

基于扩散模型的不完整数据下细粒度城市流量推断

郑雨豪, 王森章

(中南大学计算机学院, 湖南 长沙 410083)

摘要: 为了获取城市每个路段上精细的交通流数据, 需要部署大量的传感装置以及较密集的观测站, 这会增加日常运营与设备维护的成本。同时, 传统的交通流数据采集技术存在很多噪声和误差, 检测得到的数据结果并不能保证其真实可靠。因此, 如何利用粗粒度、混入噪声的传感器观测数据推断细粒度的城市交通流, 是一个重要的研究课题。针对上述问题, 提出了一种基于时空注意力的去噪扩散模型, 提供细粒度的城市交通基础数据, 以满足不同场景下的交通需求, 为交通规划与智能交通系统构建奠定基础。

关键词: 城市交通流; 细粒度推理; 时空注意力; 去噪扩散模型

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202330

Fine-grained urban flow inference based on diffusion models with incomplete data

ZHENG Yuhao, WANG Senzhang

School of Computer Science, Central South University, Changsha 410083, China

Abstract: To obtain detailed traffic flow data for each road segment of the city, it is necessary to deploy a large number of sensing devices and dense observation stations, which increases the costs of daily operations and equipment maintenance. At the same time, traditional traffic flow survey techniques are noisy and inaccurate, and the reliability of the detected data results is not guaranteed. Therefore, inferring fine-grained urban traffic flow based on coarse-grained and noise-inclusive sensor observations has become an important research topic. To address the above problems, we proposed a denoising diffusion model based on spatio-temporal attention, with the intention of providing fine-grained urban traffic base data in different scenarios of traffic demand, and laying the foundation for traffic planning and intelligent transportation system construction.

Key words: urban traffic flow, fine-grained inference, spatio-temporal attention, denoising diffusion model

0 引言

《数字交通“十四五”发展规划》为我国交通运输行业发展指明了方向, 交通要全方位向“数”融合。新一代信息技术可以监控整个交通的运行情况, 可以充分挖掘和利用信息数据的价值, 盘活现

有数据, 赋能交通部门的管理与决策^[1]。随着智能交通信息的发展, 其对更细粒度的交通流数据的分析需求越来越高。精准的交通流分析预测可以为紧急救援、日常出行等提供更实时、可靠的路径规划。

细粒度的城市流对于与智慧城市相关的应用, 如城市规划、城市更新和交通管理等, 都至关重要

收稿日期: 2023-06-01; 修回日期: 2023-07-05

通信作者: 王森章, szwang@csu.edu.cn

基金项目: 国家自然科学基金项目 (No.62172443); 湖南省自然科学基金项目 (No.2022JJ30053)

Foundation Items: The National Natural Science Foundation of China (No.62172443), The Natural Science Foundation of Hunan Province (No.2022JJ30053)

要^[2]。为了获取重要路段的交通流参数,需要大量的传感装置以及密集分布的控制站、观测站,这会增加日常运营与设备维护的成本。因此,基于粗粒度交通流数据的细粒度推断引起了广泛关注。然而,现有的传感器检测技术受到自然环境中温度、湿度、天气等不确定因素的影响,检测得到的数据结果会有一定程度的缺失和缺陷,即粗粒度交通流数据并不完整和可靠。目前大部分传统的推断方法只考虑了完整数据的情况,并未考虑到数据缺失的情况,小部分考虑了数据有噪声的情景,但并未考虑全面且效果欠佳。因此,本文将着重考虑数据缺失时,如何利用粗粒度数据来推断城市交通细粒度数据。此外,由于本文还考虑了各种交通数据的复杂情况,因此提出的方法有更好的普适性和鲁棒性。本文的主要贡献如下。

(1) 为解决粗粒度向细粒度转化时原始数据缺失的问题,提出了基于时空注意力的去噪扩散模型 STDiff。

(2) 首次将去噪扩散概率模型(DDPM)引入城市粗细粒度交通流转化任务中,通过时空特征提取网络捕获交通流的时空信息,通过特征提取网络提取天气、温度等城市环境语义信息,提升了模型的泛化性。

(3) 在TaxiBJ数据集上进行了模型效果评估,并在完整数据下的细粒度流量推断、不完整数据下的细粒度流量推断这两种情况下,与现有基线模型进行了效果对比。结果表明,STDiff不仅在基于完整交通流数据的细粒度推断上优于已有的最优模型,在不完整交通流的场景下也取得了远高于传统模型的性能。

1 国内外相关工作

细粒度城市流量监测系统是智慧城市信息基础设施的重要组成部分,为城市规划和交通管理等各种应用提供基础。为了获得细粒度的城市流量观测数据,需要在城市的不同区域部署大量的传感器,导致日常运维开销巨大。随着智慧城市在全球范围内的快速发展,人力和能源成本将成为地球进一步智能化的阻碍因素。因此,如何基于稀疏、粗粒度的传感器观测数据推断出细粒度的城市流,成为一个重要的研究问题,近年来也引发越来越多研究者的关注。

传统上,双线性、双三次和最近邻插值^[3]等统

计方法被广泛应用于细粒度城市流量推断。Xu和Zhu^[4]设计了一种用于城市流量推断的张量分解方法。Chen等人^[5]提出了一种用于检测细粒度城市事件的张量协同分解模型。但是,这些方法只考虑了城市流量数据,而忽略了天气、节假日等外部环境特征,因此性能不理想。近年来,计算机视觉中超分辨率算法取得巨大成功^[6],研究人员试图将细粒度的城市流量推断建模为一个时空数据超分辨率问题^[7],因为城市单元区域的城市流量可以被视为空间地图图像。SRCNN^[8]首次将双三次插值方法与卷积神经网络结合,实现了图像的超分辨率。为了解决SRCNN接收域小、收敛速度慢的问题,Kim等人^[9]采用深度卷积网络和残差学习策略提出了VDSR。UrbanFM^[10]首次采用数据超分辨率方法进行城市流量推断。与图像超分辨率不同,UrbnFM包含了一个提取外部特征(如天气、温度和假日)的外部因子融合网络和一个推理网络。为了进一步提高UrbanFM在高分辨率下的性能,UrbnPy^[11]采用了一个由多个组件组成的金字塔结构的网络模型,其中每个组件都是一个小尺度的原子上采样器,将原始任务分解为多个子任务。Liang等人^[12]设计了一个名为DeepLGR的通用框架,用于全市人群流分析,解决不完全粗粒度数据的细粒度流量推理问题,DeepLGR包含局部特征提取模块、全局上下文模块和特定区域的预测器。

然而,上述方法的局限性在于未考虑数据缺失对数据流推断的影响。为了弥补以往方法的不足,在多任务学习框架下,MT-CSR^[13]同时完成不完整城市交通流补全和细粒度交通流超分辨率推断的任务。它考虑了数据中的局部地理和全局语义相关性,采用了联合训练和端到端优化的策略。但MT-CSR未考虑到交通城市流的丰富语义,效果并不显著。

2 术语定义和问题描述

下面对本文将要使用的一些术语进行定义。

定义1 网格地区:根据经纬度把一个城市划分为 $I \times J$ 网格图,网格图中所有网格区域表示为 $R = r_{1,1}, \dots, r_{m,n}, \dots, r_{I,J}$,其中 $r_{m,n}$ 是网格图的第 m 行和第 n 列的单元区域。

定义2 交通流量图:设 T 是城市交通流轨迹的集合。给定网格区域 $r_{m,n}$,在时间 t 城市交通流的流入和流出被定义为:

$$X_{in,m,n}^t = \sum_{f \in T} \{f^{t-1} \notin r_{m,n} \cap f^t \in r_{m,n}\} \quad (1)$$

$$X_{out,m,n}^t = \sum_{f \in T} \{f^{t-1} \in r_{m,n} \cap f^t \notin r_{m,n}\} \quad (2)$$

其中 $f \in T$ 表示城市交通流, $f^{t-1} \notin r_{m,n}$ 表示 f 在时间 $t-1$ 不在区域 $r_{m,n}$ 内, $f^t \in r_{m,n}$ 表示 f 在时间 t 在区域 $r_{m,n}$ 内。在时间 t 所有区域的交通流入和流出被表示为城市交通流量张量 $X^t \in R^{2 \times I \times J}$ 。

定义3 粗粒度和细粒度城市交通流量图：粗粒度的城市交通流量图是通过交通传感器采集的原始交通数据构建的。给定放大因子 N , 在细粒度流中对 $N \times N$ 范围内的邻近网格区域内的交通流量进行累计, 即可获得粗粒度流。譬如, 图1显示的是 $N=2$ 时粗粒度和细粒度城市流之间的关系, 每个粗粒度城市交通流由 2×2 个较小的城市流组成。

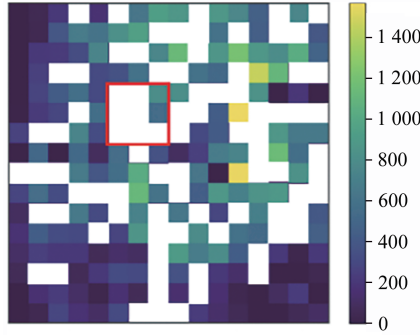
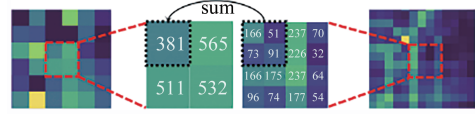


图1 粗粒度和细粒度城市流之间的关系

将时间 t 上的粗粒度和细粒度城市流量图分别表示为 $X_{cg}^t \in R^{2 \times I \times J}$ 和 $X_{fg}^t \in R^{2 \times NI \times NJ}$, 其中 NI 、 NJ 分别表示细粒度下的流量图的长和宽。



定义4 不完整的粗粒度城市空间流量图：由于传感器可能出现器械损坏、传输故障等问题, 获取的粗粒度交通流数据并不完整。将时间 t 上, 数据不完整情况下的粗粒度城市交通流量图表示为 $X_{cg,un}^t \in R^{2 \times I \times J}$ 。如图2所示, 本文希望通过左边不完整的粗粒度城市流获得右边完整细粒度的城市流。

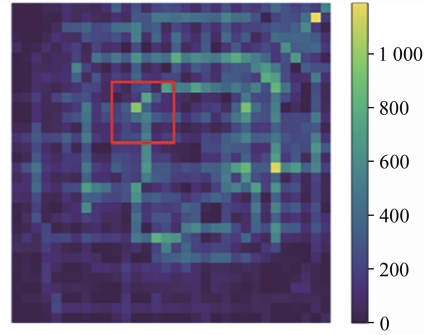


图2 不完整的粗粒度和细粒度城市流

问题描述：给定放大因子 N , 根据时间 t 和前 k 个时间的不完整粗粒度城市交通流数据 $X_{cg,un}^{t-k}, X_{cg,un}^{t-k+1}, \dots, X_{cg,un}^t \in R^{2 \times I \times J}$, 以及城市的天气、节假日、道路结构等外部语义和环境信息, 推断出一个完整的细粒度城市流空间图 $X_{fg}^t \in R^{2 \times NI \times NJ}$ 。

3 方法的提出

针对上述问题, 本文提出了一种基于时空注意力的去噪扩散模型 STDiff, 对不完整且包含噪声的城市交通流数据进行细粒度交通流推理。

STDiff 是一个以 U-Net 为骨干架构的条件 DDPM。DDPM 包含前向过程和反向过程各 N 个步骤。前向过程是预定义好的, 而反向过程是可以学习的。如图3所示, 在反向过程的每个步骤中, STDiff 包含 STTNet、Extractor 和 Embedding 模块。STTNet 能够更有效地捕捉城市流量数据复杂、动态的时空相关性。Extractor 用于提取外部语义和环境信息, 如外部因素和土地属性。Embedding 用于

嵌入 DDPM 的步骤位置。考虑到城市交通流数据在扩散过程中的独特性, 本文将外部因素和土地属性这两类特征灵活地整合到 U-Net 中, 并提出了一种改进的条件 DDPM 反向扩散策略。在每个去噪步骤中, 将 STTNet 提取的特征图和粗粒度流量图叠加到流量图输入上, 扩展反向去噪过程的输入通道, U-Net、STTNet 和 Extractor 以端到端的方式进行联合训练。

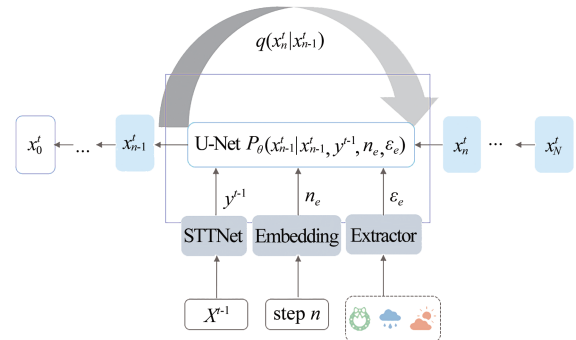


图3 模型整体框架

3.1 DDPM

深度学习在生成任务中显示出巨大的潜力，生成模型可根据某些隐含的参数随机生成观察结果。然而，目前的生成模型面临四大问题，即生成过程缓慢、数据处理类型单一、似然优化、模型降维。近些年，许多优化工作被陆续提出。

下面从去噪扩散模型的正向扩散、反向扩散、内部的U-Net结构以及扩展的条件信息嵌入来详细介绍该模型。

3.1.1 正向扩散过程

给出一个从真实数据分布中取样的数据点 $x_0 \sim q(x)$ ，DDPM 定义了一个前向扩散过程，通过 N 个步骤分步向样本添加少量的高斯噪声，产生一连串的噪声样本 $x_0, \dots, x_N$ ，步长由一个方差表 $\{\beta_n \in (0, 1)\}_{n=1}^N$ 来控制（为了方便数据表示，这里的过程数学描述没有添加表示时间的上标 t ）。

从第 $n-1$ 步到第 n 步的采样过程如下：

$$q(x_n | x_{n-1}) = \mathcal{N}(x_n; \sqrt{1 - \beta_n} x_{n-1}, \beta_n \mathbf{I}) \quad (3)$$

从本身的样本 x_0 采样得到 x_N ：

$$q(x_N | x_0) = \prod_{n=1}^N q(x_n | x_{n-1}) \quad (4)$$

随着 n 的变大，数据样本 x_0 逐渐失去可识别的特征。当 $N \rightarrow \infty$ 时， x_N 相当于一个各向同性的高斯分布。

如图4所示，从左往右是一个正向加噪过程，可以在任意的时间步长对 x_n 进行采样。

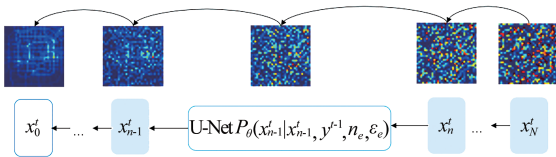


图4 正向加噪/反向去噪的马尔可夫链

在任意第 n 步中，用重参数化的技巧，以封闭的形式取样。令 $\alpha_n = 1 - \beta_n$ 以及 $\dot{\alpha}_n = \prod_{i=1}^n \alpha_i$ ，则有：

$$\begin{aligned} x_n &= \sqrt{\alpha_n} x_{n-1} + \sqrt{1 - \alpha_n} \varepsilon_{n-1} \\ &= \sqrt{\dot{\alpha}_n} x_0 + \sqrt{1 - \dot{\alpha}_n} \varepsilon \end{aligned} \quad (5)$$

所以可得：

$$q(x_n | x_0) = \mathcal{N}(x_n; \sqrt{\dot{\alpha}_n} x_0, (1 - \dot{\alpha}_n) \mathbf{I}) \quad (6)$$

3.1.2 反向扩散过程

若逆转上述过程并从 $q(x_{n-1} | x_n)$ 中取样，就能从高斯噪声输入 $x_N \sim \mathcal{N}(0, \mathbf{I})$ 中重现真实样本。注

意，如果 β_n 足够小， $q(x_{n-1} | x_n)$ 也将是高斯分布，但是不能轻易估计 $q(x_{n-1} | x_n)$ ，因为它需要使用整个数据集。因此，需要学习一个模型 p_θ 来近似条件概率，以便运行反向扩散过程。

反向过程的采样计算式如下：

$$p_\theta(x_{0:N}) = p(x_N) \prod_{t=1}^T p_\theta(x_{n-1} | x_n) p_\theta(x_{n-1} | x_n) = \mathcal{N}(x_{n-1}; \mu_\theta(x_n, n), \Sigma_\theta(x_n, n)) \quad (7)$$

图4从右往左是一个反向去噪过程，通过模型训练，学习正向传播每一步的均值 $\mu_\theta(x_n, n)$ 和协方差 $\Sigma_\theta(x_n, n)$ ，从而反向去噪。

值得注意的是，当以 x_0 为条件时，反向去噪的条件概率是可控的：

$$q(x_{n-1} | x_n, x_0) = \mathcal{N}(x_{n-1}; \tilde{\mu}(x_n, x_0), \tilde{\beta}_n \mathbf{I}) \quad (8)$$

使用贝叶斯法则，可以得到：

$$\begin{aligned} q(x_{n-1} | x_n, x_0) &= q(x_n | x_{n-1}, x_0) \frac{q(x_{n-1} | x_0)}{q(x_n | x_0)} \\ &\propto \exp\left(-\frac{1}{2} \left(\frac{(x_n - \sqrt{\alpha_n} x_{n-1})^2}{\beta_n} + \frac{(x_{n-1} - \sqrt{\dot{\alpha}_{n-1}} x_0)^2}{1 - \dot{\alpha}_{n-1}} - \frac{(x_n - \sqrt{\dot{\alpha}_n} x_0)^2}{1 - \dot{\alpha}_n} \right)\right) \\ &= \exp\left(-\frac{1}{2} \left(\left(\frac{\alpha_n}{\beta_n} + \frac{1}{1 - \dot{\alpha}_{n-1}} \right) x_{n-1}^2 - \left(\frac{2\sqrt{\alpha_n}}{\beta_n} x_n + \frac{2\sqrt{\dot{\alpha}_n - 1}}{1 - \dot{\alpha}_{n-1}} x_0 \right) x_{n-1} + C(x_n, x_1) \right)\right) \end{aligned} \quad (9)$$

其中， $C(x_n, x_1)$ 是一些不涉及 x_{n-1} 的函数，在计算式的化简过程中可以省略一些系数的细节，按照式 (9)，通过对比标准的高斯分布密度函数得到想要的参数。

按照标准的高斯密度函数，均值和方差可以被参数化，结果如下：

$$\tilde{\mu}_n = \frac{1}{\sqrt{\alpha_n}} \left(x_n - \frac{1 - \alpha_n}{\sqrt{1 - \dot{\alpha}_n}} \varepsilon_n \right) \quad (10)$$

$$\tilde{\beta}_n = \frac{1 - \dot{\alpha}_{n-1}}{1 - \dot{\alpha}_n} \cdot \beta_n \quad (11)$$

反向过程的均值和方差被参数化之后，通过构建目标函数来训练模型，学习正向加噪过程的对应参数，从而达到去噪的效果。

训练所用的损失项 L_n 的参数化是为了最小化

均值 $\tilde{\mu}$ 与正向过程的差异：

$$\begin{aligned} L_n &= \mathbb{E}_{x_0, \varepsilon} \left[\frac{1}{2 \|\Sigma(x_n, n)\|_2^2} \|\tilde{\mu}_n(x_n, x_0) - \mu_\theta(x_n, n)\|^2 \right] \\ &= \mathbb{E}_{x_0, \varepsilon} \left[\frac{(1 - \alpha_n)^2}{2 \alpha_n (1 - \dot{\alpha}_n) \|\Sigma_\theta\|_2^2} \|\varepsilon_n - \varepsilon_\theta(\sqrt{\dot{\alpha}_n} x_0 + \sqrt{1 - \dot{\alpha}_n} \varepsilon_n, n)\|^2 \right] \end{aligned} \quad (12)$$

Ho 等人^[14]发现，用忽略系数的简化方法训练扩散模型效果更好，所以对训练时的目标函数进行了微调：

$$\begin{aligned} L_n^{\text{simple}} &= \mathbb{E}_{n \sim [1, N], x_0, \varepsilon_n} \left[\|\varepsilon_n - \varepsilon_\theta(x_n, n)\|^2 \right] \\ &= \mathbb{E}_{n \sim [1, N], x_0, \varepsilon_n} \left[\|\varepsilon_n - \varepsilon_\theta(\sqrt{\dot{\alpha}_n} x_0 + \sqrt{1 - \dot{\alpha}_n} \varepsilon_n, n)\|^2 \right] \end{aligned} \quad (13)$$

从本质上来说，简化前后只是目标函数的形式略有不同，其实都是对正向过程关键参数进行学习，模仿加噪过程，从而实现有效的去噪。

3.1.3 U-Net 结构

U-Net 包含编码器和解码器，其中编码器是一个典型的卷积神经网络，用于提取输入图像的特征；解码器则通过上采样操作将编码器中提取的特征图还原为原始图像的大小，还能将特征图用于生成分割掩码。

U-Net 的特殊之处在于，编码器和解码器间存在一个跳跃连接的结构，允许解码器中的每一层访问编码器中对应层的信息，这有助于解码器更好地还原原始图像的细节信息。

3.2 STTNet 时空特征提取网络

结合交通流推断任务的具体场景，在第 3.1 节的基础上进一步考虑扩散模型的有条件生成变体，即图 3 中的 $P_\theta(x_{n-1}^t | x_{n-1}^t, y^{t-1}, n_e, \varepsilon_e)$ ，其中 n_e 表示扩散步数的嵌入， y^{t-1} 和 ε_e 是引导信息。 y^{t-1} 包含凝聚了长短时序信息的时序特征， ε_e 包含城市的所有外部因素。本文设计了特定的时空特征提取网络和外部因素提取网络，提取了交通问题所涉及的时序信息、温度、天气、道路等特征，具有很高的泛化性。

结合实际可知，某时刻同一条路段的流量规律与每周同一时刻该路段流量具有相似的模式，与近期该路段的流量也具有较强的相关性。综合考虑长程信息和短程信息，本文采用 STTNet 时空特征提取网络捕捉长短期的时空特征信息，为细粒度数据流推断提供进一步的指导。

如图 5 所示，STTNet 由 Pre-Conv Block 和 Vision Transformer (Vit) 组成，用于学习局部和全局上下文中的空间相关性。此外，将历史数据作为基础预测，直接连接到输出。

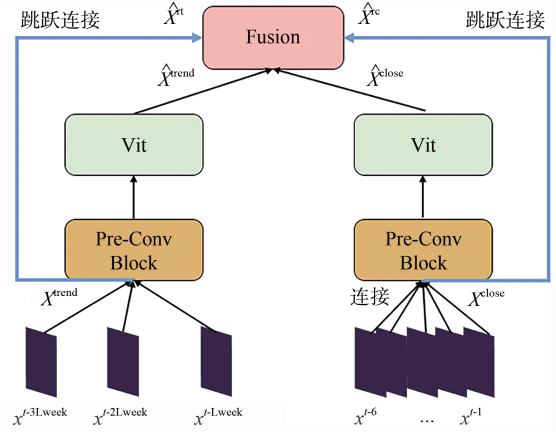


图5 STTNet网络框架

STTNet 采用对称结构来处理长程时序数据 X^{trend} 和短程时序数据 X^{close} 。长程时序数据 X^{trend} 包括当前时间 t 对应的前 3 周相同时间的流量图，短程时序数据 X^{close} 包括当前时间 t 对应前 6 天时间（近一周）的流量图。由卷积网络组成的变换块 Pre-Conv Block 对 X^{trend} 、 X^{close} 进行预处理；Vit 被用来提取出时序的特征信息 $\hat{X}^{\text{trend}}$ 、 $\hat{X}^{\text{close}}$ ；跳跃连接可以提取出预处理前的数据 $\hat{X}^{\text{rc}}$ 、 $\hat{X}^{\text{rt}}$ ；最后，在融合层（Fusion）中自适应地融合上述的 4 个分量（ $\hat{X}^{\text{trend}}$ 、 $\hat{X}^{\text{close}}$ 、 $\hat{X}^{\text{rc}}$ 、 $\hat{X}^{\text{rt}}$ ），生成最终凝聚了长短时序信息的时序特征图 y^{t-1} 。STTNet 主要由下面的几个部分组成。

3.2.1 预卷积神经网络

一个区域内的交通流量通常与邻近区域高度相关。受此启发，本文设计了一个预处理卷积网络来捕获短距离的相关性，而将长距离的特征交给 Vision Transformer 处理。具体来说，采用 3×3 的卷积核进行卷积，形成 5×5 的感受野，这样的设计保证了预卷积模块最多只捕获 5×5 个区域中的局部特征，可以很好地捕获短范围依赖性。将残差分量加入该分支中，残差连接使用 1×1 的卷积核。以往的实验经验表明，预转换块可以较好地学习短期时空依赖性，将它们细化为紧凑的特征图，从而增强 Vision Transformer 的表现效果。

3.2.2 跳跃连接

为了保留历史数据中的原始相似模式，本文通

过跳跃连接直接将输入 X^{trend} 和 X^{close} 添加到融合层, 如图5中的蓝线所示。在添加之前, 将历史数据在时间维度上聚合, 从而适应输出的形状。对两个历史序列 $X^{\text{close}} \in R^{2 \times d_c \times H \times W}$ 和 $X^{\text{trend}} \in R^{1 \times d_t \times H \times W}$ 进行如下计算:

$$\hat{X}^{\text{rc}} = f(X^{\text{close}}) \in R^{2 \times 1 \times H \times W} \quad (14)$$

$$\hat{X}^{\text{rt}} = f(X^{\text{trend}}) \in R^{2 \times 1 \times H \times W} \quad (15)$$

其中 $\hat{X}^{\text{rc}}$ 、 $\hat{X}^{\text{rt}}$ 是两个残差分量, $f(\cdot)$ 为聚集函数, 其映射变换将矩阵维度映射为 $R^{2 \times 1 \times H \times W}$ 。在本文中, $f(\cdot)$ 使用求和函数, 两个残差分量在 Fusion 中自适应地聚合。

3.2.3 Vision Transformer

在预处理卷积网络之后, 采用 Vision Transformer 来捕获全局依赖性。Vision Transformer 由 Patchify 块和连续 L 个 Transformer 编码器组成。Patchify 块包括 patchify Stem 和 Linear Projection, 在空间上将输入特征图分割为非重叠 $p \times p$ 的区域, 并将每个区域线性投影为向量, 每个向量包含一片区域的信息。此外, 定义了二维空间上的可学习的位置编码信息。然后, 将区域信息向量和区域的位置

信息融合, 送至 Transformer 编码器。编码器利用多头自注意力机制来建模长距离相关性, 经过前馈神经网络后, 再进行层归一化和残差连接处理, 得到最终向量表示。最后, 对最终向量进行 resize 和线性变换, 生成输出 $\hat{X}^{\text{trend}}$ 、 $\hat{X}^{\text{close}}$ 。

3.2.4 融合层

4 个分量 ($\hat{X}^{\text{trend}}$, $\hat{X}^{\text{close}}$, $\hat{X}^{\text{rc}}$, $\hat{X}^{\text{rt}}$) 对输出结果的影响程度是不同的, 对不同区域的影响也不同。为了衡量差异化的影响, 使用参数矩阵的融合方法聚合了长短时序信息的时序特征图 y^{t-1} :

$$y^{t-1} = w_c \cdot \hat{X}^{\text{close}} + w_t \cdot \hat{X}^{\text{trend}} + w_{\text{rc}} \cdot \hat{X}^{\text{rc}} + w_{\text{rt}} \cdot \hat{X}^{\text{rt}} \quad (16)$$

其中 $\cdot$ 表示逐元素乘法, w 是度量每个分量重要性的可学习权重参数。

3.3 Extractor 城市语义提取网络

城市交通流会受到各种城市语义信息的影响。为了进一步提取城市的语义信息, 本文设计了一个城市语义提取模块 Extractor 来学习、整合外部特征和土地特征, 并将这两类特征作为条件约束纳入反向扩散过程, 如图6所示。该模块对交通流问题中相关的天气、风速、道路结构等因素进行了完整全面的考量, 使得模型具有较高的泛化性。

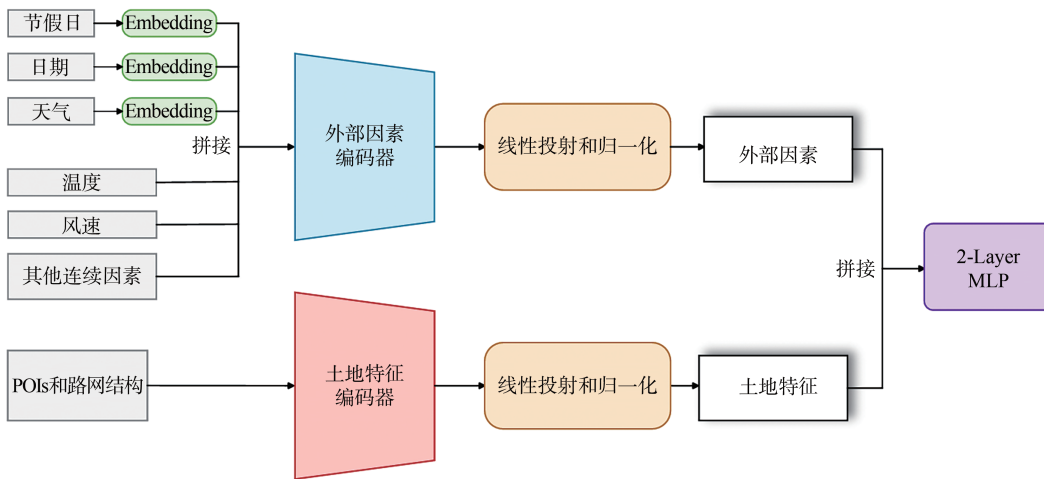


图6 Extractor网络框架

外部环境因素包括连续因素和类别因素。将连续因素直接作为特征输入, 而将类别因素输入特征嵌入层, 使用外部因素编码器和土地特征编码器分别提取二者的特征。具体来说, 外部因素编码器由两个带有 Swish 激活函数的 MLP 模块组成, 土地特征编码器将 ResNet-18 作为主体结构。通过线性投射和归一化, 将两个编码器得到的特征向量连接起

来。然后, 使用双层 MLP 对合并的特征进行处理, 得到特征向量 ε_e 并输入 U-Net 中。

4 实验分析

4.1 实验设置

本文在出租车轨迹数据集 (BJTaxi) 上实验以评估模型的性能。BJTaxi 包含从 2015 年 3 月 1 日到

2015年6月30日的北京出租车行程轨迹数据，按照7:2:1的比例将整个数据划分为不重叠的训练集、测试集和验证集。

数据缺失分为随机区域缺失和时间片缺失。对于随机区域缺失，在训练过程中的每一批数据中，以0.5的概率将区域数据随机设置为0；对于时间片缺失，先对数据集进行预处理，在整个数据集中选择 K 个持续半天的时间间隔（例如，8:00—20:00的交通流量数据），然后在选定的时间段内，将区域中的交通流值设置为0。

4.2 基线模型

将本文提出的模型与以下6个基准模型进行比较。

(1) 平均分配法 (Mean)：将粗粒度区域的流量平均分配到 $n \times n$ 的细粒度区域，其中 n 是放大系数。例如，如果一个粗粒度区域的人群流量为4，若将其划分为4个细粒度小区域，则每个细粒度区域的流量为1。

(2) 历史平均值法 (HA)：将历史城市流量的平均值作为预测值，然后将粗粒度城市流量平均分配到相应的4个细粒度区域。

(3) VDSR^[8]：利用深度为20的深度卷积网络，通过多次级联小型滤波器进行图像超分辨率，从而有效地学习图像区域的上下文特征。

(4) SRResnet^[9]：一种生成对抗网络，用于图像超分辨率。它引入了由对抗损失和内容损失组成的感知损失函数，能更好地还原数据的高频细节。

(5) UrbanPy^[11]：一种细粒度城市交通流推理模型，它采用了包含多个组件的金字塔结构。每个组件都是一个小尺度的原子上采样器，包含外部因子融合网、推理网、建议网和修正网。

(6) MTCSR^[13]：在多任务学习框架下同时完成城市交通流补全和细粒度超分。它考虑了数据的局部地理和全局语义相关性，采用了联合训练和端到端的优化策略。

4.3 模型效果对比

针对放大因子 $N=2, I=J=16$ 的情况，对北京的粗粒度出租车流进行了细粒度推断，在每个时间 t 推断出一个完整的细粒度城市交通流图 $X_{\text{fg}}^t \in R^{2 \times 32 \times 32}$ 。采用平均绝对百分比误差 (MAPE) 评估推理性能，绝对值越低，模型效果越好，定义如下：

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (17)$$

实验结果如表1所示，可以看出，大多数模型在数据完整的情况下有很好的表现，性能差距主要体现在数据缺失或退化的情况。传统的统计方法效果不佳，因为它们只是简单地将大量错误数据平均化，实验结果与地面实况有很大偏差。基于图像超分辨率的算法（如VDSR和SRResnet）可以有效地捕捉交通流数据的模式，对于缺失数据和有噪声数据的情况表现出一定的适应能力。细粒度城市交通流量推断方法利用了与交通相关的辅助信息，在数据缺失的情况下，仍能保持较好的性能。

表1 不同模型的MAPE

基线模型	Mean	HA	VDSR	SRResnet	UrbanPy	MTCSR	STDiff	提高率
完整数据	34.66%	28.73%	29.14%	25.97%	25.45%	23.12%	17.78%	23.10%
缺失20%	36.61%	33.32%	33.82%	31.82%	30.66%	28.56%	19.35%	32.34%

本文提出的模型对数据缺失或有噪声数据的情况的适应能力更强。在完整数据和数据缺失20%的情况下，STDiff的MAPE分别为17.78%和19.35%，与最佳基线模型相比，分别提高了23.10%和32.34%。值得注意的是，当数据从完整变成缺失20%时，STDiff的效果提升更加显著。这表明，随着任务难度的增加，本文提出的模型的优越性更加明显。这是因为STDiff结合了几种基线模型的优点。首先，STDiff使用了自注意模块而不是简单的卷积架构，对复杂、长期的时间相关性具有更强大

的学习能力。其次，扩散模型的设计适用于此类有误差的数据，因为恢复噪声数据的过程可视为反向去噪过程的一部分，这使得STDiff在城市交通流推断中表现更佳，在现实噪声环境中表现出更卓越的性能、更好的鲁棒性和泛化能力。

4.4 模型效果可视化

为了更直观地比较本文提出的模型和基线模型的性能，在数据不完整的情景下，对2015年5月1日12:00的BJTaxi数据进行细粒度推断，并提供一个可视化案例，如图7所示。其中Input表示完整粗

粒度数据, **Incomplete**表示对应的不完整粗粒度数据, **Ground Truth**表示对应的完整细粒度数据, 即标签, **Ours**表示本文提出的模型 **STDiff**。其余表

示各个模型输出的数据效果, 可见本文提出的模型效果十分卓越, 优于其他模型, 数据状态与 **Ground Truth**十分接近。

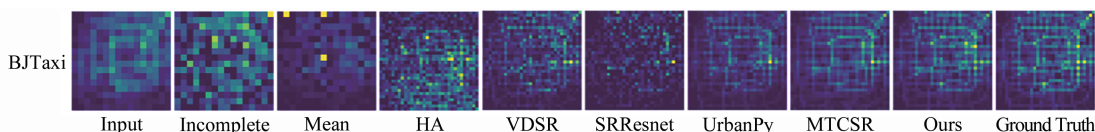


图7 可视化展示

5 结束语

解决城市交通规划中出现的问题, 根本在于对城市交通流量进行精确获取、对各路段交通概况进行正确评估。因此, 精细、可靠的交通流数据可为决策者提供最有力的帮助。本文提出的基于时空注意力的去噪扩散模型 **STDiff**, 在对不完整、粗粒度的城市交通流进行细粒度推断的任务中表现优异。由于去噪扩散模型本身包含加噪和去噪的过程, 因此适用于处理带有噪声、不完整、稀疏的数据。实验结果表明, **STDiff**的推断性能优于现有主要的基线模型。

参考文献:

- [1] SU K H, LI J, FU H B. Smart city and the applications[C]//Proceedings of 2011 International Conference on Electronics, Communications and Control (ICECC). Piscataway: IEEE Press, 2011: 1028-1031.
- [2] LEE S, KWON H, CHO H, et al. International case studies of smart cities: Orlando, United States of America[R]. Washington: Inter-American Bank, 2016.
- [3] OLIVIER R, CAO H Q. Nearest neighbor value interpolation[J]. International Journal of Advanced Computer Science and Applications, 2012, 3(4): 25-30.
- [4] XU Y N, ZHU Y M. When remote sensing data meet ubiquitous urban data: fine-grained air quality inference[C]//Proceedings of 2016 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE Press, 2017: 1252-1261.
- [5] CHEN L B, JAKUBOWICZ J, YANG D Q, et al. Fine-grained urban event detection and characterization based on tensor cofactorization[J]. IEEE Transactions on Human-Machine Systems, 2017, 47(3): 380-391.
- [6] WANG Z H, CHEN J, HOI S C H. Deep learning for image super-resolution: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3365-3387.
- [7] WANG S Z, CAO J N, YU P S. Deep learning for spatio-temporal data mining: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(8): 3681-3700.
- [8] DONG C, LOY C C, HE K M, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analy-

sis and Machine Intelligence, 2016, 38(2): 295-307.

- [9] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 1646-1654.
- [10] LIANG Y X, OUYANG K, JING L, et al. UrbanFM: inferring fine-grained urban flows[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 3132-3142.
- [11] OUYANG K, LIANG Y X, LIU Y, et al. Fine-grained urban flow inference[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(6): 2755-2770.
- [12] LIANG Y X, OUYANG K, WANG Y W, et al. Revisiting convolutional neural networks for citywide crowd flow analytics[C]//HUTTER F, KERSTING K, LIJFFIJT J, et al. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2021: 578-594.
- [13] LI J Y, WANG S Z, ZHANG J Q, et al. Fine-grained urban flow inference with incomplete data[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(6): 5851-5864.
- [14] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

[作者简介]



郑雨豪 (2002-), 男, 中南大学计算机学院本科生, 主要研究方向为深度学习、数据挖掘、生成式模型。



王森章 (1986-), 男, 中南大学计算机学院教授, 主要研究方向为时空数据挖掘和城市计算。发表高水平论文 100 余篇, 主持国家级、省部级、企业项目 10 余项, 担任多个国际期刊的审稿人, 40 余次担任领域内主流国际会议的 SPC 和 PC 等。