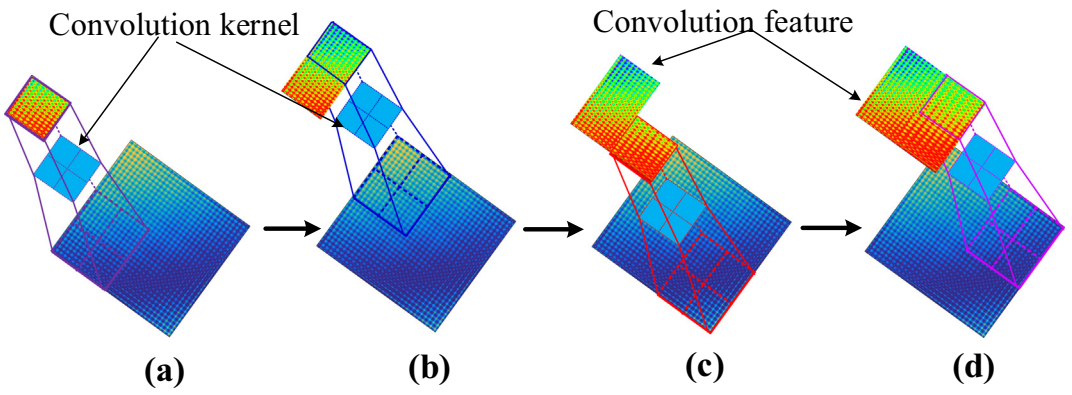

Fig. 6. The process of feature extraction using convolution layers.

According to Eq. (9), the convolution feature of $I\left(\tilde{X}_j\right)$ can be obtained, i.e.,

$$C^{\#}\left(\tilde{X}_j\right)=\begin{bmatrix} c_{11}^{j\#} & \cdots & c_{1k''}^{j\#} & \cdots & c_{1(n-k+1)}^{j\#} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{k'1}^{j\#} & \cdots & c_{k'k''}^{j\#} & \cdots & c_{k'(n-k+1)}^{j\#} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{(n-k+1)1}^{j\#} & \cdots & c_{(n-k+1)k''}^{j\#} & \cdots & c_{(n-k+1)(n-k+1)}^{j\#} \end{bmatrix}, j=1,2,...,T-n+1-h, n\in\{1,2,...,T-1\}.$$

To explain the calculation process more clearly, an example is given here. The activation function is assumed to be linear, i.e., $\Theta(x)=x$; the weight matrix of the convolution kernel is $W_C=\begin{bmatrix}1&0\\0&1\end{bmatrix}$, the two-dimensional matrix of an image is $I\left(\tilde{X}_j\right)=\begin{bmatrix}1&2&1&0&0\\0&1&3&1&0\\0&0&1&1&1\\0&0&3&1&0\\0&3&1&0&0\end{bmatrix}$, and the bias term is $b=0$. According to Eq. (9), the convolution feature of $I\left(\tilde{X}_j\right)$ can be determined, i.e.,

$$C\left(\tilde{X}_j\right)=\begin{bmatrix}2&5&2&0\\0&2&4&2\\0&3&2&1\\3&1&3&1\end{bmatrix}.$$

where the 2 in the upper left corner of $C\left(\tilde{X}_j\right)$ is determined by the sum of the products of the matrix $I\left(\tilde{X}_j\right)_{1:2,1:2}=\begin{bmatrix}1&2\\0&1\end{bmatrix}$ and the weight matrix of the convolution kernel $W_C=\begin{bmatrix}1&0\\0&1\end{bmatrix}$, i.e., $1\times1+2\times0+0\times0+1\times1=2$.

Based on the obtained $C^{\#}\left(\tilde{X}_j\right)$, to retain the main features, reduce the number of parameters, and prevent over-fitting, the pooling layer is used. Let $P^{\#}\left(\tilde{X}_j\right)$ denote the pooling result of $C^{\#}\left(\tilde{X}_j\right)$ obtained by a $f\times f$ filter with stride s, i.e.,

$$P^{\#}\left(\tilde{X}_j\right)=\begin{bmatrix} p_{11}^{j\#} & \cdots & p_{1z}^{j\#} & \cdots & p_{1Z}^{j\#} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{z1}^{j\#} & \cdots & p_{zz}^{j\#} & \cdots & p_{zZ}^{j\#} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{Z1}^{j\#} & \cdots & p_{Zz}^{j\#} & \cdots & p_{ZZ}^{j\#} \end{bmatrix}.$$

The elements in $P^{\#}\left(\tilde{X}_j\right)$ can be calculated by two methods, namely maximum pooling and average pooling, which can be determined by Eqs. (10) and (11), respectively.

$$p_{zz'}^{j\#}=\max\left\{c^{\#}_{(1+(z-1)\times s):(f+(z-1)\times s),(1+(z'-1)\times s):(f+(z'-1)\times s)}\right\} z,z'=1,2,...,Z \tag{10}$$

$$p_{zz'}^{j\#}=\mathrm{sum}\left\{c^{\#}_{(1+(z-1)\times s):(f+(z-1)\times s),(1+(z'-1)\times s):(f+(z'-1)\times s)}\right\}/(f\times f), z,z'=1,2,...,Z \tag{11}$$

To illustrate the pooling process more clearly, an example is shown in Fig. 7, where Fig. 7(a) and (b) represent the calculation process of maximum pooling and average pooling, respectively. Here, the partial matrix shaded red (■) in the matrix illustrates

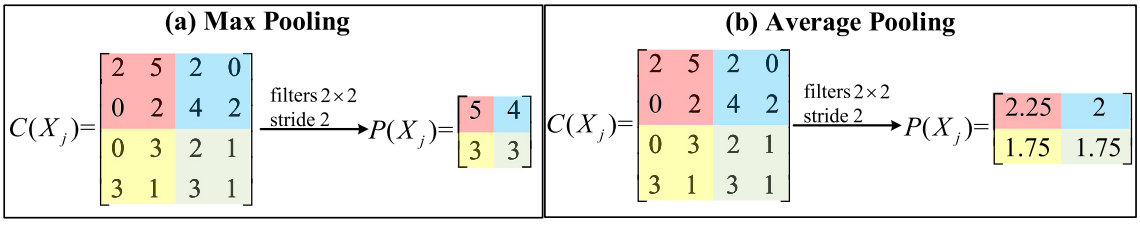

Fig. 7. An example of the pooling process.

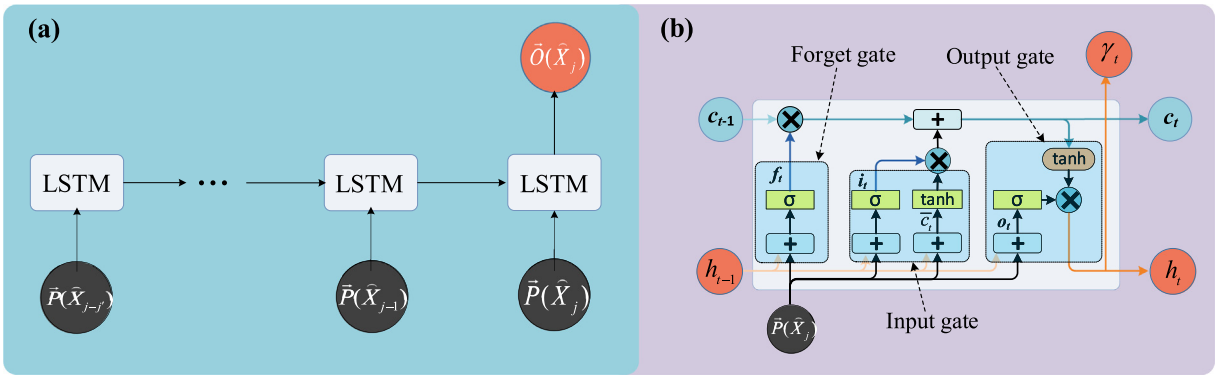

Fig. 8. The framework of the LSTM networks.

the calculation process. In Fig. 7(a), the 5 in $P^{\#}\left(\widehat{X}_j\right)$ is determined by $\max\{2,5,0,2\} = 5$. In Fig. 7(b), the 2.25 in $P^{\#}\left(\widehat{X}_j\right)$ is determined by $(2 + 5 + 0 + 2) \div 4 = 2.25$.

Model training

Based on the obtained $P^{\#}\left(\widehat{X}_j\right)$, a one-dimensional vector with $Z \times Z$ elements can be obtained, i.e., $\overrightarrow{P}^{\#}\left(\widehat{X}_j\right) = \left(p_{11}^{j\#}, ..., p_{1z}^{j\#}, ..., p_{z1}^{j\#}, ..., p_{zz}^{j\#}, ..., p_{ZZ}^{j\#}\right)$, through a flatten layer. Since $\# \in \{GAF, MTF, RP\}$, three vectors can be obtained: $\overrightarrow{P}^{GAF}\left(\widehat{X}_j\right) = \left(p_{11}^{jGAF}, ..., p_{zz}^{jGAF}, ..., p_{ZZ}^{jGAF}\right)$, $\overrightarrow{P}^{MTF}\left(\widehat{X}_j\right) = \left(p_{11}^{jMTF}, ..., p_{zz}^{jMTF}, ..., p_{ZZ}^{jMTF}\right)$ and $\overrightarrow{P}^{RP}\left(\widehat{X}_j\right) = \left(p_{11}^{jRP}, ..., p_{zz}^{jRP}, ..., p_{ZZ}^{jRP}\right)$. Then, by combining vectors $\overrightarrow{P}^{GAF}\left(\widehat{X}_j\right)$, $\overrightarrow{P}^{MTF}\left(\widehat{X}_j\right)$ and $\overrightarrow{P}^{RP}\left(\widehat{X}_j\right)$, a new one-dimensional vector with $3 \times Z \times Z$ elements can be obtained, i.e., $\vec{P}\left(\widehat{X}_j\right) = \left(p_{11}^{jGAF}, ..., p_{zz}^{jGAF}, ..., p_{ZZ}^{jGAF}, p_{11}^{jMTF}, ..., p_{zz}^{jMTF}, ..., p_{ZZ}^{jMTF}, p_{11}^{jRP}, ..., p_{zz}^{jRP}, ..., p_{ZZ}^{jRP}\right)$.

Based on $\vec{P}\left(\widehat{X}_j\right)$, LSTM networks with $j - j' + 1$ time steps can be constructed, where $\vec{P}\left(\widehat{X}_{j-j'}\right), ..., \vec{P}\left(\widehat{X}_{j-1}\right), \vec{P}\left(\widehat{X}_j\right)$ are the inputs. The structure of the constructed LSTM networks is shown in Fig. 8(a), and the structure of a module in the LSTM networks is shown in Fig. 8(b). As can be seen from Fig. 8, the LSTM networks have three gates: a "forget gate", an "input gate", and an "output gate" (Hochreiter & Schmidhuber, 1997). The forget gate is mainly to "remember" important information, and "forget" useless information, i.e., selectively forget the information from the previous module. Let f_j denote the output of the "forget gate". f_j can be calculated by

$$f_j = \sigma\left(w_f \cdot \left[h_{j-1}, x_j\right] + b_f\right) \tag{12}$$

where w_f and b_f are the weight matrix and bias vector for the "forget gate", respectively.

The "input gate" selectively memorizes the information from the current inputs: it retains ("remembers") any information that is more important for forecasting. The output of the input gate, denoted i_j, can be calculated by

$$i_j = \sigma\left(w_i \cdot \left[h_{j-1}, x_j\right] + b_i\right) \tag{13}$$

where w_i and b_i are the weight matrix and bias vector for that gate, respectively.

The cell state of the LSTM module, which gives LSTM networks their long-term memory function, is denoted c_j. Using Eq. (14), c_j is calculated from the obtained f_j and i_j:

$$\begin{cases} \bar{c}_j = \tanh\left(w_c \cdot \left[h_{j-1}, x_j\right] + b_c\right) \\ c_j = f_j * c_{j-1} + i_j * \bar{c}_j \end{cases} \tag{14}$$

where w_c and b_c are the weight matrix and bias vector for the cell state, respectively.

Finally, the "output gate" contains previously input relevant information and determines the value of the next hidden state. The output of the "output gate", denoted o_j, can be calculated from Eq. (15):

$$o_j = \sigma\left(w_o \cdot \left[h_{j-1}, \alpha_j\right] + b_o\right) \tag{15}$$

where w_o and b_o are respectively the weight matrix and bias vector for that gate.

With c_j and o_j obtained, the final output of the LSTM module can be calculated using Eq. (16):

$$\gamma_j = o_j * \tanh\left(c_j\right) \tag{16}$$

Through the above information updating process, the final output of the LSTM networks can be obtained, denoted as $\vec{O}\left(\widetilde{X}_j\right)$, $j = 1, 2, ..., T - n + 1 - h$.

With $\vec{O}\left(\widetilde{X}_j\right)$ as the input and $x_{n+j-1-h}$ as the output, the functional relationship between $\vec{O}\left(\widetilde{X}_j\right)$ and $x_{n+j-1-h}$ can be constructed through a full connection layer, i.e.,

$$x_{n+j-1-h} = \Theta\left(W_F \cdot \vec{O}\left(\widetilde{X}_j\right) + b_F\right), j = 1, 2, ..., T-n+1-h \tag{17}$$

where Θ denotes an activation function, W_F and b_F are the weight matrix and bias terms; and $\cdot$ is a multiplication factor that is applied to the matrices.

By feeding the actual tourism demand data into the above-constructed model, the relevant parameters can be determined, that is, the neural network model with actual forecasting ability can be obtained. It should be noted that since the model is trained on the normalized data, an inverse normalization is needed in the actual forecasting. Let Image(X_j) denote the output of the trained neural network with inputs $X_{j-j'}$, ..., X_j, $j, j' = 1, 2, ..., T - n + 1 - h$, $j > j'$. Then, the tourism demand at time $j + h$ can be obtained using Eq. (18):

$$y_{j+h} = \text{Image}\left(X_j\right) \times l_j \times s_{j+h-m} \tag{18}$$

where Image(X_j) reflects the local recurring patterns and temporary relations in tourism demand; l_t and s_{j+h-m} respectively reflect the global scale and seasonality in tourism demand, which are determined by the Holt-Winter model.

It can be seen from Eq. (18) that the proposed model simultaneously considers the scale change in tourism demand, the seasonality change in tourism demand, and the local recurring patterns and temporary relations in tourism demand.

Empirical study

Experimental design

Data

To verify the effectiveness of the proposed model, an empirical analysis is conducted using daily tourist volume at two well-known attractions in China, Jiuzhaigou and Mount Siguniang. For Jiuzhaigou, the time span is from Jan. 1, 2013 to Oct. 31, 2020, with 2845 observations (Fig. 9(a)). It should be noted that on Aug. 8, 2017, an earthquake in the scenic area caused great damage to the park and its surrounding areas. Thereafter, through to Sep. 27, 2019, the scenic area was initially closed and then measures were taken to limit the number of tourists. Shortly after that, due to the outbreak of COVID-19 in China, the scenic area was closed again on Jan. 28, 2020 until Mar. 31, 2020. For Mount Siguniang, the time span is from Jan. 2, 2016 to Oct. 31, 2020, with 1765 observations (Fig. 9(b)). Similarly, with the outbreak of the pandemic in China, the scenic area was closed on Jan. 28, 2020 until Mar. 31, 2020. Therefore, the data for Jiuzhaigou from Aug. 8, 2017 to Mar. 31, 2020 and the data for Mount Siguniang from Jan. 28, 2020 to Mar. 31, 2020 are not used in the experiment.

To fully verify the performance of different models, two experiments are conducted: (1) forecasting before the onset of the COVID-19 pandemic; and (2) forecasting during the pandemic. Therefore, each dataset is twice divided into a training set and a test set, i.e., four training sets and four test sets are obtained, where the time span of two test sets is before the outbreak of the COVID-19, and the time span of the other two test sets is during the pandemic. The details with respect to the time span and number of observations for the training set and test set are shown in Table 1.

It should be noted that both long-term and short-term forecasts are important in tourism decision-making. Therefore, it is necessary to compare the performance of different models over different prediction horizons, that is, to compare the performance of different models in h-step-ahead forecasting, where h is set to 1 and 3 in the experiment.

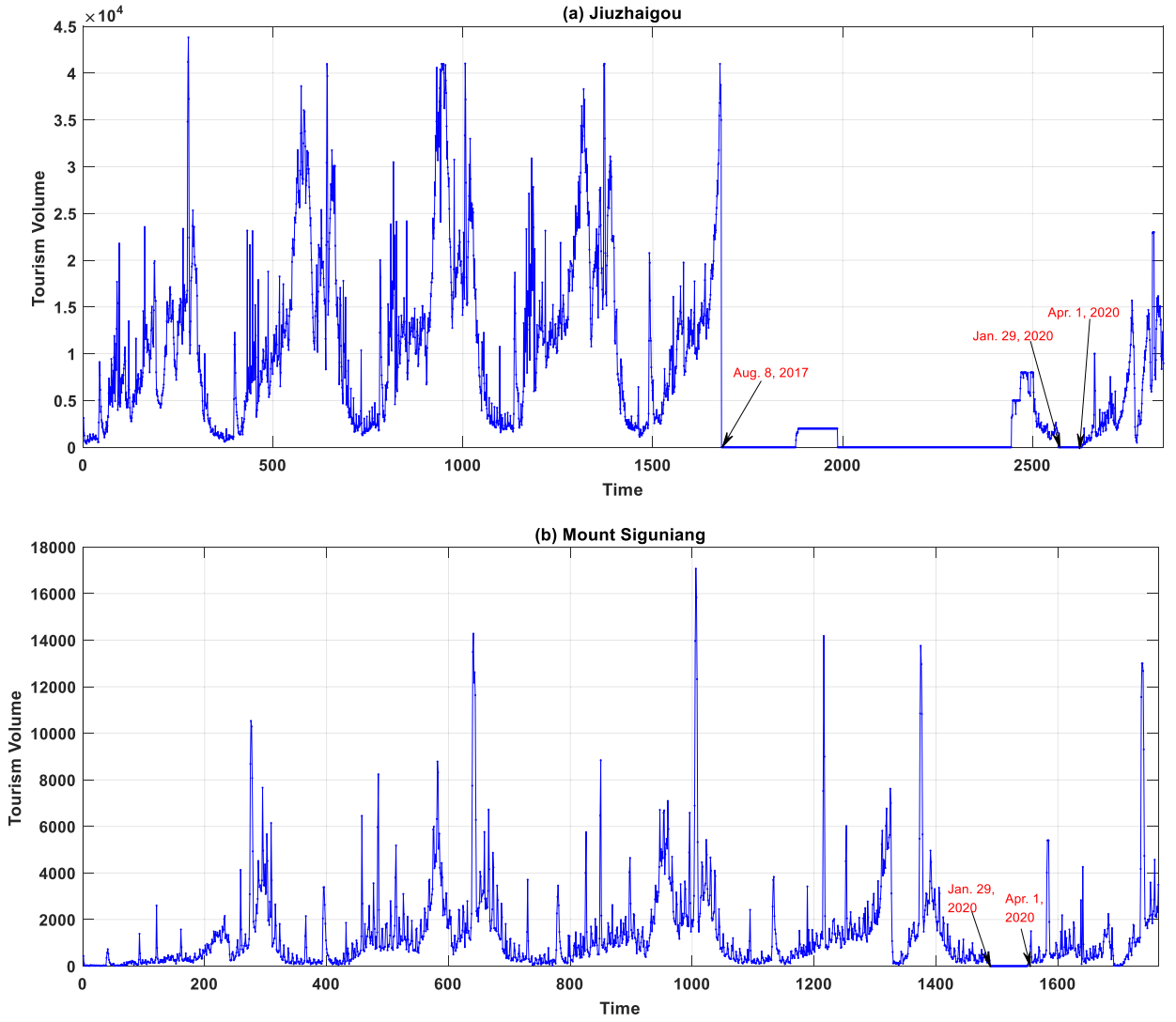

Fig. 9. The daily tourist volume at two famous tourist attractions in China.

Performance measures

Three measures of performance are adopted: mean absolute deviation (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). They are calculated in the usual way, using Eqs. (19)–(21):

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |\tau_t - \tau_t'| \tag{19}$$

Table 1
The details for the training set and test set in each dataset.

Time of test set	Dataset	Time span		Number of observations	
		Training set	Test set	Training set	Test set
Forecasting before the outbreak of the COVID-19	Jiuzhaigou	Jan. 1, 2013 to Apr. 1, 2015	Apr. 2, 2015 to Jun. 30, 2017	821	821
	Mount Siguniang	Jan. 2, 2016 to Sep. 15, 2017	Sep. 16, 2017 to May 31, 2019	623	623
Forecasting during the COVID-19	Jiuzhaigou	Jan. 1, 2013 to Aug. 7, 2017	Apr. 1, 2020 to Oct. 31, 2020	1680	214
	Mount Siguniang	Jan. 2, 2016 to Jan. 28, 2020	Apr. 1, 2020 to Oct. 31, 2020	1488	214

Table 2
Optimal n for each model with respect to different values of h.

Models	Jiuzhaigou		Mount Siguniang	
	$h = 1$	$h = 3$	$h = 1$	$h = 3$
SVM	$n = 9$	$n = 6$	$n = 9$	$n = 12$
BPNN	$n = 15$	$n = 9$	$n = 9$	$n = 9$
CNN	$n = 6$	$n = 12$	$n = 12$	$n = 15$
LSTM	$n = 12$	$n = 6$	$n = 6$	$n = 15$
CNN-LSTM	$n = 6$	$n = 6$	$n = 6$	$n = 9$
The proposed model	$n = 9$	$n = 15$	$n = 12$	$n = 12$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\tau_t - \tau_t'}{\tau_t}\right| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\tau_t - \tau_t')^2} \tag{21}$$

where τ_t and τ_t' are the actual and predicted tourist volumes at time t, respectively; T is the total number of predicted samples. Obviously, the smaller are the values of MAE, RMSE, and MAPE, the better is the performance of the given model.

Benchmark models and parameter setting

To fully illustrate the effectiveness of the proposed model, three kinds of benchmark models are selected: traditional time series models, shallow machine learning models, and deep learning models. The first group includes the autoregressive integrated moving average model (ARIMA) and seasonal autoregressive integrated moving average model (SARIMA); the second group includes support vector machine (SVM) and back propagation neural network (BPNN); and the last group includes CNN, LSTM, and "CNN-LSTM", where "CNN-LSTM" means that the CNN is first used to perform one-dimensional convolution on the time series to extract features, and then a LSTM network model is built based on the extracted features.

As mentioned in the Methodology section, when forecasting tourism demand with machine learning models, it is necessary to convert the time series into input-output sequence pairs. An important parameter in this process is the length of the input sequence, n, i.e., the tourism demands of n previous observations are used to forecast the tourism demand of the following observation. Previous studies have shown that n has a significant impact on the performance of machine learning models in tourism forecasting. Thus, to obtain more reasonable comparisons, the parameter search method is used to determine the optimal n of each machine learning model in each database with respect to different values of h, where n is set to 6, 9, 12, and 15 in the experiment. The results are shown in Table 2.

In addition, the parameters of ARIMA and SARIMA are directly determined by minimization of the Schwarz criterion and Akaike information criterion, and the important parameters of other models are determined by the exhaustive grid search technique. The results are shown in Table 3.

Experimental results

Experimental results using data for the period before the outbreak of the COVID-19

The results using data for the period before the outbreak of the COVID-19 generated by the different models are separately presented below for one-step-ahead forecasting ($h = 1$) and multi-step-ahead forecasting ($h = 3$).

(1) One-step-ahead forecasting

The one-step-ahead forecasting results of the different models for Jiuzhaigou and Mount Siguniang before the outbreak of the COVID-19 are shown in Table 4. It can be seen that the proposed model significantly outperforms the benchmark models in terms

Table 3
The optimal parameters for each model with respect to the two attractions.

Models	Jiuzhaigou	Mount Siguniang
SVM	C = 1000, Gamma = 0.01	C = 500, Gamma = 0.01
BPNN	Epoch = 1000, Batch size = 16	Epoch = 200, Batch size = 8
CNN	Epoch = 100, Batch size = 64	Epoch = 100, Batch size = 128
LSTM	Epoch = 100, Batch size = 8	Epoch = 200, Batch size = 8
CNN-LSTM	Epoch = 100, Batch size = 32	Epoch = 100, Batch size = 16
The proposed model	Number of CNN layers = 2, Number of LSTM layers = 2, Image size = 12, Epoch = 200, Batch size = 16	

Table 4
The results from different models for one-step-ahead forecasting using data for the period before the outbreak of the COVID-19.

Models	Jiuzhaigou			Mount Siguniang		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA	1930.23	18.55	3073.75	617.30	55.97	1069.83
SARIMA	1898.26	18.15	3065.32	542.52	54.43	1039.33
SVM	2405.44	34.71	3325.09	930.89	200.2	1228.82
NN	2192.14	22.27	3357.57	626.59	65.77	1149.52
CNN	2053.38	18.75	3230.2	558.84	51.84	1243.14
LSTM	2243.57	28.16	3326.95	798.04	158.22	1178.10
CNN-LSTM	2265.03	18.88	3610.66	579.14	50.78	1203.71
The proposed model	501.83	4.47	777.11	125.45	9.71	241.33

of MAE, MAPE and RMSE. Specifically, for Jiuzhaigou, the MAE, MAPE and RMSE of the proposed model are 501.83, 4.47% and 777.11 respectively, which are far lower than the best MAE, MAPE and RMSE obtained by the benchmark models (1915.49, 17.58% and 3073.36); for Mount Siguniang, the MAE, MAPE and RMSE of the proposed model are 125.45, 9.71% and 241.33 respectively, which are far lower than the best MAE, MAPE and RMSE obtained by the benchmark models (542.52, 50.78% and 997.24).

It should be noted that the CNN-LSTM model does not outperform the CNN and LSTM model in some cases. The main reason is that the number of samples and features used for training the model is relatively small without encoding the tourism demand time series into images. However, the CNN-LSTM model is required to train more parameters, almost equal to the sum of the parameters of the CNN and LSTM models. This makes the CNN-LSTM model more prone to over-fitting, resulting in relatively poor performance. Additionally, the results show that LSTM does not outperform the ARIMA model, which is inconsistent with the findings of Law et al. (2019). The main reasons are as follows: Law et al. (2019) use a large amount of search engine data as exogenous variables, which means more features can be learned and utilized by LSTM, and the advantages of LSTM can be brought into play; in contrast, this paper forecasts tourism demand only from historical time series data, which means fewer features can be learned by LSTM, and thus the predictive ability of LSTM is not fully utilized.

(2) Multi-step-ahead forecasting

The multi-step-ahead forecasting results using data for the period before the outbreak of the COVID-19 from the different models are shown in Table 5. It can be seen that the proposed model significantly outperforms the benchmark models in terms of MAE, MAPE and RMSE in the case of multi-step-ahead forecasting ($h = 3$). Specifically, for Jiuzhaigou, the values of MAE, MAPE and RMSE of the proposed model are 2681.57, 23.35% and 4471.71 respectively, which are lower than the best MAE, MAPE and RMSE obtained by the benchmark models (2897.68, 27.16% and 4497.47); for Mount Siguniang, the values of MAE, MAPE and RMSE of the proposed model are 658.63, 52.31% and 1616.89 respectively, which are lower than the best MAE, MAPE and RMSE values obtained by the benchmark models (722.68, 58.8% and 1617.28).

To verify whether the proposed model significantly outperforms the benchmark models without being affected by the outbreak of the COVID-19, the Diebold-Mariano (DM) test is carried out for the one-step-ahead and multi-step-ahead forecasting. The results are shown in Table 6, which indicates that the proposed model significantly outperforms the benchmark models with respect to the two attractions at a 99% confidence level, except that the significance level of "the proposed model vs. CNN-LSTM" on Mount Siguniang is 95%. Therefore, the proposed model is effective in relation to data for the period before the outbreak of the COVID-19.

Experimental results using data for the period during the COVID-19 pandemic

The results during the COVID-19 pandemic generated by the different models are separately presented below for one-step-ahead forecasting ($h = 1$) and multi-step-ahead forecasting ($h = 3$).

(1) One-step-ahead forecasting

Table 5
The results of different models for multi-step-ahead forecasting using data for the period before the outbreak of the COVID-19.

Models	Jiuzhaigou			Mount Siguniang		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA	2947.30	27.56	4549.09	957.40	104.86	1889.47
SARIMA	2897.68	27.50	4497.47	897.43	94.85	1878.16
SVM	3210.64	40.49	4505.03	1279.56	281.83	1768.86
NN	3245.77	31.91	4733.27	899.33	101.7	1754.59
CNN	3038.67	28.68	4686.88	758.05	68.01	1753.31
LSTM	3534.96	46.95	4747.98	1018.88	198.7	1619.69
CNN-LSTM	2923.79	27.16	4680.49	722.68	58.8	1617.28
The proposed model	2681.57	23.35	4471.71	658.63	52.31	1616.89

Table 6
The DM test results for one-step-ahead and multi-step-ahead forecasting using data for the period before the outbreak of the COVID-19.

h	Attractions	Benchmark models						
		ARIMA	SARIMA	SVM	NN	CNN	LSTM	CNN-LSTM
1	Jiuzhaigou	-17.36^{***}	-16.67^{***}	-19.84^{***}	-18.57^{***}	-16.91^{***}	-20.40^{***}	-18.38^{***}
	Mount Siguniang	-16.00^{***}	-8.26^{***}	-14.00^{***}	-15.49^{***}	-12.09^{***}	-13.91^{***}	-14.51^{***}
3	Jiuzhaigou	-3.77^{***}	-3.58^{***}	-10.69^{***}	-6.86^{***}	-4.48^{***}	-12.39^{***}	-3.53^{***}
	Mount Siguniang	-6.67^{***}	-7.37^{***}	-13.00^{***}	-7.68^{***}	-3.36^{***}	-12.78^{***}	-2.04^{**}

Note: The ** and *** indicate the statistical significance at the 5% and 1% levels, respectively.

The one-step-ahead forecasting results of the different models for Jiuzhaigou and Mount Siguniang during the COVID-19 pandemic are shown in Table 7. It can be seen that the proposed model significantly outperforms the benchmark models in terms of MAE, MAPE and RMSE. Specifically, for Jiuzhaigou, the MAE, MAPE and RMSE values of the proposed model are 309.67, 5.89% and 462.23 respectively, which are far lower than the best MAE, MAPE and RMSE values obtained by the benchmark models (1142.18, 26.18% and 1993.77); for Mount Siguniang, the MAE, MAPE and RMSE values of the proposed model are 76.18, 13.51% and 154.85 respectively, which are far lower than the best MAE, MAPE and RMSE values obtained by the benchmark models (850.25, 285.11% and 1361.14).

(2) Multi-step-ahead forecasting

The multi-step-ahead forecasting results during the COVID-19 pandemic from the different models are shown in Table 8. It can be seen that the proposed model significantly outperforms the benchmark models in terms of MAE, MAPE and RMSE in the case of multi-step-ahead forecasting ($h = 3$). Specifically, for Jiuzhaigou, the MAE, MAPE and RMSE values of the proposed model are 2051.84, 41.40% and 3379.93 respectively, which are lower than the best MAE, MAPE and RMSE values obtained by the benchmark models (2148.90, 50.99% and 3474.18); for Mount Siguniang, the MAE, MAPE and RMSE values of the proposed model are 816.53, 97.86% and 1698.71 respectively, which are lower than the best MAE, MAPE and RMSE values obtained by the benchmark models (825.58, 194.31% and 1704.01).

To verify whether the proposed model significantly outperforms the benchmark models when applied to the data relating to the period during the COVID-19 pandemic, the DM test is again carried out for the one-step-ahead and multi-step-ahead forecasting. The results are shown in Table 9. For one-step-ahead forecasting, the proposed model significantly outperforms the benchmark models with respect to the two attractions at a 99% confidence level. For multi-step-ahead forecasting, the proposed model significantly outperforms the benchmark models with respect to the two attractions at least at a 95% confidence level. Therefore, the proposed model is also effective when applied to data relating to the period during the COVID-19 pandemic.

As can be seen from Tables 7 and 8, most of the models tested here perform worse when applied to data relating to the period during the COVID-19 pandemic, but the proposed model nevertheless achieves a relatively high forecasting accuracy. The main reasons for this are as follows. The differences in tourism demand before and during the COVID-19 pandemic are mainly reflected in changes in seasonality and scale, and in local recurring patterns and temporal correlations.

In terms of seasonality and scale, the pure neural networks are sensitive to the scale variance of tourism demand. The scale of tourism demand before and during the COVID-19 pandemic changed in a non-periodic manner, which significantly lowers the accuracy of these models. However, some traditional models, such as the auto regressive model and the Holt-Winter model, are good at capturing changes in the seasonality and scale of tourism demand, but these models are incapable of nonlinear fitting. It should be pointed out that the proposed model is essentially a hybrid model composed of Holt-Winter and deep learning. Specifically, the Holt-Winter model is first used to normalize the tourism demand data to eliminate the influence of seasonality and scale on deep learning, using Eq. (8). In this way, the noise in the data can be smoothed out, and

Table 7
The results from different models for one-step-ahead forecasting using data for the period during the COVID-19 pandemic.

Models	Jiuzhaigou			Mount Siguniang		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA	1305.92	32.23	2049.22	953.10	206.12	2052.50
SARIMA	1142.18	26.18	1969.56	894.71	197.34	1874.88
SVM	1925.15	68.61	2497.89	1149.13	337.60	2363.91
NN	1160.09	26.91	1993.77	930.22	338.06	1824.20
CNN	1399.32	32.84	2433.79	845.20	194.31	1704.01
LSTM	1590.27	43.32	2502.00	1023.56	246.70	2002.73
CNN-LSTM	1212.65	27.50	2089.63	825.58	280.46	1762.00
The proposed model	309.67	5.89	462.23	76.18	13.51	154.85

Table 8
The results of different models for multi-step-ahead forecasting using data for the period during the COVID-19 pandemic.

Models	Jiuzhaigou			Mount Siguniang		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA	2255.18	54.47	3566.38	1435.87	509.00	2759.33
SARIMA	2148.90	50.99	3474.18	1332.72	467.87	2562.67
SVM	2492.13	81.79	3573.06	1357.51	736.88	2448.94
NN	2198.35	58.70	3492.81	1268.87	494.40	2487.50
CNN	2249.24	54.81	3796.02	1325.05	514.01	2703.36
LSTM	2647.54	82.94	3840.26	850.25	285.11	1361.14
CNN-LSTM	2159.90	57.78	3639.37	1268.37	558.43	2641.22
The proposed model	2051.84	41.40	3379.93	816.53	97.86	1698.71

the essential skeleton of the tourism demand data with respect to scale and seasonality can be determined, as reflected in the second and third terms of Eq. (18), i.e., l_j and n. Thus, the impact of seasonality and scale variance on the forecasting results can be effectively reduced.

In terms of local recurring patterns and temporal correlations, the normalized and deseasonalized data are first encoded into images using three methods (or from three perspectives), and then CNNs and LSTM networks are used to extract deep-seated features of the images to fully capture the local recurring patterns and temporal correlations. In this way, the impact of changes in local recurrence patterns and temporal correlations on the forecasting results can be reduced. For these two reasons, the impact of the COVID-19 on the forecasting results might be reduced to a certain extent. Therefore, the proposed model can achieve a relatively high forecasting accuracy with the data relating to the period during the COVID-19 pandemic.

Conclusions

The main purpose of this study is to improve the accuracy of tourist demand forecasting by leveraging the advantages of deep learning in image processing. To this end, a model based on deep learning with time series imaging is proposed, which consists of three parts: sequence image generation, image feature extraction, and model training. The effectiveness of the proposed model is verified by an experimental study using tourism demand data from two tourist attractions. The study makes four major contributions.

First, a novel deep learning model for tourism demand forecasting is proposed. The proposed model simultaneously considers the scale change in tourism demand, the seasonality change in tourism demand, and the local recurring patterns and temporary relations in tourism demand. In tourism scenarios, the scale change in tourism demand is often highly related to some key events, and the seasonality change of tourism demand is often highly related to the seasonal nature of tourism. Moreover, compared with the demand in other fields, such as most types of product demand, tourism demand is affected by specific complex factors (e.g., weather), which leads to the changes in the local recurring patterns and temporary relations of tourism demand. By normalizing the tourist demand data through the Holt-Winter model, the essential skeleton of the tourism demand data with respect to scale and seasonality can be captured well. By encoding time series into images, various deep-seated features, which reflect the local recurring patterns and temporary relations of tourism demand, embedded in the tourism demand data can be fully captured. Therefore, the factors that affect tourism demand, i.e., key events, seasonal factors, and other complex factors that lead to local small-scale changes in tourism demand, are considered at the same time in the proposed model, which ensures that it can capture systematic changes in tourism demand and improve forecasting accuracy. The essence of social science is to consider the impact of multi-level and multi-dimensional factors on human behavior as far as possible. Thus, capturing the impact of different factors on tourism demand not only increases the accuracy of forecasts but also allows a better understanding of the factors that drive changes in tourism demand from social science perspective. This enriches the methodology of deep learning in tourism demand forecasting.

Table 9
The DM test results for one-step-ahead and multi-step-ahead forecasting using data for the period during the COVID-19 pandemic.

h	Attractions	Benchmark models						
		ARIMA	SARIMA	SVM	NN	CNN	LSTM	CNN-LSTM
1	Jiuzhaigou	-10.43^{***}	-7.84^{***}	-9.77^{***}	-8.05^{***}	-7.57^{***}	-10.68^{***}	-7.70^{***}
	Mount Siguniang	-3.80^{***}	-3.70^{***}	-3.96^{***}	-3.88^{***}	-3.90^{***}	-4.32^{***}	-3.71^{***}
3	Jiuzhaigou	-3.48^{***}	-2.89^{***}	-5.74^{***}	-3.43^{***}	-2.46^{**}	-7.00^{***}	-2.53^{**}
	Mount Siguniang	-2.13^{**}	-2.03^{**}	-4.45^{***}	-3.37^{***}	-2.94^{***}	-3.03^{***}	-3.89^{***}

Note: The ** and *** indicate the statistical significance at the 5% and 1% levels, respectively.

Second, this study presents a new idea: encoding tourism demand time series as "color" images by using GAF, MTF, and RP, simultaneously. This is the first attempt to encode tourism demand data into image data. The study thus provides a good foundation for the application to tourism demand forecasting of other deep learning techniques with good performance in image processing.

Third, a combined deep learning model based on CNNs and LSTM is proposed for automatically mining the important features embedded in the images. The combined model can give full play to the advantages of CNNs in image processing and LSTM in sequence data processing. This is a meaningful attempt to combine CNNs and LSTM to develop a tourism demand forecasting model.

Finally, the validity of the proposed model is verified by a systematic experimental study. The feasibility of the proposed model is confirmed by analyses of data pertaining to the periods before and during the COVID-19 pandemic.

Additionally, with this proposed model, more accurate forecasts can be obtained. These can provide a scientific basis for tourism-related decision-making, such as the planning of tourist attractions and transportation facilities, the adjustment of the supply and demand of tourism products, the employment of temporary employees, and the pricing of tourism products.

The study's limitations may serve as directions for further investigations. The proposed model is univariate, and does not consider other exogenous variables. Nonetheless, several exogenous variables could be used to improve the accuracy of the forecasts (e.g., search engine data), and how to introduce other exogenous variables into the proposed model would be a promising direction for future research. Furthermore, compared with a shallow learning algorithm, the model proposed here proved to be computationally intensive. Nevertheless, as technology advances, this limitation will be lessened.

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.

Acknowledgements

This work was partly supported by the Humanities and Social Science Fund of Ministry of Education of China (project No. 20YJC630002), the China Postdoctoral Science Foundation (project Nos. 2020T130318 and 2019M661000), the Liberal Arts Development Fund of Nankai University (project Nos. ZX20210067 and ZB21BZ0106), the National Natural Science Foundation of China (project Nos. 71971124, 71871049 and 71932005), the Fundamental Research Funds for the Central Universities of China (project No. 63202074), and the One Hundred Talents Program of Nankai University (project No. 63213023)

References

Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2018). Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research*, *58*(3), 383–397.
Athanasopoulos, G., Hyndman, R. J., Song, H. Y., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*(3), 822–844.
Bi, J. W., Liu, Y., & Fan, Z. P. (2020). A deep neural networks based recommendation algorithm using user and item basic data. *International Journal of Machine Learning and Cybernetics*, *11*(4), 763–777.
Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2020). Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry. *Tourism Management*, *77*, Article 104006.
Bi, J. W., Liu, Y., & Li, H. (2020). Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research*, *83*, Article 102923. https://doi.org/10.1016/j.annals.2020.102923.
Chan, F., Lim, C., & Mcaleer, M. (2005). Modelling multivariate international tourism demand and volatility. *Tourism Management*, *26*, 459–471.
Cho, V. (2001). Tourism forecasting and its relationship with leading economic indicators. *Journal of Hospitality and Tourism Research*, *25*(4), 399–420.
Claveria, O., Monte, E., & Torra, S. (2015). Tourism demand forecasting with neural network models: Different ways of treating information. *International Journal of Tourism Research*, *17*(5), 492–500.
Claveria, O., Monte, E., & Torra, S. (2016). Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters*, *23*, 428–431.
Fildes, R., Wei, Y., & Ismail, S. (2011). Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting*, *27*(3), 902–922.
Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, *23*(5), 499–510.
Hatami, N., Gavet, Y., & Debayle, J. (2018). Classification of time-series images using deep convolutional neural networks. *Tenth international conference on machine vision, ICMV 2017*.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.
Kon, S. C., & Turner, W. L. (2005). Neural network forecasting of tourism demand. *Tourism Economics*, *11*, 301–328.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.
Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand forecasting. *Annals of Tourism Research*, *83*, Article 102925.
Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, *21*, 331–340.
Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, *75*, 410–423.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
Li, C., Ge, P., Liu, Z., & Zheng, W. (2020). Forecasting tourist arrivals using denoising and potential factors. *Annals of Tourism Research*, *83*, Article 102943.
Li, C., Wang, Y., Lv, X., & Li, H. (2021). To buy or not to buy? The effect of time scarcity and travel experience on tourists' impulse buying. *Annals of Tourism Research*, *86*, Article 103083.
Li, G., Song, H., & Witt, S. F. (2006). Time varying parameter and fixed parameter linear AIDS: An application to tourism demand forecasting. *International Journal of Forecasting*, *22*, 57–71.
Li, G., Wu, D. C., Zhou, M., & Liu, A. (2019). The combination of interval forecasts in tourism. *Annals of Tourism Research*, *75*, 363–378.
Li, S., Chen, T., Wang, L., & Ming, C. (2018). Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tourism Management*, *68*, 116–126.
Li, X., Kang, Y., & Li, F. (2020). Forecasting with time series imaging. *Expert Systems with Applications*, *160*, Article 113680.

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management*, *59*, 57–66.
Martínez-Arellano, G., Terrazas, G., & Ratchev, S. (2019). Tool wear classification using time series imaging and deep learning. *The International Journal of Advanced Manufacturing Technology*, *104*(9–12), 3647–3662.
Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5–6), 237–329.
Radenović, F., Tolias, G., & Chum, O. (2018). Fine-tuning CNN image retrieval with no human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(7), 1655–1668.
Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.
Sencheong, K., & Turner, L. W. (2005). Neural network forecasting of tourism demand. *Tourism Economics*, *11*, 301–328.
Shahrabi, J., Hadavandi, E., & Asadi, S. (2013). Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series. *Knowledge-Based Systems*, *43*, 112–122.
Shen, S., Li, G., & Song, H. (2009). Effect of seasonality treatment on the forecasting performance of tourism demand models. *Tourism Economics*, *15*(4), 693–708.
Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research*, *74*, 134–154.
Smyl, S., Ranganathan, J., & Pasqua, A. (2018). M4 forecasting competition: Introducing a new hybrid ES-RNN model. URL: https://eng.uber.com/m4-forecasting-competition.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, *29*(2), 203–220.
Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research*, *75*, 338–362.
Song, H., Witt, S. F., & Jensen, T. C. (2003). Tourism forecasting: Accuracy of alternative econometric models. *International Journal of Forecasting*, *19*, 123–141.
Turner, L. W., & Witt, S. F. (2001). Factors influencing demand for international tourism: Tourism demand analysis using structural equation modelling, revisited. *Tourism Economics*, *7*, 21–38.
Wang, Z., & Oates, T. (2015). *Imaging time-series to improve classification and imputation.* arXiv preprint arXiv:1506.00327.
Wong, K. K., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management*, *28*(4), 1068–1078.
Wu, D. C., Cao, Z., Wen, L., & Song, H. (2020). Scenario forecasting for global tourism. *Journal of Hospitality and Tourism Research*, *45*(1), 28–51.
Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management*.
Zhang, G., & Guo, J. (2020). A novel ensemble method for hourly residential electricity consumption forecasting by imaging time series. *Energy*, *203*.
Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, *26*(7), 3142–3155.
Zhang, Y., Li, G., Muskat, B., & Law, R. (2020). Tourism demand forecasting: A decomposed deep learning approach. *Journal of Travel Research*, *60*(5), 981–997.
Zhang, Y., Li, G., Muskat, B., Law, R., & Yang, Y. (2020). Group pooling for deep tourism demand forecasting. *Annals of Tourism Research*, *82*. https://doi.org/10.1016/j.annals.2020.102899.

Jian-Wu Bi, Ph.D., is a Postdoctoral Fellow in College of Tourism and Service Management, Nankai University, Tianjin, China. His research interests are tourism information technology and tourism forecasting.
Hui Li, Ph.D., is a Professor in College of Tourism and Service Management, Nankai University, Tianjin, China. His research focuses on tourism management analysis and prediction.
Zhi-Ping Fan, Ph.D., is a Professor in School of Business Administration, NEU. His research interests include decision analysis and operations research.