Soft Computing (2022) 26:695–707
https://doi.org/10.1007/s00500-021-06521-7

DATA ANALYTICS AND MACHINE LEARNING

# Spatial–temporal attention fusion for traffic speed prediction

Anqin Zhang[1] · Qizheng Liu[1] · Ting Zhang[1]

Accepted: 29 October 2021 / Published online: 19 November 2021

**Abstract**
Accurate vehicle speed prediction is of great significance to the urban traffic intelligent control system. However, in terms of traffic speed prediction, the modules that integrate temporal and spatial features in the existing traffic speed prediction methods are effective in short-term prediction, but the medium-term or long-term prediction errors are relatively large. In order to reduce the errors of existing methods in short-term prediction and predict the medium-term and long-term traffic speed, this paper proposes a traffic speed prediction method that combines attention and Spatial–temporal features, referred to as ASTCN. Specifically, unlike previous methods, ASTCN can use the temporal attention convolutional network (ATCN) to separately extract temporal features from the traffic speed features collected by each sensor, and use the spatial attention mechanism to extract spatial features and then perform spatial–temporal feature fusion. Experiments on three real-world datasets show that the proposed ASTCN model outperforms the state-of-the-art baselines.

**Keywords** Traffic speed prediction · Temporal attention convolutional network · Spatial attention mechanism · Spatial–temporal features

## 1 Introduction

Transportation plays a vital role in everyday life. According to a 2015 survey, the average driving time of American drivers is about 48 min per day (https://aaafoundation.org/american-driving-survey-2014-2015/). The intelligent control of urban traffic is very important, and traffic speed prediction has been paid more and more attention to the intelligent control of traffic. Traffic speed prediction is using the known road network structure and historical time step traffic speed data to predict the traffic speed at future time steps. The time step length of traffic speed prediction can be divided into three types, short-term prediction (within 30 min), medium-term prediction (30–60 min) and long-term prediction (over 60 min). In the past four decades, due to the increasing demand for urban traffic intelligent control system technology, traffic intelligent control system can not only provide drivers with accurate information but also can be used for signal optimization and vehicle coordinated control. Therefore, traffic speed prediction has always been hot research (Guo et al. 2020). If it can predict accurately in advance, the traffic management department can guide the vehicles more reasonably and improve the operating efficiency of the road network. However, due to complex temporal and spatial features, accurate traffic speed prediction is a challenging problem.

Traffic speed prediction is a classic problem of spatial–temporal data prediction. The traffic data are recorded at a fixed point in time and a fixed location with the continuous spatial distribution. Obviously, observations made at adjacent locations and adjacent time points are dynamically related to each other, as shown in Fig. 1. The correlation of road network traffic data shows strong dynamics in both spatial and temporal dimensions. Therefore, the key to solving the problem of dynamic prediction based on the existing conditions is how to effectively extract the temporal and spatial features and effectively integrate them to predict the traffic speed. How to mine nonlinear and complex spatial–temporal data, discover its inherent spatial–temporal patterns and make accurate traffic speed predictions is a very challenging problem.

In Fig. 1, it can be seen that with the time going by, the speed of traffic at each intersection will be affected by the traffic conditions of the previous time step of the

✉ Anqin Zhang
aqzhang@fudan.edu.cn

[1] College of Computer Science and Technology, ShangHai University of Electric Power, ShangHai 200090, China

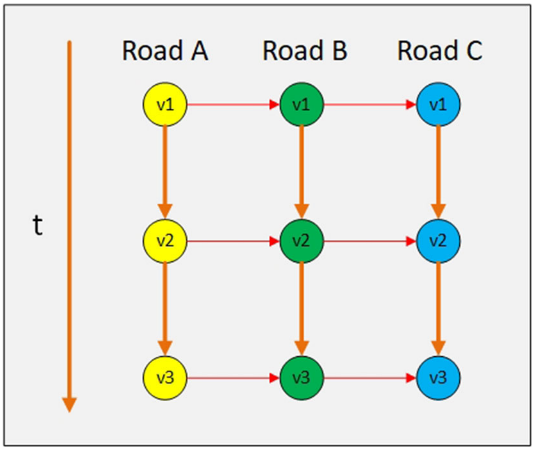


**Fig. 1** Complex spatial–temporal correlation of traffic data

intersection (the brown thick arrow in the vertical direction) and the traffic conditions of the adjacent intersection (the red thin arrow in the transverse direction). In short, the correlation of road network traffic data shows a strong dynamic in both spatial and temporal dimensions.

In order to extract spatial and temporal feature of road network traffic data, we propose the ASTCN method, the main contributions are as follows:

1. In traffic speed prediction, the length of historical time steps and future time steps are regarded as significant factors, and temporal attention convolutional networks are used to extract the temporal features from traffic speed data observed by each observation device.
2. The revised attention mechanism is used to extract spatial features.
3. The spatial–temporal feature fusion (FST) module is used to fuse spatial–temporal features.

The rest of this paper is as follows: Sect. 2 gives a description and some definitions of the traffic speed prediction problem. Section 3 introduces the architecture of ASTCN for traffic speed prediction. Section 4 is the experiment, and Sect. 5 is the conclusion and future work.

## 2 Related work

With the development of traffic, many information collection devices have been deployed to the road network, so that we can directly use the information collected by these devices to predict the traffic speed. Many researchers have made great efforts to solve these problems. In the early days, the time series analysis model was used for traffic prediction. However, in practical applications, they are difficult to deal with unstable nonlinear data. Based on the learning adaptability and capability to solve complex computations, classifiers are always the best suited for the pattern recognition problems (Kumar et al. 2019). Later, traditional machine learning methods were developed to model more complex data, but they are still difficult to consider the spatial–temporal correlation of high-dimensional traffic data at the same time. Deep learning (DL) is the most effective, supervised and stimulating machine learning approach in big data analysis (Dargan et al. 2019). It can automatically identify patterns and features in complex data through unsupervised/supervised learning. In recent years, many researchers have been using some deep learning methods to process high-dimensional spatial–temporal data, that is, convolutional neural network (CNN) is used to extract spatial features of grid data effectively; graph convolution neural network (GCN) is used to describe the spatial correlation of graph-based data. ChebNet (Defferrard et al. 2016) is a powerful GCN, which uses Chebyshev extension to reduce the complexity of Laplacian computation. GraphSAGE (Hamilton et al. 2017) samples a fixed number of neighborhoods for each node in the graph and aggregates its neighborhood and its own elements. GAT (Velickovic et al. 2018) is a powerful variant of GCN defined in the vertex domain, which uses the attention layer to dynamically adjust the importance of neighbor nodes. Najjar et al. (Najjar et al. 2017) proposed a deep learning-based mapping approach that leverages open data to learn from raw satellite imagery robust deep models able to predict accurate city-scale road safety maps at an affordable cost. Brewer et al. (Brewer et al. 2021) leveraged satellite imagery to estimate road quality and concomitant information about travel speed.

In order to make full use of spatial features, some researchers use a convolutional neural network (CNN) to capture the adjacent relationship between traffic networks, and use the recurrent neural network (RNN) on the time axis. By combining long short-term memory (LSTM) network (Hochreiter and Schmidhuber 1997) with one-dimensional CNN, Wu, and Tan (Wu and Tan 2016), a feature-level fusion structure CLTFP for short-term traffic prediction is proposed. Later, Shi et al. (Shi et al. 2015) proposed the convolutional LSTM, which is an extended all connected LSTM (FC-LSTM) embedded in the convolution layer. Zhang et al. (Zhang et al. 2018) designed an ST-RESNET model based on a residual convolution unit to predict crowd flow. Yao et al. (Yao et al. 2018) proposed a traffic volume prediction method combining CNN with long short-term memory (LSTM), which combined spatial and temporal correlation modeling. Yu et al. (Yu et al. 2018) proposed a new deep learning framework spatial–temporal graph convolution network (STGCN) to solve the problem of time series prediction in the field of transportation. Li et al. (Li et al. 2018) proposed the diffusion convolution recurrent neural network (DCRNN), which introduced graph convolution network into spatial–temporal network data prediction, and used diffusion graph convolution network to describe information diffusion process in the spatial network. Guo et al. (Guo and Yuan 2020) proposed a deep learning traffic prediction framework based on graph attention network (GAT) and time convolution network (TCN), called graph attention temporal convolution network (GATCN). Zhao et al. (Zhao et al. 2020) proposed a new traffic prediction method based on neural network, the temporal graph convolution network (T-GCN) model. Song et al. (Song et al. 2020) proposed a new spatial–temporal synchronous graph convolutional network (STSGCN). Guo et al. (Guo et al. 2019) proposed a deep spatial–temporal 3D convolutional neural network (ST-3DNet), which introduced three-dimensional convolution into this field. Wu et al. (Wu et al. 2019) designed an adaptive matrix to consider the change of influence between nodes and their neighbors. Bai et al. (Bai et al. 2019) attempted to simultaneously model spatial–temporal correlation by using gating residual GCN module with two attention mechanisms. Kong et al. (Kong et al. 2020) proposed an end-to-end deep learning-based dual path framework, spatial–temporal graph attention network (STGAT). Zheng et al. (Zheng et al. 2020) proposed an attention-based encoder–decoder framework.

However, the above methods are effective in the short-term forecast, and the error is large in the medium-term or long-term forecast. On the basis of the above background, in order to address these problems, we propose the ASTCN method that can capture the complex temporal and spatial features from traffic data and can then be used for traffic speed prediction tasks based on road network.

## 3 Problem setup

This section introduces the transportation network structure, the description of the traffic speed prediction problem and the structure of the input and output data.

### 3.1 Transportation network structure

In this paper, we use an undirected graph $G = (V, E, A)$ to define the transportation network, where V is a finite set of $|V| = N$ vertices, corresponding to the number of observation devices in the transportation network; E is the set of edges, indicating the connectivity between observation points; and A represents the weighted adjacency matrix of G. If the observation device i and the observation device j are directly connected, the value of $A_{ij}$ is the cost (distance or time, etc.) paid from the observation device i to the observation device j, otherwise the value of $A_{ij}$ is 0. The adjacency matrix is calculated according to the connection relationship among the observation devices in the road network. As shown in Fig. 2, the circles in the figure represent the observation devices, in which the number on the edge represents the weight of the edge.

In Fig. 2, there are six roads in the figure, and each road has an observation device (circle 1–6). If the observation devices are connected by edges, it indicates that they can reach each other directly. The value of the edge indicates the weight. If there is no edge connection, it indicates that they cannot reach each other directly.

The traffic speed observed by the observation device in the road network is represented by a two-dimensional matrix $X \in \mathbb{R}^{P \times V}$, where $P$ corresponds to the number of observation timestamps of the observation equipment; $V$ is a finite set of $|V| = N$ vertices; the size of $N$ corresponds to the number of vertices in the adjacency matrix of the road network; and $X_t^n$ is the speed observed by the observation device $n$ at time $t$.

### 3.2 Traffic speed forecast

Traffic speed prediction is a typical time series prediction problem. Traffic speed prediction is based on the current and historical situation of the road network, plus some objective conditions (such as road network structure, weather conditions, emergencies and other factors) to predict the traffic speed in the future.

Therefore, the traffic speed prediction problem can be regarded as learning the mapping function $f$ on the premise of knowing the road network structure G and the traffic speed matrix X, and then calculating the traffic speed at time T, as shown in Formula 1.

$$[X_{t+1}, \ldots, X_{t+T}] = f(G, [X_{t-n}, \ldots, X_{t-1}, X_t)]) \quad (1)$$

where $n$ is the length of the historical time step and $T$ is the length of the time step to be predicted.

### 3.3 The structure of input data and output data

The input data of ASTCN traffic speed prediction model include weighted adjacency matrix and historical step traffic speed matrix. The output data structure of this model is the traffic speed matrix of prediction time step.

The error value of the model is calculated by comparing the predicted result of the model with the real data.

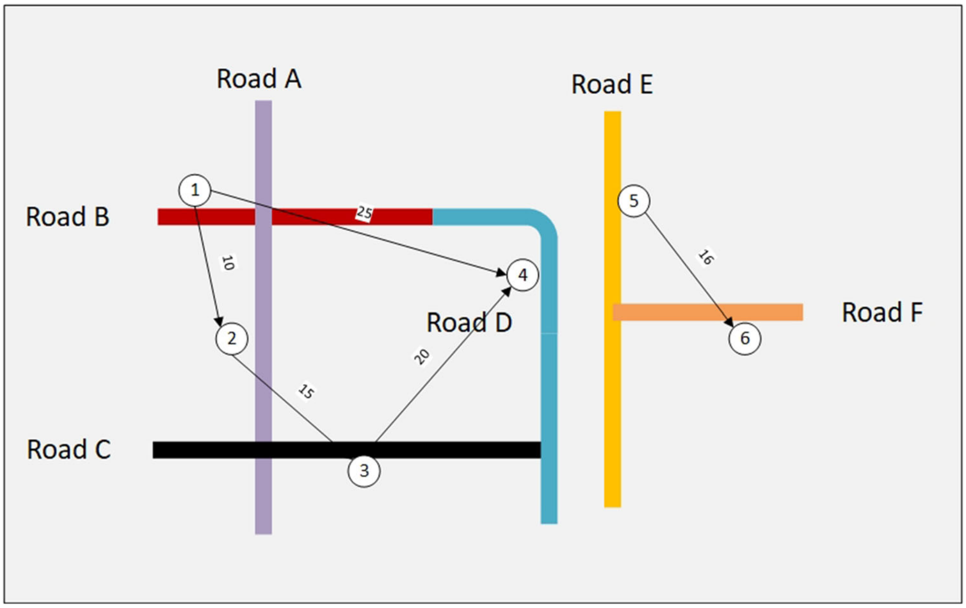


**Fig. 2** Simple road network structure

# 4 Methodology

This section introduces the ASTCN network structure and its details, including spatial–temporal convolution block and fully connected output layer. The main ideas of ASTCN are in the following:

ASTCN contains spatial–temporal convolution blocks and fully connected output layer. Spatial–temporal convolution block is used to extract spatial and temporal features and fuse these extracted features.

## 4.1 Model framework

In this part, we elaborate the structure of ASTCN. As shown in Fig. 3, ASTCN contains two spatial–temporal convolution blocks and a fully connected output layer. Each spatial–temporal convolution block contains temporal attention convolutional network (ATCN), spatial attention network and spatial–temporal feature fusion module (FST). We add an attention mechanism to extract temporal features on the basis of temporal convolutional network (TCN), which is named ATCN. And we use spatial–temporal feature fusion module (FST) to fuse the extracted temporal and spatial features.

## 4.2 Spatial–temporal convolutional block

Spatial–temporal convolutional block can capture the dynamic spatial–temporal correlation in road network. And it includes spatial attention network, temporal attention convolution network and spatial–temporal feature fusion module.

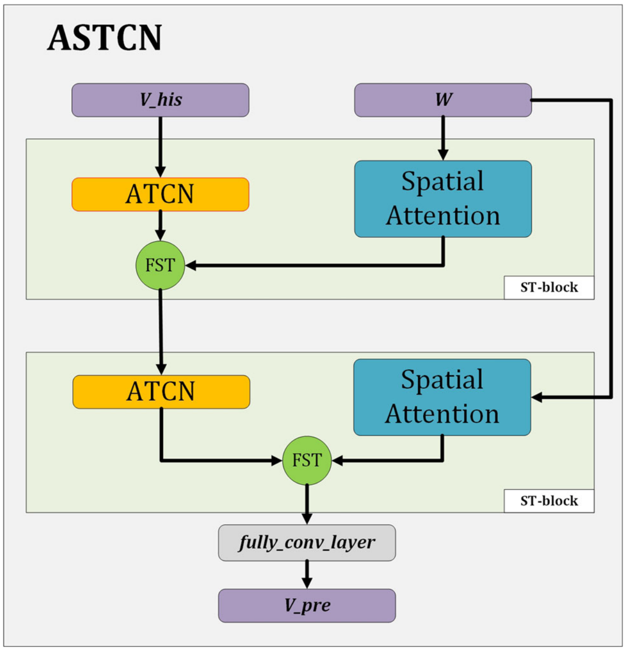


**Fig. 3** The structure of ASTCN model

### 4.2.1 Temporal attention convolutional network

Different historical time steps have different effects on the prediction results. In traffic speed prediction, the length of historical time step is regarded as a significant dependent variable, and the length of future time step is a significant indicator to measure the accuracy of the model. In Sect. 4, we do a comparative experiment with different lengths of historical time steps.

In the temporal dimension, the traffic speed at the current moment is dynamically affected by the traffic speed at the historical moment. Here we use the temporal attention convolutional network to extract the temporal features from traffic speed data observed by each observation device. In this module, we dynamically extract the temporal correlation, as shown in Formula 2.

$$T = \sum_{i}^{N} \sigma(W_1 \times TCN(X^i)) \tag{2}$$

$$\sigma(X) = \max(0, X) = \begin{cases} X, & X > 0 \\ 0, & X \leq 0 \end{cases} \tag{3}$$

Among them,$W_1 \in \mathbb{R}^{t_{pre} \times t_{pre}}$ is a learnable weight matrix, $t_{pre}$ is the historical time step input in the experiment, $X^i$ is the speed set observed by the observation device i and N is the number of observation devices in the road network, $\sigma(\cdot)$ represents an activation function. Here, the activation function is a ReLU function, as shown in Formula 3. $TCN(\cdot)$ is a temporal convolutional network. The specific formula 4 is as follows:

$$TCN\left(X^i\right) = \sigma(conv1d(\sigma(conv1d(X^i)))) \tag{4}$$

where $conv1d(\cdot)$ represents one-dimensional convolution, $\sigma(\cdot)$ is the $ReLU$ activation function and $X^i$ represents the traffic speed series observed by the observation device $i$.

The architecture in temporal convolutional network (TCN) (Bai et al. 2018) is a causal convolution, that is, no information is leaked from the future to the past during model training. At the same time, this architecture can use sequences of any length and map them to sequences of the same length, in similar to RNN. We can give TCN an input sequence $x_0, x_1, \ldots, x_n$ and then hope that TCN will output the related results $y_0, y_1, \ldots, y_n$ and generate a mapping relationship, which is named f function: $Y_0, \ldots, Y_n = f(X_0, \ldots, X_n)$. The value of $Y_j$ here only depends on $X_0, \ldots, X_j$ and has nothing to do with any $X_{j+1}, \ldots, X_n$. The goal of structural learning for sequence modeling is to find a $f$ function mapping that minimizes the expected loss between the actual output and the prediction.

In addition to causal convolution, TCN also has a principle that the length of the input sequence and the output sequence is the same. TCN uses a one-dimensional fully connected network to meet this principle, that is, the number of neurons in each hidden layer in the network is the same as the number of input layers, and zero padding with a length of core size-1 is added to maintain the same length of subsequent layer and previous layer. We can use $TCN = 1DFCN + causalconvolutions$ to briefly describe the characteristics of TCN.

In this paper, the experiment only needs to input the traffic speed sequence of 24 historical time steps to predict the traffic speed of 24 future time steps. Therefore, the length of the historical time step that needs to be input is relatively short, so we do not use the expansion convolution of TCN. The TCN structure used in this article is shown in Fig. 4.

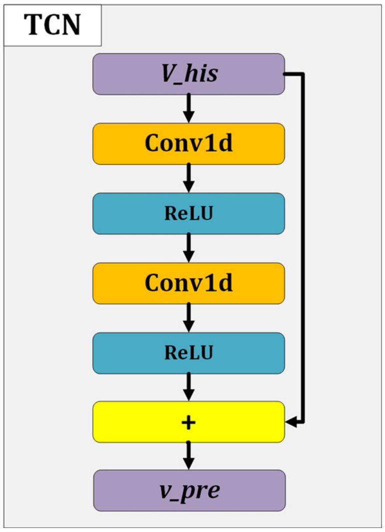


**Fig. 4** The structure of temporal convolutional network (TCN) module

#### 4.2.2 Spatial attention network

In the spatial dimension, the traffic speed of the current location is affected by the dynamics of the neighboring locations. Here we use a revised attention mechanism to capture the dynamic correlation between different nodes in the spatial dimension. In this module, we dynamically capture the spatial correlation as shown in formula 5. We use two learnable weight matrices $W_2, W_3$ to multiply the road network weight matrix $A$ to obtain a tensor with the same dimension as the input tensor of the fully connection output layer.

$$S = \sigma((W_2 \times A) \times W_3) \tag{5}$$

Among them, $A \in \mathbb{R}^{N \times N}$ is the standardized road network weighted adjacency matrix, $W_2 \in \mathbb{R}^{I \times B \times t_{his} \times N}$ and $W_3 \in \mathbb{R}^{I \times O}$ are the learnable weight matrices, $N$ is the number of observation devices in the road network, I is the input dimension of the convolution, $B$ is the number of each batch of data in the experiment, $t_{his}$ is the historical time step input in the experiment, $O$ is the output dimension of the convolution, $\sigma(\cdot)$ is the $ReLU$ activation function.

#### 4.2.3 Spatial–temporal feature fusion module (FST)

In order to make full use of the temporal and spatial features extracted by the above method in ASTCN model, we need to fuse the temporal features and spatial features.

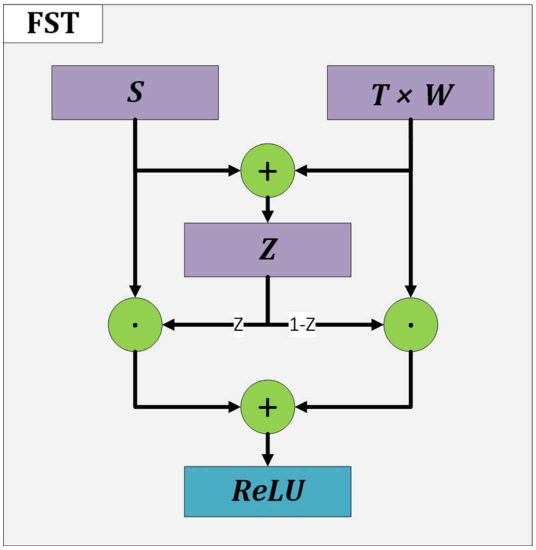


**Fig. 5** The spatial–temporal feature fusion method (FST) used in this paper

**Table 1** Dataset description

| Datasets | Number of sensors | Time range |
|---|---|---|
| PEMS04 | 307 | 1/1/2018–2/28/2018 |
| PEMS08 | 170 | 7/1/2016–8/31/2016 |
| LOS | 207 | 3/1/2012–3/7/2012 |

Zheng et al. (Zheng et al. 2020) designed a gated fusion to adaptively fuse the spatial and temporal features. In this paper, we modify this method by adding a learnable weight matrix $W$, which is used to make the tensor dimension of the temporal feature $T$ consistent with the spatial feature S.

The traffic speed of a road at a specific time is related to its previous traffic speed and the traffic speed of adjacent roads. In this paper, we propose a spatial–temporal feature fusion method, the specific method is shown in Fig. 5.

$$Z = (T \times W_4) + S \tag{6}$$

$$H = \sigma(Z \cdot S + (1 - Z) \cdot (T \times W_4)) \tag{7}$$

Among them, $W_4 \in \mathbb{R}^{I \times O}$ is a learnable weight matrix, where the temporal characteristic matrix $T$ and the weight matrix $W_4$ multiplication are to make the tensor dimension consistent with the spatial feature $S$, $\sigma(\cdot)$ is the *ReLU* activation function. We add them together to get the spatial–temporal features and then proceed to the next operation.

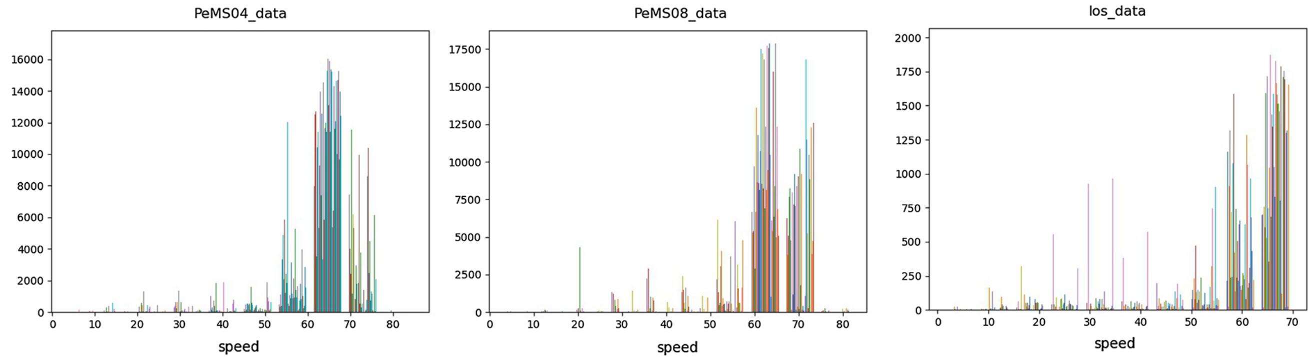
### 4.3 ASTCN training algorithm

The training process of the ASTCN is shown in Algorithm 1.

```
Algorithm 1. ASTCN training algorithm

Dataset input:

W = road network weighted adjacency matrix

X = matrix of historical traffic speed

Finding the most appropriate mapping function f:

    $X_{pre} = f(W, X_{his})$

Calculating:

$X_1 = X_{his}$

St-block × 2:

For i = 1,2 do:

  Extracting temporal features:

  For n = 1, 2, ……, N do:

    $T_n = \sigma(W_1 \times TCN(X_i))$
    $T = T + T_n$

  End for

  Extracting spatial features:

  $S = \sigma((W_2 \times A) \times W_3)$

  Fusing the spatial-temporal features:

  $H = \sigma((T \times W_4) + S)$
  $X_{i+1} = H$

End for

$X_{pre} = fullyOutput(X_3)$
return $X_{pre}$
```

## 5 Experiment

In this section we describe datasets, baseline methods, evaluation metrics and comparison results.



**(a)** PEMS04 **(b)** PEMS08 **(c)** LOS

**Fig. 6** The distribution of datasets

### 5.1 Datasets

We evaluated the traffic prediction performance of ASTCN on three real datasets. The three real datasets are PEMS04, PEMS08 (Song et al. 2020) and LOS (Hochreiter and Schmidhuber 1997).

PEMS04 and PEMS08 are collected by Caltrans Performance Measurement System. The Caltrans Performance Measurement System collects datasets in real time every 30 s. And the traffic data are aggregated from the original data every 5 min. The system deployed more than 39,000 detectors on highways in major metropolitan areas in California. And the geographic information of the observation device has been recorded in the dataset. The LOS dataset is collected in real time from Los Angeles highways through loop detectors. This dataset is similar to PEMS in that the traffic speed is collected every 5 min. In this experiment, 80% of these three datasets is used as the training set and the remaining 20% is used as the test set.

And the three datasets are composed of adjacency matrix and speed feature matrix. The specific details and the traffic speed distributions of the three datasets are shown in Table 1 and Fig. 6, respectively.

In this paper, each dataset is composed of an adjacency matrix dataset and a traffic speed dataset. Among them, the adjacency matrix data represent the distance of each observation device, and each column of the traffic speed matrix corresponds to the traffic speed collected by each observation device in the adjacency matrix. We standardize the adjacency matrix by formula 7 and use formula 8 (Najjar et al. 2017) to normalize the traffic speed matrix.

$$A' = \left(\frac{1}{\sqrt{\sum_{i,j}^{j=N} A_{i,j}}}\right)^{T} \cdot A \cdot \frac{1}{\sqrt{\sum_{i,j}^{j=N} A_{i,j}}} \tag{7}$$

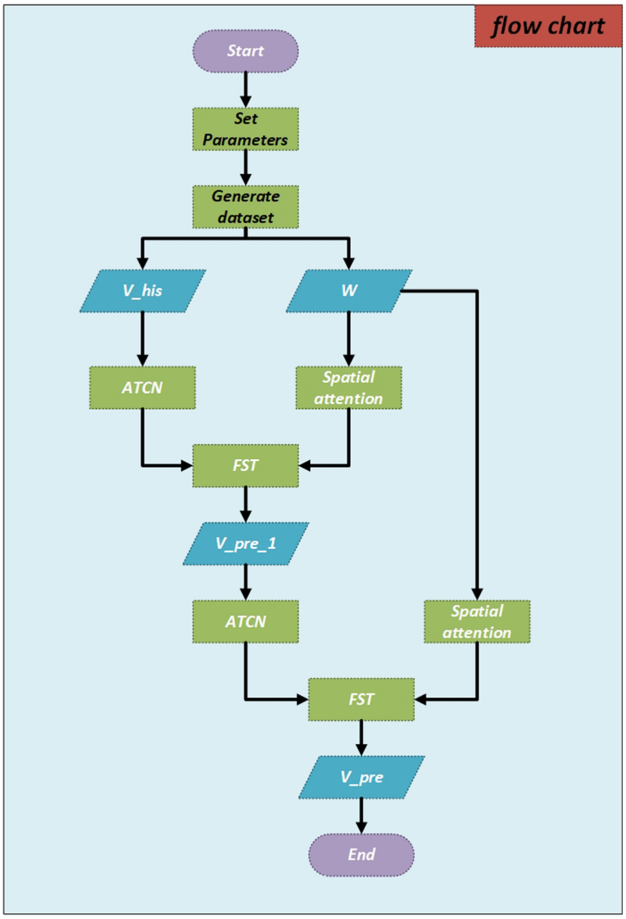
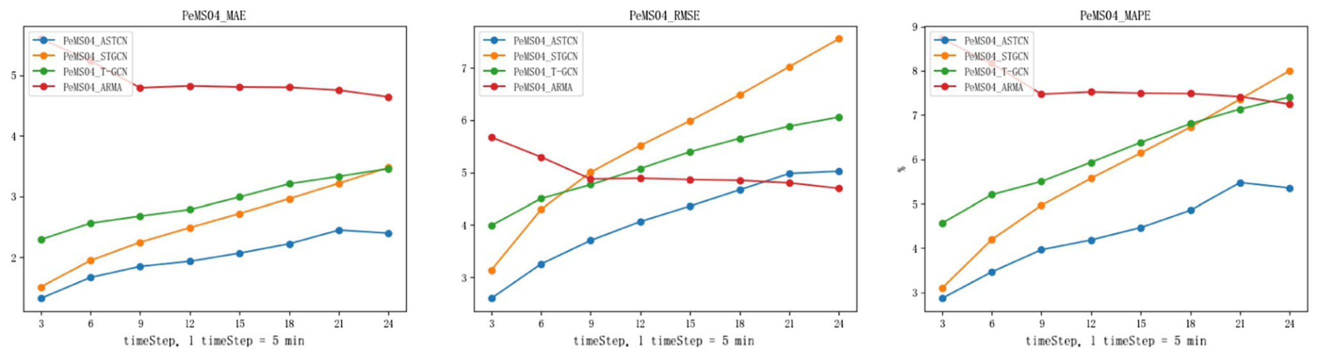


**Fig. 7** The flowchart of ASTCN

Among them, $A \in \mathbb{R}^{N \times N}$ is the adjacency matrix; $\sqrt{\sum_{i,j}^{j=N} A_{i,j}} \in \mathbb{R}^{1 \times N}$ represents the sum of each column of matrix $A$; $A'$ is the normalized adjacency matrix.

$$X' = \frac{X - mean(X)}{std(X)} \tag{8}$$



**(a)** MAE **(b)** RMSE **(c)** MAPE

**Fig. 8** The comparison results of prediction error of three methods on pems04 dataset

**Table 2** The prediction error results of the ASTCN model and other baseline methods on PEMS04

| Future time step | Metric | STGCN | T-GCN | ASTCN | ARMA |
|---|---|---|---|---|---|
| 3 | MAE | 1.552 | 2.294 | **1.322** | 5.606 |
| | RMSE | 3.185 | 3.992 | **2.600** | 5.673 |
| | MAPE | 3.098 | 4.566 | **2.869** | 8.724 |
| 6 | MAE | 2.015 | 2.564 | **1.668** | 5.242 |
| | RMSE | 4.299 | 4.502 | **3.256** | 5.296 |
| | MAPE | 4.195 | 5.201 | **3.465** | 8.167 |
| 9 | MAE | 2.327 | 2.680 | **1.850** | 4.795 |
| | RMSE | 4.928 | 4.767 | **3.707** | 4.876 |
| | MAPE | 4.965 | 5.502 | **3.967** | 7.473 |
| 12 | MAE | 2.583 | 2.787 | **1.935** | 4.826 |
| | RMSE | 5.422 | 5.075 | **4.067** | 4.891 |
| | MAPE | 5.575 | 5.934 | **4.188** | 7.523 |
| 15 | MAE | 2.828 | 2.993 | **2.070** | 4.808 |
| | RMSE | 5.901 | 5.395 | **4.360** | 4.865 |
| | MAPE | 6.144 | 6.382 | **4.467** | 7.495 |
| 18 | MAE | 3.076 | 3.210 | **2.224** | 4.802 |
| | RMSE | 6.391 | 5.651 | **4.669** | 4.851 |
| | MAPE | 6.728 | 6.806 | **4.855** | 7.486 |
| 21 | MAE | 3.313 | 3.332 | **2.450** | 4.756 |
| | RMSE | 6.879 | 5.884 | 4.981 | **4.802** |
| | MAPE | 7.360 | 7.136 | **5.478** | 7.415 |
| 24 | MAE | 3.546 | 3.461 | **2.400** | 4.645 |
| | RMSE | 7.364 | 6.058 | 5.024 | **4.696** |
| | MAPE | 8.004 | 7.407 | **5.356** | 7.245 |

Bold values are the best compared to other statistics in the same metrics

where $X \in \mathbb{R}^{P \times N}$ is the traffic speed matrix, P is the total number of minutes of the datasets divided by 5, which corresponds to the observation time step of the observation devices; N corresponds to the number of observation devices, $X^{'}$ is the standardized traffic speed matrix, mean $(X)$ and STD $(X)$ correspond to the mean and standard deviation of the historical time series, respectively.

## 5.2 Baseline method

During the verification test stage, the ASTCN model we proposed will be compared with the following two methods in terms of traffic speed prediction.

*STGCN*: For predicting future traffic speed data, Spatial–Temporal graph convolutional network (Yu et al. 2018) mainly uses graph convolutional network and two-dimensional convolution to extract spatial and temporal features, respectively.

*T-GCN*: The temporal graph convolutional network (Zhao et al. 2020) uses graph convolutional network and GRU to extract spatial and temporal features, respectively, which captures the spatial and temporal features from traffic data for application in predicting future traffic data.

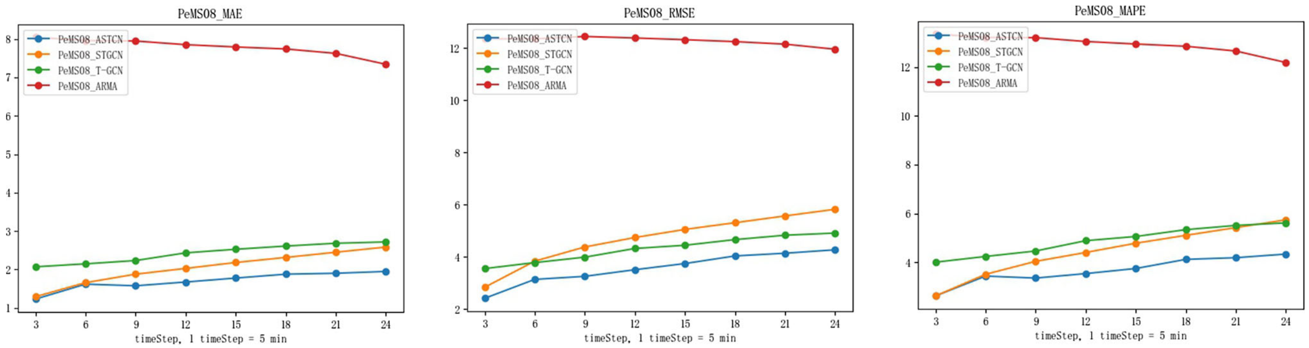
*ARMA:* Auto-regressive and moving average model is a well-known time series analysis method for predicting the future values.

## 5.3 Evaluation metrics

In this paper, we use three metrics to evaluate the prediction performance of different traffic speed prediction models. They are the mean absolute error (MAE), the root mean square error (RMSE) and mean absolute percentage error (MAPE), which are represented by Formula 9, Formula 10 and Formula 11.

$$MAE = \frac{\sum_{i=1}^{n} |\widehat{y_i} - y_i|}{n} \tag{9}$$

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}{n}\right)} \tag{10}$$



**(a)** MAE **(b)** RMSE **(c)** MAPE

**Fig. 9** The comparison results of prediction error of three methods on pems08 dataset

**Table 3** The prediction error results of the ASTCN model and other baseline methods on PEMS08

| Future time step | Metric | STGCN | T-GCN | ASTCN | ARMA |
|---|---|---|---|---|---|
| 3 | MAE | 1.268 | 2.070 | **1.230** | 8.044 |
| | RMSE | 2.852 | 3.554 | **2.416** | 12.351 |
| | MAPE | **2.622** | 4.006 | 2.624 | 13.347 |
| 6 | MAE | 1.633 | 2.147 | **1.619** | 7.958 |
| | RMSE | 3.897 | 3.783 | **3.141** | 12.350 |
| | MAPE | 3.509 | 4.240 | **3.437** | 13.213 |
| 9 | MAE | 1.876 | 2.233 | **1.576** | 7.946 |
| | RMSE | 4.482 | 3.991 | **3.257** | 12.441 |
| | MAPE | 4.040 | 4.464 | **3.355** | 13.203 |
| 12 | MAE | 2.078 | 2.433 | **1.674** | 7.851 |
| | RMSE | 4.918 | 4.324 | **3.511** | 12.383 |
| | MAPE | 4.403 | 4.886 | **3.538** | 13.046 |
| 15 | MAE | 2.292 | 2.526 | **1.777** | 7.791 |
| | RMSE | 5.324 | 4.447 | **3.747** | 12.317 |
| | MAPE | 4.782 | 5.058 | **3.747** | 12.942 |
| 18 | MAE | 2.509 | 2.610 | **1.878** | 7.739 |
| | RMSE | 5.707 | 4.664 | **4.038** | 12.244 |
| | MAPE | 5.105 | 5.338 | **4.121** | 12.850 |
| 21 | MAE | 2.714 | 2.683 | **1.902** | 7.624 |
| | RMSE | 6.055 | 4.831 | **4.142** | 12.147 |
| | MAPE | 5.422 | 5.512 | **4.189** | 12.655 |
| 24 | MAE | 2.907 | 2.717 | **1.949** | 7.340 |
| | RMSE | 6.395 | 4.914 | **4.273** | 11.952 |
| | MAPE | 5.741 | 5.616 | **4.338** | 12.184 |

Bold values are the best compared to other statistics in the same metrics

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\widehat{y_i} - y_i}{y_i + 10^{-10}}\right| \tag{11}$$

The range of MAE, RMSE and MAPE is $[0, +\infty)$. The three metrics are 0 when the real value and the predicted value are equal, which is a perfect model. A value of MAPE exceeding 100% is indicated as an inferior model.

## 5.4 Model parameters and flowchart

The hyperparameters of the ASTCN model mainly include: learning rate, batch size and training epoch. In the experiment, we manually adjust and set the learning rate to 0.001, the batch size to 50 and the training epoch to 50.

We use the temporal attention convolutional network to extract the temporal correlation, which includes four-layer one-dimensional convolution neural network. And the number of neurons of one-dimensional convolutional neural networks is 1024.

The flowchart of ASTCN is shown in Fig. 7.

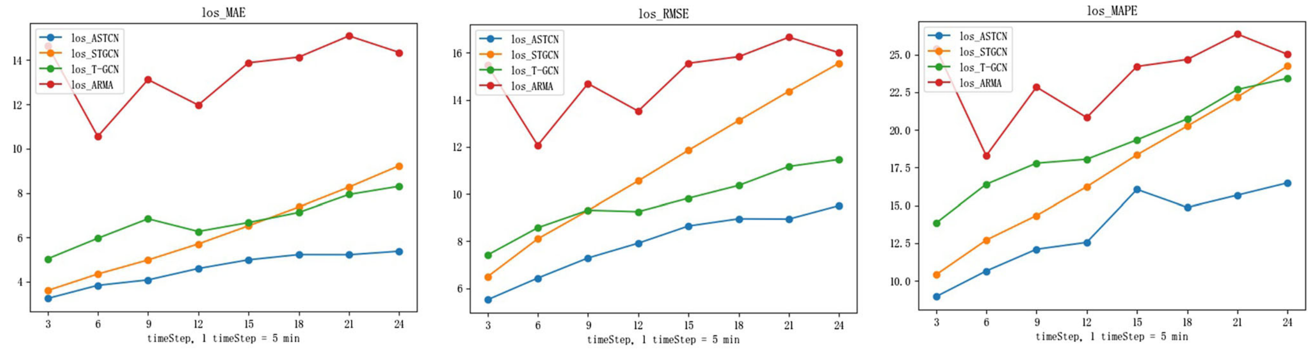
## 5.5 Experimental result

### 5.5.1 PEMS04

Figure 8 and Table 2 show the comparison of the three methods for 24 time steps future predictions, which include ASTCN, STGCN, T-GCN and ARMA, in three evaluation metrics on PEMS04 dataset. And the three metrics are MAE, RMSE and MAPE.

The above experimental results show that in PEMS04 dataset, the prediction error of ASTCN model is lower than that of STGCN and T-GCN models. For example, when the prediction time step length is 12, the prediction error MAE of ASTCN model is 1.935, but the prediction error MAE of the other baseline model are 2.583, 2.787 and 4.826, respectively. There is one exception, when the prediction time step lengths are 21 and 24, the prediction error RMSE of ARMA model is 4.802 and 4.696, respectively, which are lower than that of ASTCN.

In summary, the ASTCN model performs better than the other three methods on the pems04 dataset.



**(a)** MAE **(b)** RMSE **(c)** MAPE

**Fig. 10** The comparison results of prediction error of three methods on LOS dataset

**Table 4** The prediction error results of the ASTCN model and other baseline methods on LOS

| Future time step | Metric | STGCN | T-GCN | ASTCN | ARMA |
|---|---|---|---|---|---|
| 3 | MAE | 3.773 | **2.032** | 3.241 | 14.664 |
| | RMSE | 6.812 | 7.415 | **5.515** | 15.466 |
| | MAPE | 10.405 | 13.830 | **8.954** | 25.415 |
| 6 | MAE | 4.514 | 5.962 | **3.834** | 10.552 |
| | RMSE | 8.406 | 8.573 | **6.436** | 12.065 |
| | MAPE | 12.705 | 16.395 | **10.656** | 18.284 |
| 9 | MAE | 4.995 | 6.843 | **4.080** | 13.120 |
| | RMSE | 9.313 | 9.313 | **7.290** | 14.686 |
| | MAPE | 14.329 | 17.790 | **12.100** | 22.838 |
| 12 | MAE | 5.484 | 6.270 | **4.595** | 11.971 |
| | RMSE | 10.102 | 9.253 | **7.918** | 13.518 |
| | MAPE | 16.230 | 18.052 | **12.554** | 20.817 |
| 15 | MAE | 6.041 | 6.671 | **4.989** | 13.882 |
| | RMSE | 10.994 | 9.822 | **8.647** | 15.552 |
| | MAPE | 18.345 | 19.337 | **16.051** | 24.209 |
| 18 | MAE | 6.539 | 7.126 | **5.224** | 14.138 |
| | RMSE | 11.833 | 10.363 | **8.952** | 15.829 |
| | MAPE | 20.247 | 20.730 | **14.870** | 24.668 |
| 21 | MAE | 6.963 | 7.932 | **5.219** | 15.096 |
| | RMSE | 12.551 | 11.168 | **8.941** | 16.657 |
| | MAPE | 22.183 | 22.679 | **15.674** | 26.353 |
| 24 | MAE | 7.337 | 8.301 | **5.376** | 14.350 |
| | RMSE | 13.217 | 11.464 | **9.512** | 15.998 |
| | MAPE | 24.231 | 23.421 | **16.489** | 25.007 |

Bold values are the best compared to other statistics in the same metrics

### 5.5.2 PEMS08

Figure 9 and Table 3 show the comparison of the three methods for 24 time steps future predictions, which include ASTCN, STGCN, T-GCN and ARMA, in three evaluation metrics on PEMS08 dataset. And the three metrics are MAE, RMSE and MAPE.

The above experimental results show that in PEMS08 dataset, when the prediction time step length is 3, although the prediction error MAPE of STGCN model is 2.622, which is lower than that of ASTCN, on the whole, when the prediction time step lengths are 3, 6, 9, 12, 15, 18, 21, 24, the prediction errors of ASTCN model are lower than that of STGCN, T-GCN and ARMA models.

### 5.5.3 LOS

Figure 10 and Table 4 show the comparison of the three methods for 24 time steps future predictions, which include ASTCN, STGCN, T-GCN and ARMA, in three evaluation metrics on LOS dataset. And the three metrics are MAE, RMSE and MAPE.

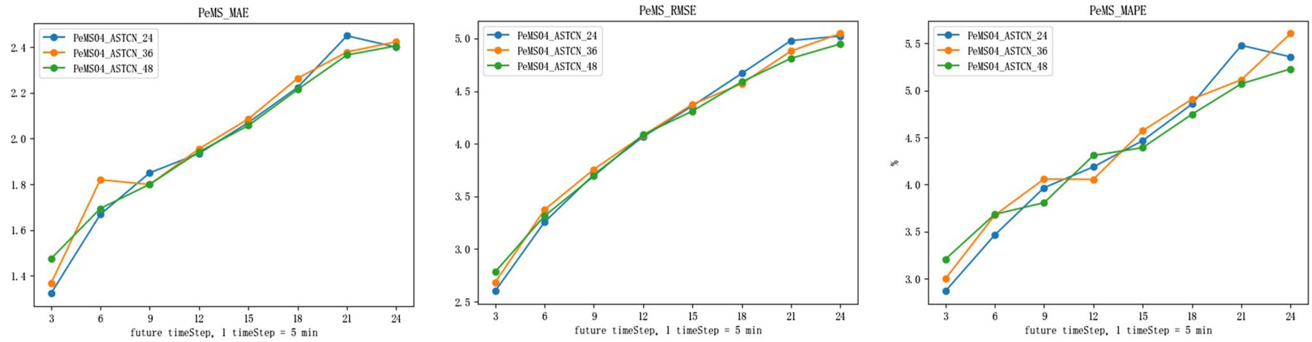
The LOS dataset is real urban traffic data, and there are many and miscellaneous factors affecting vehicle speed, so the prediction error of ASTCN model is higher than that of PEMS dataset.

For example, in LOS dataset, when the prediction time step length is 15, the prediction error RMSE of ASTCN model is significantly lower than that of the other three models. Therefore, there is no denying that ASTCN model outperforms STGCN, T-GCN and ARMA model in traffic speed prediction.

## 5.6 Choosing historical Time step

In order to choose a more appropriate length of the historical time step, we designed a comparative experiment, which sets the length of the historical time step to 24 (2 h), 36 (3 h) and 48 (4 h), respectively, to compare the error of the prediction results on PEMS04 dataset.

Figure 11 and Table 5 show the comparison of the three lengths of historical time steps for 24 time steps future predictions in three evaluation metrics on PEMS04 dataset. And the three metrics are MAE, RMSE and MAPE.



**(a)** MAE **(b)** RMSE **(c)** MAPE

**Fig. 11** The comparison results of prediction error of three different lengths of historical time steps on PEMS04 dataset

**Table 5** The prediction error results of the different lengths of historical time steps on PEMS04

| Future time step | Metric | Historical time steps | | |
|---|---|---|---|---|
| | | 24 | 36 | 48 |
| 3 | MAE | **1.322** | 1.366 | 1.474 |
| | RMSE | **2.600** | 2.683 | 2.783 |
| | MAPE | **2.869** | 3.000 | 3.208 |
| 6 | MAE | **1.668** | 1.820 | 1.694 |
| | RMSE | **3.256** | 3.373 | 3.314 |
| | MAPE | **3.465** | 3.675 | 3.685 |
| 9 | MAE | 1.850 | **1.799** | 1.800 |
| | RMSE | 3.707 | 3.757 | **3.693** |
| | MAPE | 3.967 | 4.059 | **3.808** |
| 12 | MAE | **1.935** | 1.955 | 1.942 |
| | RMSE | **4.067** | 4.079 | 4.087 |
| | MAPE | 4.188 | **4.056** | 4.309 |
| 15 | MAE | 2.070 | 2.086 | **2.057** |
| | RMSE | 4.360 | 4.375 | **4.311** |
| | MAPE | 4.467 | 4.572 | **4.393** |
| 18 | MAE | 2.224 | 2.264 | **2.215** |
| | RMSE | 4.669 | **4.571** | 4.589 |
| | MAPE | 4.855 | 4.907 | **4.747** |
| 21 | MAE | 2.450 | 2.379 | **2.366** |
| | RMSE | 4.981 | 4.882 | **4.812** |
| | MAPE | 5.478 | 5.113 | **5.071** |
| 24 | MAE | **2.400** | 2.424 | 2.406 |
| | RMSE | 5.024 | 5.048 | **4.949** |
| | MAPE | 5.356 | 5.604 | **5.228** |

Bold values are the best compared to other statistics in the same metrics

In above comparative experiment, we found that with the increase of the length of the historical time step, the medium- and long-term prediction error of traffic speed decreased, but the short-term prediction error of traffic speed increased. When future time step length of the predicted traffic speed is 3 and 6, the prediction error of the model with historical time step length of 24 is lower than that of the model with historical time step length of 36 and 48, but when future time step length of the predicted traffic speed is more than 12, the prediction error of the model with historical time step length of 24 is higher than that of the model with historical time step length of 36 and 48. Therefore, in the short-term prediction of traffic speed, we should set the historical time step length to 24, that is, 2 h; in the medium- and long-term prediction of traffic speed, the length of historical time step should be set to 48, that is, 4 h.

### 5.7 Model interpretation

In order to better understand the ASTCN model, we chose an observation device in pems04 dataset, in this test set and visualized the prediction results and actual traffic speed. Figure 12 shows the visualization results with the predicted horizon of 15 min, 30 min, 45 min, 60 min, 75 min, 90 min, 105 min and 120 min. With the increase of prediction time step, the worse the prediction effect is, which accords with the actual situation.

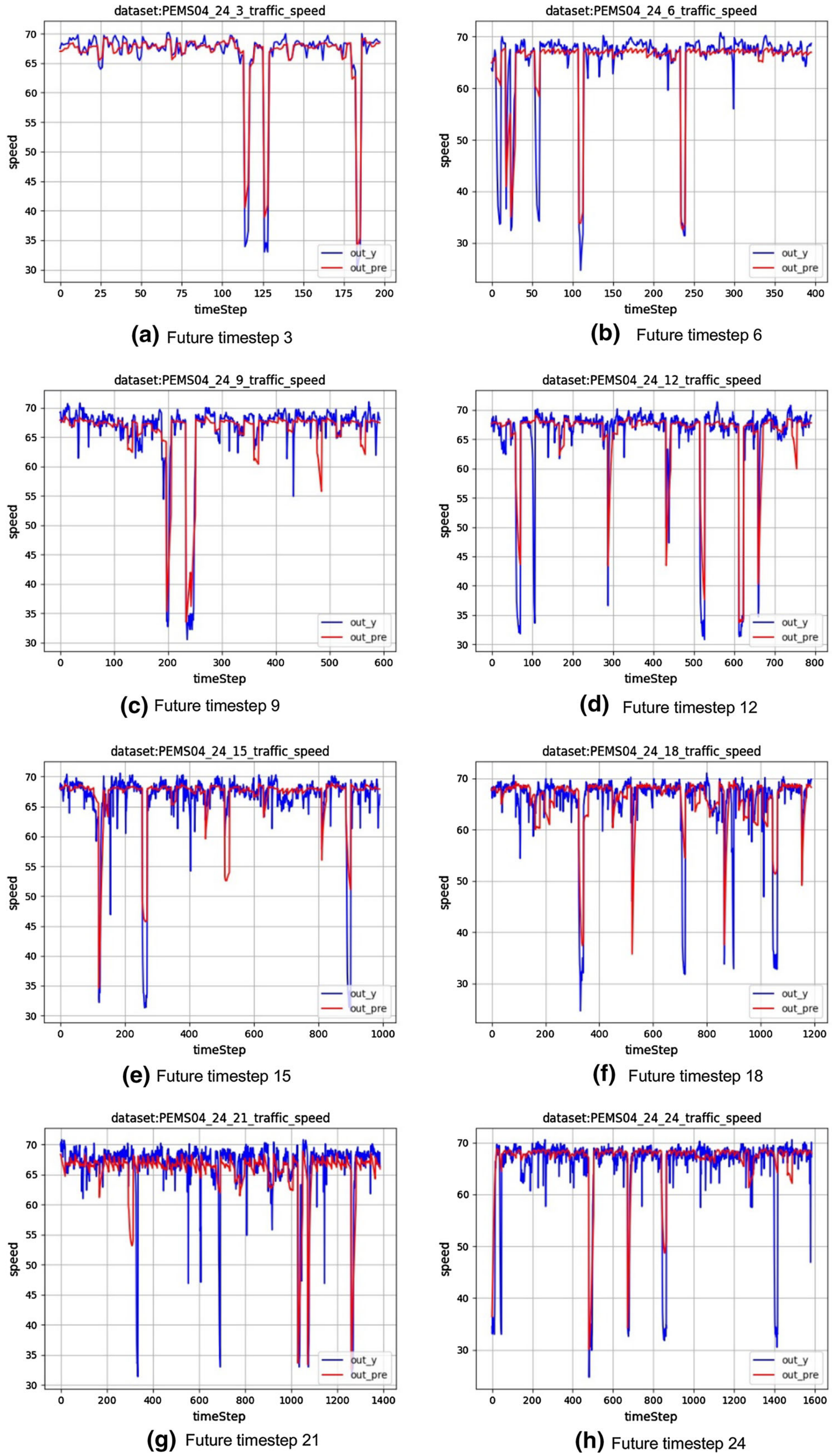
In Fig. 12, the “out_y” denotes the test set data and the “out_pre” denotes the prediction result. The titles of these pictures, for example “PEMS04_24_15_traffic_speed”, the first number 24 is the historical time step and the second number 15 is the predicting time step.

## 6 Conclusion

Transportation plays a vital role in our everyday life. However, due to complex temporal and spatial features, accurate traffic speed prediction is a challenging problem, and the existing traffic forecasting methods are effective in the short-term forecast, but the errors of these methods are large in the medium-term or long-term forecast.

In order to increase the accuracy of existing methods in short-term prediction and predict the medium-term and

**Fig. 12** The visualization results for prediction horizon of 15, 30, 45, 60, 75, 90, 105, 120 min



**(a)** Future timestep 3

**(b)** Future timestep 6

**(c)** Future timestep 9

**(d)** Future timestep 12

**(e)** Future timestep 15

**(f)** Future timestep 18

**(g)** Future timestep 21

**(h)** Future timestep 24

long-term traffic speed, we propose the ASTCN method. ASTCN introduces temporal attention convolution network, spatial attention network and spatial–temporal feature fusion module. ATCN is the TCN with attention mechanism, and TCN contains one-dimensional convolution and causal convolution, which related to time, so temporal features can be extracted using ATCN. And the revised attention mechanism and the improved gate fusion method are used to extract the spatial features and fuse the extracted temporal and spatial features, respectively. The experiments of ASTCN on three real datasets show that, with the verification of three indicators (MAPE, RMSE and MAE), ASTCN has better performance than baseline methods (STGCN, T-GCN and ARMA) in traffic speed prediction, not only in short-term prediction, but also in medium-term and long-term prediction.

Since ASTCN is a general spatial–temporal prediction framework, we can also apply it to other spatial–temporal prediction tasks (precipitation forecast, etc.). In the future, in traffic forecasting, we can regard the traffic state diagram as an image and use the ORB and SIFT in (Chhabra et al. 2018) to extract the main features in the traffic state diagram and predict the traffic data.

**Author contributions** A.Z. contributed to conceptualization and methodology. Q.L contributed to experiment, writing original draft and visualization. T.Z. contributed to suggestions of experiment.

**Data availability** The datasets generated during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** We declare that we have no conflicts of interest.

## References

Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. abs/1803.01271

Bai L, Yao L, Kanhere SS, Wang X, Sheng Q (2019) STG2Seq: spatial-temporal graph to sequence model for multi-step passenger demand forecasting. abs/1905.10069

Brewer E, Lin J, Kemper P, Hennin J, Runfola DM (2021) Predicting road quality using high resolution satellite imagery: a transfer learning approach. PLoS ONE 16:e0253370

Chhabra P, Garg NK, Kumar M (2018) Content-based image retrieval system using ORB and SIFT features. Neural Comput Appl 32:2725–2733. https://doi.org/10.1007/s00521-018-3677-9

Dargan S, Kumar M, Ayyagari MR, Kumar G (2019) A survey of deep learning and its applications: a new paradigm to machine learning. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-019-09344-w

Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. NIPS

Guo G, Yuan W (2020) Short-term traffic speed forecasting based on graph attention temporal convolutional networks. Neurocomputing 410:387–393

Guo S, Lin Y, Li S, Chen Z, Wan H (2019) Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting. IEEE Trans Intell Transp Syst 20:3913–3926

Guo G, Li P, Hao L (2020) Adaptive fault-tolerant control of platoons with guaranteed traffic flow stability. IEEE Trans Veh Technol 69:6916–6927. https://doi.org/10.1109/TVT.2020.2990279

Hamilton WL, Ying Z, Leskovec J (2017) Inductive representation learning on large graphs. NIPS

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780

https://aaafoundation.org/american-driving-survey-2014-2015/

Kong X, Xing W, Wei X, Bao P, Zhang J, Lu W (2020) STGAT: spatial-temporal graph attention networks for traffic flow forecasting. IEEE Access 8:134363–134372

Kumar M, Jindal MK, Sharma RK, Jindal SR (2019) Performance evaluation of classifiers for the recognition of offline handwritten Gurmukhi characters and numerals: a study. Artif Intell Rev 53:2075–2097

Li Y, Yu R, Shahabi C, Liu Y (2018) Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv: Learning

Najjar A, Kaneko S, Miyanaga Y (2017) Combining satellite imagery and open data to map road safety. AAAI

Shi X, Chen Z, Wang H, Yeung D, Wong W, Woo W (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. NIPS

Song C, Lin Y, Guo S, Wan H (2020) Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. AAAI

Velickovic P, Cucurull G, Casanova A, Romero A, Lio' P, Bengio Y (2018) Graph attention networks. ArXiv, abs/1710.10903.

Wu Y, Tan H (2016) Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. abs/1612.01022

Wu Z, Pan S, Long G, Jiang J, Zhang C (2019) Graph WaveNet for deep spatial-temporal graph modeling. IJCAI

Yao H, Wu F, Ke J, Tang X, Jia Y, Lu S, Gong P, Ye J, Li ZJ (2018) Deep multi-view spatial-temporal network for taxi demand prediction. AAAI

Yu T, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. IJCAI

Zhang J, Zheng Y, Qi D, Li R, Yi X, Li T (2018) Predicting citywide crowd flows using deep spatio-temporal residual networks. Artif Intell 259:147–166

Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2020) T-GCN: a temporal graph convolutional network for traffic prediction. IEEE Trans Intell Transp Syst 21:3848–3858

Zheng C, Fan X, Wang C, Qi J (2020) GMAN: a graph multi-attention network for traffic prediction. abs/1911.08415