A Hybrid Deep Learning-Based Traffic Forecasting Approach Integrating Adjacency Filtering and Frequency Decomposition

JUN CAO, XUEFENG GUAN, NA ZHANG, XINGLEI WANG, AND HUAYI WU
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan 430079, China
Corresponding author: Xuefeng Guan (guanxuefeng@whu.edu.cn)

ABSTRACT
Traffic forecasting in urban area has attracted substantial attention in recent years due to its significant assistance for traffic dispatching and trip planning. However, this task is very challenging due to the complex dependencies inherent in the traffic process. In our study, recent and periodic dependencies are identified and used to describe the corresponding near-term and long-term effects of historical traffic data on future traffic states. Subsequently, during the recent dependency modeling, each road is found to correlate with its adjacent roads through traffic flow diffusion, then the correlation intensities are quantified and used to choose strongly correlated roads to build a critical road sequence. While for the periodic dependency modeling, since the historical speed series of a road segment exhibits multi-frequency attributes (i.e., low-frequency daily period and high-frequency stochastic fluctuation), wavelet transform is conducted to decompose the original speed series into low and high-frequency sub-series. On these bases, we propose a hybrid traffic speed forecasting model, flow and wavelet-integrated spatio-temporal network (FW-STN). In the FW-STN, the recent features are captured by the convolutional neural network (CNN) with near-term traffic data from the derived critical road sequence, and the periodic features are captured by the long short-term memory (LSTM) with the low and high-frequency sub-series. Both the recent and periodic features are then fused to conduct the final prediction. Experimental results on real traffic data show that the proposed approach outperforms nine state-of-the-art methods (with improvements of 3% ~15% in mean average percentage error).

INDEX TERMS
Traffic forecasting, dependencies, spatial correlation, traffic flow, multi-frequency, wavelet transform.

I. INTRODUCTION
Traffic forecasting in urban road networks plays an important role in Intelligent Transportation Systems (ITS) [1]. Accurate and reliable traffic forecasting can help to improve transportation dispatching and guide urban travel. For example, traffic controllers can arrange vehicle diversions and allocate resources before the predicted congestion occurs, while drivers can choose their driving paths based on the predicted traffic states of road segments in order to reduce travel time. These practical capabilities have made traffic forecasting a hot topic for researchers in this field. Particularly, in recent years, the evolution of big traffic data and deep learning has inspired the development of many more powerful and efficient forecasting methods [2].

Traffic forecasting aims to make accurate estimations of future traffic states based on historical traffic data, as well as additional context information [3]. The traffic state of a road segment is affected by various factors, including the traffic states of surrounding roads, date-time (e.g., holidays, rush hours), weather, etc. Complex dependencies thus exist between these factors and future traffic states, making the forecasting a very challenging one.

In this study, two different types of dependencies, recent and periodic, are identified and defined separately. In more detail, recent dependency emphasizes the near-term traffic
interactions among locally connected roads. During the process of traffic diffusion on the urban road network, the future traffic states of the target road are closely correlated to those on spatially connected road segments at recent time intervals (e.g., from about 30 minutes ago to the current time).

Moreover, the intensity of this correlation varies among different connected roads. As illustrated in Fig. 1, most of the vehicles on the target road come from or drive into the directly connected primary roads rather than the low-traffic residential roads. Under these conditions, the target road traffic will be more strongly affected by these adjacent primary roads; for example, congestion on a downstream primary road will very likely cause subsequent congestion on the target road, as most vehicles are driving through it towards the congested road. Therefore, traffic flow between roads can be treated as an effective indicator to quantify this correlation intensity.

Periodic dependencies exist among long-term historical traffic data series due to daily commute routines (e.g., people going out to work in the morning and coming back home in the evening). Future traffic states can be quite similar to past traffic states around the same time interval of previous days or weeks.

Periodic traffic data series also exhibit multi-frequency characteristics. Fig. 2 illustrates a real traffic speed series of one road segment in a week. In Fig. 2, there is a distinct low-frequency trend component with daily periodicity. Moreover, affected by the complex traffic process and variable external conditions, traffic speeds also fluctuate rapidly at high-frequencies around the low-frequency component. The periodic dependencies among different frequencies are different and should therefore be modeled independently. Wavelet transform (WT) is usually used to decompose the raw data into separate sub-series with different frequencies. Previous studies have shown that wavelet-based models can effectively improve accuracy in traffic prediction problems [4], [5].

Motivated by the above, we conduct adjacency filtering to select strongly correlated adjacent roads that diffuse large-part traffic flow to the target road, and construct a critical road sequence for recent dependency modeling. For periodic dependency modeling, wavelet transform is employed to decompose the original speed series into sub-series with low-frequency and high-frequency. Furthermore, we propose a hybrid forecasting model, flow and wavelet-integrated spatio-temporal network (FW-STN), to predict the short-term traffic speed of individual road segment. FW-STN integrates both the convolutional neural network (CNN) and long short-term memory (LSTM), in which CNN captures the recent features from the defined critical road sequence, and LSTM learns the periodic features from the decomposed sub-series. These two types of derived features are fused for final prediction.

The main contributions of this paper can be summarized as follows:

1) We propose to quantify the correlation intensities between roads with traffic flow interaction, and hereby construct a critical road sequence. This sequence can filter and provide highly correlated traffic information for later prediction.

2) We propose to learn the periodic dependencies of traffic process with different-frequency sub-series decomposed by wavelet transform. This frequency decomposition and separate modeling prove to improve the accuracy of traffic prediction.

3) We build a novel short-term traffic forecasting model, FW-STN, which can well capture the identified recent and periodic spatio-temporal features in the historical traffic data.

4) We conduct extensive experiments on real-word traffic data. The results indicate that the proposed model outperforms nine baseline methods and achieves the state-of-the-art performance.

II. RELATED WORK

A. STATISTICAL AND DEEP LEARNING-BASED TRAFFIC FORECASTING

Time series-based statistical methods were early applied to traffic forecasting problems with the aim of finding the temporal patterns of stationary traffic dynamics. These approaches include Autoregressive Integrated Moving Average (ARIMA) [6], [7] and its variations, as well as Kalman filtering [8]. Subsequently, due to the rapid increase in traffic data collected from sensors deployed in vehicles and traffic infrastructure, data-driven machine
learning methods gradually became mainstream methods for traffic prediction. These methods include nearest neighbor (KNN) [9], [10], support vector regression (SVR) [9], gradient boosting regression tree (GBRT) [11], artificial neural networks (ANNs) [12]–[14], and some hybrid models [15], [16]. Overall, machine learning methods have achieved desirable performances as regards modeling non-linearity in traffic dynamics.

Recently, deep learning-based forecasting approaches have been widely developed. Deep neural networks such as deep belief networks (DBN) [17] and the stacked autoencoder (SAE) model [18] were utilized early on to capture the deep feature representation of traffic observations. However, these methods are unable to adequately model the complex spatial and temporal correlations involved. To deal with this, convolutional and recurrent neural networks have been widely implemented.

CNNs can learn structural and hierarchical features, and have also achieved breakthroughs in image, video, and sound recognition tasks [19]. A collection of studies has utilized CNNs to capture the spatial features of traffic or crowd flow dynamics [20]–[22]. CNNs usually assimilate input with a regular structure, such as a 1D sequence or 2D matrix. In these CNN-based models, ring roads are most studied, as they can be directly arranged into a 1D sequence or as one dimension of a 2D matrix. For network-structured road networks, the latest researches utilize graphs to describe the topological structure and apply graph convolutional networks (GCNs) [23], [24] to extract the spatial features; some examples include DCRNN [25], ST-GCNN [26], GCRN [27], and ST-MGCN [28].

Recurrent neural networks (RNNs) such as LSTM and gated recurrent unit (GRU) are also being widely utilized to capture the temporal features of traffic dynamics. In particular, LSTM can capture both long-term and short-term memories, and has thus seen widespread utilization in the modeling of temporal correlations [29]–[31].

Moreover, many hybrid approaches have been built by combining convolutional and recurrent neural networks in order to simultaneously model spatial and temporal correlations; examples include DNN-BTF [3], DMVST-Net [32], and STDN [33]. Motivated by the success of convolutional and recurrent neural networks in traffic forecasting, the present paper utilizes both CNN and LSTM to capture the spatio-temporal features among traffic dynamics.

B. SPATIAL CORRELATION MODELING FOR INDIVIDUAL ROAD SEGMENT

Spatial correlation has been widely exploited in recent traffic prediction studies. The key issues associated with these models include how to determine correlated roads and capture spatial characteristics. The most direct approach is to utilize the topological relationships between roads, i.e., the upstream and downstream ones. Li et al. [34] found that integration with data from both the nearest upstream and downstream junctions significantly improved model performances, while Yao et al. [35] utilized speed data of first-order upstream and downstream road links to estimate the traffic speed of the center road.

In addition to the topological strategy, some statistical methods have also been proposed to quantify the correlations between roads. For instance, Zou et al. [36] utilized auto-correlation and cross-correlation functions to examine the temporal and spatial correlation of traffic data from adjacent traffic links. Min and Wynter [37] estimated the reachability between link pairs during certain time steps and built a spatial correlation matrix to determine which other links have an impact on the target link. Cai et al. [9] defined a concept of equivalent distance to measure the correlations; this distance metric integrated both spatial properties (i.e., physical distance and connective grade) and temporal properties (i.e., correlation coefficient of traffic data series).

However, these topological or statistical methods neglect the real traffic interactions between road segments. The dynamic traffic states of roads are directly generated by the volume change during the complicated diffusion of traffic flow. Traffic flow between connected roads directly results in correlations between their traffic states, and can also be used to indicate the correlation intensity: that is, the larger the traffic flow, the stronger the correlation. Thus, traffic information from adjacent roads with large traffic flow into/from the target road is preferable for speed forecasting.

After the correlated roads have been selected, spatial characteristics are captured by means of various methods, including linear modeling methods [36], [38] and nonlinear machine learning methods [9], [35], [39]. Among the deep learning-based methods, CNN is a powerful means of capturing spatial characteristics via convolution operation on the local units. However, it is difficult to map complex urban road networks into regular structures for CNN while also preserving the topological relations between roads. Therefore, in the proposed approach, we not only select strongly correlated roads, but also re-arrange them into a 1D sequence ordered by topological relationships.

C. WAVELET TRANSFORM INTEGRATION IN TRAFFIC FORECASTING

Wavelet transform (WT) is often used to extract multi-frequency information in the analysis of non-stationary data, such as audio signals and images. Wavelet transform provides good feature localization in both the time and frequency domains [40]. WT has been applied in traffic forecasting to extract different frequency components of the original traffic data series and help with modeling the associated complex temporal characteristics.

Some studies utilize WT to remove the stochastic noise from the original data. Xie et al. [4] and Mousavizadeh Kashi and Akbarzadeh [41] employed WT to conduct data decomposition and preserved only the low-frequency components of original data series, then made predictions using a Kalman filter and ANN, respectively. In these works, the high-frequency components have been regarded as noise.
Other studies retain all decomposed components and make predictions on each one independently. The obtained prediction results of all frequency components are then reconstructed as the final forecasting result. For example, Sun et al. [42] decomposed flow data using WT and made predictions for each frequency component via SVM. Diao et al. [43] utilized discrete wavelet transform (DWT) for decomposition, making predictions by employing a tracking model for the low-frequency component and a Gaussian process model for the high-frequency component. Zhang et al. [44] employed motif-based graph convolutional recurrent neural network (Motif-GCRNN) and Autoregressive Moving Average (ARMA) to model the low-frequency and high-frequency components, respectively. Moreover, some other works have further eliminated the noise information of the high-frequency components via thresholding before making predictions [5, 45–47].

WT applies convolution operations to the original data series. The calculation of each data point covers the surrounding data so that the local frequency information can be obtained. Some studies, however, have failed to notice this characteristic when doing decomposition, instead directly decomposing the whole experimental data series, which inevitably causes information leakage. As shown in Fig. 3, for a given time interval $t + 1$, the decomposed data shortly before $t + 1$ (i.e., decomposed recent data in Fig. 3) contains the original information at $t + 1$. These data are then input into the model to predict the traffic state at $t + 1$; however, the output information has already been leaked, meaning that unauthentic prediction results are obtained [48]. To avoid this leakage issue, the decomposition should be conducted on a data series that ends before the output time interval [49].

![FIGURE 3. Example of information leakage.](image)

Furthermore, the border distortion of wavelet transform can cause decomposition errors at both ends of data series [49]. For example, if one decomposition ends at $t$ and another one ends at $t + 1$, these two decompositions will produce different results at the recent time intervals (e.g., $t, t − 1, t − 2$), and then unstable inputs to forecasting model will cause prediction errors, especially during the online prediction scenarios. Therefore, in this study, we do not utilize decomposed recent data, but instead only employ the decomposed data from previous days to model the periodic multi-frequency characteristics. The data around the predicted time interval from previous days will be far from the borders of the decomposition series, meaning that the effects of border distortion can be avoided.

### III. METHODOLOGY

As presented in Fig. 4, the proposed FW-STN framework consists of three parts. Part (a) captures the recent features of near-term traffic dynamics. A critical road sequence is firstly constructed; this sequence contains strongly correlated adjacent roads as well as the predicted one itself. With this sequence and the recent time intervals as orthogonal axes, two traffic matrices are built from traffic speed and volume observations, respectively. These matrices are then input into CNN for feature extraction. Part (b) captures the periodic features among long-term historical traffic speed series. WT-based frequency decomposition is utilized to decompose the original speed series into two sub-series with low and high frequency, respectively. The multi-frequency periodic features are then captured by periodic LSTM networks with speed sub-series. Finally, part (c) fuses the recent and periodic features to create the final feature, which is then input into a fully connected layer for traffic prediction.

#### A. THE CAPTURE OF RECENT FEATURES

1) CRITICAL ROAD SEQUENCE CONSTRUCTION

We construct the critical road sequence in the form \([u_0, \ldots, u_2, u_7, p, d_1, d_2, \ldots, d_n]\): here, \(p\) is the predicted road, while \(u\) and \(d\) are the selected upstream and downstream roads of \(p\), respectively. The sequence is constructed by means of a filter-and-sort strategy. The filter process selects strongly correlated roads by flow rates, while the sort process then arranges the selected roads into an ordered sequence by means of adjacency relations.

Firstly, flow rate is defined to measure the proportion of traffic flow between a road segment and its adjacent roads, and used to indicate the relative intensities of traffic interaction between these roads. For a given road \(i\), \(j\) and \(k\) denote one of its upstream and downstream roads, respectively. The inflow rate \(I_i\) is the inbound flow ratio from \(j\) to \(i\), while the outflow rate \(O_i\) is the outbound flow ratio from \(i\) to \(k\). For the road whose order is more distant than one, flow rate can be obtained by the accumulated multiplication of flow rates among the shortest path between the current road and target road. For instance, in the road network presented in Fig. 5, the inflow rate \(I_0\) can be calculated by \(I_1 \times I_7 \times I_0\).

In the filter process, the inflow and outflow rates of all adjacent roads within \(n\)-order are firstly calculated and sorted in descending order, respectively. The top-ranking \(n\) upstream and \(n\) downstream roads are then selected. The sort process arranges the selected roads by adjacency order and distance deviations. More specifically, the selected roads are first sorted by adjacency order to the predicted road. Upstream roads are sorted in descending adjacency order, and downstream roads in ascending adjacency order. If multiple possible road sequences are obtained with the same order (as shown in Fig. 5: \([1, 2, 7, 0, 8, 9, 11]\) and \([1, 2, 7, 0, 8, 11, 9]\)), the expected sequence order is determined by minimizing the weighted sum of the distance deviations.
Here, the distance between two roads refers to the minimum number of intersections required to travel from one road to another. These distances can be deviated when arranging net-structured roads into a 1D sequence, and the topological relations among roads will thus be altered. We aim to minimize the overall deviations of distances brought about by sorting. Assuming that there are \( m \) adjacent road pairs among all selected roads, the distance of the \( i \)-th road pair is \( d_{in}^i \) in the road network and \( d_{rs}^i \) in the sorted road sequence. We then find the road sequence \( rs \) such that:

\[
rs = \text{arg min} \sum_{i=1}^{m} \max(I_i, O_i) \times \left| d_{in}^i - d_{rs}^i \right|
\]

where \( \max(I_i, O_i) \) is the biggest flow rate of the two roads. The flow rate-weighting aims to impose a far higher penalty on the distance deviation of adjacent roads with a large flow rate. The derived road sequence \( rs \) can preserve the particular adjacency relations with the predicted road by means of adjacency order sorting, as well as the overall adjacency relations among all selected roads by distance deviation sorting.

2) TRAFFIC MATRIX CONSTRUCTION AND FEATURE EXTRACTION

Two traffic matrices, namely the speed matrix and volume matrix, are built to capture the recent features. According to the fundamental diagrams (FD) of traffic flow, there are strong correlations between the speeds and volumes of an urban road [50]. We therefore add traffic volume information into the model. In each input traffic matrix, the critical road sequence is set as the vertical axis, while the recent time intervals from \( t-p \) to \( t \) are set as the horizontal axis. The architecture of the CNN is illustrated in Fig. 6; here, there are \( K \) convolution layers used to extract spatio-temporal patterns, along with a fully connected layer to reduce feature dimensions.

The transformation in each convolution layer is defined as:

\[
X^k = f(W^k \ast X^{k-1} + b^k)
\]

where \( X^k \) is the output of \( k \)-layer, \( \ast \) denotes the convolution operation, \( W^k \) and \( b^k \) are learnable parameters in each layer, \( f \) is the activation function.
FIGURE 6. The architecture of CNN in FW-STN.

The fully connected layer is defined as:

\[ c_t = f(W_{fc}X^K + b_{fc}) \]  

where \( X^K \) is the final output of \( K \) convolution layers, \( W_{fc} \) and \( b_{fc} \) are the parameters of the FC layer, and \( c_t \) is the obtained recent feature representation.

B. THE CAPTURE OF PERIODIC FEATURES

1) FREQUENCY DECOMPOSITION

Due to its translation-invariance characteristics, stationary wavelet transform (SWT) [51] is selected to conduct the frequency decomposition. In real forecasting scenarios, new observed traffic data is continuously imported into the speed series. This decomposition also needs to be translated forwards to process newly incoming data. Translation-variant WT methods, such as DWT, produce different decomposition results for the same time intervals along the translated decomposition. Because the speed data at a certain time interval is usually input in multiple different forecasting steps, this decomposition inconsistency can bring about nondeterminacy when preparing input data.

SWT can decompose original data series with multiple levels, as shown in Fig. 7. At the level \( r \), two new series are produced—namely, the approximation coefficient series \( a^r \) and detail coefficient series \( d^r \)—by applying low-pass filters \( H^r \) and high-pass filters \( G^r \) to the approximation coefficient series of the former level. This approximation coefficient series contains the low-frequency components, while the detail coefficient series contains the high-frequency components of the original data series. Following multi-level decomposition, what remains is one approximation coefficient series and multiple detail coefficient series. These series carry distinct components of the original data series with different frequency distributions. With a higher decomposition level, the approximation series will contain lower-frequency components and occupy the more distinct trend part of the original data series. However, too high a level can also cause the approximation series to deviate substantially from the original data series.

Following the decomposition process, the reconstruction of the coefficient series can be accomplished by means of inverse stationary wavelet transform (ISWT). Each coefficient series, or the combination of certain series, can be reconstructed as a new data series that contains the specified frequency components of the original data series.

Based on SWT and ISWT, we conduct frequency decomposition as shown in Fig. 8 (a). After SWT is conducted at a certain level, the approximation coefficient series is independently reconstructed as a low-frequency sub-series for capturing the trend periodic features. The detail coefficient series is difficult to model due to its stochastic characteristics. Rather than reconstructing and modeling each detail series independently, which can be especially difficult, we instead reconstruct all detail coefficient series as one high-frequency sub-series to capture the overall stochastic periodic features. Fig. 8 (b) provides an example of the decomposition results with true traffic speed data within one week. The low-frequency sub-series exhibits a smooth changing trend and distinct periodicity, while the high-frequency sub-series stochastically vibrates on a small scale.

FIGURE 7. The SWT process.

FIGURE 8. Frequency decomposition process and results.

2) FEATURE CAPTURING BY PERIODIC LSTM NETWORKS

Two periodic LSTM networks with the same architecture are built to capture the periodic features among low-frequency and high-frequency speed sub-series, respectively. As shown in Fig. 9, in the LSTM network for low-frequency sub-series, some speed data from the previous \( c \) days before the predicted day \( d \) is selected from the sub-series and input into \( c \) LSTMs.
For each day, speed data from the time interval $t + 1 - q$ to $t + 1 + q$ are input into LSTM, where $t + 1$ is the predicted time interval. The output feature of each LSTM is denoted as $h_{d-\varepsilon}(1 \leq \varepsilon \leq c)$. We further concatenate the output of each LSTM from previous days, then input the concatenated feature into a fully connected layer and obtain the output feature $h_{L_{t}}$. As shown in Fig. 4, the output features from two sub-series, $h_{L_{t}}$ and $h_{H_{t}}$, have the same dimensions and are added together in an element-wise fashion to enable the final periodic features $h_{t}$ to be obtained.

C. FEATURE FUSION AND PREDICTION

The recent features $c_{t}$ and periodic features $h_{t}$ are fused via concatenation to form the final features $g_{t}$, which is then input into a fully connected layer activated by sigmoid for obtaining the predicted speed value at the next time interval $t + 1$, shown in Eq. (4). The mean squared error (MSE) is used as the loss function to train the whole network, which includes the CNN, periodic LSTM networks, and FC.

$$\hat{y}_{t+1} = \text{sigmoid}(Wg_{t} + b)$$ (4)

IV. EXPERIMENTS

A. DATASETS

The proposed approach is verified using a taxi GPS dataset collected from 1st October 2016 to 30th November 2016 (61 days) in Chengdu, China, provided by the on-demand ride service platform Didi Chuxing. The recorded information includes driver ID, order ID, recording timestamp and car position. The sample rate of GPS points is three seconds.

Traffic speed and volume data are estimated with this GPS dataset. All GPS points are first matched to road segments via the map matching algorithm [52]. The path distance along roads, divided by time intervals between two continuous GPS points, is calculated as a travelling speed record for the matched road segment. All speed records during the 61-day period are aggregated into 5 min intervals; the traffic speed of each road segment at each 5 min time interval is then obtained by averaging all speed records with it. The traffic volume of each road segment at each 5 min time interval is calculated as the number of taxis passing through this road segment during this time interval. Missing speed and volume data at some low-traffic time intervals are imputed by means of temporal-neighboring interpolation, which fills missing data by averaging the data from neighboring time intervals in the same day and the same time intervals in neighboring days.

Training and testing samples are then collected with these traffic data. Because the proposed approach captures periodic features from days in the past, data from first 10 days are reserved for sampling model input, data from subsequent 44 days are used as training data, and the final 7 days are used as testing data. In addition, in each day, the time period 0:00-8:00 is omitted since taxis are very rare during this time, and first 10 time intervals in the rest ones are reserved for sampling model input. In the end, 182 samples are collected in each day and 9282 samples are obtained in total 51 days.

The experimental road network contains 114 unidirectional road segments, as highlighted in Fig. 10. The blue lines represent road segments, while the red dots denote the intersections between road segments. The adjacent road segments of 114 predicted road segments are derived from the bigger road network covering the selected network.

B. EXPERIMENT SETUP AND COMPARISON BASELINES

1) MODEL IMPLEMENTATION

We implement our model in Python with Keras 2.2.4 and TensorFlow 1.13.1 backend. In the recent part, $K$ is set to 3 (number of convolutional layers), while each convolutional layer is activated via ReLU. We set all convolution kernel sizes to $3 \times 3$ with 64 filters. The critical road sequence within the first-order (i.e., three roads in total), and three time intervals before the predicted one are set to construct the traffic matrices. The output of the CNN is flattened and input into a fully connected layer with an output dimension of 64.
In the periodic part, we set the decomposition level of SWT to 4 and use the commonly utilized haar wavelet as the mother wavelet. The decomposed traffic data from the previous seven days and three time intervals around the predicted time interval for each day are input into periodic LSTM network. Each LSTM is built with one layer and the dimension of hidden representation is 64; the output dimension of periodic features is also 64. The dimension of the final fused spatio-temporal features is 128. The FW-STN model is optimized via the Adam optimizer and the learning rate is set to 0.00005. Finally, the batch size in our experiment is set to 128.

2) BASELINE MODELS
We compared our model with the following baseline models and tuned the parameters for all methods. For all compared methods, traffic data is not processed with frequency decomposition.

1) Historical average (HA): The average traffic speed at the predicted time interval for all previous days is set as the predicted result.

2) ARMA: The traffic speed series before the predicted time interval is used to train the ARMA model and predict the traffic speed at the next time interval.

3) SAE [18]: This method is built using three autoencoder layers with 400 units. The input of FW-STN is flattened and concatenated as a vector to be input into SAE.

4) LSTM: The input of recent part from FW-STN is reshaped and input into an LSTM to capture the recent features.

5) CNN: The recent part of FW-STN is directly connected to a fully connected layer with the predicted traffic speed as output.

6) GCRN [27]: The selected roads are utilized to build the graph of the road network. Traffic speeds of the most recent three time intervals are input into GCN and then LSTM.

7) DCRNN [25]: A same road graph is built as in GCRN. The input and output time intervals in DCRNN are set as three and one.

8) DMVST-Net [32]: We build DMVST-Net with the same architecture, except that 2D convolution is replaced by 1D convolution on the critical road sequence within the third-order. The embedding input of each road segment is obtained based on the experimental road network.

9) STDN [33]: We build STDN with the same architecture and replace 2D convolution with 1D convolution, as well as applying traffic volume gating to the traffic speed when capturing spatial features.

3) METRICS
Mean Average Percentage Error (MAPE) and Rooted Mean Square Error (RMSE) are chosen to evaluate the forecasting accuracy of each road segment; these metrics are defined as follows:

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_{t+1}^{i} - y_{t+1}^{i}}{y_{t+1}^{i}} \right|
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_{t+1}^{i} - y_{t+1}^{i})^2}
\]

where \( n \) is the number of testing samples, \( \hat{y}_{t+1}^{i} \) and \( y_{t+1}^{i} \) are the predicted and true traffic speed, respectively. The mean values of MAPE and RMSE in all predicted road segments are calculated to facilitate evaluation of the overall model performance.

**TABLE 1. Performance comparison between different forecasting approaches.**

| Method       | MAPE  | RMSE  |
|--------------|-------|-------|
| HA           | 19.08% | 5.531 |
| ARMA         | 19.33% | 5.612 |
| SAE          | 17.35% | 5.174 |
| LSTM         | 18.57% | 5.377 |
| CNN          | 18.31% | 5.354 |
| GCRN         | 18.03% | 5.694 |
| DCRNN        | 17.92% | 5.660 |
| DMVST-Net    | 17.92% | 5.292 |
| STDN         | 17.34% | 5.128 |
| FW-STN       | 16.83% | 5.081 |

C. PERFORMANCE COMPARISON
Table 1 presents the comparison of different forecasting models. The results indicate that the proposed model, FW-STN, achieves the lowest MAPE (16.83%) and RMSE (5.081) among all comparison methods. More specifically:

1) HA and ARMA obtain worse performances than deep learning-based methods, as they are unable to model the complex non-linearity of traffic dynamics.

2) Among deep learning-based methods, LSTM, CNN, GCRN, DCRNN and DMVST-Net only capture the recent features of traffic dynamics; they thus perform worse than SAE, STDN and FW-STN, which capture the recent and periodic features simultaneously. This indicates the importance of periodic information.

3) The network-wide forecasting method, GCRN and DCRNN, achieves good MAPE performances among five deep learning-based and recent feature-captured methods. The other three, LSTM, CNN and DMVST-Net, are all built for individual road segment. However, their RMSE performances are obviously poor. Since they are optimized for achieving better average performances of all roads but not the individual one, the performances vary more greatly among different roads.

4) The proposed model, FW-STN, outperforms all other methods. Compared to SAE and STDN, FW-STN
further extracts significant traffic information by constructing the critical road sequence and capturing the multi-frequency periodic features via frequency decomposition. These strategies help FW-STN to better model the spatio-temporal characteristics of urban road traffics.

D. MODEL EVALUATION AND ANALYSIS

In this section, the effectiveness of critical road sequence and frequency decomposition is evaluated. Subsequently, the impact of temporal hyperparameters on the model performance is analyzed.

1) EFFECTIVENESS OF CRITICAL ROAD SEQUENCE

We compare model performances with different selection methods to determine the correlated roads. In addition to the constructed critical road sequence (CRS), the topology relation (TR), time series correlation coefficient (TS), and equivalent distance (ED) proposed by Cai et al. [9] are also utilized for comparison. All of these methods select roads within $n$-order to the predicted road. More specifically:

1) TR selects all candidate roads within $n$-order;
2) TS selects $n$ upstream and $n$ downstream roads with the biggest correlation coefficients of the speed series;
3) ED selects $n$ upstream and $n$ downstream roads with the biggest distance values.

For CRS, TS and ED, there are $2n + 1$ roads selected in total. By contrast, the total number of roads selected by TR is road-specific and depends on the local road distributions. All methods sort the selected roads in the same way as CRS. We evaluate the performances of these methods within different adjacency orders, and the results are presented in Fig. 11.

![FIGURE 11. Performance comparison between different selection methods.](image)

The CRS method obtains the best performance within the first-order among all compared results, and is always superior to other three methods (especially to TR and TS) within different orders. CRS filters out many low-correlated roads compared to TR; non-critical traffic information is thus discarded, while model complexity is also reduced. CRS also outperforms TS, indicating that traffic flow interaction is a more effective indicator than the time series correlation coefficient for selecting strongly correlated roads.

Within the second and third-order, ED obtains very close accuracies to CRS. ED quantifies the spatial correlation using a hybrid metric that includes topological, physical and time series measurements. As the adjacency order is extended, the number of candidate roads also increases rapidly. In this case, ED can select strongly correlated roads under different correlation modes, and provides mixed traffic information. Despite this, CRS is still capable of picking out the most critical adjacent roads for prediction purposes, meaning that it obtains the best performance within only the first-order.

In addition, the overall model performance of all tested roads deteriorates for all methods when the input adjacency order is extended. More weakly correlated or distant roads generally have weaker influence in the traffic state of the predicted road; as a result, inputting a large number of their traffic observations may introduce irrelevant information to the model and thereby degrade the model performance.

2) EFFECTIVENESS OF FREQUENCY DECOMPOSITION

To demonstrate the effectiveness of the frequency decomposition and feature fusion in the periodic part, we compared the performances of three models: no frequency decomposition (F-STN), input with low-frequency sub-series only (FW-L-STN), and input with high-frequency sub-series only (FW-H-STN). The results are illustrated in Fig. 12.

![FIGURE 12. Performance comparison of different periodic feature capture strategies.](image)

As shown in Fig. 12, FW-STN captures periodic features from both low-frequency and high-frequency sub-series, and thereby achieves the best performance. F-STN performs worse than FW-STN, which validates the positive effect of frequency decomposition. Despite this, the accuracy of F-STN is still higher than both FW-L-STN and FW-H-STN. F-STN contains complete periodic information, while the other two only input low-frequency or high-frequency information. The absence of either could result in model performance deterioration.
In addition, the performance of FW-L-STN is superior to that of FW-H-STN. The low-frequency sub-series contains
the main component and approximates the general changing
trend of the original speed series; by contrast, the high-
frequency sub-series only contains the detailed component
with small-scale random variations. The low-frequency sub-
series is thus more important than the high-frequency sub-
series for periodic features capture.

We further evaluate the impact of decomposition level on
model performance. As shown in Fig. 13, the model achieved
the best performance at the level 4. A lower decomposition
level may not sufficiently separate the low-frequency and
high-frequency component, leaving them mixed with each
other; moreover, higher decomposition levels can excessively
smooth the low-frequency sub-series and cause it to deviate
from the original speed series, as well as rendering the high-
frequency sub-series more complicated.

![FIGURE 13. Performance comparison of different decomposition levels.](image)

3) TEMPORAL HYPERPARAMETER EVALUATION
In the next step, temporal hyperparameters are also tuned to
verify their impact on model performance: these include the
number of historical days $c$, the number of time intervals in
each historical day $2q + 1$ and the number of recent time
intervals $p$. The results are presented in Fig. 14.

As shown in Fig. 14 (a), model performance is improved
as the historical days $c$ increase until reaching 7 days, and
then drops. This can be attributed to the rotation of work-
ing and weekend days in human travel routines. Historical
traffic data from one previous week contains both weekly
and daily periodic traffic information and leads to the best
model performance. For longer period of historical traffic
information, the model performance could deteriorate due to
the repetition of daily periodic information and the increase
of model complexity.

Moreover, illustrated from Fig. 14 (b) and (c), the
model obtains the best performance when being input with
decomposed traffic speeds of three historical time intervals
(i.e., 15 min) in each previous day, along with original traffic
speeds and volumes in three recent time intervals. The peri-
odic traffic data reveals the empirical traffic states around
the predicted time interval in historical days, while the recent
traffic data reflects the real-time traffic dynamics before
the predicted time interval. Since road traffic changes dynami-
cally, short and close traffic observations can provide impor-
tant trend information towards the predicted time interval;
however, a too-long input may bring in interference informa-
tion to the model.

![FIGURE 14. Performance evaluation of temporal hyperparameters.](image)
V. CONCLUSION

In this paper, we propose a hybrid model, namely a flow and wavelet-integrated spatio-temporal network (FW-STN), for short-term traffic speed forecasting of individual road segment. This model can capture the recent features of near-term traffic dynamics as well as the periodic features of long-term historical speed series. In the recent part, we construct a critical road sequence consisting of adjacent roads that diffuse large-part traffic flow with the predicted road in order to model the local spatial correlations. In the periodic part, SWT-based frequency decomposition is utilized to decompose the original speed series into two sub-series, with low and high frequency, to model the multi-frequency characteristics of speed series. Experiments on a real traffic dataset demonstrate that the proposed approach outperforms nine existing baseline methods. Performance analysis further proves the effectiveness of critical road sequence construction for spatial correlation modeling and frequency decomposition for periodic features capture.

In future work, we will explore the hybrid selection method with multiple metrics, including traffic flow, time series correlation coefficient, etc. Not only locally connected roads but also distant roads can be captured for various correlation modes. In addition, approaches for forecasting multiple road segments at one time, such as the multitask-based deep learning approach [17], are also necessary, as they may be more applicable in large-scale road network forecasting.

REFERENCES

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Data-driven intelligent transportation systems: A survey,” IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1624–1639, Dec. 2011.
[2] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, “Road traffic forecasting: Recent advances and new challenges,” IEEE Intell. Transp. Syst. Mag., vol. 10, no. 2, pp. 93–109, Apr. 2018.
[3] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, “A hybrid deep learning based traffic flow prediction method and its understanding,” Transp. Res. C, Emerg. Technol., vol. 90, pp. 166–180, May 2018.
[4] Y. Xie, Y. Zhang, and Z. Ye, “Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition,” Comput-Aided Civil Infrastructure Eng., vol. 22, no. 5, pp. 326–334, Jul. 2007.
[5] J. Wang and Q. Shi, “Short-term traffic speed forecasting hybrid model based on Chaos–Wavelet analysis-support vector machine theory,” Transp. Res. C, Emerg. Technol., vol. 27, pp. 219–232, Feb. 2013.
[6] B. M. Williams and L. A. Hoel, “Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results,” J. Transp. Eng., vol. 129, no. 6, pp. 664–672, Nov. 2003.
[7] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, “Predicting Taxi–Passenger demand using streaming data,” IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 1393–1402, Sep. 2013.
[8] J. Guo, W. Huang, and B. M. Williams, “Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification,” Transp. Res. C, Emerg. Technol., vol. 43, pp. 50–64, Jun. 2014.
[9] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, “A spatiotemporal correlactive k-nearest neighbor model for short-term traffic multistep forecasting,” Transp. Res. C, Emerg. Technol., vol. 62, pp. 21–34, Jan. 2016.
[10] B. L. Smith, B. M. Williams, and R. Keith Oswald, “Comparison of parametric and nonparametric models for traffic flow forecasting,” Transp. Res. C, Emerg. Technol., vol. 10, no. 4, pp. 303–321, Aug. 2002.
[11] Y. Zhang and A. Haghighi, “A gradient boosting method to improve travel time prediction,” Transp. Res. C, Emerg. Technol., vol. 58, pp. 308–324, Sep. 2015.
