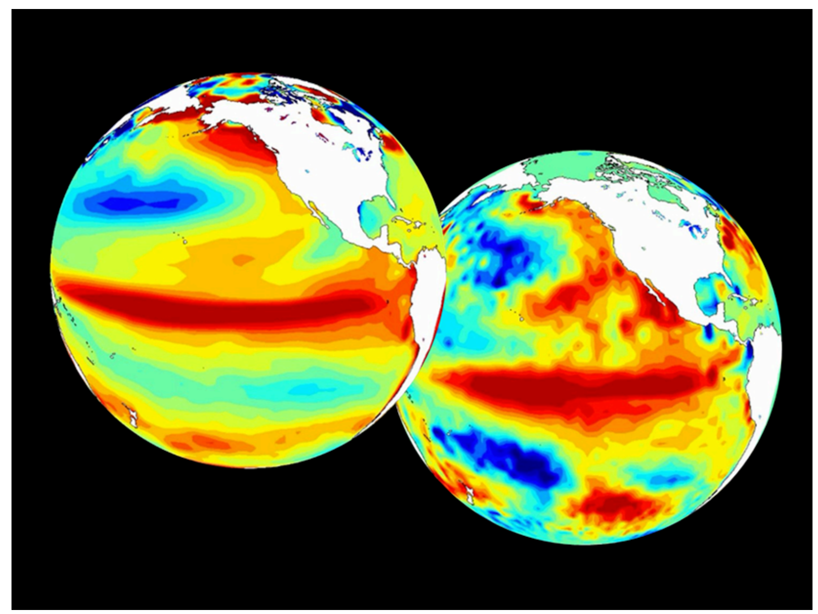
MSDR: Multi-Step Dependency Relation Networks for Spatial Temporal Forecasting

0. Motivation

Spatial temporal forecasting plays an important role in life.

Traffic prediction



Climate forecast

Urban monitor system

Spatial temporal forecasting plays an important role in life.

Challenging!

- the spatial and temporal dependencies are non-linear and dynamic

- rather difficulty to capture the shifting long-range dependency

Current works:

- Statistical and traditional machine learning models
- Deep learning techniques
  - RNN
  - LSTM (RNN variant)
  - CNN
  - Combine GNN

Rely on RNN or its variants

(i) lead to severe information oblivion in modeling the long-range temporal dependency

(ii) introduce noises into historical in- formation due to the error accumulation by steps

Observation: the temporal representation of the current input will depend on multiple previous hidden states, which in turn have varying degrees of correlation with features of the current moment

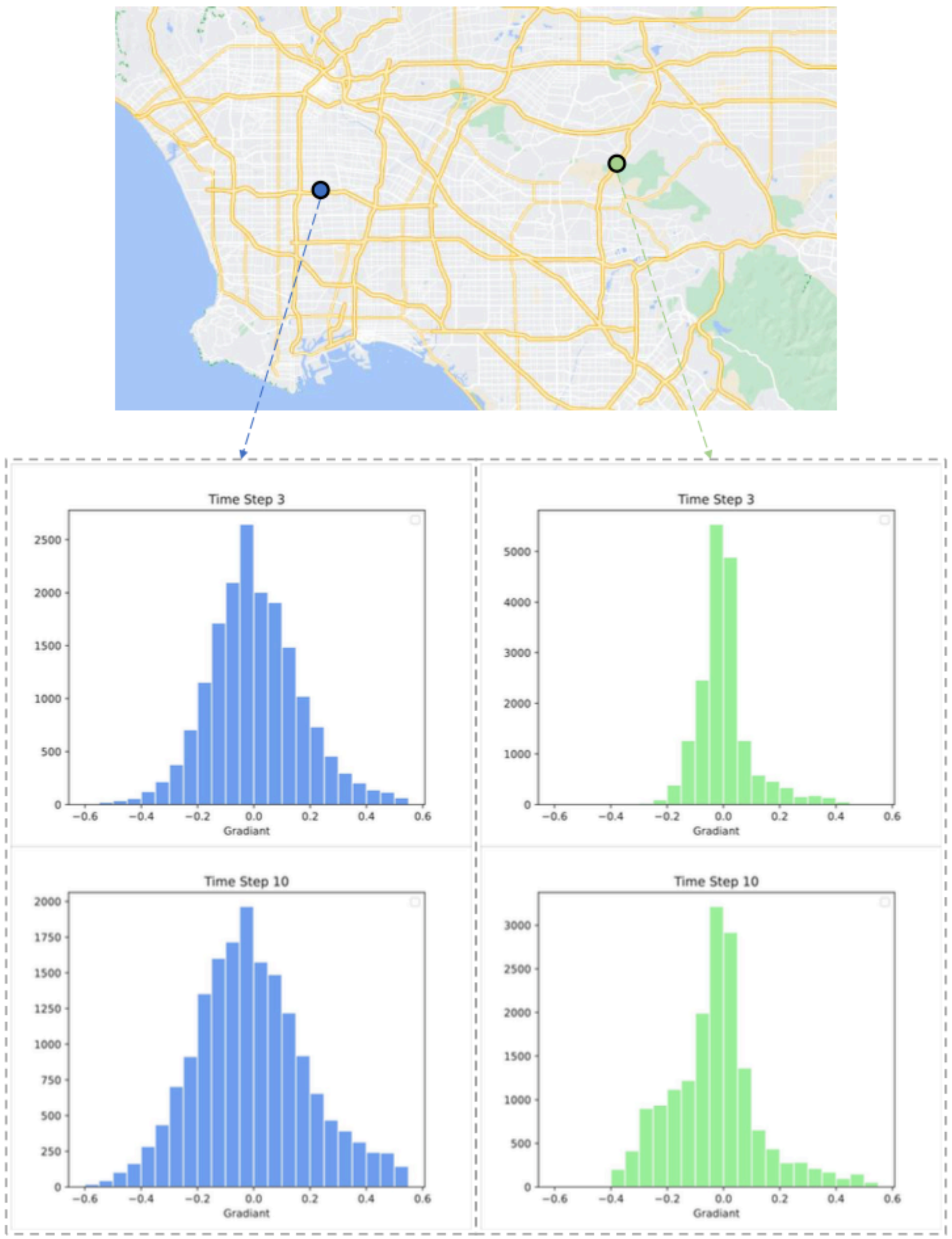


Figure 1: Example: Information from Multiple Time-steps in Traffic Network

- **Different place**
- **Different time**
- **Different previous steps**

1. Problem Definition

sections are summarized in Table 1.

Spatial Networks The spatial network is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V}$ denotes the set of traffic locations scattered throughout the zone and $|\mathcal{V}| = N$. $\mathcal{E}$ corresponds to the set of edges connected between locations. $\mathcal{A}$ is the adjacency matrix on the graph that records the non-Euclidean distances between locations.

Spatial Temporal Forecasting Problem Let $\boldsymbol{X}_t = \{\boldsymbol{x}_{1,t}, ..., \boldsymbol{x}_{N,t}\} \in \mathbb{R}^{N \times f}$ represent the features of all locations at the t-th interval, where f is the input feature dimension. The goal of spatial temporal forecasting can be described as Equation (1):

$$[\boldsymbol{X}_{t-\tau+1}, ..., \boldsymbol{X}_t] \xrightarrow{\mathcal{H}(\cdot)} [\hat{\boldsymbol{Y}}_{t+1}, ..., \hat{\boldsymbol{Y}}_{t+\lambda}], \tag{1}$$

where $[\boldsymbol{X}_{t-\tau+1}, ..., \boldsymbol{X}_t]$ denotes a feature sequence of τ historical times. The goal is to learn a mapping function $\mathcal{H}(\cdot)$ that can predict the spatial temporal features at λ future time steps.

Spatial Temporal Forecasting —> Traffic Flow Prediction / Traffic Demand Prediction

2. Methodology

2.1 MSDR: the new RNN variant

The representation of current time step could be influenced by multiple previous time steps.

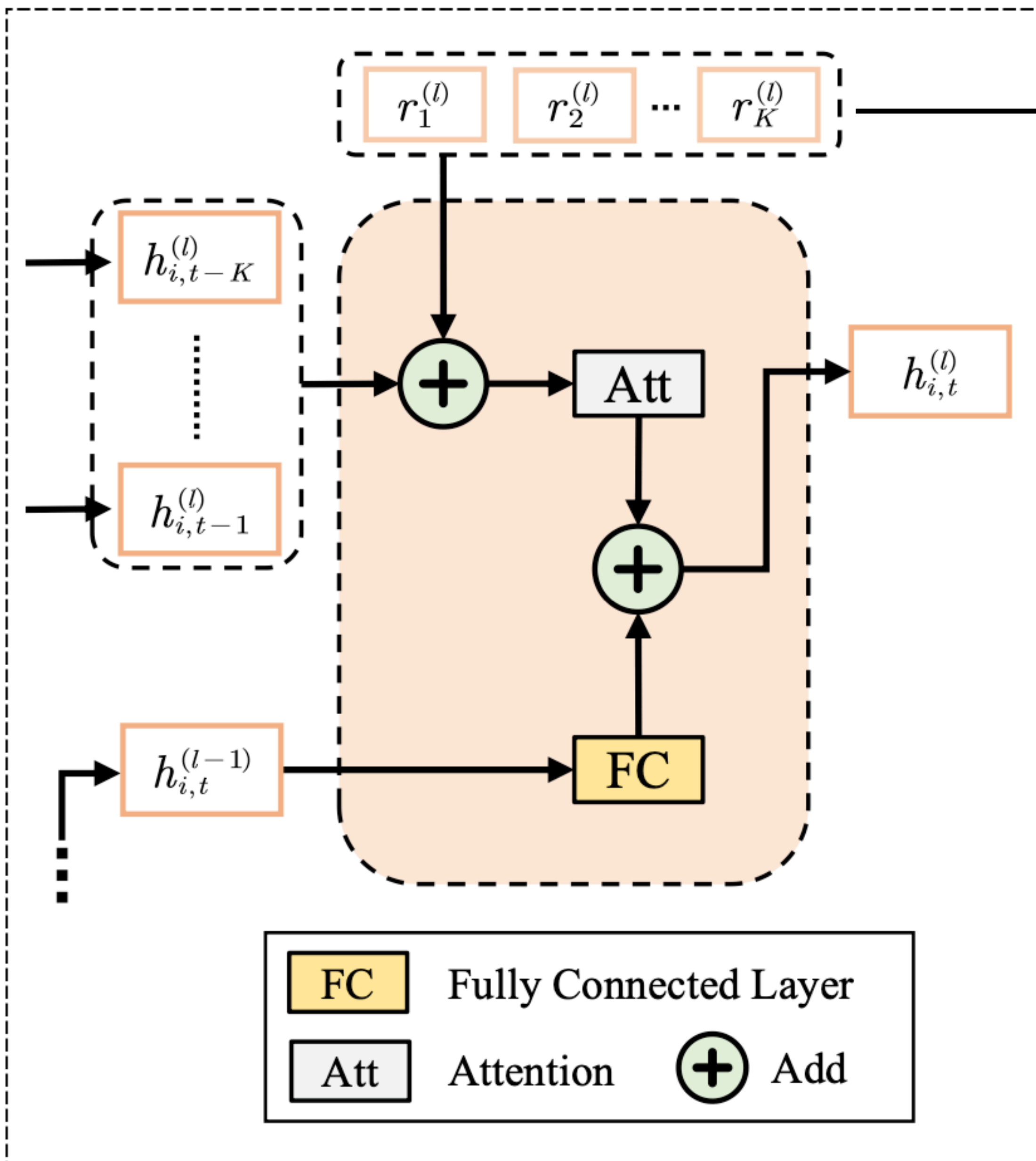


Figure 2: Computation Process of Temporal Explicit Dependency

Dependency relation (Trainable)

h: hidden states

$$\boldsymbol{g}_{i,t}^{(l)} = \sum_{k=1}^{K} \alpha_k^l (\boldsymbol{h}_{i,t-k}^{(l)} + \boldsymbol{r}_k^{(l)}). \tag{2}$$

Attention coefficient

$$\alpha_k^{(l)} = \frac{exp((\boldsymbol{h}_{i,t-k}^{(l)} + \boldsymbol{r}_k^{(l)})\boldsymbol{W}_\alpha^{(l)} + \boldsymbol{b}_\alpha^{(l)})}{\sum_{k=1}^{K} exp((\boldsymbol{h}_{i,t-k}^{(l)} + \boldsymbol{r}_k^{(l)})\boldsymbol{W}_\alpha^{(l)} + \boldsymbol{b}_\alpha^{(l)})}, \tag{3}$$

$$\boldsymbol{s}_{i,t}^{(l)} = \boldsymbol{h}_{i,t}^{(l-1)} \boldsymbol{W}^{(l)} + \boldsymbol{b}^{(l)}, \tag{4}$$

$$\boldsymbol{h}_{i,t}^{(l)} = \boldsymbol{s}_{i,t}^{(l)} + \boldsymbol{g}_{i,t}^{(l)}. \tag{5}$$

2.1 MSDR: the new RNN variant

The representation of current time step could be influenced by multiple previous time steps.

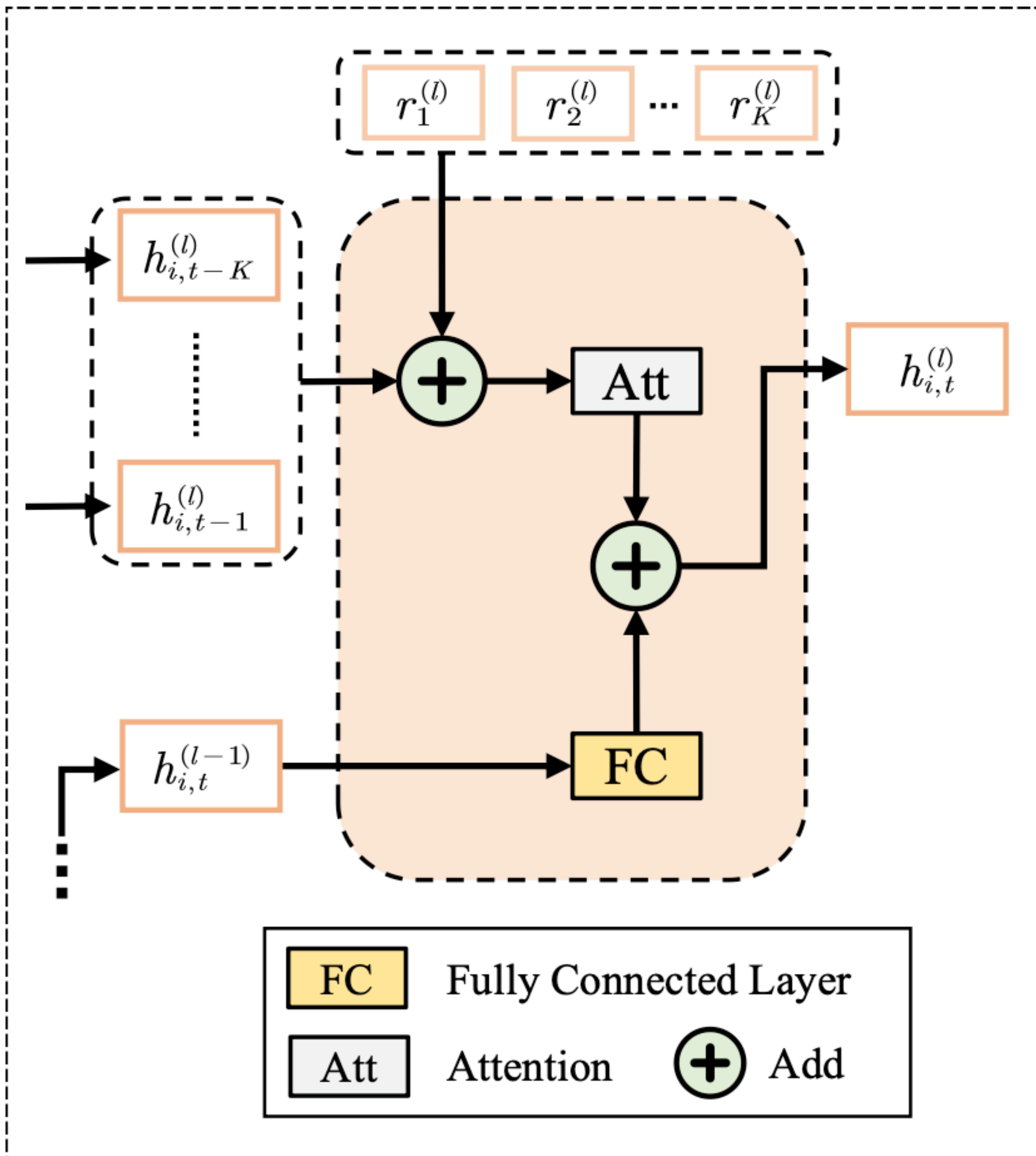


Figure 2: Computation Process of Temporal Explicit Dependency

Advantage:

Explicitly capture the dependency between multiple previous time steps and current time step

Easy to integrate MSDR into many existing spatial temporal forecasting frameworks

2.1 Spatial Dependency

Data-based Spatial dependency: defined by the k-th latest hidden embedding

$$idx_k = argmax(\boldsymbol{h}_{i,t-k}^{(l)}), \tag{6}$$

$$r_k^{(l)} = L(\boldsymbol{R}_{idx_k}^{(l)}, k). \tag{7}$$

Explicit Spatial dependency: involve location information

$$\boldsymbol{r}_{i,k}^{(l)} = L(\boldsymbol{R}_i^{(l)}, k) \tag{8}$$

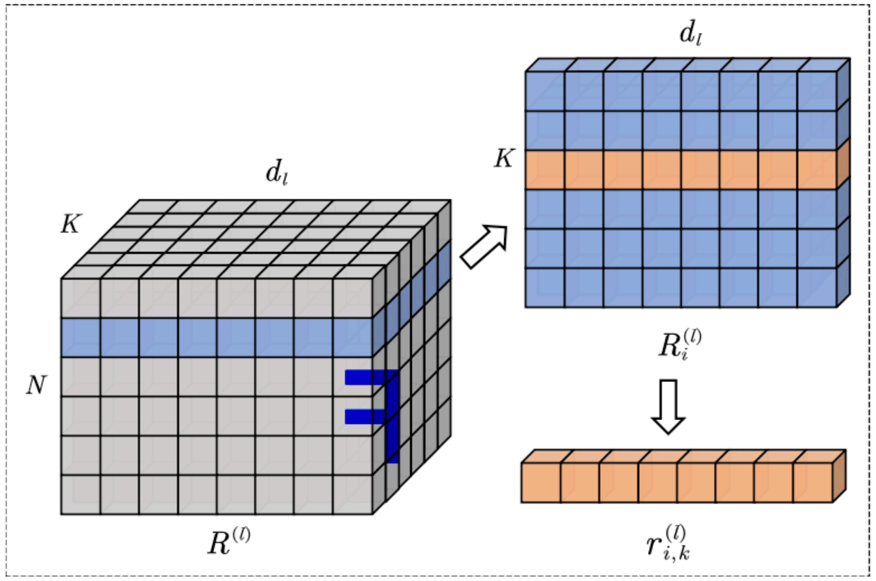


Figure 3: Selection Process of Explicit Spatial Dependency

2.2 Overall Framework

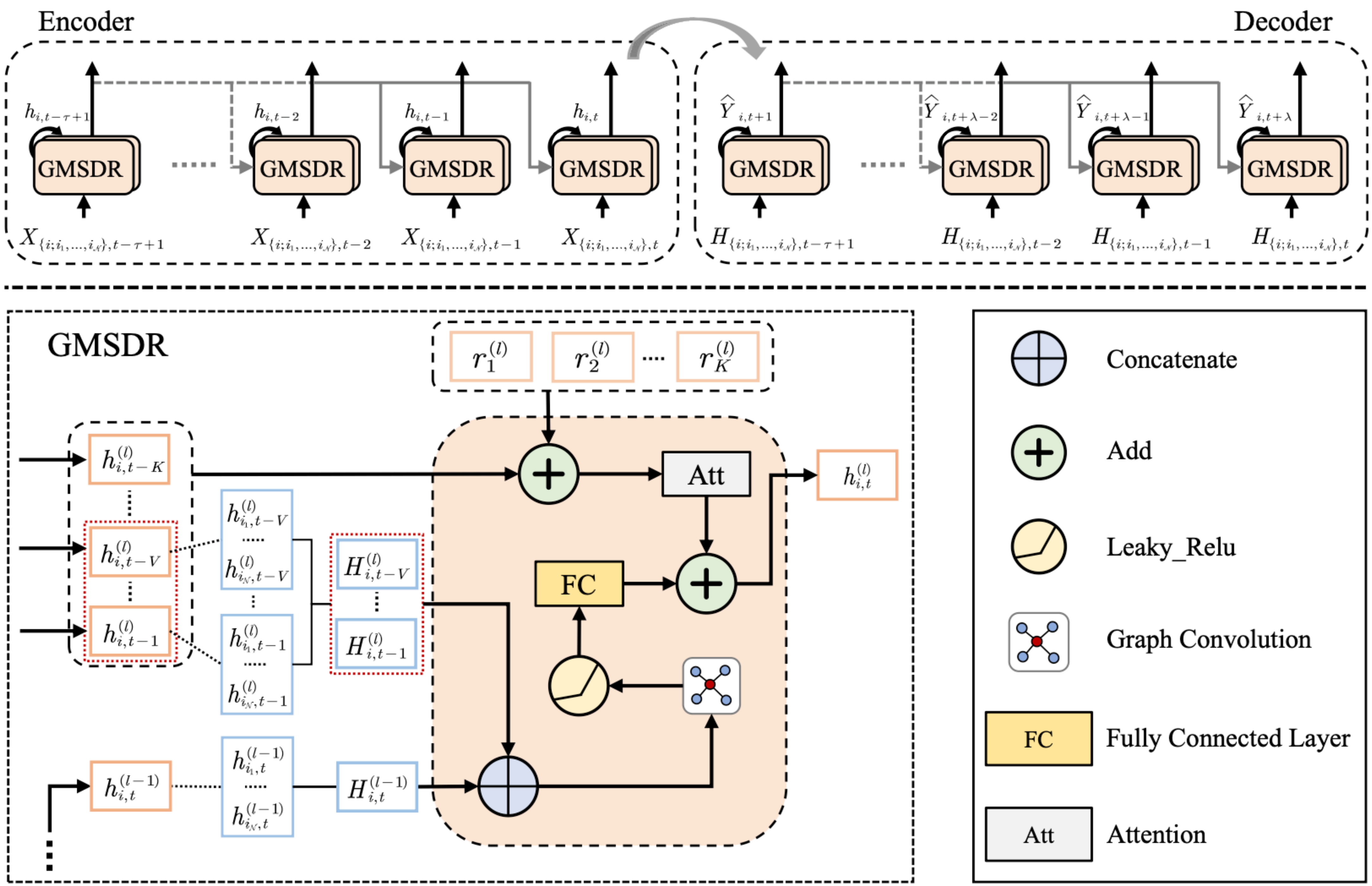


Figure 4: Overall Framework

3. Experiment

Result on Traffic Demand Prediction and Traffic Flow Prediction

Table 2: Main Results of Traffic Demand Prediction

Method	NYC Citi Bike			NYC Taxi		
	RMSE	MAE	PCC	RMSE	MAE	PCC
HA	5.2003	3.4617	0.1669	29.7806	16.1509	0.6339
FC-LSTM	3.8139	2.3026	0.5675	18.0708	10.2200	0.8645
DCRNN	3.2094	1.8954	0.7227	14.7926	8.4274	0.9122
STGCN	3.6042	2.7605	0.7316	22.6489	18.4551	0.9156
STG2Seq	3.9843	2.4976	0.5152	18.0450	9.9415	0.8650
GraphWaveNet	3.2943	1.9911	0.7003	13.0729	8.1037	0.9322
CCRNN	2.8382	1.7404	0.7934	9.5631	5.4979	0.9648
GMSDR	**2.7218**	**1.6760**	**0.8107**	**8.6533**	**4.9831**	**0.9711**

Table 3: Main Results of Traffic Flow Prediction

Method	PEMS03			PEMS08		
	MAE	MAPE(%)	RMSE	MAE	MAPE(%)	RMSE
FC-LSTM	21.33 ± 0.24	23.33 ± 4.23	35.11 ± 0.50	22.20 ± 0.18	14.20 ± 0.59	34.06 ± 0.32
DCRNN	18.18 ± 0.15	18.91 ± 0.82	30.31 ± 0.25	17.86 ± 0.03	11.45 ± 0.03	27.83 ± 0.05
STGCN	17.49 ± 0.46	17.15 ± 0.45	30.12 ± 0.70	18.02 ± 0.14	11.40 ± 0.10	27.83 ± 0.20
ASTGCN(r)	17.69 ± 1.43	19.40 ± 2.24	29.66 ± 1.68	18.61 ± 0.40	13.08 ± 1.00	28.16 ± 0.48
GraphWaveNet	19.85 ± 0.03	19.31 ± 0.49	32.94 ± 0.18	19.13 ± 0.08	12.68 ± 0.57	31.05 ± 0.07
STSGCN	17.48 ± 0.15	16.78 ± 0.20	29.21 ± 0.56	17.13 ± 0.09	10.96 ± 0.07	26.80 ± 0.18
STFGNN	16.77 ± 0.09	16.30 ± 0.09	28.34 ± 0.46	16.64 ± 0.09	10.60 ± 0.06	26.22 ± 0.15
GMSDR	**15.78 ± 0.10**	**15.33 ± 0.11**	**26.82 ± 0.08**	**16.36 ± 0.07**	**10.28 ± 0.08**	**25.58 ± 0.12**

Result about different methods of relation dependency

Table 4: Different Strategies on Traffic Demand Prediction

	Metric	E-Spatial	D-Spatial	Simple
Bike	RMSE	**2.722**	2.838	2.961
	MAE	**1.677**	1.740	1.832
	PCC	**0.811**	0.793	0.761
Taxi	RMSE	**8.653**	8.932	9.561
	MAE	**4.983**	5.141	5.498
	PCC	**0.971**	0.965	0.943

Table 5: Different Strategies on Traffic Flow Prediction

	Metric	E-Spatial	D-Spatial	Simple
PEMS03	MAE	**15.78**	16.69	17.72
	MAPE(%)	**15.33**	15.81	16.63
	RMSE	**26.82**	28.52	29.36
PEMS08	MAE	**16.36**	17.54	18.33
	MAPE(%)	**10.28**	11.47	12.24
	RMSE	**25.58**	27.45	28.52

Simple: relation dependency is the same for every step and location

D-spatial: Data-based Spatial dependency

E-spatial: Explicit Spatial dependency

Hyper-parameters

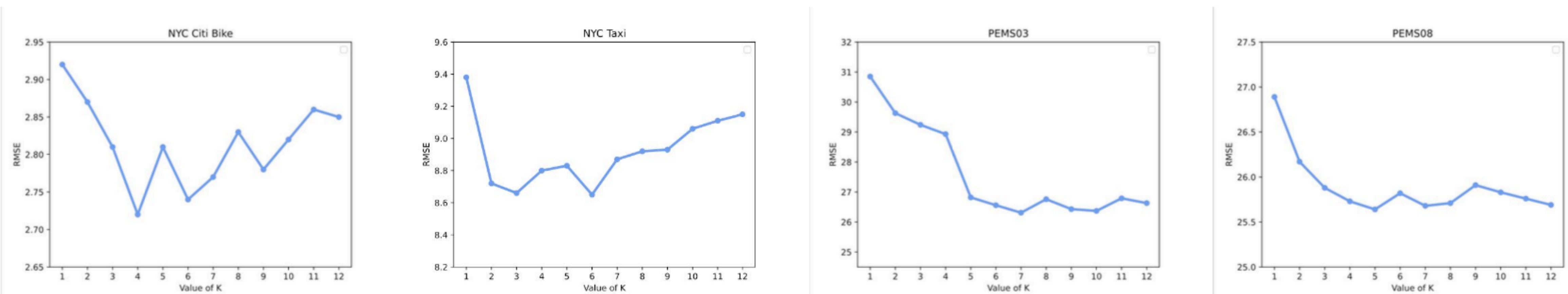
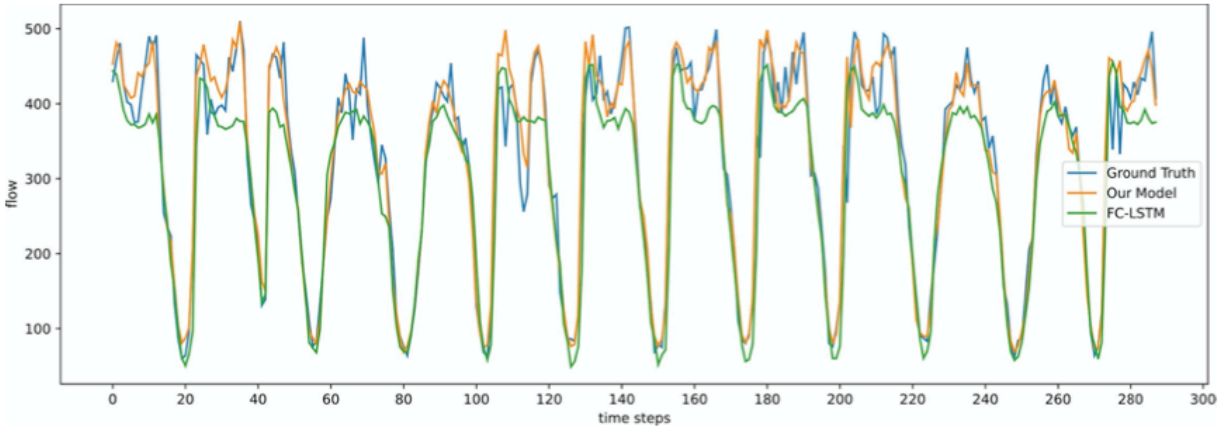
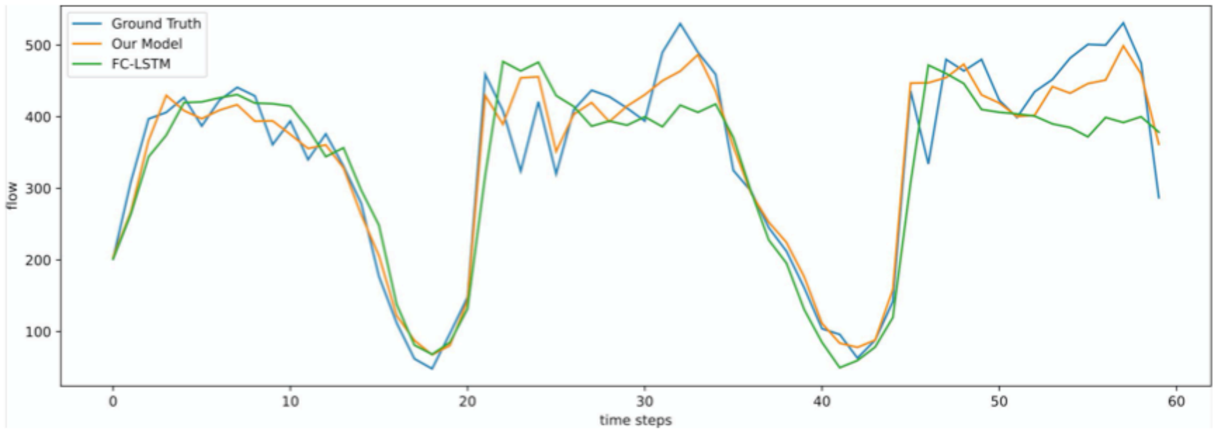


Figure 5: Parameter Analysis: Varying K

Case study



(a) Traffic Flow with Multiple Peaks



(b) Traffic Flow with Smooth Data

Figure 6: Case Study

Thanks