




Article

# Graph Attention Informer for Long-Term Traffic Flow Prediction under the Impact of Sports Events

Yaofeng Song , Ruikang Luo , Tianchen Zhou , Changgen Zhou \* and Rong Su \*

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore; yaofeng.song@ntu.edu.sg (Y.S.); ruikang001@e.ntu.edu.sg (R.L.); zhou0429@e.ntu.edu.sg (T.Z.)

\* Correspondence: 0127000107@ahcme.edu.cn (C.Z.); rsu@ntu.edu.sg (R.S.)

**Abstract:** Traffic flow prediction is one of the challenges in the development of an Intelligent Transportation System (ITS). Accurate traffic flow prediction helps to alleviate urban traffic congestion and improve urban traffic efficiency, which is crucial for promoting the synergistic development of smart transportation and smart cities. With the development of deep learning, many deep neural networks have been proposed to address this problem. However, due to the complexity of traffic maps and external factors, such as sports events, these models cannot perform well in long-term prediction. In order to enhance the accuracy and robustness of the model on long-term time series prediction, a Graph Attention Informer (GAT-Informer) structure is proposed by combining the graph attention layer and informer layer to capture the intrinsic features and external factors in spatial-temporal correlation. The external factors are represented as sports events impact factors. The GAT-Informer model was tested on real-world data collected in London, and the experimental results showed that our model has better performance in long-term traffic flow prediction compared to other baseline models.

**Keywords:** ITS; deep learning; traffic flow prediction; attention; sports events



**Citation:** Song, Y.; Luo, R.; Zhou, T.; Zhou, C.; Su, R. Graph Attention Informer for Long-Term Traffic Flow Prediction under the Impact of Sports Events. *Sensors* **2024**, *24*, 4796. <https://doi.org/10.3390/s24154796>

Academic Editor: Jordi Mongay Batalla

Received: 16 June 2024  
Revised: 17 July 2024  
Accepted: 23 July 2024  
Published: 24 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Traffic flow prediction is a core component of an Intelligent Transportation System (ITS) [1]. An ITS integrates people, vehicles, and roads into a comprehensive consideration so that they can work closely together to achieve a synergistic effect [2]. Traffic flow prediction is a rather complex process with multiperiod features, which is affected by a variety of factors such as traffic patterns, abnormal events, bad weather, and data collection [3]. For the government, traffic flow prediction mainly serves at the macro level, such as in urban master planning, urban transportation development policy, and urban transportation professional planning [4]. Accurate and efficient traffic flow prediction helps to alleviate urban traffic congestion and improve urban transportation efficiency [5,6]. For individual travelers, traffic flow prediction is an important reference for path planning [7], which is helpful for people to plan their travel routes in advance to avoid getting into traffic congestion.

In modern society, traffic congestion is a common problem [8,9]. Traffic congestion not only causes inconvenience to people's traveling, but it also may produce a series of problems for some specific occasions, such as sports events. There have been some reports of teams not being able to get to the stadium on time due to traffic congestion, thus causing games to be postponed. For fans, the delay has forced them to readjust their follow-up plans, and for those who traveled from afar, this may cause more inconvenience. During large-scale sports events, urban traffic will see a surge in urban road traffic and public transport passenger flow [10]. Compared to normal urban traffic planning, the traffic organization during large-scale sports events has unique travel and service characteristics [11]: it requires safe, smooth, reliable, and high-level services for the event traffic and must ensure the normal operation of the host city's traffic system at the same time.

With the arrival of the era of big data, the development of big data's enabling of transportation has attracted wide attention [12]. Based on a better traffic flow prediction result, the traffic police can conduct timely traffic diversion to ease congestion caused by excessive traffic. While companies such as cabs, online car-hailing and bike-sharing can conduct vehicle scheduling to safeguard the public's need for transportation. Short-term prediction provides accurate forecasting from minutes to hours, thus benefiting applications such as real-time traffic control and traveler information systems. This includes practices like adjusting traffic signal timings based on real-time data, rapid response to incidents such as accidents and breakdowns to minimize disruptions, and providing real-time updates and rerouting suggestions through navigation systems. Meanwhile, long-term prediction offers forecasts over weeks to months, which is helpful for developing traffic management plans for special events to handle temporary surges. In this case, how to effectively use big data to make good traffic flow prediction, especially under the impact of special events, has become a new challenge.

In terms of the development process of traffic flow prediction, the methods can be categorized into classical methods and deep learning methods [13]. Among the classical methods, models based on statistical methods have better interpretability, such as the Historical Average Model (HA) [14], the Autoregressive Integrated Moving Average (ARIMA) [15], and the Kalman filter [16]. Models based on traditional machine learning methods are more flexible, such as the Support Vector Machine (SVM) [17], K-Nearest Neighbor (KNN) [18], and Hidden Markov Model (HMM) [19]. However, classical methods are difficult to characterize the deep nonlinear relationships in spatiotemporal series, and they are not ideal for predicting data with high data perturbation. In recent years, with the development of deep learning, more and more researchers tend to consider deep learning methods, such as the Convolutional Neural Network (CNN) [20], Recurrent Neural Network (RNN) [21], Graph Convolutional Network (GCN) [22], Gate Recurrent Unit (GRU) [23], Long Short-Term Memory (LSTM) [24], and their variants. These models perform well in short-term prediction, but their accuracy and stability are not satisfactory for long-term prediction.

Compared to the above analysis, we propose a novel deep neural network model, which can effectively combine the spatiotemporal information of traffic maps and accurately realize long-term traffic flow prediction. The contributions are as follows:

- To the best of our knowledge, this is the first study to consider the impact of major sports events in traffic flow prediction.
- The Graph Attention Informer (GAT-Informer) structure is firstly proposed to address the long-term traffic flow prediction problem by combining the Graph Attention Network and the Informer Network.
- In addition to using the classical dataset, another dataset used to verify and test GAT-Informer is newly collected real-world data containing real sports events. This dataset has not been applied to other published articles.

The rest of the work is organized as follows: Section 2 issues the problem and describes the development of traffic flow prediction methods. Section 3 first gives the quantitative criteria for the impact of sports events on traffic flow; it then follows with the description of the model structure. Section 4 presents the detailed experimental results and the visualization of model performance comparisons with analysis. Section 5 summarizes the conclusions and the outlook for future research.

## 2. Related Research

### 2.1. Problem Statement

Traffic flow is the number of vehicles passing through a spatial unit (e.g., a roadway segment or a traffic detector) in a given time period [25]. Traffic flow prediction can be categorized into short-term prediction and long-term prediction according to the prediction length. Long-term traffic flow prediction corresponds to transportation planning, while short-term traffic flow prediction corresponds to traffic guidance, traffic management, and

control. There is no clear criterion for which category it belongs to, and it is generally believed that a prediction length of 5–30 min is short-term, while more than 30 min is long-term [26,27]; some researchers also define 1 day and above as long-term [28]. In this paper, the prediction length of 30 min or less is regarded as a short-term prediction. For a large-scale sports event, the traffic department needs to do traffic planning in advance so as to establish an efficient and coordinated departmental communication and traffic command system to ensure the effective implementation of the various stages of planning and orderly operation.

## 2.2. Classical Prediction Model

Without considering traffic maps, traffic flow prediction problems can be reduced to a time series analysis. ARIMA is the most common time series prediction model in statistical modeling [29]. It contains Autoregressive (AR), Integrated (I), and Moving Average (MA) models. The AR model captures historical trends and makes forecasts accordingly, but it is not ideal to deal with data with sudden changes. The I model performs linear subtraction of data with equal period intervals to smooth the data. The MA model makes up for the shortcomings of the AR model and is used to deal with noisy terms. Many variants of ARIMA have been proposed to continuously optimize the model implementation over time. Seasonal ARIMA [30] supports time series data with seasonal components, Subset ARIMA [31] improves the accuracy of short-term prediction tasks, and FARIMA [32] improves the accuracy of long-term prediction tasks. However, these models require data with high smoothness and a linear relationship, which makes it difficult to deal with data containing special events such as sports events.

## 2.3. Deep Learning-Based Prediction Model

In order to utilize the advantages of each neural network to realize a better prediction performance, a combination of two or more models is generally used. With the proposal of an attention mechanism in the Transformer [33], many researchers applied it to traffic flow prediction. DNN-BTF [7] applies an attention mechanism to automatically learn the importance of past traffic flow information and uses a CNN to mine the spatial features, as well as an RNN to mine the temporal features of traffic flow. GCN-DHSTNet [34] dynamically learns the spatiotemporal features of traffic crowd flow data on a citywide scale and simultaneously predicts the traffic flow in each area. STGCN [26] is the first to apply graph convolution to the traffic prediction problem. Based on this, ASTGCN [35] introduces an attention mechanism to capture spatiotemporal correlations. GMAN [36] predicts the traffic conditions at different locations at different time steps on a road network graph. DMSTGCN [37] devises a dynamic graph constructor and a dynamic graph convolution method to propagate the hidden states of the nodes based on dynamic spatial relationships. STFGCN [38] efficiently learns the hidden spatiotemporal dependencies using a novel fusion operation of various spatial and temporal graphs to process different time steps in parallel.

In conclusion, motivated by related works and existing problems, the Graph Attention Informer (GAT-Informer) network for long-term traffic prediction, which combines both the spatial and temporal correlations of each detector in the traffic communication network while satisfying the high accuracy requirement of long-term time series prediction is proposed in this paper. In the next section, the structure of the entire model will be illustrated in detail.

## 3. Methodology

### 3.1. Problem Formulation

At time  $t$ , the input data contain the traffic flow information of each detector,  $\mathcal{X}_t = [x_1, x_2, \dots, x_n]$ , where  $\mathcal{X}_t \in R^{n \times F}$ ,  $x_a \in R^F$  refers to the  $a$ th input node with a feature dimension  $F$ , and  $n$  is the number of total detectors. A graph is the basic structure of a

Graph Neural Network (GNN) [39], and the traffic map is a special type of graph. It can be defined as follows:

$$G = (V, E, A) \quad (1)$$

where  $V$  is the set of nodes,  $E$  is the set of edges, and  $A$  is the adjacency matrix.

Node set  $V$  with  $n$  nodes contains the traffic flow information for each node, and different nodes are connected by the edge set  $E$  with a link adjacency matrix  $A$ . We consider a directed graph, and the node adjacency is described in the adjacency matrix  $A$  indicating the connectivity between nodes.

The main purpose of the graph-based traffic flow prediction is to predict the future  $N$  step traffic flow, which can be described as follows:

$$Y = f(X; G) \quad (2)$$

$$X = [\mathcal{X}_{t-M}, \mathcal{X}_{t-M+1}, \dots, \mathcal{X}_t]^T \quad (3)$$

$$Y = [\mathcal{Y}_{t+1}, \mathcal{Y}_{t+2}, \dots, \mathcal{Y}_{t+N}]^T \quad (4)$$

where  $X$  is the historical traffic flow data,  $Y$  is the future traffic flow value,  $f$  is the relationship function,  $M$  is the time step of historical data usage, and  $N$  is the prediction horizon.

The effect of sports events is a dynamic factor which varies by time. An impact factor  $\sigma$  is utilized to denote the effect of sports events in the temporal domain. At time step  $t$ , the impact factor is described as  $\sigma_t \in R^n$  and the  $M$  steps  $\sigma \in R^{M \times n}$ . With the consideration of sports events effect, the prediction can be further expressed as:

$$Y = f(X, \sigma; G) \quad (5)$$

$$\sigma = [\sigma_{t-M}, \sigma_{t-M+1}, \dots, \sigma_t]^T \quad (6)$$

In conclusion, the issue of traffic flow prediction under the impact of sports events is to find the relationship function  $f$ , which will be realized by our GAT-Informer model. In the next subsection, the mathematical expression of the external factor that we considered in this paper will be introduced in detail.

### 3.2. Definition of Sports Events Impact Factor

The impact of sports events is generally temporary, and it may cause a spike in traffic flow around the arena. It usually occurs before the start of the game and after the end of the game, which indicates a large number of people entering the vicinity of the arena and a large number of people leaving the vicinity of the arena, respectively. Hence, we can assume that sports events have a weak impact on traffic flow on other days. In this paper, we consider both of the temporal and spatial effects of sports events, since they are also geographically limited.

Based on these assumptions, we use a function with a shape similar to the Gaussian distribution to represent the extent to which sports events affect traffic flow during this special period. In order to make the curve smoother and cover more time periods, we choose the random variable  $Z$  to obey the Gaussian distribution with mathematical expectation  $\mu$  and variance  $\text{Var}(Z)$ , which is written as  $Z \sim N(\mu, \text{Var}(Z))$ . Furthermore, the probability density function is amplified by  $\text{Var}(Z)$  to enhance its amplitude and enlarge the influence of impact factor  $\sigma$ . In order to better express the continuous impact on traffic flow before and after the game, a maximum operation is applied to these two distribution functions. Thus, the sports event's impact factor  $\sigma$  can be calculated as follows:

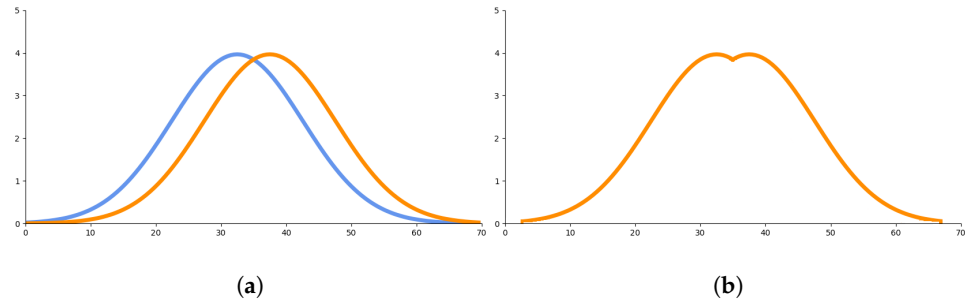
$$\sigma = \mathbf{max}[f(z), g(z)] \quad (7)$$

$$f(z) = \frac{10}{\sqrt{2\pi}} \mathbf{exp}\left(\frac{-[z - (t_{start} - \Delta t)]^2}{200}\right) \quad (8)$$

$$g(z) = \frac{10}{\sqrt{2\pi}} \exp\left(-\frac{[z - (t_{end} + \Delta t)]^2}{200}\right) \quad (9)$$

where  $t_{start}$  is the start time of the game;  $t_{end}$  is the end time of the game;  $z$  is the sampling time point. As most people will choose to enter the stadium ahead of the start time and stay for a while after the game before leaving, we use  $\Delta t$  to represent this situation.

The sports events impact factor  $\sigma$  is illustrated in Figure 1.



**Figure 1.** Calculation procedure for  $\sigma$  in one game. (a) shows the original  $f(z)$  (blue curve) and  $g(z)$  (orange curve); (b) shows the value  $\sigma$  after taking maximum.  $\sigma$  in this figure only keeps the value between 30 h before the start of the game and 30 h after the end of the game for illustration.

Given that the geographical impact of sports events is significant, we introduce a geographical factor  $\beta$  that follows an amplitude-enhanced two-dimensional Gaussian distribution. For node  $i$ , this factor can be expressed as follows:

$$\beta_i = \exp\left(-\frac{[(x_i - x_0)^2 + (y_i - y_0)^2]}{2\tau^2}\right) \quad (10)$$

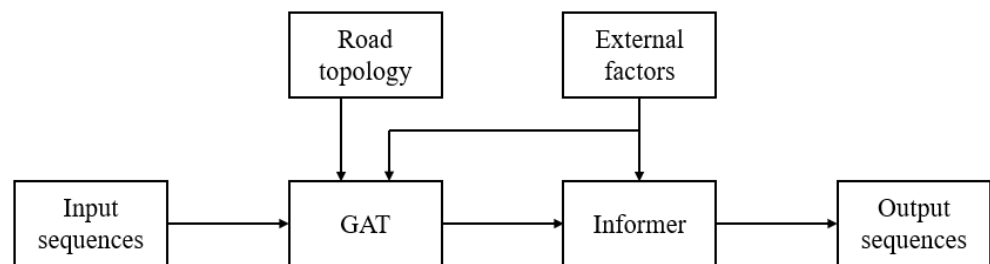
where  $(x_i, y_i)$  represents the geographical location of node  $i$ , and  $(x_0, y_0)$  denotes the arena location where the sports event is held. The  $X$  and  $Y$  directions share the same variance  $\tau^2$ .

Under the geographical effect factor  $\beta$ ,  $\sigma_i$  for node  $i$  can be further derived as follows:

$$\sigma_i = \beta_i \max[f(z), g(z)] \quad (11)$$

### 3.3. GAT-Informer Architecture

In this subsection, the structure of the GAT-Informer model is introduced in detail. GAT-Informer contains two layers: the Graph Attention layer and the Informer layer. The Graph Attention layer is used to fuse neighbor node information of the input data, so as to capture their spatial features, and then output the time series data with fused neighbor features. The Informer layer is used to process this time series data to achieve long-term prediction. The overview of the GAT-Informer architecture is illustrated in Figure 2. External factors will be introduced into both the GAT layer and the Informer layer for external factor fusion.



**Figure 2.** GAT-Informer architecture.

### 3.3.1. Graph Attention Layer

Graph Attention Network (GAT) [40] is a special neural network designed for processing graph-structured data. Unlike traditional neural networks, GAT takes into full consideration the relationships between data during processing, which enables the ability to capture the correlations between data more accurately without manual presetting. When there is a sports event, the adjacency relationships of the traffic network will change, thus rendering the pre-established road adjacency matrix less effective. That is why GAT stands out from other graph neural networks. It is capable of recognizing dynamic node adjacencies and capturing real-time correlations by calculating attention scores for adjacent nodes.

The input data of the Graph Attention layer contains three parts: a set of node features, the adjacency matrix, and the external impact factor for the consideration of sports events in the spatial information extraction. Node features can be described as an  $F \times n$  matrix  $\mathbf{H} = [h_1, h_2, \dots, h_n]$ ,  $h_i \in R^F$ , where  $n$  is the number of detectors in the traffic map, and  $F$  is the number of input features for each node. Since GAT is a layered structure, the input node features of the first GAT layer would be the traffic flow observation  $X$ . The distance between the detectors can be calculated from the longitude and latitude information in the traffic map, and then the adjacency matrix can be obtained by combining the connection relationship between them. The traffic map with spatial connectivity of the detectors is shown in Figure 3. This traffic map shows the connectivity and relative position of the nodes, but it does not represent the true spatial layout and shape of the road.

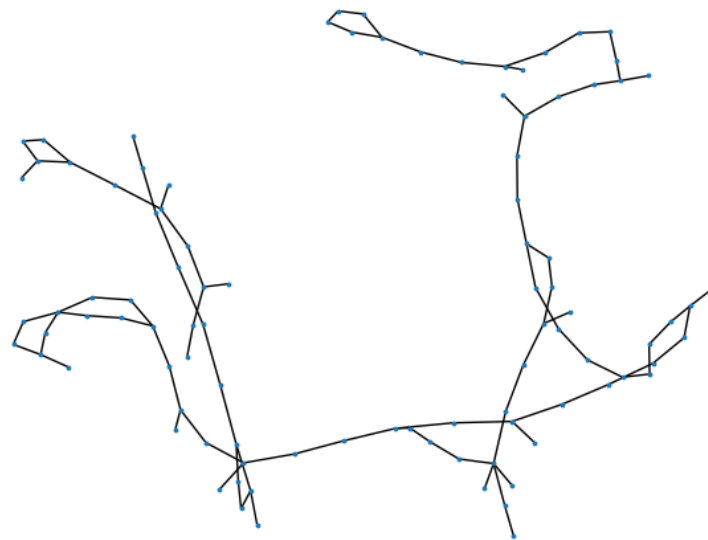


Figure 3. Traffic map.

The entire framework of Graph Attention layer is shown in Figure 4.

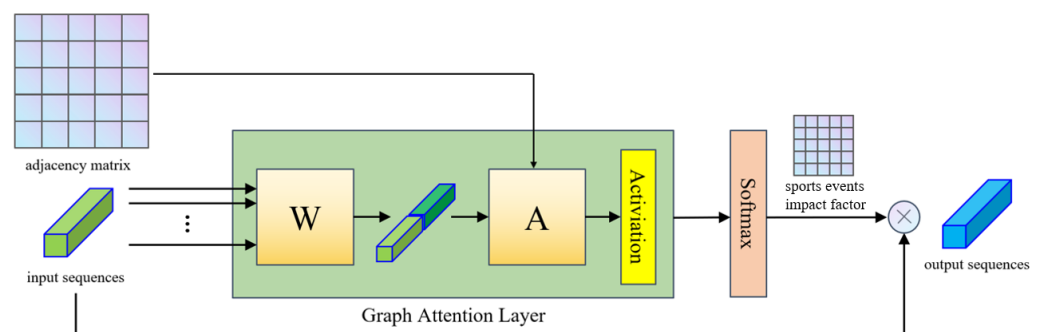


Figure 4. Graph Attention layer architecture.

As discussed above, the input to this layer contains an  $F \times n$  matrix (set of node features) and an  $n \times n$  matrix (adjacency matrix).  $\mathbf{W}$  and  $\mathbf{A}$  are both linear transformation matrices with shared parameters and are used for dimension transform. For node  $i$  and the neighbouring node  $j$ , attention score  $e_{ij}$  can be calculated as follows:

$$e_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{A}([\mathbf{W} \cdot h_i] \parallel [\mathbf{W} \cdot h_j])))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(\mathbf{A}([\mathbf{W} \cdot h_i] \parallel [\mathbf{W} \cdot h_k])))} \quad (12)$$

where  $\parallel$  means concatenation.  $N_i$  refers to neighboring nodes for node  $i$ .

Although the adjacency matrix is not explicitly present in the core formulas, its role is implicitly crucial. The adjacency matrix defines the relationships between nodes, and only adjacent nodes would be considered when calculating  $e_{ij}$ . In this paper, we employ a layered structure of GAT to perform feature aggregation, and each layer of GAT aggregates information from the direct neighbors of the nodes.

The aggregation of node features is enhanced using a multihead attention mechanism and the sports events impact factor  $\sigma$ . Final output of node  $i$  can be expressed as follows:

$$\mathbf{h}'_i = \varphi \left( \frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} e_{ij}^k \mathbf{W}^k (\mathbf{h}_j + \bar{\mathbf{h}}_j \sigma_j) \right) \quad (13)$$

where  $e_{ij}^k$  are normalized attention scores for the  $k$ th head of attention.  $\mathbf{W}^k$  is the corresponding linear input transformation, and  $\varphi(x)$  denotes a nonlinear activation function.  $\bar{\mathbf{h}}_j$  refers to the average value of  $\mathbf{h}_j$ .

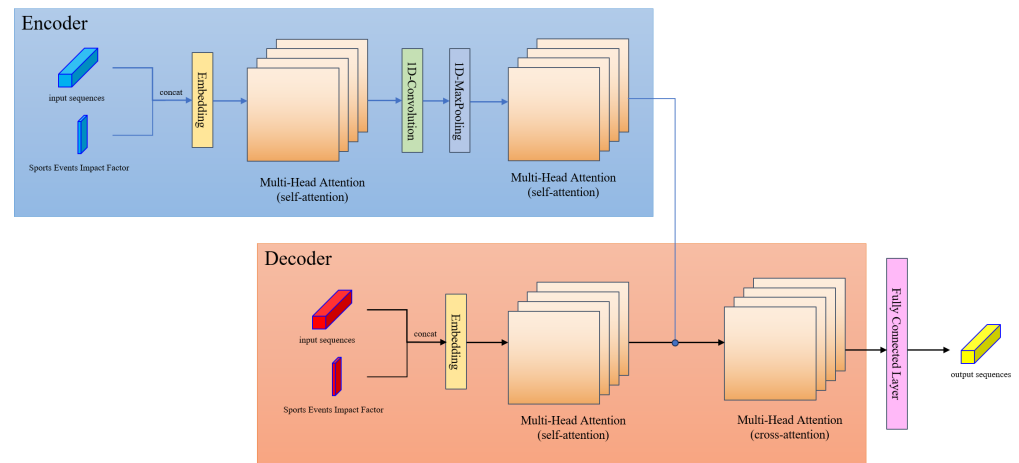
During the feature aggregation, signal from nodes affected by sports events would be enhanced on the basis of their distance to the arena. Final output matrix  $\tilde{\mathbf{H}}$  remains in the form  $F \times n$  and contains all aggregated nodes.

### 3.3.2. Informer Layer

Informer is a variant of the Transformer, which achieves higher efficiency than the Transformer in long-sequence time series forecasting [41]. In the scenario of traffic prediction considering sports events, the forecasting needs to cover a long time span, and the Informer is famous for its ability of long-term prediction and a lower computational cost compared to the Transformer, which enables it to better capture the temporal dependency in a longer prediction horizon. The Informer has an encoder–decoder structure, and the architecture of the Informer layer is shown in Figure 5. The input of the encoder contains two parts: one is the same sequences as the output sequences from the Graph Attention layer  $\tilde{\mathbf{H}}$ ; the other one is the sports events impact factor defined in Section 3.2. For node  $i$ , the time series impact factor  $\sigma_i$  would be regarded as a temporal indicator and concatenated with the Informer input sequence  $\mathbf{h}_i \in R^M$ , which can be expressed as follows:

$$\mathbf{h}_i^* = [\mathbf{h}_i \mid \sigma_i] \quad (14)$$

By concatenating these two together, each time period will be given a label indicating whether the traffic flow was affected by sports events at that time and the extent to which it was affected. The input of the decoder is similar to the input of the encoder; it also consists of two parts: the main difference is that the input sequences are no longer taken exclusively from the dataset but contain a portion of the data to be predicted, which are initialized by 0. The advantage of this configuration is that the real data in the dataset can guide the predict data, thus improving the accuracy of prediction. The main function of the Informer layer is to assign values to these predict data to obtain the output sequences.



**Figure 5.** Informer layer architecture. The blue input sequences are the input of encoder, which are the same as the output of Graph Attention layer. The red input sequences are the input of decoder, and they contain two parts: label data and predict data.

The core idea of the Informer layer is a ProbSparse attention mechanism. The general attention mechanism can be expressed as follows:

$$\mathcal{A}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q} \cdot \mathbf{K}^T}{\sqrt{\mathbf{d}}}\right) \cdot \mathbf{V} \quad (15)$$

where  $\mathbf{Q}$  denotes query,  $\mathbf{K}$  denotes key,  $\mathbf{V}$  denotes value,  $\mathbf{d}$  is the dimension of  $\mathbf{K}$ .

In order to reduce time complexity of this algorithm, ProbSparse attention mechanism optimizes the choice of  $\mathbf{Q}$  and  $\mathbf{K}$ . Due to the potential sparsity of the probability distribution of self-attention, only a small number of dot products contribute to the primary attention. Hence,  $\mathbf{Q}$  is sampled by keeping the active ones and replacing the others with uniform distribution; then, Equation (15) will be updated as follows:

$$\mathcal{A}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\hat{\mathbf{Q}} \cdot \mathbf{K}^T}{\sqrt{\mathbf{d}}}\right) \cdot \mathbf{V} \quad (16)$$

where  $\hat{\mathbf{Q}}$  has the same size as  $\mathbf{Q}$ , but we replace the inactive values with mean values.

ProbSparse attention mechanism is only used in self-attention in the encoder. For self-attention in the decoder, since the input sequences contain the data need to be predicted, and due to the prediction causal, future values cannot be used to calculate current values; this will result in the original number of  $\mathbf{Q}$  and  $\mathbf{K}$  being already reduced. Hence, there is no need to do further sampling.

The encoder has two layers and contains two self-attention blocks. It forms a basic unit of attention–convolution–maxpooling–attention. If the input sequence is excessively lengthy, it can be extended by adding multiple groups of convolution–maxpooling–attention units after the last attention block. This can effectively reduce the feature dimensions and avoid overfitting. The decoder has one layer and contains only one self-attention block. The output of this block will conduct a crossattention with the output of the encoder.

### 3.3.3. Loss Function

In order to minimize the error during model training, MSE loss is used as the loss function, which can be written as follows:

$$L(pred, true) = \frac{1}{n} * \sum (pred - true)^2 \quad (17)$$

where  $n$  is the number of samples.

#### 4. Experiment and Analysis

Our GAT-Informer model was evaluated on two datasets: one is the PeMS04 dataset, which is widely used for traffic flow prediction, and the other one is a newly collected London dataset. We chose five models commonly used for time series prediction for performance comparison. In evaluation using the PeMS04 dataset, the impact of sports events on traffic flow will be disregarded; the target is to compare the performance of these models on solving time series prediction problem. In the London dataset, the sports events impact factor will be included.

##### 4.1. Dataset and Preprocessing

###### 4.1.1. PeMS04 Dataset

The PeMS04 dataset records data from 307 detectors on 29 roads in San Francisco Bay Area collected every 5 min between 1 January 2018 and 28 February 2018. The shape of the data is (16,992, 307, 3), where the three-dimensional features are flow, occupy, and speed. Only the flow feature is used in our experiments, thus constructing a (16,992, 307) traffic flow sequence.

###### 4.1.2. London Dataset

The London dataset records data from 96 detectors on three motorways around London collected every 15 min between 1 August 2022 and 31 October 2022. These three motorways contain M1, M4, and M25. The M1 motorway is a major north–south motorway in England, which connects London and Leeds. The M4 motorway is one of the most important motorways in the United Kingdom, thus running from west London to southwest Wales. The M25 motorway, also known as the London Orbital Motorway, is a ring-shaped high-speed road that encircles London. In order to better study the impact of sports events on traffic flow, the section of M4 motorway and M25 motorway around London Heathrow Airport were selected. The location of detectors used in this dataset is shown in Figures 6 and 7. The dataset contains a (8740, 96) traffic flow sequence.

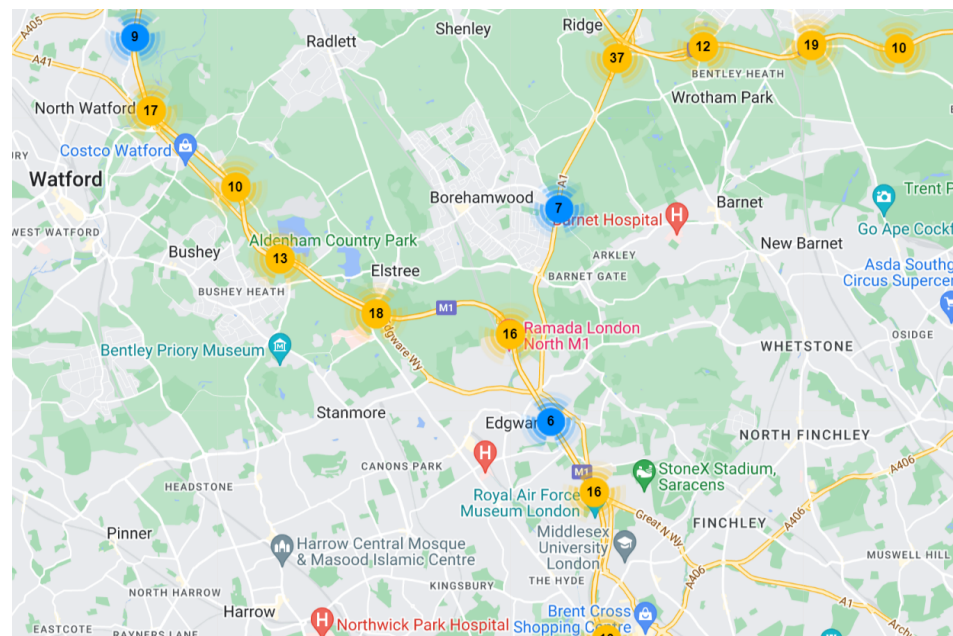
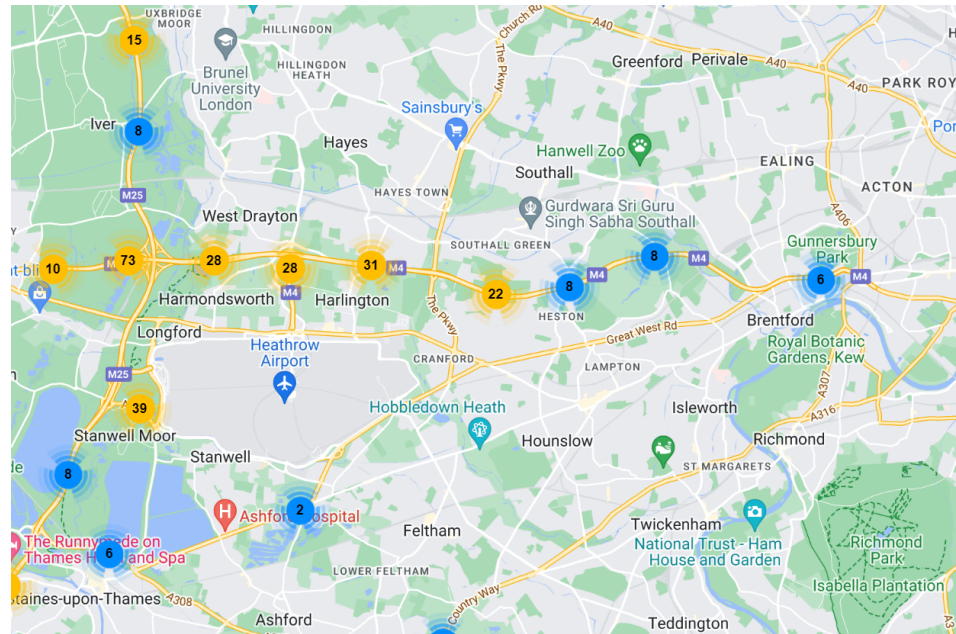
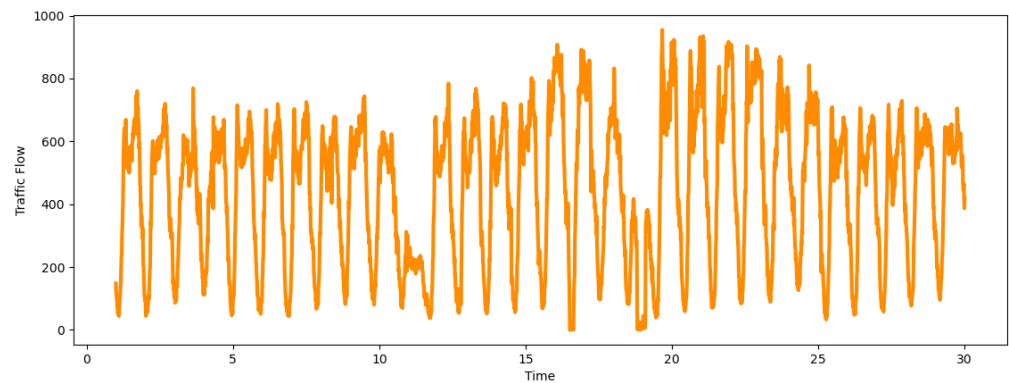


Figure 6. Part of M1 motorway.



**Figure 7.** Part of M4 motorway and M25 motorway.

The Laver Cup was held from 23 September 2022 to 25 September 2022 in the O2 arena in London, and this match gained a high level of global attention. Our dataset covers this time period to study the impact that sports events have on traffic flow. Figure 8 shows the traffic flow data recorded by our 2178A detector on the M4 motorway between 1 September 2022 and 30 September 2022. Obviously, there was an uptick in traffic flow around match days compared to other times of this month. Based on the definition in Section 3.2, the sports events impact factor will be calculated for this special period and used as the input of our GAT-Informer model.



**Figure 8.** Traffic flow data recorded on M4/2178A from 1 September 2022 to 30 September 2022.

#### 4.1.3. Data Preprocessing

Both the PeMS04 dataset and London dataset were min-max normalized in order to accelerate the model convergence, avoid numerical instability, and enhance the model performance. The normalization can be expressed as follows:

$$f(X) = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (18)$$

## 4.2. Experimental Verification

### 4.2.1. Evaluation Metrics

Three commonly used metrics were chosen to evaluate the model performance during model training, including the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE). Each of the metrics is defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (20)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (21)$$

where  $y_i$  and  $\hat{y}_i$  represent the true traffic flow and the predicted one at time  $i$ , respectively;  $n$  is the prediction length;  $\mathbf{Y}$  is the set of  $y_i$ ; and  $\bar{\mathbf{Y}}$  is the average of  $\mathbf{Y}$ .

### 4.2.2. Experiment Settings

In the experiment, the collected data were divided into 60% for training, 20% for evaluation, and the remaining 20% for testing. For spatial information extraction, two layers of GAT with three heads of multihead attention were used. For temporal information extraction, three layers of encoder and one layer of decoder with eight heads of multihead attention were employed. The loss function of the MSE was utilized and the loss backwards from decoder outputs to the entire network. The batch size was set to be 32, and the model was optimized using the Adam optimizer with a dropout rate being set to 0.05. The learning rate was adapted from  $e^{-4}$ , and the number of total training epochs was 50 with early stopping. The experiments were conducted on a computer with an Intel Core i7-12800H 2.40 GHz CPU (Santa Clara, CA, USA) and a RTX 3060 GPU (NVIDIA, Santa Clara, CA, USA). During the experiment, our proposed model will be evaluated across different prediction horizons, together with baseline models, in order to have a comprehensive evaluation in short-term prediction and long-term prediction.

### 4.2.3. Experimental Results

The traffic flow prediction can be viewed as a time series forecasting problem containing multiple features. For the PeMS04 dataset, we selected five baseline models, and these models have been proven to perform well in dealing with time series problems.

- **GRU:** The GRU is a variant model of RNN. It has been proven to be effective in the short-term time series forecasting problems, and the model also stands out for its ability to alleviate the gradient explosion and vanishing problem. However, the GRU is less effective in the scenario of long-term prediction.
- **T-GCN:** The Integrated GCN and GRU capture both spatial and temporal correlation, and they can be used for both short-term and long-term traffic prediction [42]. However, similar to recurrent-based temporal information, their extraction capacity is weak in long-term prediction.
- **Informer:** The Informer model addresses the computational cost and memory usage problem of the Transformer. The model is widely used in short-term or long-term traffic prediction, but it cannot extract information from adjacent nodes.
- **ASTGCN:** The ASTGCN model is designed for traffic prediction that excels in capturing spatial and temporal dependencies [35]. The novel spatial-temporal attention mechanism enables the model to achieve high performance in traffic prediction. However, this model is constrained by the static map and cannot incorporate external factors.

- **ASTGNN:** The ASTGNN model is an attention-based spatial–temporal model, which stands out for the novel design of a special self-attention mechanism [43]. In both of the short-term and long-term prediction scenarios, the model can achieve satisfying performance.

We deployed the above baseline models to predict traffic flow data for the next 15 min, 30 min, and 45 min horizons based on the 15 min time interval of the PeMS04 dataset. The experimental results are shown in Table 1.

**Table 1.** Average performance comparison of different baseline models for traffic flow prediction on PeMS04 dataset.

Horizon	Criteria	GRU	T-GCN	Informer	ASTGNN	ASTGCN	GAT + Informer
15 min	MAE	28.04	23.69	26.27	25.69	22.40	<b>21.78</b>
	RMSE	45.41	39.42	41.51	40.73	<b>35.11</b>	35.24
	MAPE	20.62	17.42	19.32	18.89	16.47	<b>16.01</b>
30 min	MAE	30.07	28.90	28.52	27.02	23.05	<b>22.62</b>
	RMSE	47.96	44.76	46.40	41.49	35.87	<b>35.36</b>
	MAPE	21.82	20.34	20.14	19.23	17.00	<b>16.98</b>
45 min	MAE	30.98	30.21	29.44	27.81	<b>23.91</b>	24.21
	RMSE	48.94	45.46	43.46	42.51	<b>36.42</b>	36.81
	MAPE	22.11	21.25	20.97	19.87	16.95	<b>16.63</b>

In the 15 min horizon, all the metrics of T-GCN were slightly better than Informer, thus indicating the necessity of graph neural network in solving the traffic flow prediction problem. At the same time, although the T-GCN performed better than the Informer in short-term predictions, its advantage could not be sustained in long-term predictions. This highlights the effectiveness of multihead attention in long-term prediction scenarios. In the 45 min horizon, the RMSE of the GAT-Informer was 13.40% smaller, the MAE was 12.94% smaller, and the MAPE was 14.54% smaller than the ASTGNN, which outperformed most of the baseline models. Moreover, as the prediction horizon lengthened, the performance of models experienced varying degrees of decay. Compared to other baseline models, the proposed GAT-Informer exhibited the least decay, thus maintaining high prediction performance. The MAE of the GRU increased by 10.48% from a 15 min prediction to 45 min prediction, while the MAE of the GAT-Informer only increased by about 6.56%. However, the difference between the performance of the GAT-Informer and the ASTGCN is not obvious in the PeMS04 dataset experiment.

Based on the above experimental results, the performance of The Informer, ASTGCN, and the proposed GAT-Informer were evaluated on the London dataset. They were used to predict the future traffic flow data of 15 min, 30 min, 45 min, 60 min, 90 min, and 120 min based on the 15 min time interval. For the impact factor of the sports event,  $\Delta t$  defined in Equations (8) and (9) was set to 0.5 h, and  $t_{start}$  and  $t_{end}$  followed the schedule of the Laver Cup held in 2022, London. Sports events will affect nodes within a 10-kilometer radius. The experimental results are shown in Table 2.

According to the evaluation metrics in Table 2, GAT-Informer maintained the best performance. In long-term (45 min, 60 min, 90 min, and 120 min) traffic flow prediction, GAT-Informer’s performance was only marginally attenuated; the model still maintained a high fitting degree and had the smallest error between the predict values and true values among all the baseline models. When the input of the model removed the sports events impact factor, the performance of the model decreased. In the 15 min horizon, regardless of the impact of sports events, the RMSE increased by 1.95%, the MAE increased by 6.35%, and the MAPE increased by 6.35%. In the 120 min horizon, regardless of the impact of sports events, the RMSE increased by 1.31%, the MAE increased by 1.02%, and the MAPE increased by 1.03%. Based on the model evaluation, the fusion of sports events is proven to be effective in both short-term prediction or long-term prediction. Although the model performance deteriorated with longer prediction horizons, our proposed model demonstrated the least performance degradation. Meanwhile, the GAT-Informer model considering sports events outperformed the ASTGCN in both short-term and long-term predictions, despite both

models achieving similar performance on the PeMS04 dataset. Compared to other baseline models, the proposed GAT-Informer model demonstrates superior capability in extracting information from sports events, thereby enhancing prediction performance and maintaining high prediction accuracy when the prediction horizon lengthens.

**Table 2.** Average performance comparison of different baseline models for traffic flow prediction on London dataset.

Horizon	Criteria	Informer	ASTGCN	GAT + Informer	GAT + Informer *
15 min	RMSE	82.61	80.29	<b>75.86</b>	77.34
	MAE	48.72	45.42	<b>42.19</b>	44.87
	MAPE	12.18	11.36	<b>10.55</b>	11.22
30 min	RMSE	86.21	84.65	<b>81.45</b>	83.56
	MAE	53.65	49.98	<b>46.22</b>	48.08
	MAPE	13.41	12.50	<b>11.56</b>	12.02
45 min	RMSE	92.07	87.65	<b>85.20</b>	89.77
	MAE	55.80	53.37	<b>51.52</b>	53.40
	MAPE	13.95	13.34	<b>12.88</b>	13.35
60 min	RMSE	94.83	93.82	<b>89.82</b>	91.26
	MAE	60.74	58.27	<b>56.01</b>	57.71
	MAPE	15.19	14.57	<b>14.00</b>	14.43
90 min	RMSE	112.96	104.16	<b>100.63</b>	103.09
	MAE	72.44	65.46	<b>62.63</b>	63.17
	MAPE	18.11	16.37	<b>15.66</b>	15.79
120 min	RMSE	117.88	107.06	<b>105.30</b>	106.68
	MAE	76.11	67.37	<b>65.85</b>	66.52
	MAPE	19.03	16.84	<b>16.46</b>	16.63

\* means the dataset removes the sports events impact factor.

#### 4.2.4. Evaluation of Arrival and Departure Redundancy

A redundancy  $\Delta t$  was set in Equations (8) and (9) to represent the early arrival and delayed departure of the audience. Although the start and end times of a game are predetermined, varying redundancies of  $\Delta t$  can influence the impact factor curve, thereby affecting traffic flow predictions. In this case study, we evaluated the effect of different  $\Delta t$  values on the prediction performance of GAT-Informer. The model was evaluated in prediction horizons of 30 min, 60 min, 90 min, and 120 min, with  $\Delta t$  varying from 10 min to 40 min, and the experiment results are shown in Table 3.

**Table 3.** Average performance comparison of different redundancy  $\Delta t$  values.

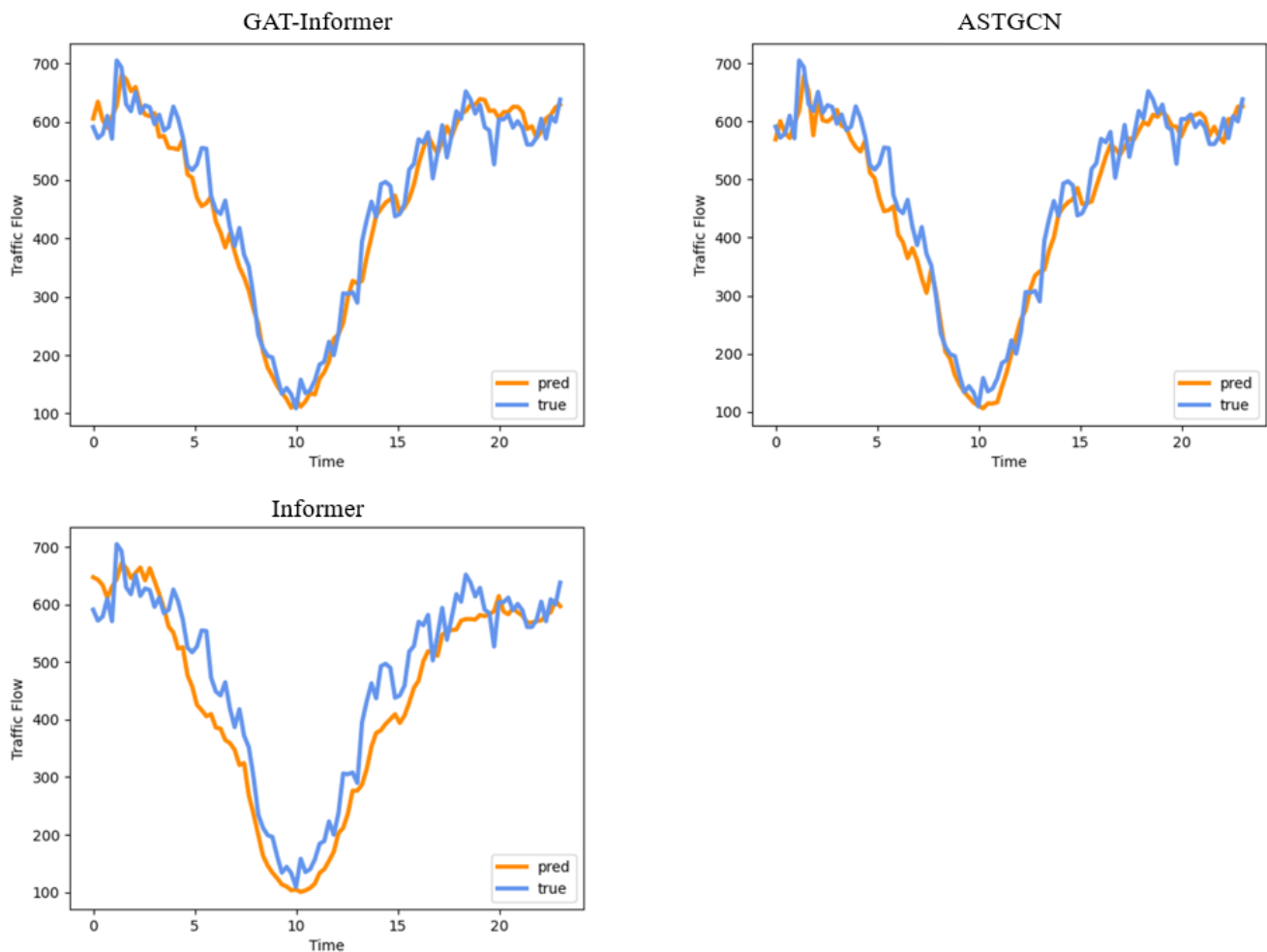
Horizon	Criteria	$\Delta t = 10$ min	$\Delta t = 20$ min	$\Delta t = 30$ min	$\Delta t = 40$ min
30min	MAE	49.25	48.46	<b>48.08</b>	49.09
	RMSE	84.98	84.40	<b>83.56</b>	84.65
	MAPE	12.31	12.12	<b>12.02</b>	12.27
60min	MAE	56.20	56.51	57.71	<b>56.35</b>
	RMSE	92.11	91.39	<b>91.26</b>	92.17
	MAPE	14.05	14.13	14.43	<b>14.09</b>
90min	MAE	64.04	<b>63.12</b>	63.17	64.20
	RMSE	100.88	<b>100.52</b>	103.09	101.02
	MAPE	16.01	<b>15.78</b>	15.79	16.05
120min	MAE	68.87	<b>65.87</b>	66.52	66.62
	RMSE	109.05	<b>106.81</b>	105.30	106.66
	MAPE	18.58	<b>16.47</b>	16.63	16.66

Table 3 shows that varying the redundancy  $\Delta t$  did not lead to significant performance changes. However, in general, a redundancy of 20 min slightly improved the prediction performance compared to other settings. This suggests that the impact factor  $\sigma$  influences the model based on the audience arrival and departure patterns. Among different settings, a redundancy of 20 min best represents these patterns in the London dataset.

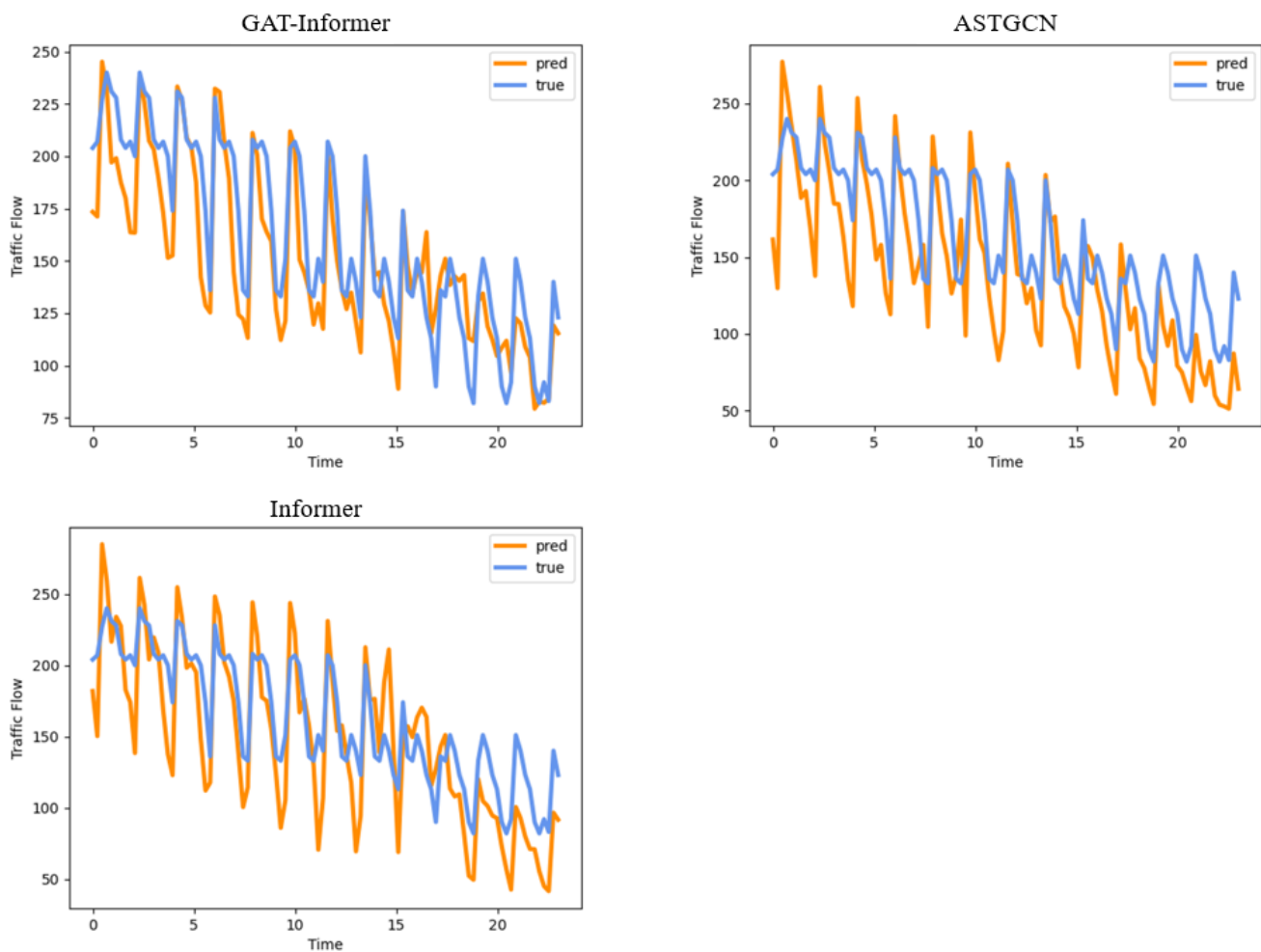
#### 4.2.5. Results Visualization

In order to visualize the prediction performance of each model in Table 2, we randomly selected the prediction results of a continuous 24 h time interval by two different detectors based on a 15 min horizon and a 120 min horizon. The results are shown in Figure 9 and Figure 10, respectively. By comparing the difference between the true values and the predicted values, we can intuitively compare the performance of these three models in terms of short-term prediction and long-term prediction.

From these curves, we can find that our GAT-Informer model had a very good model fit, with a quite small error between the true values and predicted values, thus keeping basically the same curve when dealing with short-term traffic flow prediction. When dealing with long-term traffic flow prediction, although GAT-Informer did not perform as well as the short-term, it was still better than the other baseline models. Its prediction accuracy of peak traffic flow was good, but it had some errors in valley traffic flow. For traffic management, it is often more important to control the peak traffic flow, which may have a greater impact on vehicle travel time.



**Figure 9.** The 15 min horizon: 24 h prediction results on a randomly selected detector.



**Figure 10.** The 120 min horizon: 24 h prediction results on a randomly selected detector.

## 5. Conclusions

In this paper, the GAT-Informer model was proposed for long-term traffic flow prediction considering the impact of sports events. The model contains the Graph Attention layer and Informer layer. Graph Attention layer was used to fuse the information of neighbor nodes so as to fully captured the spatial features. The Informer layer was used to capture the dynamic time correlation between different times based on its ProbSparse attention mechanism. The GAT-Informer and other baseline models were tested on the PeMS04 dataset and London dataset. The experiment results showed that the GAT-Informer had better prediction performance in different horizons and metrics.

The detectors may fail temporarily due to failure, blackout, or other reasons; hence, the traffic flow data recorded by them may have missing data. Future plans include introducing fuzzy representation into the deep learning model to reduce the impact of data uncertainty and thus improve the matching of the model to various datasets.

**Author Contributions:** Conceptualization, Y.S. and R.L.; Methodology, Y.S., R.L. and T.Z.; Software, Y.S.; Validation, R.L.; Formal analysis, Y.S. and R.L.; Investigation, T.Z.; Data curation, T.Z.; Writing—original draft, Y.S. and R.L.; Writing—review & editing, T.Z.; Visualization, T.Z.; Supervision, C.Z. and R.S.; Project administration, R.S.; Funding acquisition, C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is supported under the RIE2020 Industry Alignment Fund-Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from the industry partner(s), and A\*STAR under its Industry Alignment Fund (LOA Award I1901E0046).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset used in this paper can be found at <https://webtris.highwaysengland.co.uk/> (accessed on 25 September 2023).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Qureshi, K.N.; Abdullah, A.H. A survey on intelligent transportation systems. *Middle-East J. Sci. Res.* **2013**, *15*, 629–642.
2. Nagy, A.M.; Simon, V. Survey on traffic prediction in smart cities. *Pervasive Mob. Comput.* **2018**, *50*, 148–163. [CrossRef]
3. Yang, H.F.; Dillon, T.S.; Chen, Y.P.P. Optimized structure of the traffic flow forecasting model with a deep learning approach. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2371–2381. [CrossRef] [PubMed]
4. Xie, P.; Li, T.; Liu, J.; Du, S.; Yang, X.; Zhang, J. Urban flow prediction from spatiotemporal data using machine learning: A survey. *Inf. Fusion* **2020**, *59*, 1–12. [CrossRef]
5. Zhang, J.; Wang, F.Y.; Wang, K.; Lin, W.H.; Xu, X.; Chen, C. Data-driven intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1624–1639. [CrossRef]
6. Ben-Akiva, M.E.; Gao, S.; Wei, Z.; Wen, Y. A dynamic traffic assignment model for highly congested urban networks. *Transp. Res. Part C Emerg. Technol.* **2012**, *24*, 62–82. [CrossRef]
7. Wu, Y.; Tan, H.; Qin, L.; Ran, B.; Jiang, Z. A hybrid deep learning based traffic flow prediction method and its understanding. *Transp. Res. Part C Emerg. Technol.* **2018**, *90*, 166–180. [CrossRef]
8. Jain, V.; Sharma, A.; Subramanian, L. Road traffic congestion in the developing world. In Proceedings of the 2nd ACM Symposium on Computing for Development, Atlanta, GA, USA, 11–12 March 2012; pp. 1–10.
9. Luo, R.; Zhang, Y.; Zhou, Y.; Chen, H.; Yang, L.; Yang, J.; Su, R. Deep learning approach for long-term prediction of electric vehicle (ev) charging station availability. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 3334–3339.
10. Bovy, P. Mega sports event transportation and main mobility management issues. In Proceedings of the Transport and Exceptional Public Events, Report of the One Hundred and Twenty Second Round Table on Transport Economics, Paris, France, 7–8 March 2002; European Conference of Ministers of Transport: Brussels, Belgium, 2003.
11. Kim, W.; Jun, H.M.; Walker, M.; Drane, D. Evaluating the perceived social impacts of hosting large-scale sport tourism events: Scale development and validation. *Tour. Manag.* **2015**, *48*, 21–32. [CrossRef]
12. Zhu, L.; Yu, F.R.; Wang, Y.; Ning, B.; Tang, T. Big data analytics in intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 383–398. [CrossRef]
13. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 4927–4943. [CrossRef]
14. Pan, B.; Demiryurek, U.; Shahabi, C. Utilizing real-world transportation data for accurate traffic prediction. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 595–604.
15. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
16. Okutani, I.; Stephanedes, Y.J. Dynamic prediction of traffic volume through Kalman filtering theory. *Transp. Res. Part B Methodol.* **1984**, *18*, 1–11. [CrossRef]
17. Feng, X.; Ling, X.; Zheng, H.; Chen, Z.; Xu, Y. Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2001–2013. [CrossRef]
18. Cai, P.; Wang, Y.; Lu, G.; Chen, P.; Ding, C.; Sun, J. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transp. Res. Part C Emerg. Technol.* **2016**, *62*, 21–34. [CrossRef]
19. Qi, Y.; Ishak, S. A Hidden Markov Model for short term prediction of traffic conditions on freeways. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 95–111. [CrossRef]
20. Zhang, W.; Yu, Y.; Qi, Y.; Shu, F.; Wang, Y. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transp. A Transp. Sci.* **2019**, *15*, 1688–1711. [CrossRef]
21. Ramakrishnan, N.; Soni, T. Network traffic prediction using recurrent neural networks. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 187–193.
22. Yu, B.; Lee, Y.; Sohn, K. Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). *Transp. Res. Part C Emerg. Technol.* **2020**, *114*, 189–204. [CrossRef]
23. Fu, R.; Zhang, Z.; Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association Of Automation (YAC), Wuhan, China, 11–13 November 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 324–328.

24. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [[CrossRef](#)]
25. Gomes, B.; Coelho, J.; Aidos, H. A survey on traffic flow prediction and classification. *Intell. Syst. Appl.* **2023**, *20*, 200268. [[CrossRef](#)]
26. Yu, B.; Yin, H.; Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv* **2017**, arXiv:1709.04875.
27. Xu, M.; Dai, W.; Liu, C.; Gao, X.; Lin, W.; Qi, G.J.; Xiong, H. Spatial-temporal transformer networks for traffic flow forecasting. *arXiv* **2020**, arXiv:2001.02908.
28. Qu, L.; Li, W.; Li, W.; Ma, D.; Wang, Y. Daily long-term traffic flow forecasting based on a deep neural network. *Expert Syst. Appl.* **2019**, *121*, 304–312. [[CrossRef](#)]
29. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [[CrossRef](#)]
30. Williams, B.M.; Hoel, L.A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *J. Transp. Eng.* **2003**, *129*, 664–672. [[CrossRef](#)]
31. Lee, S.; Fambro, D.B. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transp. Res. Rec.* **1999**, *1678*, 179–188. [[CrossRef](#)]
32. Shu, Y.; Jin, Z.; Zhang, L.; Wang, L.; Yang, O.W. Traffic prediction using FARIMA models. In Proceedings of the 1999 IEEE International Conference on Communications (Cat. No. 99CH36311), Vancouver, BC, Canada, 6–10 June 1999; IEEE: Piscataway, NJ, USA, 1999; Volume 2, pp. 891–895.
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
34. Ali, A.; Zhu, Y.; Zakarya, M. Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction. *Neural Netw.* **2022**, *145*, 233–247. [[CrossRef](#)] [[PubMed](#)]
35. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 922–929.
36. Zheng, C.; Fan, X.; Wang, C.; Qi, J. Gman: A graph multi-attention network for traffic prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 1234–1241.
37. Han, L.; Du, B.; Sun, L.; Fu, Y.; Lv, Y.; Xiong, H. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Virtual, 14–18 August 2021; pp. 547–555.
38. Li, M.; Zhu, Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 4189–4196.
39. Jiang, W.; Luo, J. Graph neural network for traffic forecasting: A survey. *Expert Syst. Appl.* **2022**, *207*, 117921. [[CrossRef](#)]
40. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
41. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 11106–11115.
42. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3848–3858. [[CrossRef](#)]
43. Guo, S.; Lin, Y.; Wan, H.; Li, X.; Cong, G. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. *IEEE Trans. Knowl. Data Eng.* **2021**, *34*, 5415–5428. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.