北京工業大學
BEIJINGUNIVERSITYOFTECHNOLOGY

A Survey on Graph Neural Networks for Time Series: Forecasting, Classification, Imputation, and Anomaly Detection

July, 2023

主要内容

Main Contents

与其他深度学习方法相比，图神经网络(Graph Neural Networks, GNN)方法可以明确地建模跨时间和变量间的关系

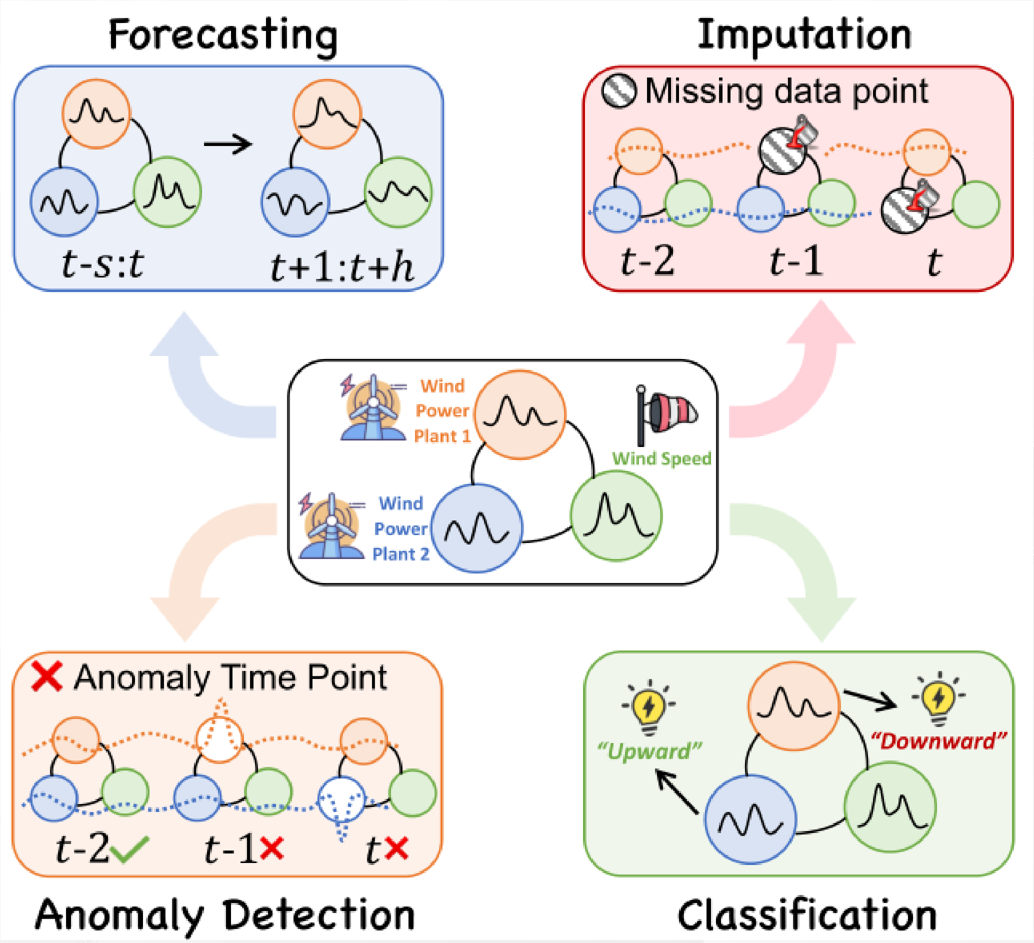
主题

全面回顾图神经网络在时间序列领域的应用：

- 预测
- 分类
- 异常检测
- 数据插补（Imputation）

关键词： 时序预测、图神经网络、分类、插补、异常检测

背景



t ---- 时间步:

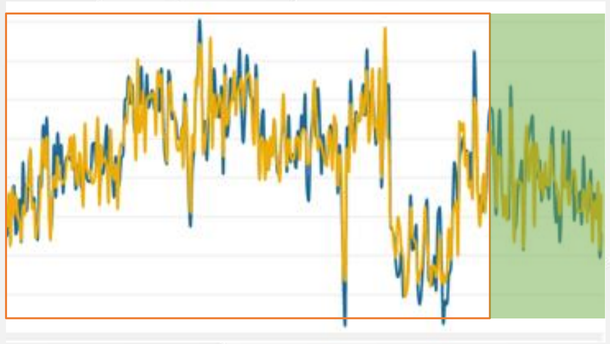
不规则采样 $\mathbf{X} = \{(t_1, x_1), (t_2, x_2), ..., (t_T, x_T)\}$

规则采样 $\mathbf{X} = \{x_1, x_2, ..., x_T\}$

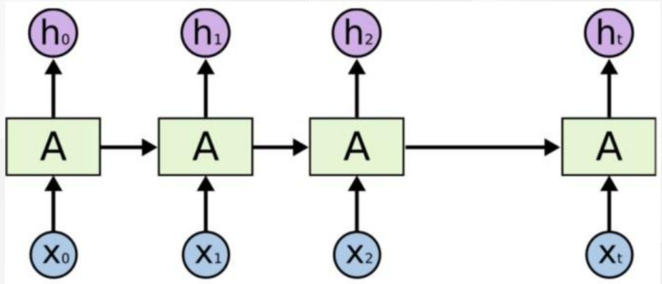
等间隔、相对顺序、不等于真实时间

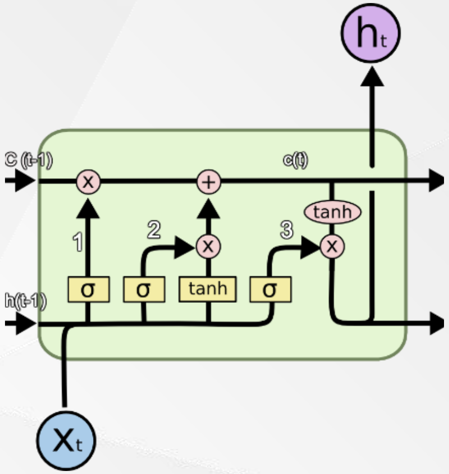
第一天 第二天 第三天 ..... 第n天

Forecasting



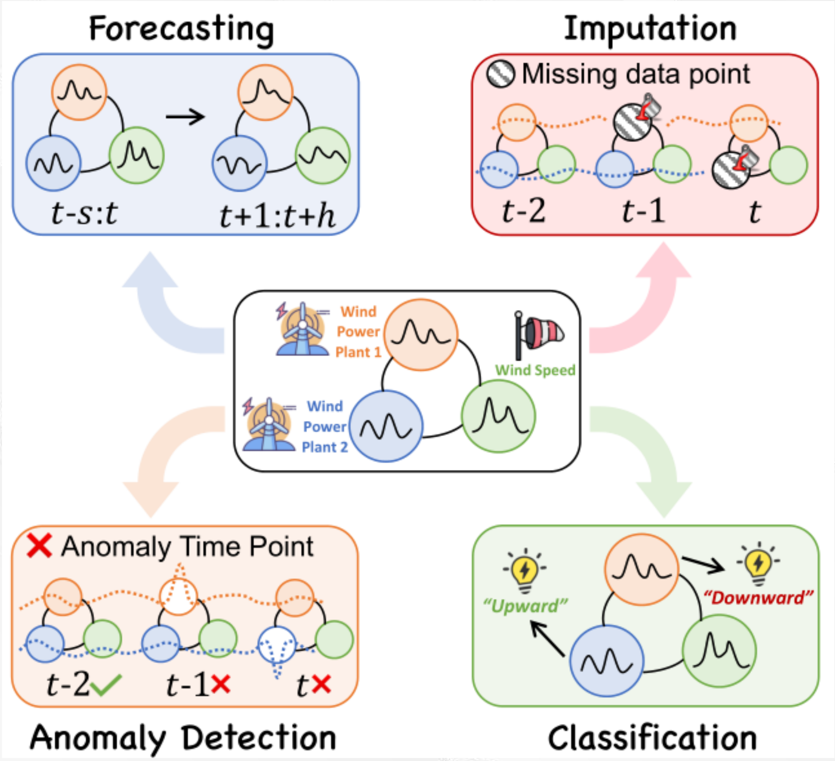
输入 输出

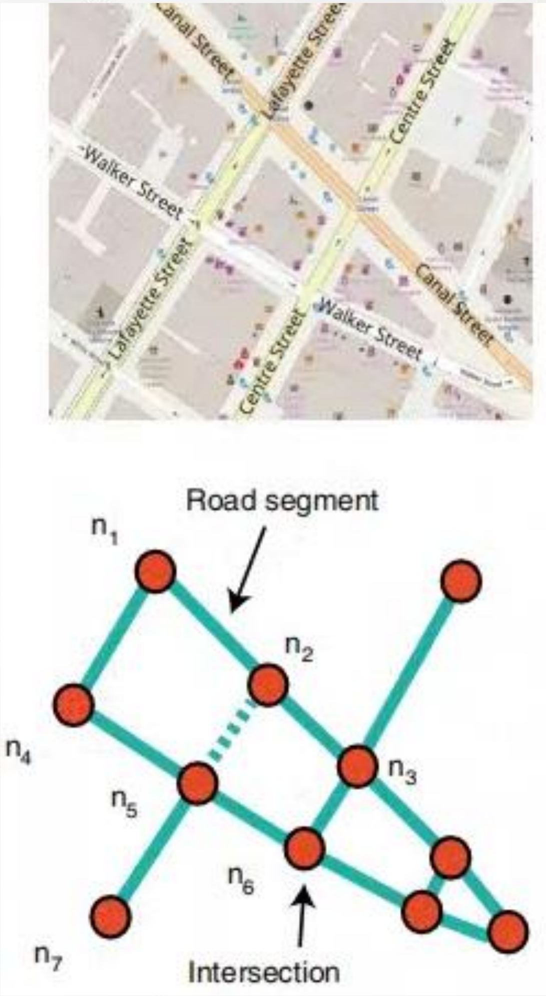




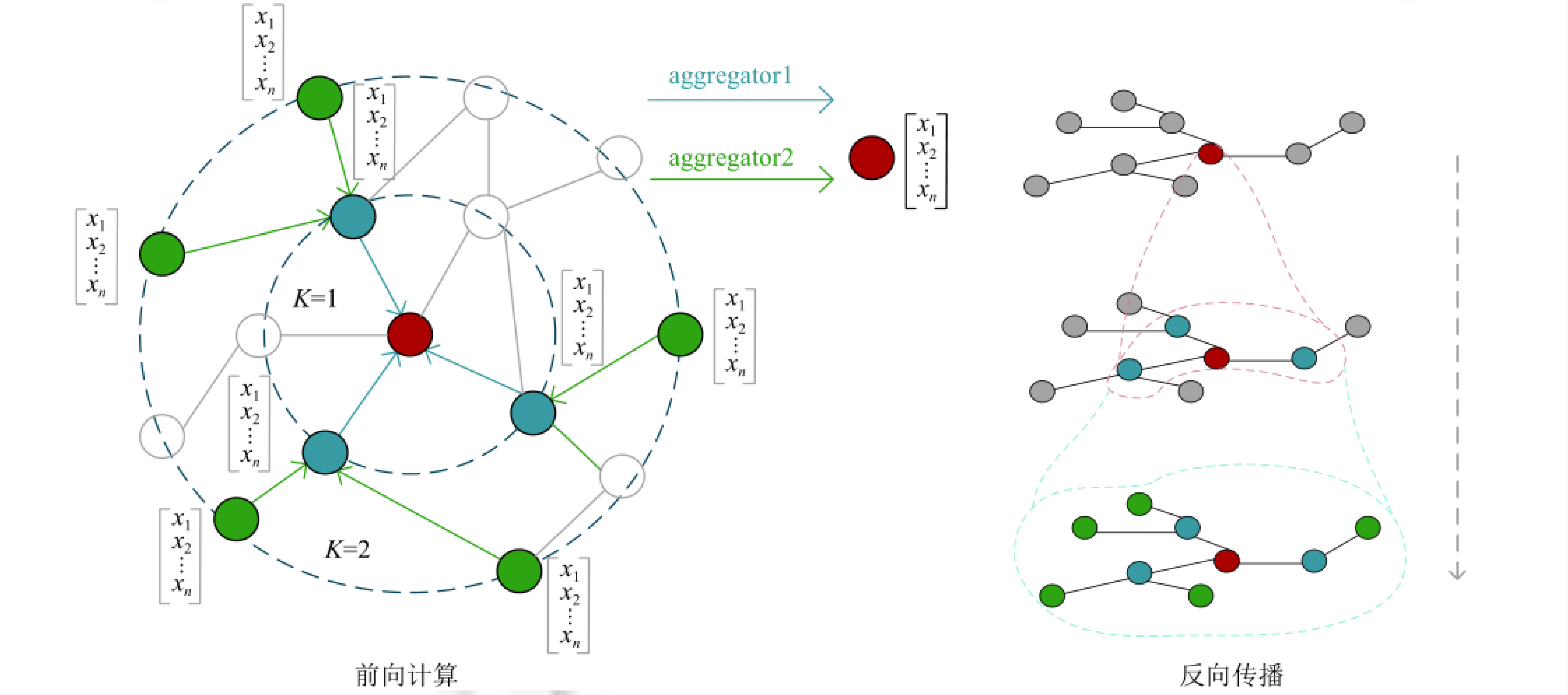
RNN & LSTM

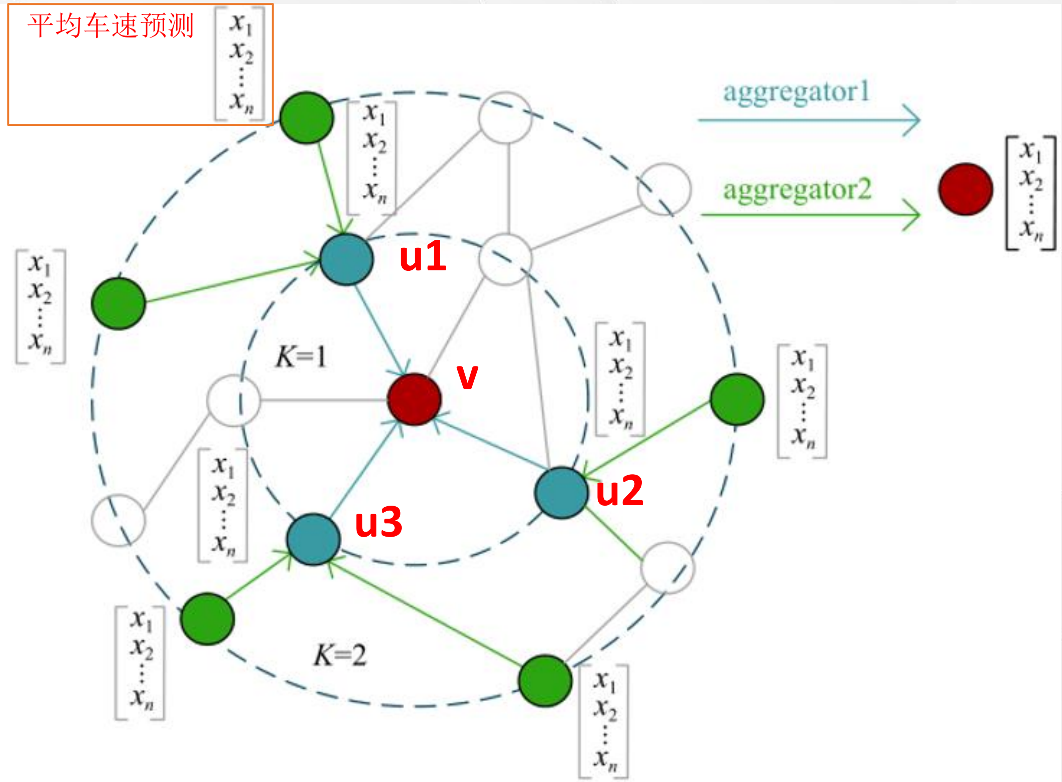
*“这些方法没有明确地对非欧几里得空间中时间序列之间存在的**空间关系**进行建模，**限制了**表达能力。”*





图神经网络的典型计算过程: 本文结合一种广泛应用的图神经网络模型 GraphSAGE, 简要介绍图神经网络的典型计算过程. GraphSAGE 是一种用于学习顶点表示的图神经网络算法, 通过对顶点邻域进行采样和聚合来生成顶点的嵌入. 其中图 $G=(V,E)$, 顶点特征 $H=\{h_v,\forall v\in V\}$, 层数为 K, 权重矩阵为 $W^{(k)},\forall k\in\{1,2,\cdots,K\}$, 非线性函数为 σ, 顶点 v 的邻居表示为 $N(v)$, 聚合操作为 $Agg^{(k)},\forall k\in\{1,2,\cdots,K\}$. 如图 1 所示.





前向计算

为防止过拟合，计算时会临时删除掉一部分邻居节点

基于空间的方法

核心公式

$$h_{N(v)}^{k-1} \leftarrow Aggregate\left(\left\{h_u^{k-1}, \forall u \in N(v)\right\}\right).$$

$$\mathrm{Agg}^{\mathrm{sum}} = \sigma(\mathrm{SUM}\{W\boldsymbol{h}_j + \boldsymbol{b}, \forall v_j \in N(v_i)\})$$

$$h_v^k \leftarrow \sigma\left(W^k Concat\left(h_v^{k-1}, h_{N(v)}^{k-1}\right)\right).$$

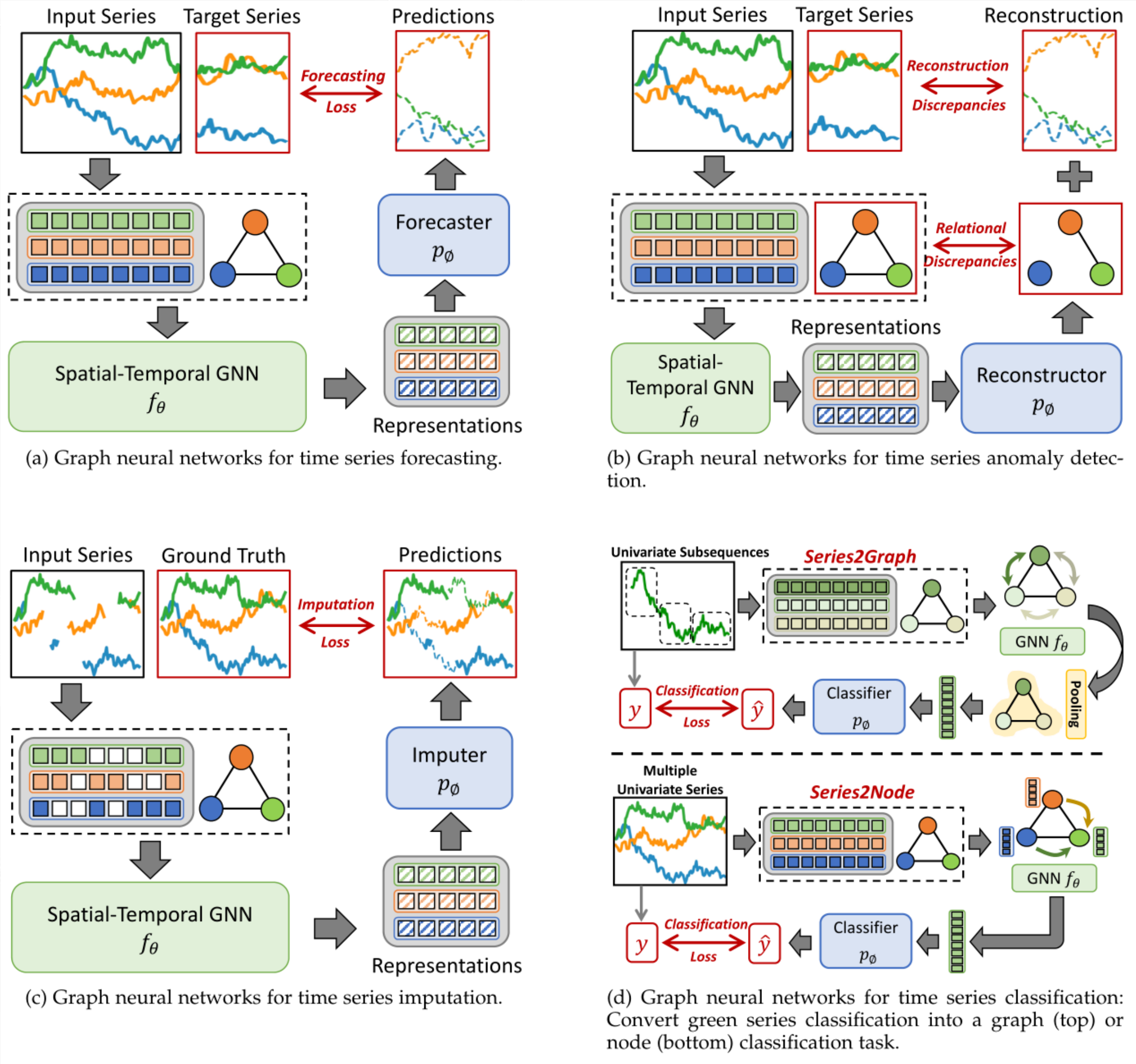
基于谱图论的方法

$$(x * g)_{HG} = \Phi g_\theta(\Lambda)\Phi^\top x,$$

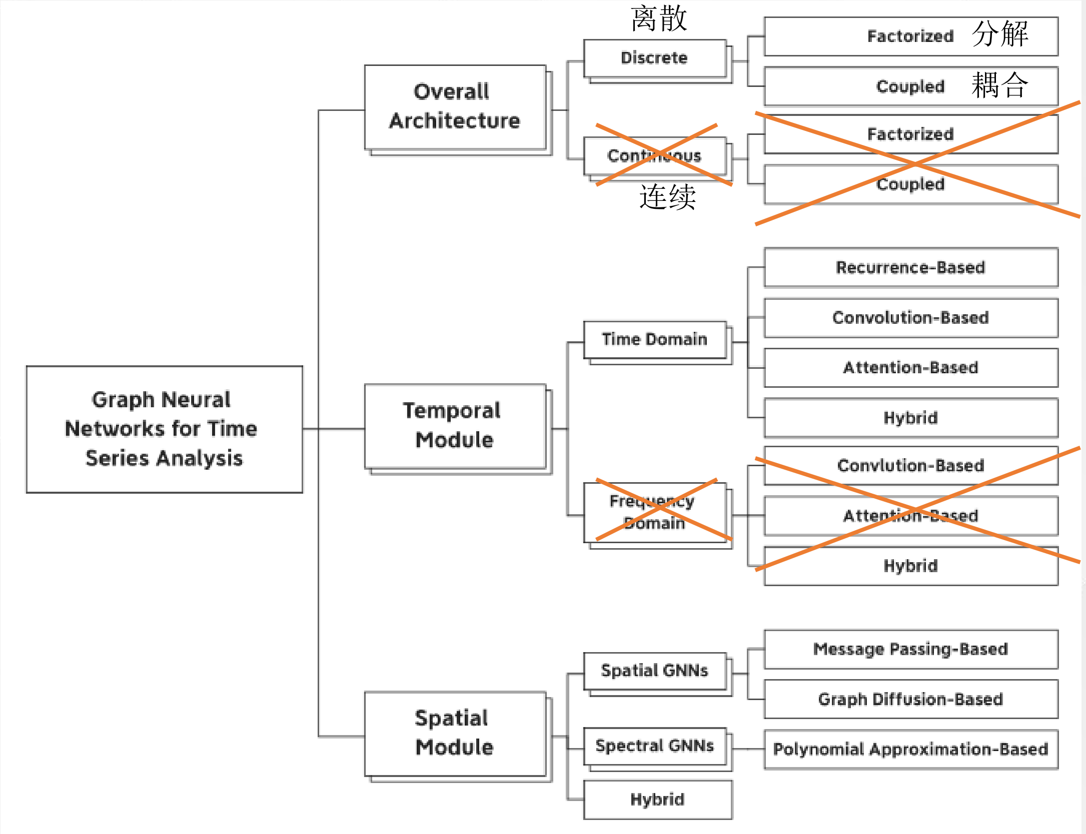
$$f(H^{(l)}, A) = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)})$$

Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\ -1&3&-1&0&-1&0\\ 0&-1&2&-1&0&0\\ 0&0&-1&3&-1&-1\\ -1&-1&0&-1&3&0\\ 0&0&0&-1&0&1 \end{pmatrix}$

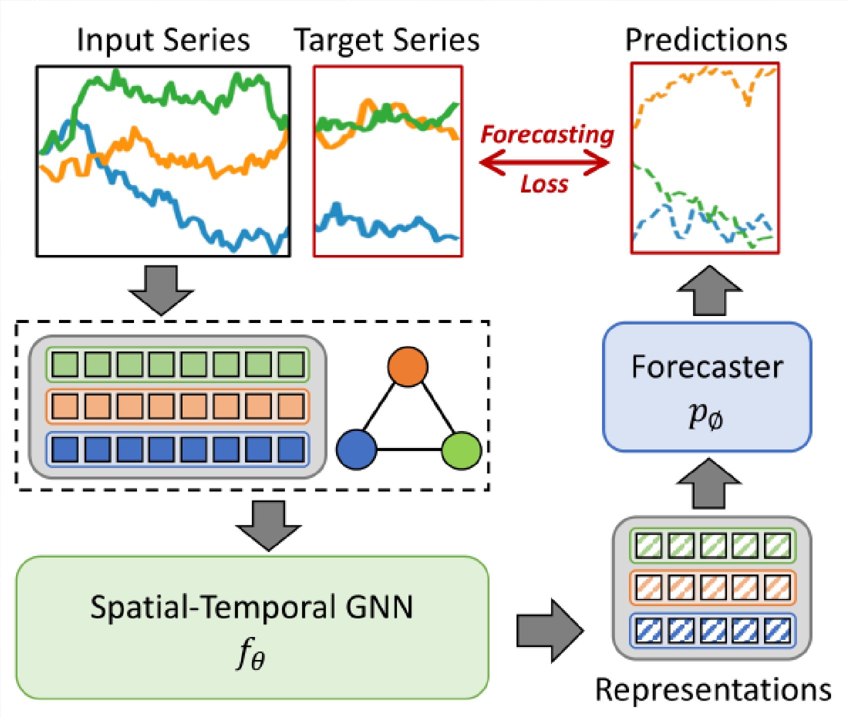
方法



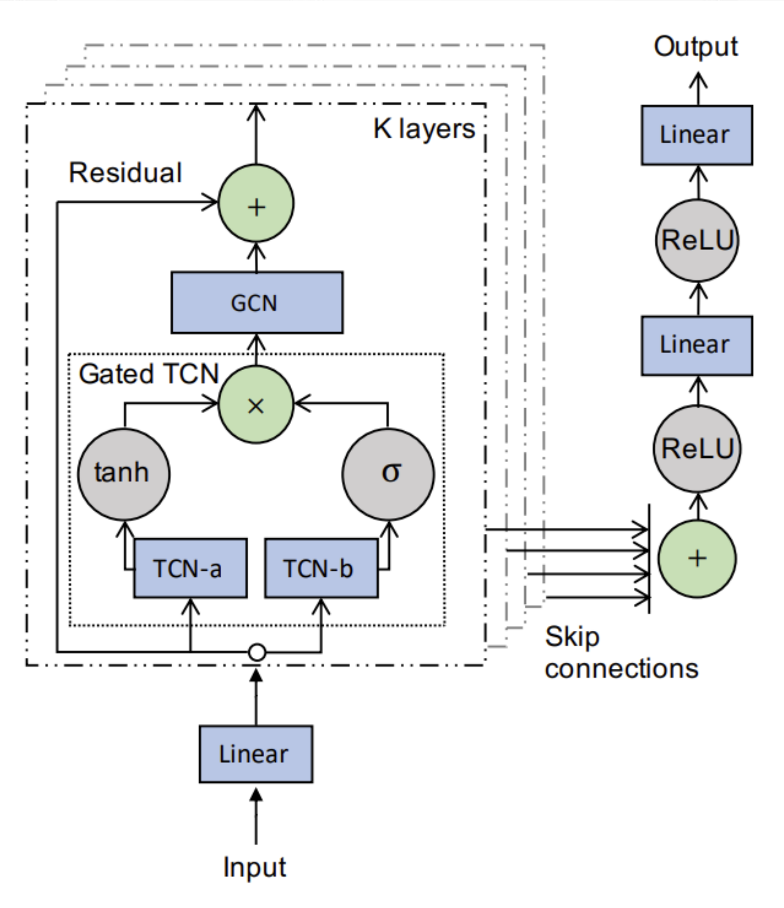
四种任务的架构



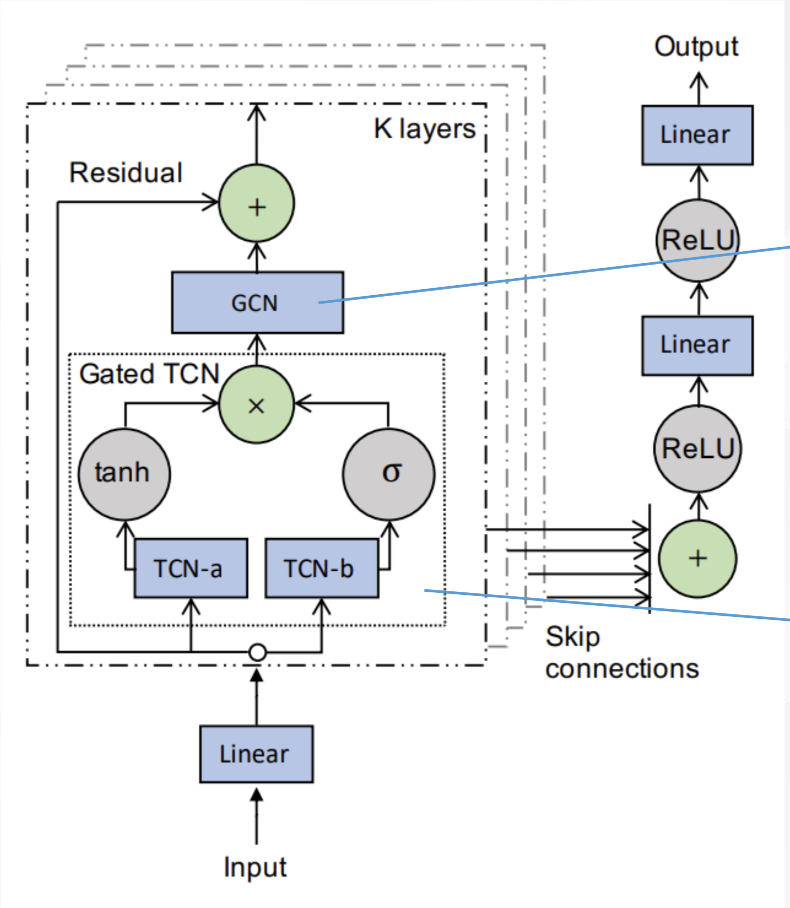
四种任务通用的处理方法



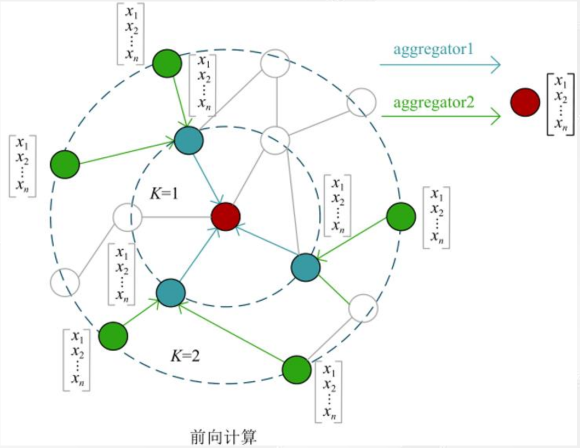
(a) Graph neural networks for time series forecasting.



Graph WaveNet for Deep Spatial-Temporal Graph Modeling, IJCAI 2019



GWNET: Graph WaveNet for Deep Spatial-Temporal Graph Modeling, ijcai 2019



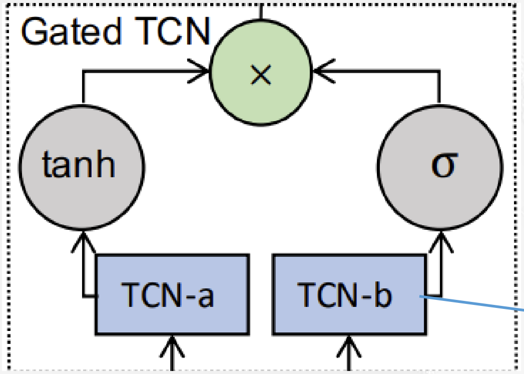
空间层

$$\mathbf{Z} = \tilde{\mathbf{A}}\mathbf{X}\mathbf{W}.$$

$$\mathbf{Z} = \sum_{k=0}^{K} \mathbf{P}_f^k \mathbf{X}\mathbf{W}_{k1} + \mathbf{P}_b^k \mathbf{X}\mathbf{W}_{k2} + \tilde{\mathbf{A}}_{apt}^k \mathbf{X}\mathbf{W}_{k3}.$$

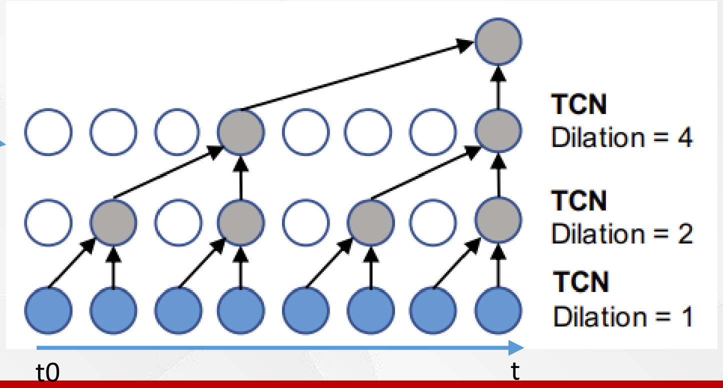
$$\mathbf{P} = \mathbf{A}/rowsum(\mathbf{A}).$$

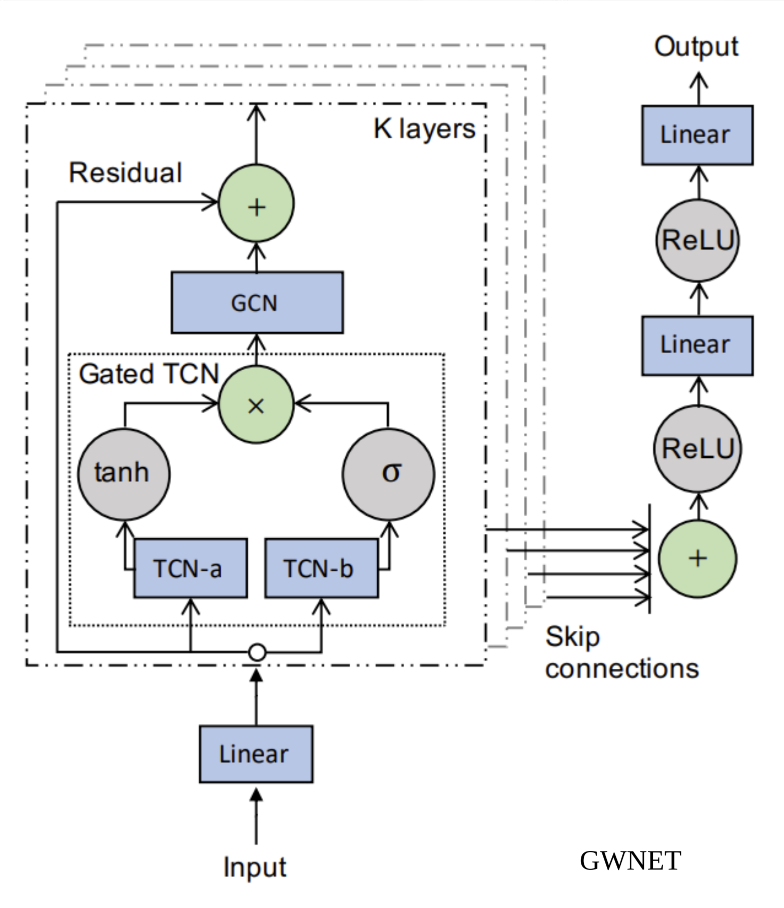
Diffusion-Convolutional Neural Networks, NIPS 2016



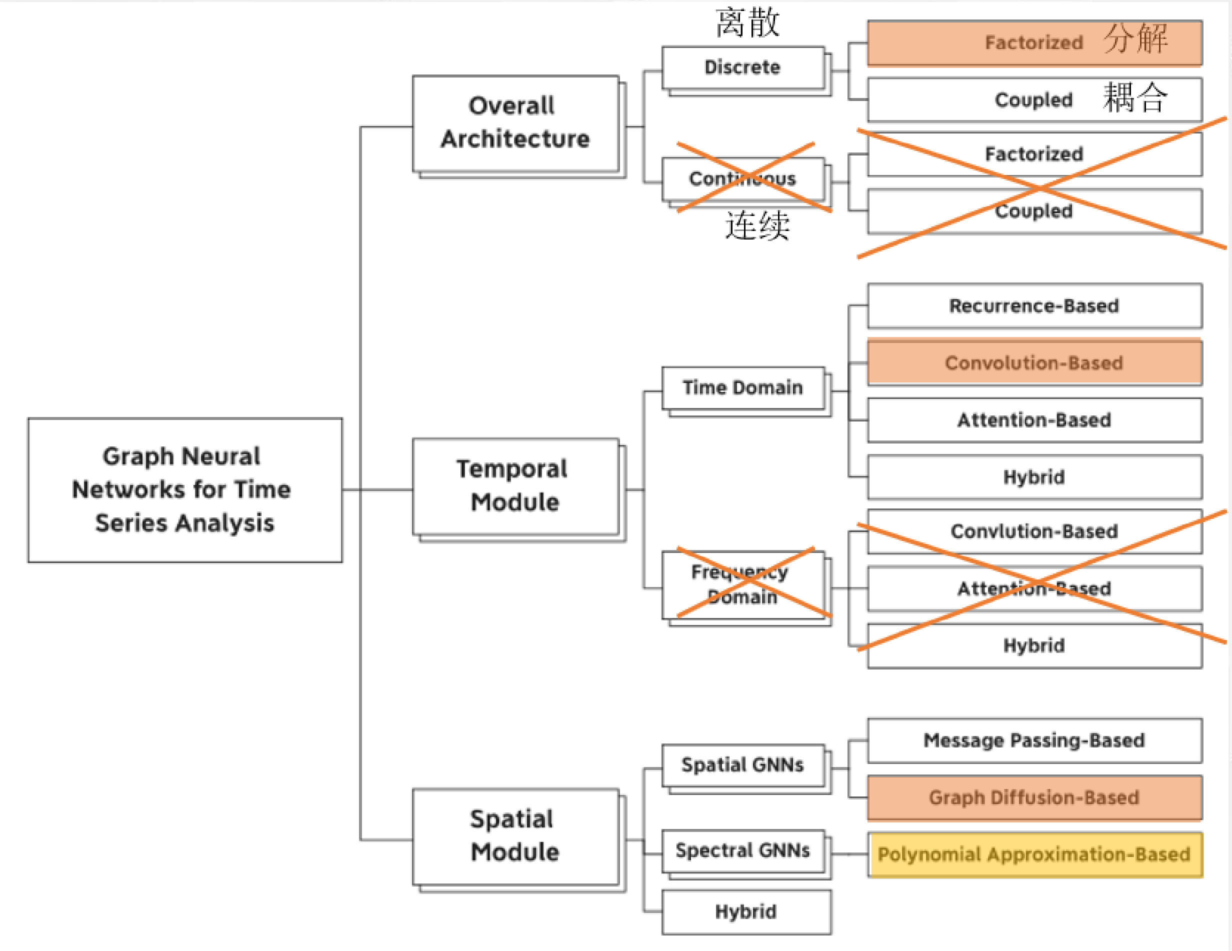
时间层

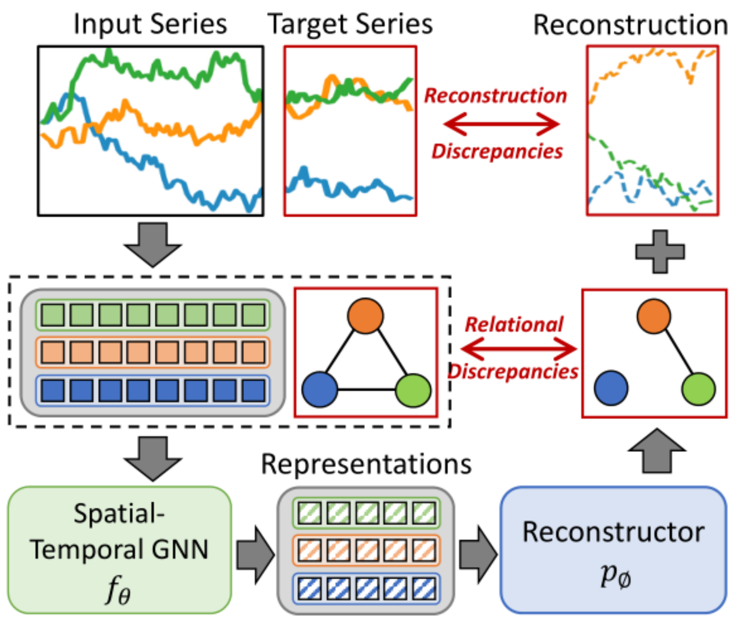
$$\mathbf{h} = g(\mathbf{\Theta_1} \star \mathcal{X} + \mathbf{b}) \odot \sigma(\mathbf{\Theta_2} \star \mathcal{X} + \mathbf{c}),$$





是否将空间和时间卷积或注意力操作整合到一个模块中





(b) Graph neural networks for time series anomaly detection.

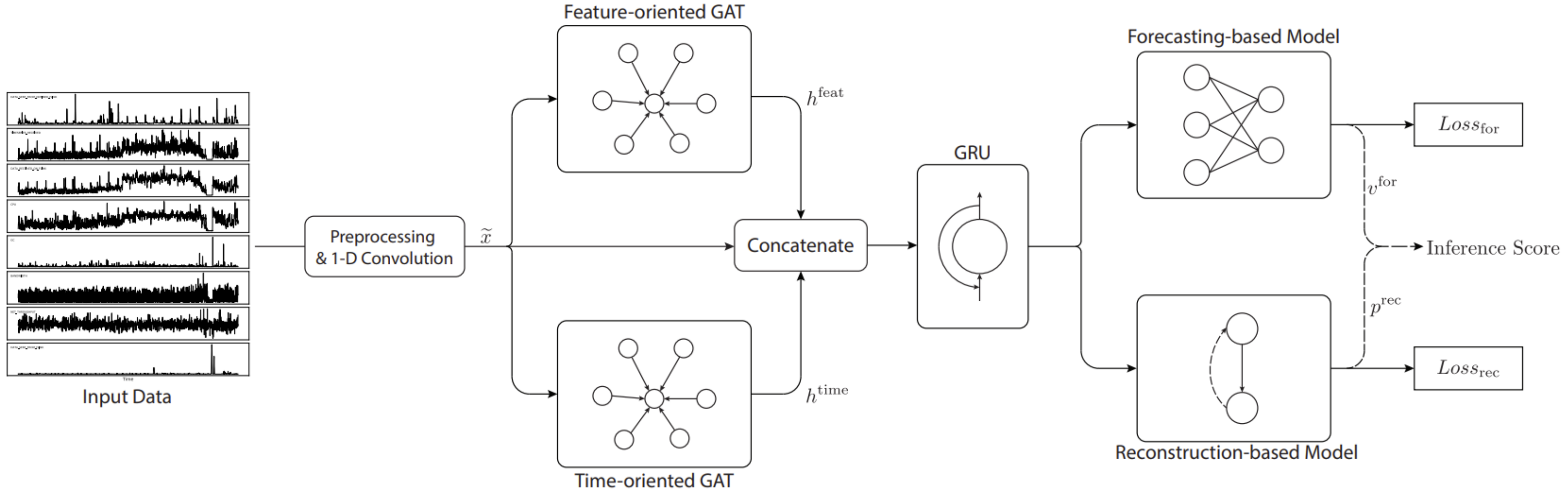


Fig. 2. The architecture of MTAD-GAT for multivariate time-series anomaly detection

Multivariate Time-series Anomaly Detection via Graph Attention Network, ICDM 2020

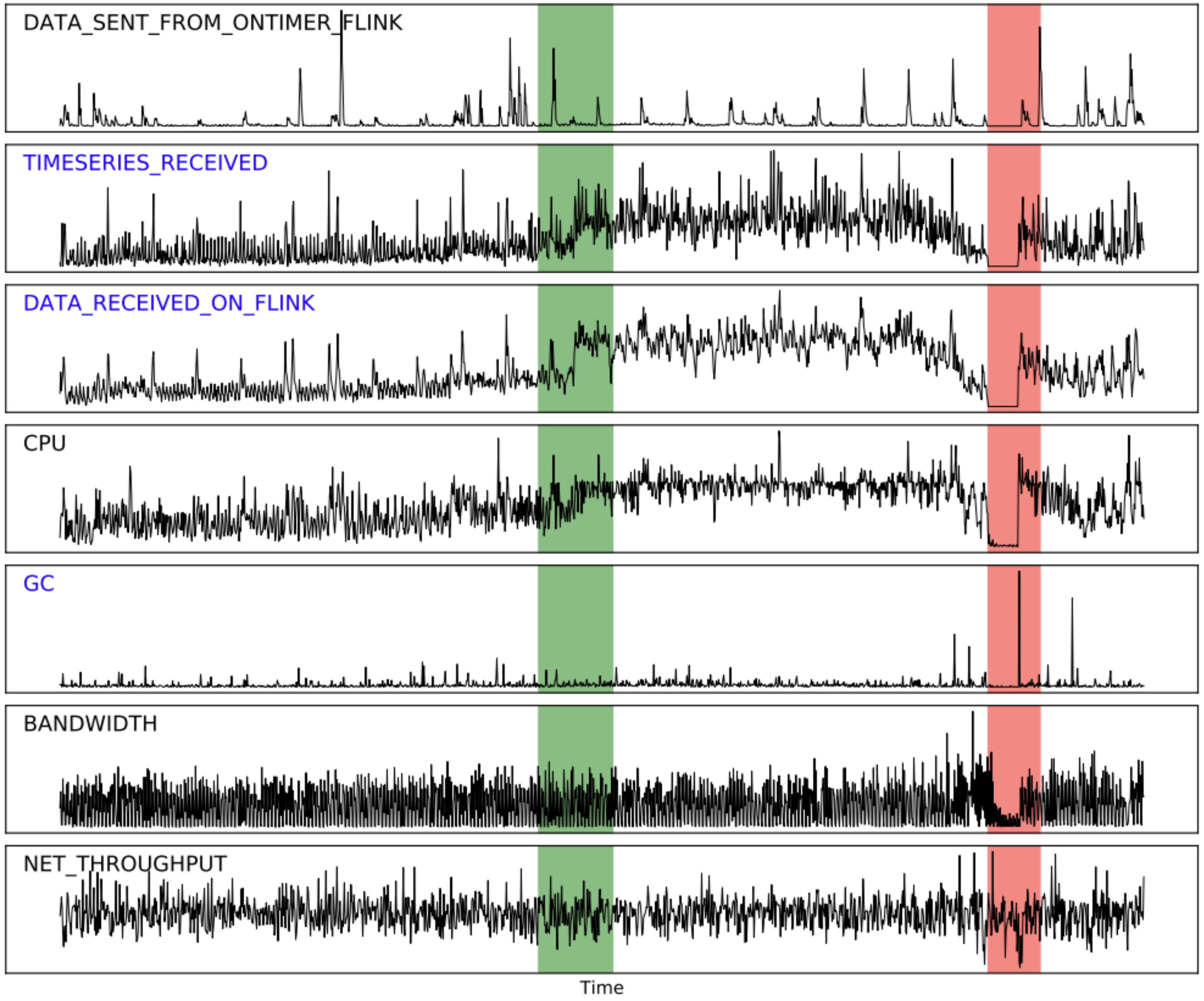


Fig. 1. An example of multivariate time-series input. Green indicates normal values and red indicates anomalies.

k个对象

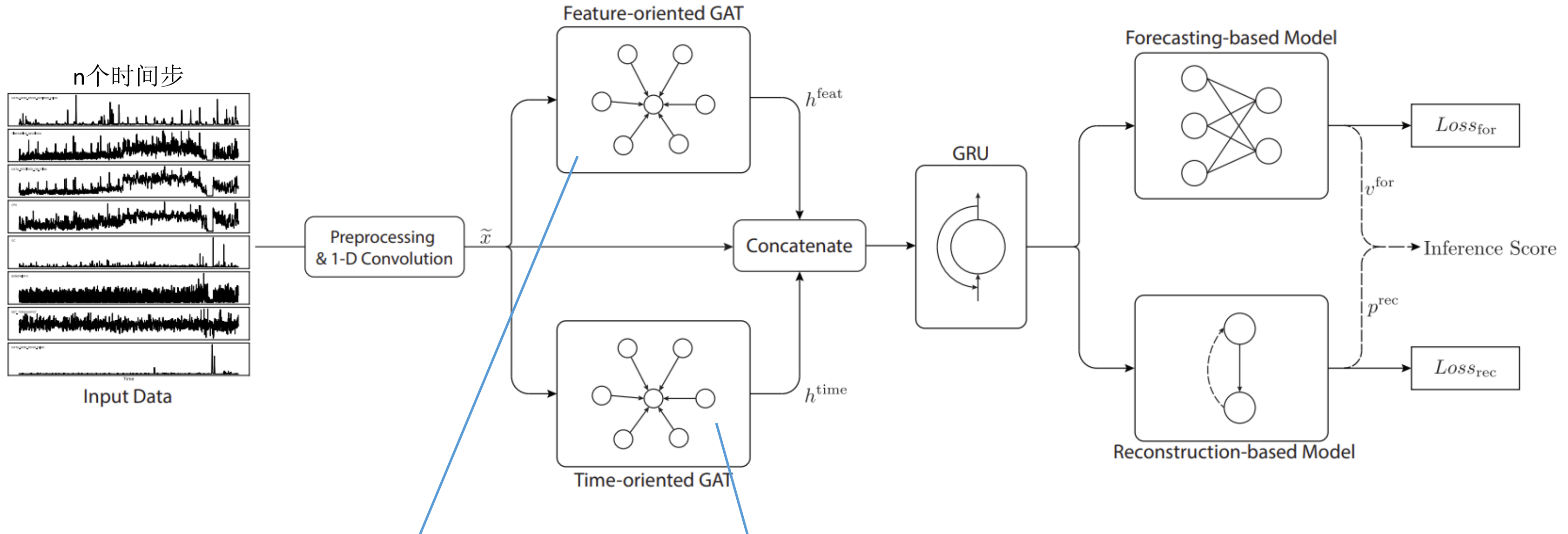
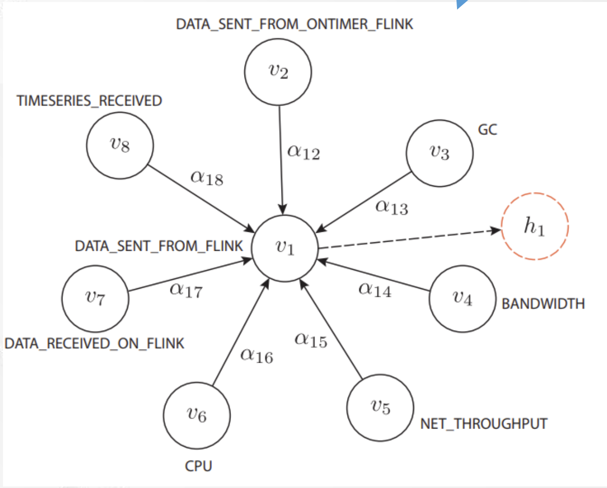
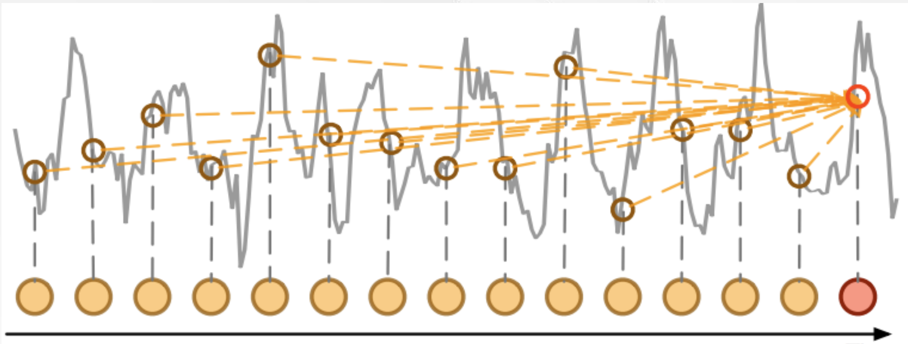


Fig. 2. The architecture of MTAD-GAT for multivariate time-series anomaly detection

考虑不同时间序列的相关性





(a) Full Attention

虽然名字叫Time-oriented GAT， 但实际更类似与Transformer

Concretely, a node x_t represents the feature vector at timestamp t, and its adjacent nodes include all other timestamps in the current sliding window. **This is much like a Transformer**

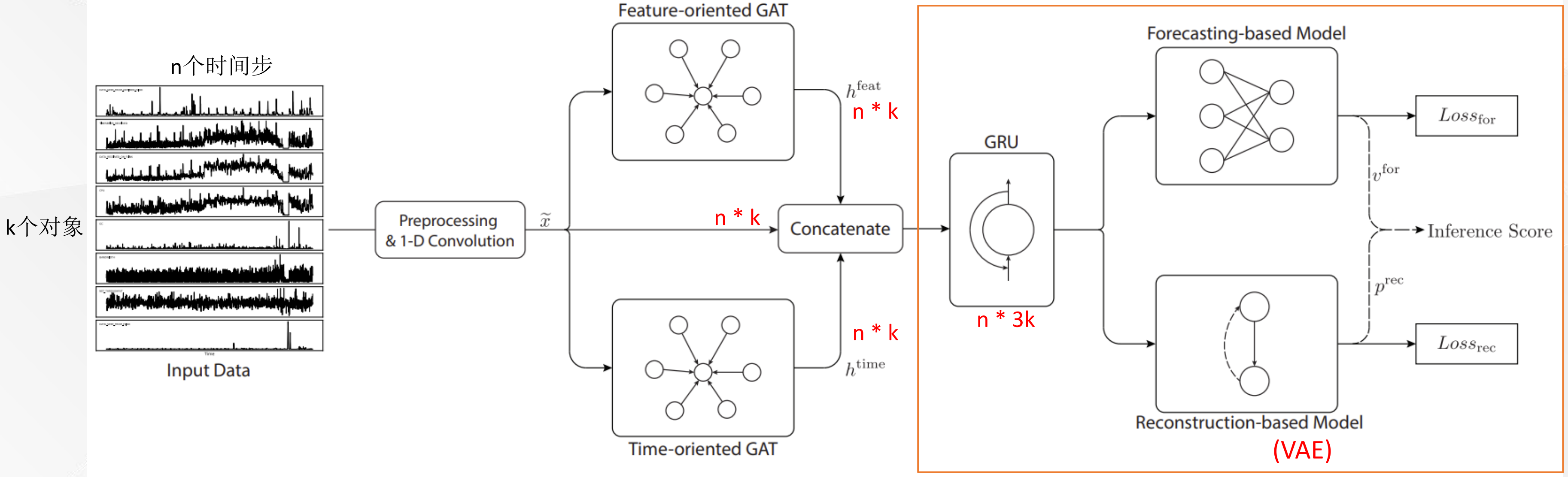


Fig. 2. The architecture of MTAD-GAT for multivariate time-series anomaly detection

$$Loss_{for} = \sqrt{\sum_{i=1}^{k}(x_{n,i} - \hat{x}_{n,i})^2}.$$

$$Loss_{rec} = -\mathrm{E}_{q_\phi(z|x)}[log p_\theta(x|z)] + D_{KL}(q_\phi(z|x)||p_\theta(z))$$

$$score = \sum_{i=1}^{k} s_i = \sum_{i=1}^{k} \frac{(\hat{x}_i - x_i)^2 + \gamma \times (1 - p_i)}{1 + \gamma}$$

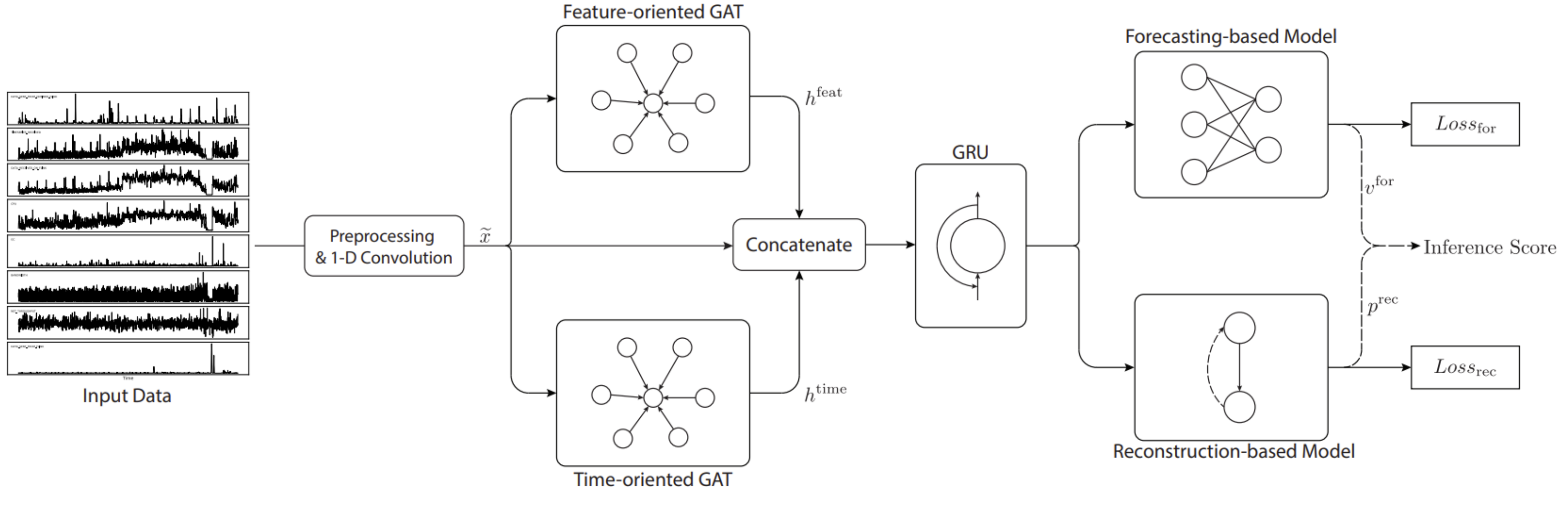
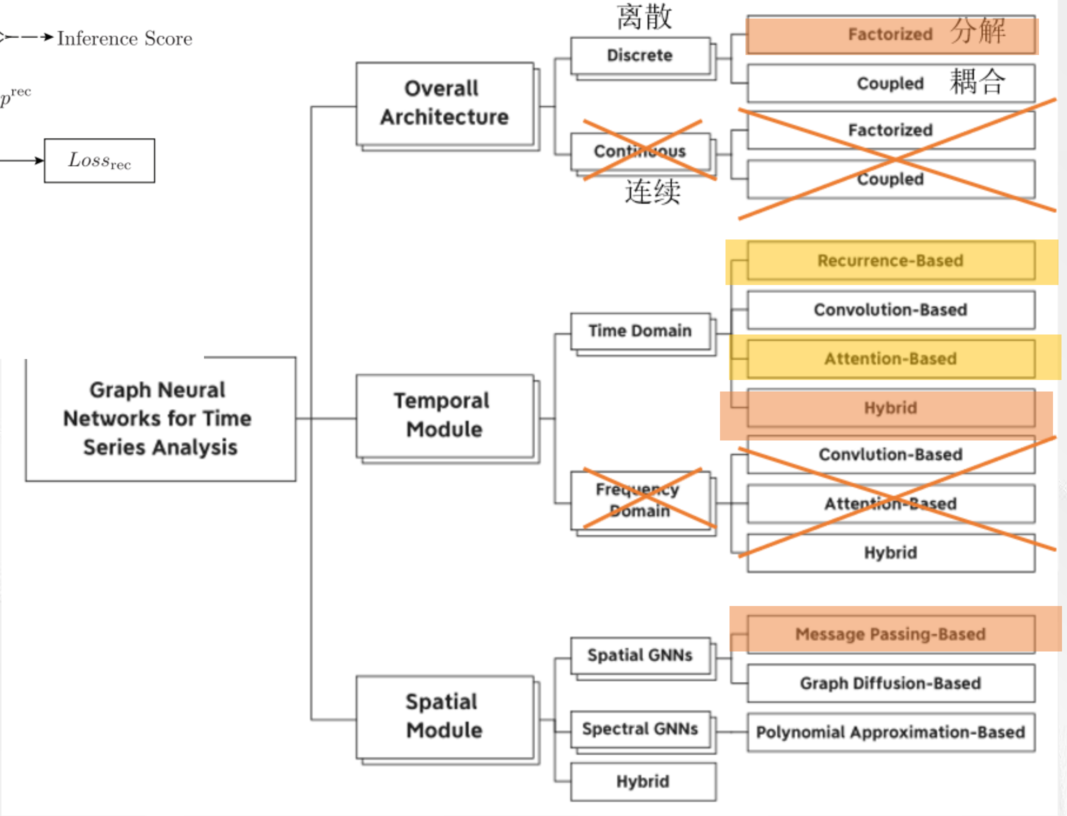
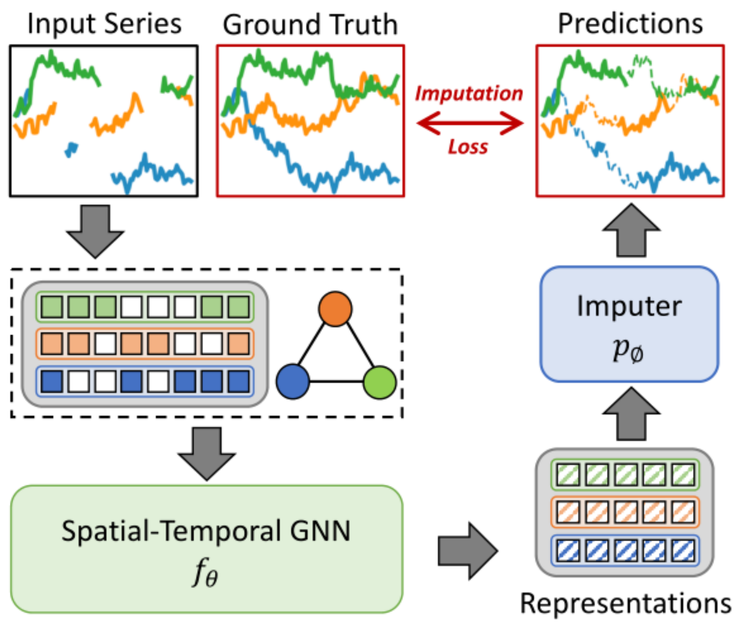


Fig. 2. The architecture of MTAD-GAT for multivariate time-series anomaly detection





(c) Graph neural networks for time series imputation.

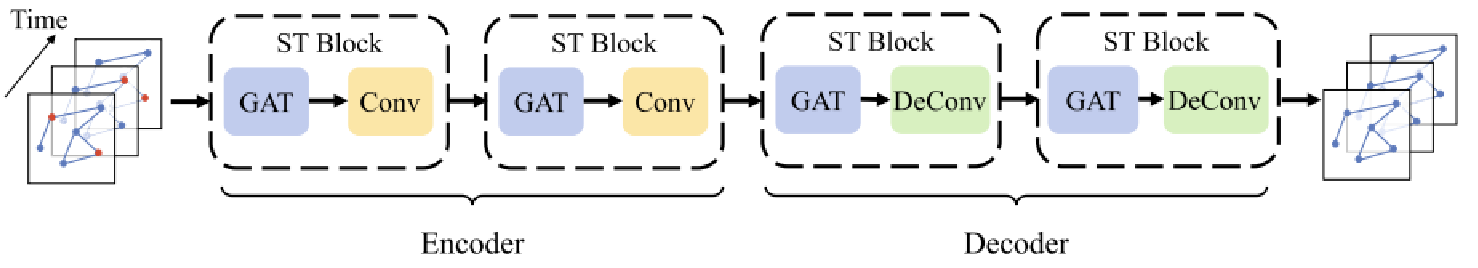


Fig. 2. Structure of the proposed graph attention convolution network.

GACN: Spatial-Temporal Traffic Data Imputation via Graph Attention Convolutional Network, ICANN 2021

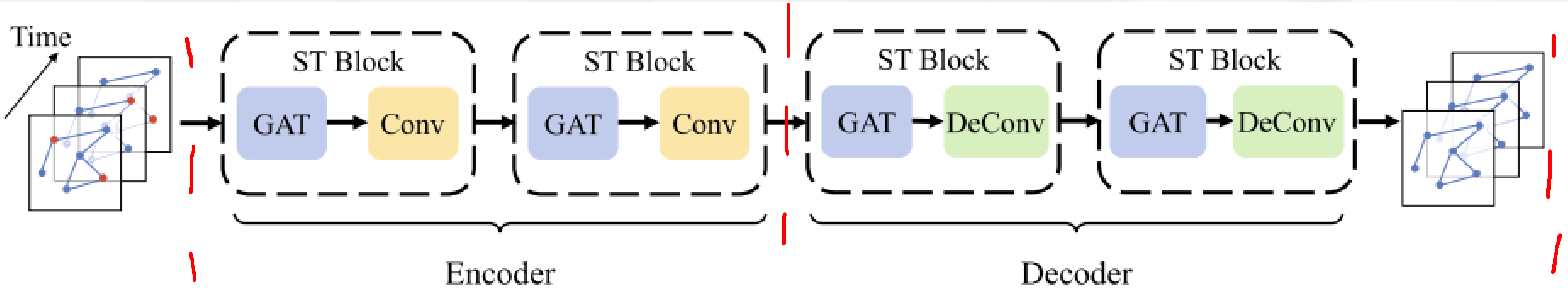


Fig. 2. Structure of the proposed graph attention convolution network.

x_1
x_2
x_3
x_n

x_1
x_n

loss

重构 || 填补

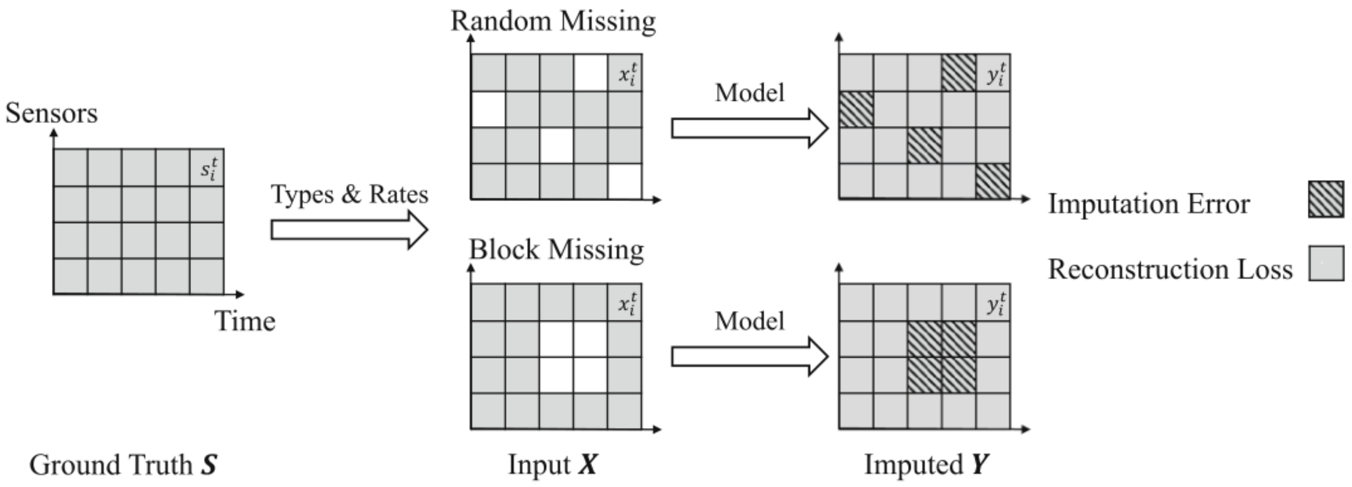


Fig. 1. Diagram of traffic data imputation problem.

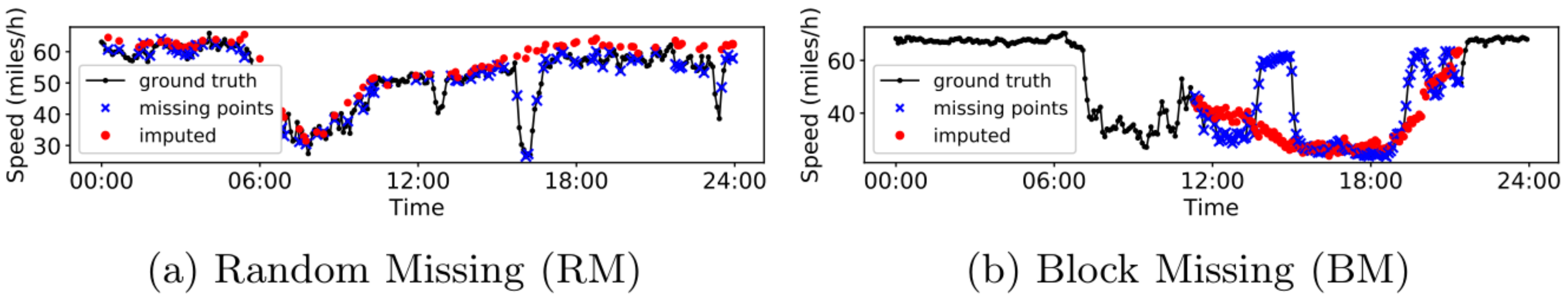


Fig. 4. Imputation results of the proposed GACN (30% missing rate).

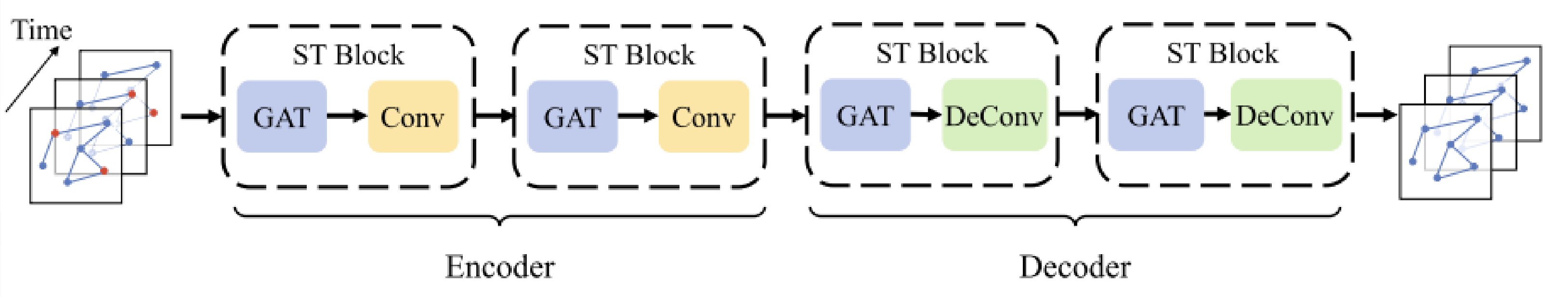
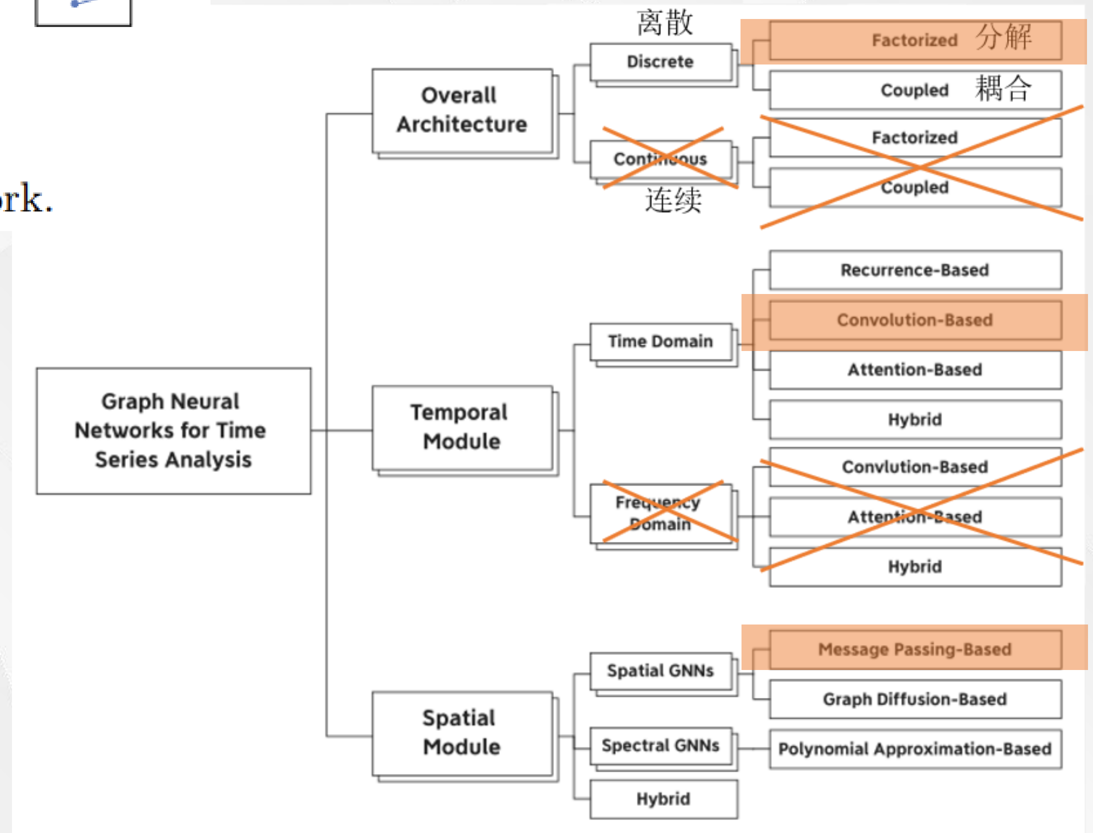
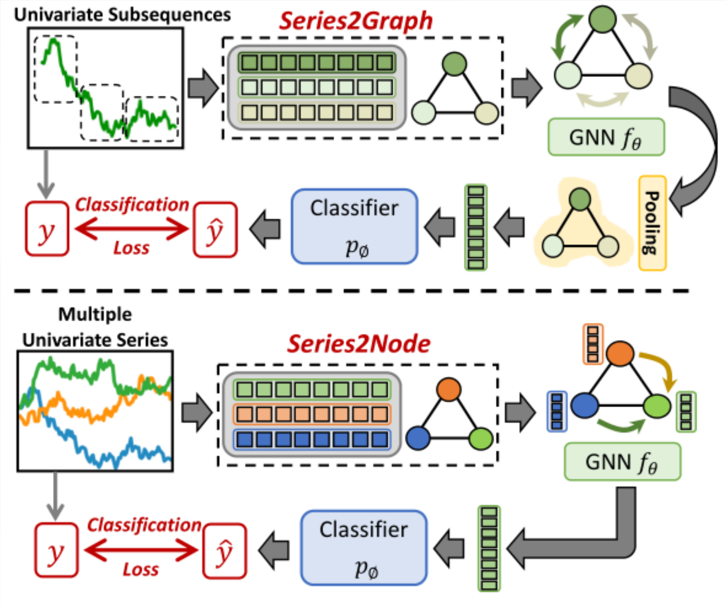


Fig. 2. Structure of the proposed graph attention convolution network.





(d) Graph neural networks for time series classification: Convert green series classification into a graph (top) or node (bottom) classification task.

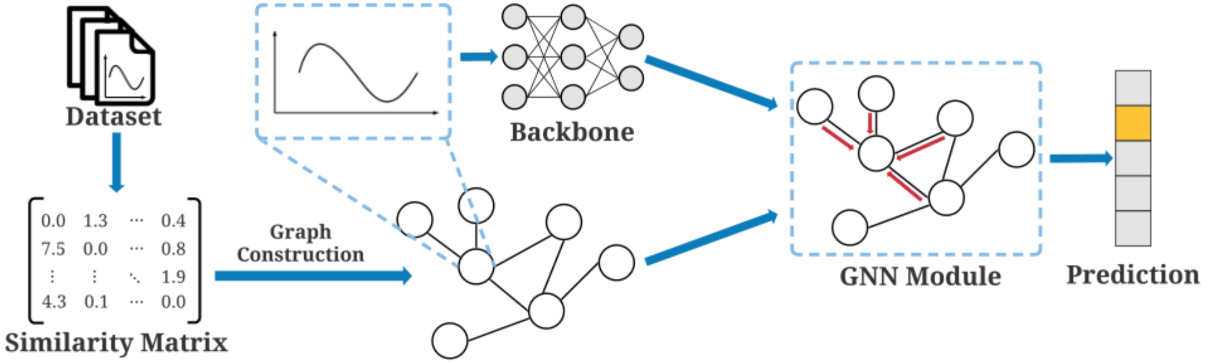


Figure 2: An overview of SimTSC framework. The graph is constructed based on the pair-wise similarities (e.g., DTW distances) of the time-series. Each time-series is processed by a backbone (e.g., ResNet) for feature extraction. The GNN module will aggregate the features and produce the final representation for classification.

SimTSC: Towards Similarity-Aware Time-Series Classification, SDM 2022

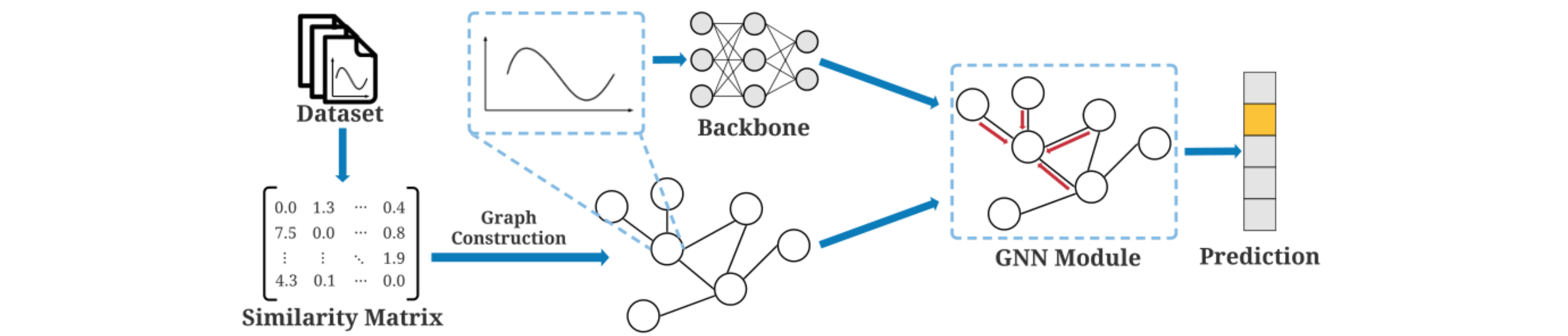


Figure 2: An overview of SimTSC framework. The graph is constructed based on the pair-wise similarities (e.g., DTW distances) of the time-series. Each time-series is processed by a backbone (e.g., ResNet) for feature extraction. The GNN module will aggregate the features and produce the final representations for classification.

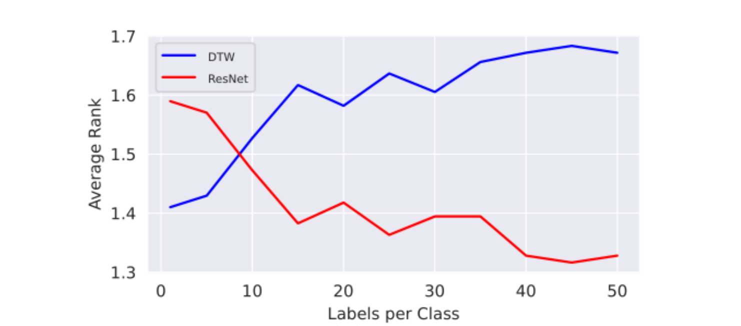
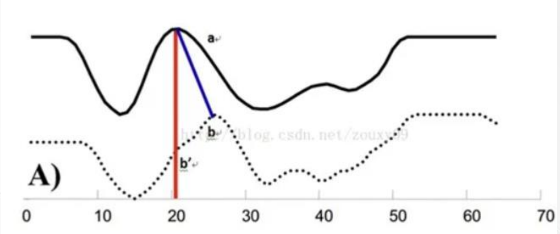
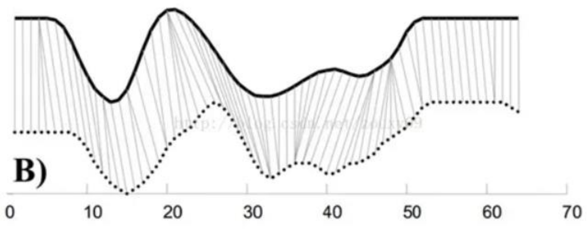


Figure 1: Average ranks (↓) of ResNet and DTW on the full 128 UCR datasets, where different numbers of labels per class is given (see Section 4.1 for more details).





核心概念：

DTW

Dynamic Time Wrap
（动态时间规整）

采用动态规划（dynamic programming）方法对时间序列进行**规整**，进而进行两序列的**相似度度量**

Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping, KDD 2012, Best Paper

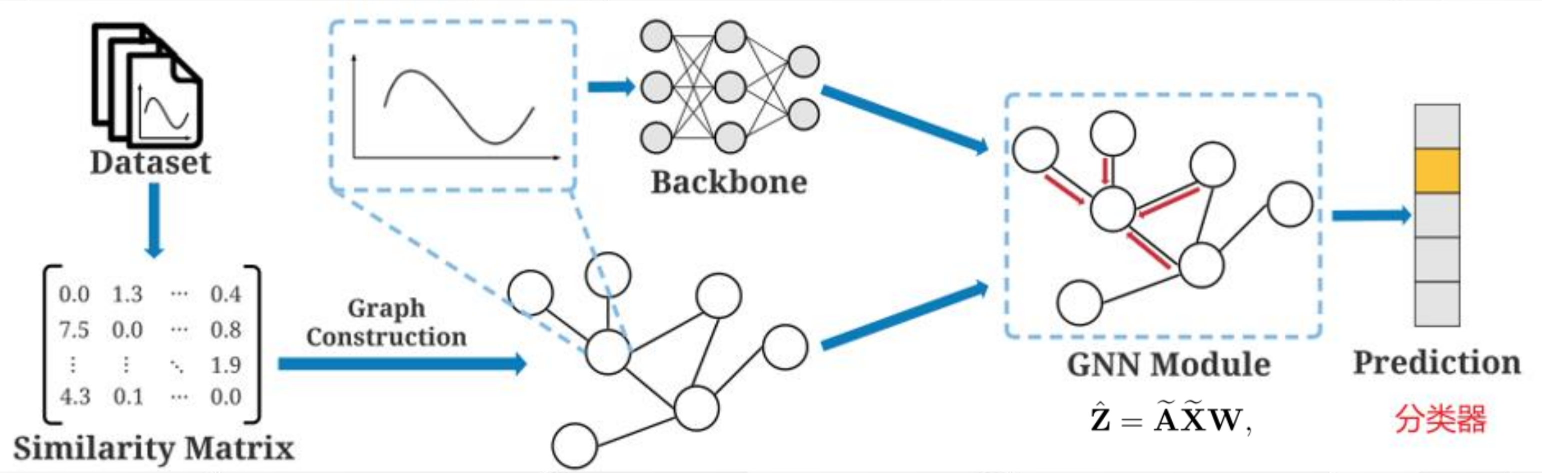
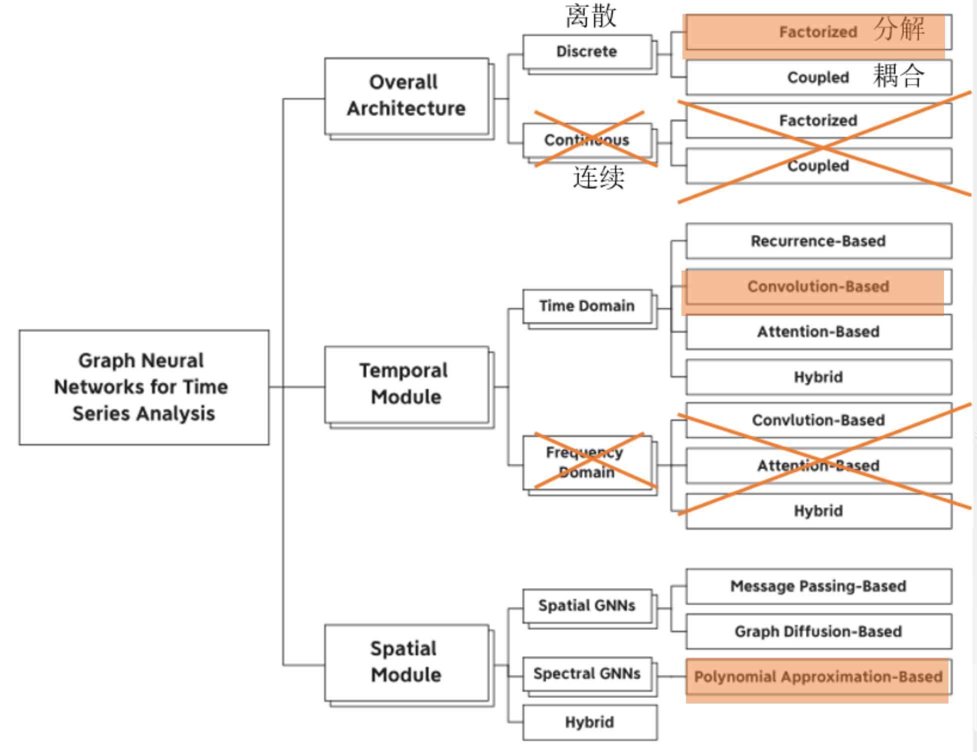
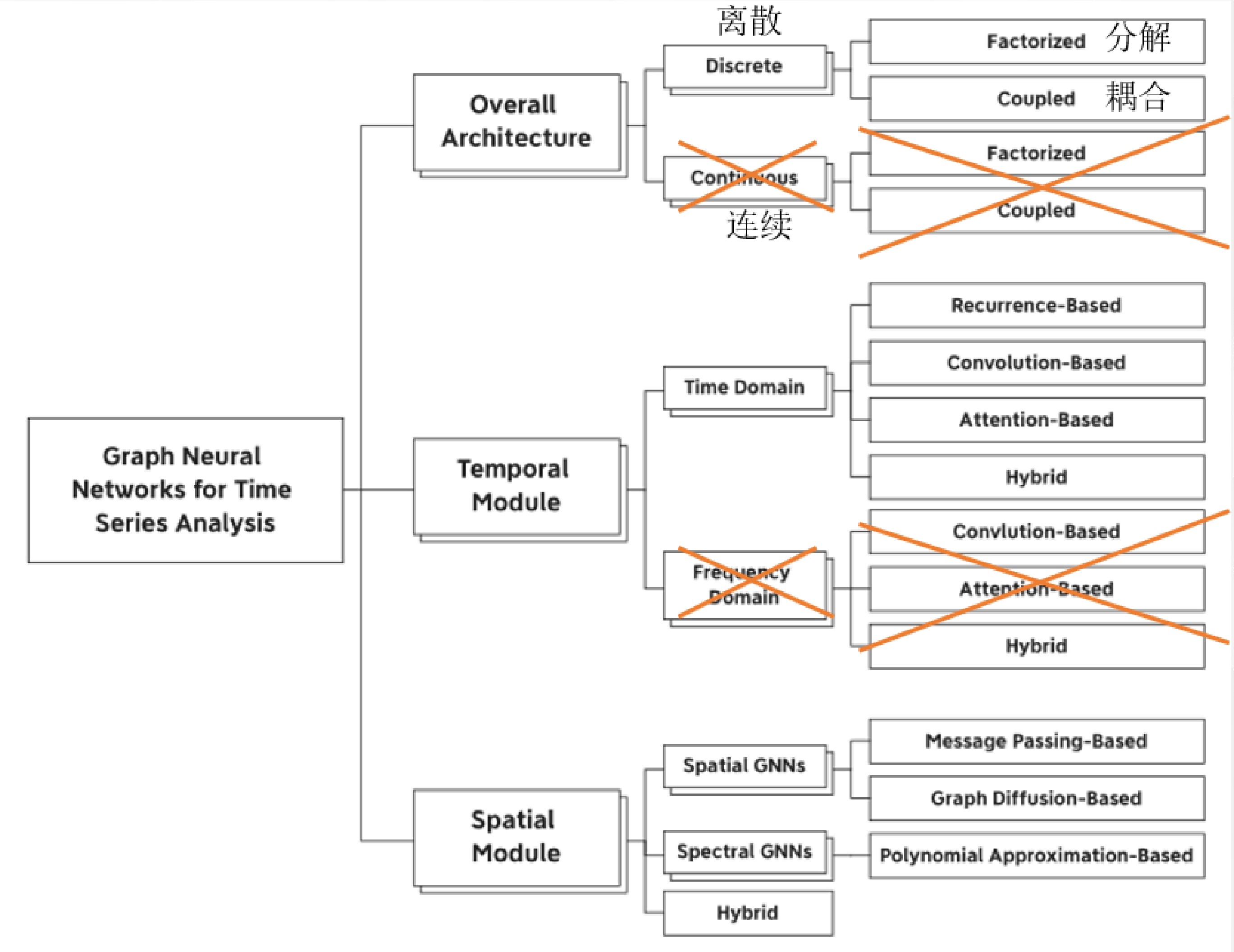


Figure 2 shows an overview of **Sim**ilarity-Aware **T**ime-**S**eries **C**lassification (SimTSC), which consists of three modules: (1) a graph construction module that connects the time-series based on a similarity measure (e.g., DTW), (2) a backbone that performs feature extraction with deep neural networks (e.g., ResNet), and (3) a GNN module that aggregates the features of neighboring time-series (e.g., GCN). The graph construction is unsupervised so that it can flexibly adapt to all the three settings defined in Section 2.1.



小结



Graph Wavenet, STGCN

DCRNN

RNN, LSTM, GRU

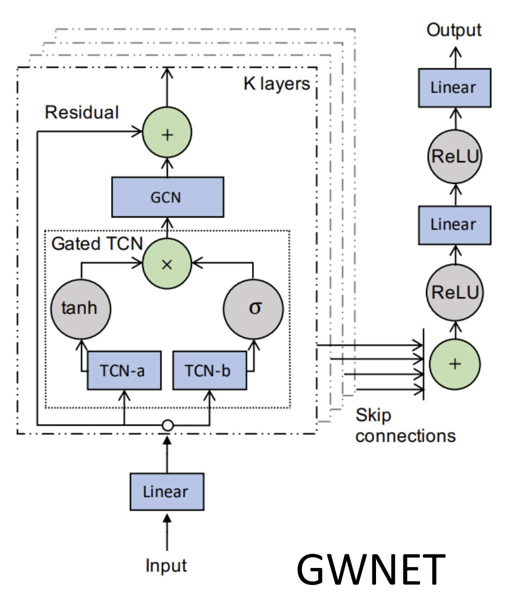
TCN

Transformer

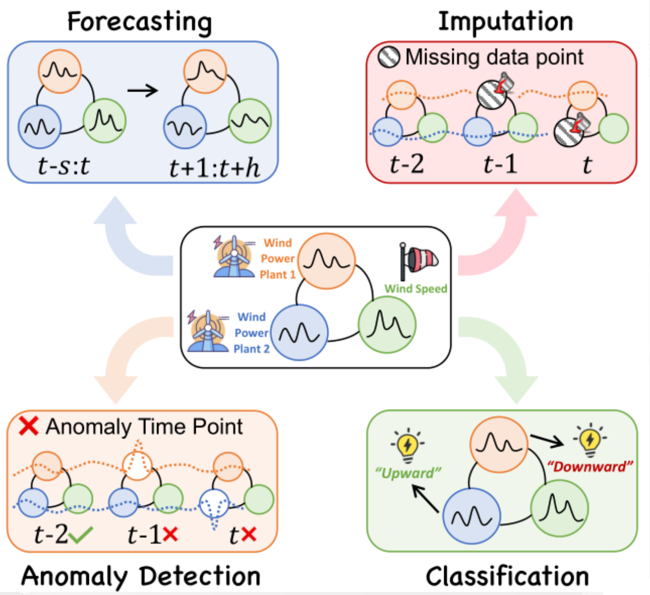
Graph SAGE, GAT

Diffusion CNN

GCN



GWNET



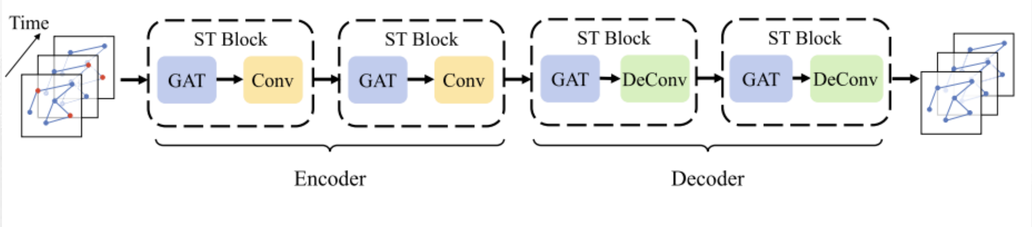


Fig. 2. Structure of the proposed graph attention convolution network.

GACN

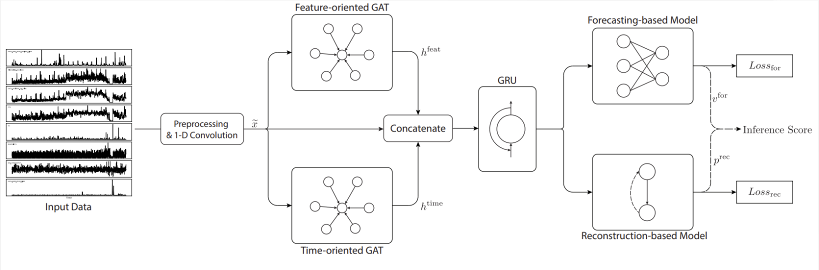
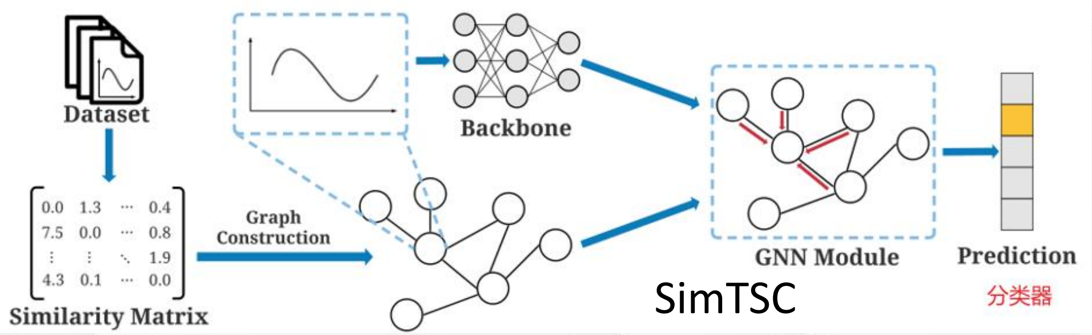


Fig. 2. The architecture of MTAD-GAT for multivariate time-series anomaly detection

MTAD-GAT



SimTSC

实验与展望

应用领域

- Smart Transportation
- On-Demand Services
- Environment & Sustainable Energy
- Internet-of-Things (IoT s)
- Healthcare
- Fraud Detection

未来方向

- Pre-training, Transfer Learning, and Large Models
- Robustness and Interpretability
- Uncertainty Quantification
- Privacy Enhancing
- Scalability
- AutoML and Automation

谢谢