engineering proceedings

MDPI

Proceeding Paper

Urban Traffic Flow Prediction Using LSTM and GRU †

Hung-Chin Jang * and Che-An Chen

Department of Computer Science, National Chengchi University, Taipei City 116302, Taiwan; an5566x@gmail.com
* Correspondence: jang@cs.nccu.edu.tw
† Presented at the IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability, Tainan, Taiwan, 2–4 June 2023.

Abstract: For smart cities, the issue of how to solve traffic chaos has always attracted public attention. Many studies have proposed various solutions for traffic flow prediction, such as ARIMA, ANN, and SVM. With the breakthrough of deep learning technology, the evolutionary models of RNN, such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Units) models, have been proven to have excellent performance in traffic flow prediction. By using LSTM and GRU models, we explore more features and multi-layer models to increase the accuracy of traffic flow prediction. We compare the prediction accuracy of LSTM and GRU models in urban traffic flow prediction. The data collected in this study are divided into three categories, namely "regular traffic flow data", "predictable episodic event data", and "meteorological data". The regular traffic flow data source is the "Vehicle Detector (VD) data of Taipei Open Data Platform". Predictable episodic event data are predictable as non-routine events such as concerts and parades. We use a crawler program to collect this information through ticketing systems, tourism websites, news media, social media, and government websites and the meteorological data from the Central Meteorological Bureau. Through these three types of data, the accuracy in predicting traffic flow is enhanced to predict the degree of traffic congestion that may be affected.

Keywords: traffic flow prediction; deep learning; LSTM (Long Short-Term Memory); GRU (Gated Recurrent Units)

check for updates

Citation: Jang, H.-C.; Chen, C.-A. Urban Traffic Flow Prediction Using LSTM and GRU. *Eng. Proc.* **2023**, *55*, 86. https://doi.org/10.3390/engproc2023055086

Academic Editors: Teen-Hang Meen, Kuei-Shu Hsu and Cheng-Fu Yang

Published: 2 January 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. Motivation

There are many factors causing traffic congestion. When the number of vehicles exceeds the area's load capacity and the traffic system's design is poor, unsolvable traffic jams occur. Most of the solutions include adjusting the time phase of traffic lights, widening the road, restricting the entry and exit of vehicles, charging, and scheduling driving lanes. However, these solutions often encounter many obstacles or require high costs in implementation. As an extension of the work of Jang [1], in which a "smart traffic control platform" is proposed, we proposed the time-phased control of traffic lights that is adaptively adjusted according to the current traffic conditions for effectively relieving congestion and reducing driving time. Reference [1] proposed a symptomatic solution to resolve the traffic jam, while this study proposes a fundamental solution through the prediction of the possible traffic flow in the future. Then, combined with a smart traffic control platform, it becomes feasible to relieve traffic congestion. An accurate and instant traffic flow prediction is essential to the Intelligent Transportation System (ITS). ITS uses various computer science and communication technologies to manage the entire transportation system effectively. With the Internet of Things, existing data can be effectively collected and analyzed through many IoT devices to provide feedback to the transportation infrastructure.

Since the 20th century, many researchers have proposed various solutions for traffic flow prediction, such as the Autoregressive Integrated Moving Average model (ARIMA), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs). With the successful application of deep learning in ImageNet classification [2], many researchers

have attempted to use deep neural networks for traffic-related predictions instead of shallower neural networks such as SVM. Researchers have confirmed that the LSTM model has good performance for the time-series-based traffic flow data by using the data from Caltrans Performance Measurement System (PeMS) in California, USA. Few studies have used the data on urban street traffic flow. Thus, based on the vehicle detector (VD) data from the Songshan District, 2017, provided by the Taipei Municipal Transportation Bureau, we propose a prediction model suitable for urban road sections with relatively low error through the deep learning method, combining it with the weather data in Songshan District and the relevant data of large-scale events. The evolutionary model of RNN in deep learning systems such as LSTM and GRU is used to obtain more features and multi-layer models to increase the accuracy. Finally, we compare the performance of LSTM and GRU models in urban traffic flow prediction.

2. Methodology

We used the data on regular traffic flow, predictable episodic events, and weather and the LSTM model for module A and model B, and the GRU model for module C to predict. The effectiveness of different models for traffic prediction in urban areas is then compared with other results.

- Module A: the LSTM module that does not include the data of "predictable episodic event" (dataset: VD data);
- Module B: the LSTM module that includes the data of "predictable episodic events" (datasets: VD data, weather, and events held by the Taipei Arena);
- Module C: the GRU module that includes the data of "predictable episodic events" (data set: VD data, weather, and events held by the Taipei Arena).

Regarding the role of module A, we demonstrated in a previous study [3] that the data of "predictable episodic event" are beneficial for predicting traffic flow. Therefore, we focused on the role of modules B and C.

2.1. Data Collection

The datasets for predicting urban traffic flow are divided into three categories: regular traffic flow data (VD data), predictable episodic event data, and weather data.

2.1.1. Regular Traffic Flow Data

The regular traffic flow data source is "Vehicle Detector (VD) data of Taipei City Government Data Open Platform". Five vehicle detectors are deployed on the main roads around Taipei Arena: VMCN800, VKRM820, VKWNV20, VKWN800, and VL7PX00. The data are provided by the Traffic Control Center of Taipei City for the entire year of 2017.

2.1.2. Predictable Episodic Events Data

Predictable and non-routine events include concerts and parades. Such information can be found through ticketing systems, tourism websites, news media, social media, and government websites. In Ref. [4], we implemented a crawler program with Python 3.6.5 to collect the event data of the Taipei Arena to predict the degree of traffic congestion.

2.1.3. Weather Data

On the subscription system of the Central Meteorological Administration, the hourly observation data of all observation stations in Taiwan since 2010 have been published. The data include temperature, humidity, rainfall, and other weather data. We wrote a crawler to download a CSV file containing all the required weather data.

2.2. Data Preprocessing

2.2.1. Filter Data Field

The vehicle detector (VD) data, weather data, and predictable episodic event data contain several fields that are not used, so these fields must be removed first. The remaining data fields after filtering are as follows.

- Vehicle detector (VD) data: fields of traffic flow, vehicle speed, and vehicle occupancy (%);
- Weather data: fields of temperature, relative humidity, rainfall, week, and hour;
- Episodic event data: field of event.

2.2.2. Feature Scaling

After the data field is normalized, the convergence speed of learning can be sped up. Depending on the type of data, there are two methods of normalization.

- For general data fields such as traffic flow and rainfall, we use the following Z-score formula to perform feature scaling and shift the data range to a smaller range.

$$X_{normalization} = \frac{(x - \mu)}{\sigma} \quad (1)$$

  where x is the value of the data field, μ is the mean of the parent data, and σ is the standard deviation.
- For the time periodic data field (hour, week), x, we make the scaled value have circularity. The range of numerical data can be scaled through normalization to $[-1, 1]$ (Equation (2)).

$$X_{normalization} = \sin(x \times 2 \times \pi/24) \quad (2)$$

2.3. Deep Learning

2.3.1. Build Up Training Data

Training data consist of these three data types, and we input different lengths of data, as shown in Figure 1. The 24 h input data are an example. The VD data from the 0th hour to the 24th hour are used as the input data (X_train [5]), and the output data are the next time point (Y_train [5]) of the last input data. That is, the traffic flow data after 30 min are the traffic flow data of the 24.5th hour; the VD data from the 0.5th hour to the 24.5th hour are used as the input data (X_train [4]), and the output data are the 25th-hour traffic flow data (Y_train [4]), and so on. The last input data are the VD data from the first hour to the I+24th hour, and the output data are the traffic flow data of the I+24.5th hour.

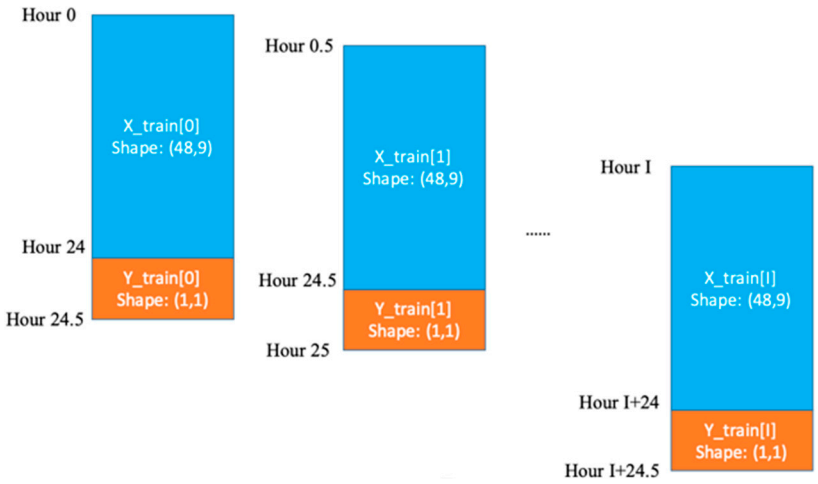


Figure 1. Training data arrangement.

2.3.2. Build Up Models

We propose module B (LSTM model) and module C (GRU model) in this study. Both modules adopt three architectures: the basic Vanilla architecture, the stacked architecture [4,5], and the encoder–decoder architecture [6].

- Vanilla architecture: an LSTM or GRU model with only a single layer;
- Stacked architecture: an LSTM or GRU model with multiple layers;
- Encoder–decoder architecture: convert the input sequence into a fixed-length vector through the encoder, then convert the fixed-length vector into the output sequence through the decoder.

(a) LSTM model

In practice, LSTM is divided into four types: one-to-one, one-to-many, many-to-one, and many-to-many, according to the number of input and output data. The input for this simulation is VD data for the 24 h prior to the target time, plus episodic event data and weather data. The output is the traffic flow for 30 min from the target time. It means that 48 pieces of data are input and one piece of data is output, which is a many-to-one LSTM.

- Vanilla LSTM: single-layer LSTM with a Dense layer to output traffic flow;
- Stacked LSTM: a four-layer LSTM with a Dense layer to output the traffic flow. However, simply using this module may encounter overfitting. We adjust the parameters of the dropout layer or increase the number of dropout layers to reduce the overfitting, depending on the situation;
- Encoder–decoder LSTM is used to customize how many layers of LSTM the encoder and decoder have.

(b) GRU model

In implementing the GRU model, we adopt the same design concept as the LSTM model, divided into Vanilla, stacked, and Encoder–decoder architectures.

- Vanilla GRU: a single-layer GRU with a Dense layer to output traffic flow;
- Stacked GRU: a three-layer GRU with a Dense layer to output traffic flow. Moreover, increase the number of dropout layers to prevent overfitting;
- Encoder–decoder GRU: a layer of GRU is used as encoder and decoder, respectively.

2.3.3. Model Training

In two different modules, we used 80% of the data for the entire year as the training data and 20% of the data as the validation data. Since the model trained from the training data expressed the behavior of the training data, it did not express other behaviors shown by the test data. Therefore, we used the validation data to find the best model, and at the same time, we used Adam Optimizer to speed up the training process and set the learning rate to 0.001.

2.4. Performance Evaluation

In the simulation, the remaining 20% of the data in the entire year was used as test data. We used MAPE, MSE, and MAE to evaluate the difference between the actual and predicted values and compare Modules B and C's performance. MAPE and MAE considered the magnitude of the error between the predicted value and the actual value, regardless of a positive or negative error. MSE took the square of the error between the predicted and actual values, so the MSE value was prone to being affected when there were extreme values. Therefore, to avoid extreme error conditions, we observed them through the MAPE and MAE metrics. Among the three metrics of MAPE, MAE, and MSE, the smaller the value, the smaller the error between the predicted value and the actual value, and the model has a better predictive ability. The MAE and MSE were relative evaluation metrics. The smaller the value, the better the performance. Regarding the MAPE evaluation metric, we referred to the classification of the effectiveness of the evaluation metric proposed by Lewis [7].

3. Experimental Simulation

The experiment consisted of three simulations: simulation 1 was the Vanilla architecture, simulation 2 was the stacked architecture, and simulation 3 was the Encoder–decoder architecture. Each simulation used LSTM and GRU models and MAPE, MSE, and MAE metrics to evaluate the prediction accuracy. We collected data from the five VD stations deployed around Taipei Arena according to different driving directions, including VKRM820 in the east lane, VKRM820 in the west lane, VKWN800 in the north lane, VKWN800 in the south lane, and VKWNV20 in the east lane, VKWNV20 in the west lane, VMCN800 in the south lane, and VL7PX00 in the north lane, as shown in Figure 2. The parameter settings of the three simulations are listed in Table 1.



Figure 2. Five VD stations deployed around Taipei Arena.

Table 1. Simulation parameter setting.

Parameter	Value	Parameter	Value
Learning rate	0.001	Optimizer	Adam
Dropout rate	0.5	Batch size	128
Recurrent dropout rate	0.3	Timestep	96
Loss function	MAPE		

3.1. Simulation 1: Vanilla Architecture

In the simulation, we used the basic Vanilla architecture, i.e., LSTM or GRU, with only a single layer. Tables 2 and 3 show the evaluation of the Vanilla LSTM and Vanilla GRU prediction in the east lane of VKRM820 under various units. The overall performance of the LSTM model was better than that of the GRU model, but the training time was longer. When there were only a single layer of LSTM and GRU, and the unit parameters were 4, 8, 16, 32, and 64. It was prone to encountering the problem of learning bottlenecks (training loss and validation loss are almost unchanged) or low-level fitting. When the unit parameters were 128 and 256, more epoch training reduced the loss value with its limit. According to Lewis' MAPE accuracy classification, its prediction was ranked "inaccurate".

Table 2. Prediction of VANILLA LSTM in the east lane of VKRM820.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	68.563	60.414	56.783	54.975	48.175	51.994	50.421
MSE	0.303	0.198	0.188	0.146	0.138	0.116	0.118
MAE	0.428	0.299	0.279	0.218	0.198	0.159	0.157
Time	29 min 48 s	19 min 33 s	28 min 39 s	28 min 10 s	20 min 57 s	59 min 40 s	1 h 40 s

Table 3. Prediction of VANILLA GRU in the east lane of VKRM820.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	70.198	79.199	72.030	66.403	68.731	57.508	50.012
MSE	0.336	0.229	0.197	0.175	0.174	0.124	0.132
MAE	0.454	0.345	0.297	0.257	0.255	0.177	0.185
Time	19 min 46 s	16 min 5 s	15 min 36 s	15 min 23 s	15 min 40 s	46 min 8 s	48 min 40 s

Tables 4 and 5 present the result of the Vanilla LSTM and Vanilla GRU prediction in the west lane of VKRM820 under various units. The overall performance of the LSTM model was better than that of the GRU model, but the training time was longer. The evaluations of the MAPE, MSE, and MAE were significantly worse than that of the VKRM820 eastward lane, and according to Lewis' MAPE accuracy classification, it is ranked "inaccurate" prediction.

Table 4. Prediction of VANILLA LSTM in the west lane of VKRM820.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	78.214	70.654	64.325	60.794	65.512	61.249	63.350
MSE	0.703	0.466	0.371	0.283	0.239	0.236	0.224
MAE	0.592	0.473	0.402	0.330	0.298	0.292	0.281
Time	29 min 2 s	28 min 46 s	27 min 59 s	27 min 55 s	27 min 58 s	57 min 57 s	1 h 23 s

Table 5. Prediction of VANILLA GRU in the west lane of VKRM820.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	89.821	83.798	80.621	75.071	70.644	68.073	69.716
MSE	0.715	0.602	0.595	0.566	0.463	0.259	0.244
MAE	0.645	0.561	0.531	0.500	0.433	0.309	0.297
Time	15 min 57 s	15 min 44 s	14 min 59 s	15 min 14 s	14 min 56 s	45 min 40 s	48 min 36 s

Tables 6 and 7 reveal the Vanilla LSTM and Vanilla GRU prediction in the north lane of VKWN800 under various units. The overall performance of the LSTM model was better than that of the GRU model, but the training time was longer. The evaluations of the MAPE, MSE, and MAE were significantly worse than that of the VKRM820 eastward lane. Even if increasing the number of units can effectively reduce the loss value, it has its limit. According to Lewis' MAPE accuracy classification, it is ranked "inaccurate" prediction.

Table 6. Prediction of VANILLA LSTM in the north lane of VKWN800.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	77.510	80.423	75.476	69.035	64.242	54.430	55.740
MSE	0.779	0.657	0.592	0.511	0.446	0.310	0.324
MAE	0.565	0.471	0.420	0.368	0.317	0.189	0.199
Time	29 min 41 s	28 min 42 s	28 min 6 s	28 min 34 s	29 min 3 s	57 min 42 s	1 h 13 s

Table 7. Prediction results of VANILLA GRU in the north lane test set of VKWN800.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	100.787	98.284	98.909	94.709	91.722	73.398	92.598
MSE	1.168	0.760	1.109	0.792	0.667	0.511	0.620
MAE	0.831	0.590	0.802	0.609	0.473	0.350	0.521
Time	15 min 57 s	15 min 44 s	15 min 13 s	15 min 9 s	15 min 13 s	45 min 50 s	48 min 26 s

Tables 8 and 9 show the Vanilla LSTM and Vanilla GRU prediction in the south lane of VKWN800 under various units. The overall performance of the LSTM model was better than that of the GRU model, but the training time was longer. In the Vanilla LSTM of 8 units, the three evaluations of MAPE, MSE, and MAE were better than that of 16, 32, and 64 units. This phenomenon is opposite to the previous VKRM820 in the east and west lanes. The evaluations of the MAPE, MSE, and MAE were significantly worse than that of the VKRM820 eastward lane. Even if increasing the number of units effectively reduced the loss value, it showed its limit. According to Lewis' MAPE accuracy classification, it is an "inaccurate" prediction.

Table 8. Prediction of VANILLA LSTM in the south lane of VKWN800.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	101.627	97.693	132.424	144.599	120.241	89.442	129.854
MSE	0.958	0.354	0.757	0.719	0.672	0.254	0.209
MAE	0.848	0.447	0.720	0.687	0.649	0.315	0.284
Time	28 min 36 s	28 min 46 s	28 min 19 s	28 min 19 s	27 min 48 s	57 min 33 s	1 h 1 min 23 s

Table 9. Prediction of VANILLA GRU in the south lane of VKWN800.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	101.278	107.800	103.922	108.222	105.009	114.851	237.453
MSE	0.964	0.938	0.962	0.938	0.957	0.789	0.886
MAE	0.851	0.838	0.849	0.838	0.847	0.752	0.803
Time	15 min 57 s	15 min 30 s	15 min 10 s	15 min 5 s	15 min 27 s	46 min 23 s	48 min 19 s

Tables 10 and 11 show the evaluation of the Vanilla LSTM and Vanilla GRU prediction in the east lane of VKWNV20 under various units. The simulation showed that the overall performance of the LSTM model was better than that of the GRU model, but the gap was not significant. The evaluations of the MAPE, MSE, and MAE were significantly worse than that of the VKRM820 eastward lane. Even if increasing the number of units effectively reduced the loss value, the effect was not as good as the VKRM820 in the east lane and VKWN800 in the north lane. According to Lewis' MAPE accuracy classification, it is an "inaccurate" prediction.

Table 10. Prediction of VANILLA LSTM in the east lane of VKWNV20.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	137.058	126.985	113.882	113.211	113.736	120.333	135.323
MSE	3.724	3.703	3.717	3.689	3.675	3.505	3.410
MAE	0.514	0.465	0.508	0.471	0.463	0.366	0.343
Time	33 min 9 s	29 min 21 s	30 min 45 s	32 min 10 s	28 min 59 s	59 min 28 s	59 min 37 s

Table 11. Prediction of VANILLA GRU in the east lane of VKWNV20.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	105.374	147.958	112.256	112.211	122.547	137.412	140.778
MSE	3.777	3.663	3.756	3.741	3.690	3.668	3.657
MAE	0.595	0.481	0.575	0.570	0.514	0.449	0.449
Time	15 min 34 s	15 min 35 s	15 min 34 s	15 min 30 s	15 min 10 s	47 min 6 s	48 min 50 s

Tables 12 and 13 present the Vanilla LSTM and Vanilla GRU prediction in the west lane of VKWNV20 under various units. The simulation shows that the overall performance of the LSTM model was better than that of the GRU model, but the gap was not significant. The evaluations of the MAPE, MSE, and MAE were significantly worse than that of the VKRM820 westward lane. Even if increasing the number of units effectively reduced the loss value, it had its limit. According to Lewis' MAPE accuracy classification, it is an "inaccurate" prediction.

Table 12. Prediction of VANILLA LSTM in the west lane of VKWNV20.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	89.634	100.658	105.359	107.994	108.121	140.375	146.721
MSE	1.697	1.481	1.471	1.453	1.444	1.410	1.370
MAE	0.567	0.464	0.445	0.398	0.392	0.358	0.343
Time	29 min 44 s	30 min 5 s	28 min 7 s	28 min 28 s	29 min 14 s	58 min 49 s	58 min 35 s

Table 13. Prediction of VANILLA GRU in the west lane of VKWNV20.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	100.886	111.012	104.821	104.347	107.894	114.547	117.554
MSE	1.731	1.431	1.487	1.425	1.418	1.378	1.362
MAE	0.651	0.452	0.468	0.415	0.391	0.344	0.332
Time	15 min 48 s	15 min 57 s	15 min 34 s	15 min 32 s	15 min 49 s	47 min 14 s	49 min 29 s

Tables 14 and 15 present the Vanilla LSTM and Vanilla GRU prediction in the south lane of VMCN800 under various units. The simulation shows that the overall performance of the LSTM model was still better than that of the GRU model. The evaluations of the MAPE, MSE, and MAE were significantly better than that of the VKRM820 eastward lane, and according to Lewis' MAPE accuracy classification, it is a "reasonable" prediction.

Table 14. Prediction of VANILLA LSTM in the south lane of VMCN800.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	39.129	28.168	34.247	28.623	28.921	24.401	22.992
MSE	0.119	0.096	0.122	0.110	0.115	0.094	0.080
MAE	0.223	0.184	0.188	0.178	0.177	0.157	0.135
Time	28 min 37 s	28 min 55 s	27 min 59 s	27 min 58 s	27 min 53 s	56 min 53 s	1 h 1 min 16 s

Table 15. Prediction of VANILLA GRU on the south lane of VMCN800.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	41.525	30.118	30.101	20.017	25.460	23.634	28.544
MSE	0.128	0.123	0.161	0.148	0.105	0.094	0.106
MAE	0.256	0.233	0.242	0.222	0.175	0.158	0.182
Time	19 min 12 s	15 min 30 s	14 min 58 s	15 min 18 s	15 min 23 s	45 min 56 s	48 min 17 s

Tables 16 and 17 reveal the Vanilla LSTM and Vanilla GRU prediction in the north lane of VL7PX00 under various units. The simulation shows that the overall performance of the LSTM model was still better than that of the GRU model. The MAPE, MSE, and MAE evaluations were better than that of the VKRM820 eastward lane. Overall, increasing the number of units effectively reduced the loss of value. According to Lewis' MAPE accuracy classification, Vanilla LSTM was a "reasonable" prediction when units are 8, 16, 32, 64, 128, and 256, while Vanilla GRU is a "reasonable" prediction when units are 64, 128, and 256.

Table 16. Prediction of VANILLA LSTM in the north lane of VL7PX00.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	65.353	49.848	47.432	44.103	40.602	40.502	41.524
MSE	0.436	0.249	0.221	0.190	0.159	0.154	0.149
MAE	0.507	0.325	0.286	0.243	0.206	0.175	0.174
Time	29 min 21 s	29 min 14 s	28 min 49 s	28 min 26 s	28 min 53 s	58 min 28 s	1 h 1 min 12 s

Table 17. Prediction of VANILLA GRU in the north lane of VL7PX00.

	4 Units	8 Units	16 Units	32 Units	64 Units	128 Units	256 Units
MAPE	68.616	57.033	55.862	50.743	45.745	39.649	43.304
MSE	0.473	0.290	0.285	0.237	0.208	0.176	0.185
MAE	0.541	0.366	0.349	0.301	0.258	0.200	0.210
Time	19 min 35 s	15 min 39 s	15 min 17 s	15 min 29 s	15 min 37 s	45 min 43 s	48 min 33 s

3.2. Simulation 2: Stacked Architecture

In simulation 2, a stacked architecture was used in an LSTM or GRU that stacks multiple layers. In the architectures of stacked LSTM and stacked GRU, the LSTM model generally outperformed the GRU model, except for the VL7PX00. However, even if the stacked architecture was used for training, the training results were different from those obtained in Refs. [6,8]. Therefore, we suspect that a dataset problem caused the training results to be inaccurate. Although the dataset used by Ref. [6] is different from ours, the MAPE of the final prediction is the lowest, so we adopt their architecture to verify whether it is a problem with the architecture or the dataset.

3.3. Simulation 3: Encoder–Decoder Architecture

In simulation 3, the encoder–decoder architecture was used to verify whether our dataset has problems. Although Ref. [6] only uses the LSTM model for simulations, we tested the effectiveness of the GRU model. Due to the space limit, we skip the details of simulation data. In the architecture of the Encoder–decoder LSTM and Encoder–decoder GRU, the LSTM model was no longer better than the GRU model. In the south lane of VKWN800, the performance of the GRU model was even better, and the training time was short. However, even referring to the architecture of [6], we were still unable to achieve as good as their MAPE below 10, and the predictions made by many stations were still ranked "inaccurate".

4. Conclusions

Three types of architecture were simulated in this study. Simulation 1 was conducted for traffic flow prediction for five stations through the basic Vanilla architecture. The simulation results showed that only a single layer of LSTM or GRU was not complex enough to make accurate predictions. Even if the number of neurons in LSTM or GRU increased or the number of epochs increased, the improvement in predictions was limited. In simulation 2, multiple layers of LSTM or GRU models were stacked through the stacked architecture. In the stacked architecture, the predictions of either the LSTM model or the GRU model were better than that of the LSTM model and GRU model of the Vanilla architecture, but the training time was long. However, even with the stacked architecture, the MAPE was still far from the MAPE of [6], so we verified whether it was a problem with the dataset by simulating the encoder–decoder architecture. In the encoder–decoder architecture of simulation 3, the predictions obtained using LSTM were generally not different from the stacked architecture, but the training time was short. The GRU model performed better than the stacked architecture, and the training time was short. We could not obtain an excellent performance like the MAPE of [6], which was below 10%, so there were data missing in the dataset. The LSTM outperformed the GRU in the three simulations, but the training time was long. In the encoder–decoder architecture of simulation 3, the GRU model occasionally outperformed the LSTM model.

Author Contributions: Conceptualization, H.-C.J. and C.-A.C.; methodology, H.-C.J. and C.-A.C.; software, C.-A.C.; validation, H.-C.J. and C.-A.C.; formal analysis, H.-C.J. and C.-A.C.; investigation, H.-C.J. and C.-A.C.; resources, H.-C.J. and C.-A.C.; data curation, C.-A.C.; writing—original draft preparation, C.-A.C.; writing—review and editing, H.-C.J.; visualization, C.-A.C.; supervision, H.-C.J.; project administration, H.-C.J.; funding acquisition, H.-C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study is sponsored by the Ministry of Science and Technology, Taiwan (Grant No. MOST 109-2221-E-004-010-).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The regular traffic flow data source is "Vehicle Detector (VD) data of Taipei City Government Data Open Platform". Five vehicle detectors are deployed on the main roads around Taipei Arena: VMCN800, VKRM820, VKWNV20, VKWN800, and VL7PX00. The data are provided by the Traffic Control Center of Taipei City for the entire year of 2017.

Acknowledgments: The authors gratefully thank the reviewers for their precise and constructive remarks, which significantly helped improve the manuscript.

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

References

1. Jang, H.C.; Lin, T.K. Traffic-aware traffic signal control framework based on SDN and cloud-fog computing. In Proceedings of the 2018 IEEE 88th Vehicular Technology Conference (VTC 2018-Fall), Chicago, IL, USA, 27–30 August 2018.
2. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the NIPS'12, 25th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 3–6 December 2012; Volume 1, pp. 1097–1105.
3. Jang, H.C.; Chang, Y.H. Traffic flow forecast for traffic with forecastable sporadic events. In Proceedings of the 12th International Conference on Ubi-Media Computing (Ubi-Media 2019), Bali, Indonesia, 6–9 August 2019.
4. Du, X.; Zhang, H.; Nguyen, H.V.; Han, Z. Stacked LSTM deep learning model for traffic prediction in vehicle-to-vehicle communication. In Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), Toronto, ON, Canada, 24–27 September 2017.
5. Chen, Y.Y.; Lv, Y.; Li, Z.; Wang, F.Y. Long short-term memory model for traffic congestion prediction with online open data. In Proceedings of the IEEE 19th International Conference on Intelligent Transportation System, Rio de Janeiro, Brazil, 1–4 November 2016.
6. Shao, H.; Soong, B.H. Traffic flow prediction with long short-term memory networks (LSTMs). In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22–25 November 2016.
7. Lewis, C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*; Butterworth Scientific: London, UK, 1982.
8. Kang, D.; Lv, Y.; Chen, Y.Y. Short-term traffic flow prediction with LSTM recurrent neural network. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017.

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.