

# 基于深度学习的 PM<sub>2.5</sub> 短期预测模型

赵文芳<sup>1,2</sup>, 林润生<sup>2</sup>, 唐 伟<sup>3</sup>, 周 勇<sup>3</sup>

(1. 北京城市气象研究院, 北京 100089)

(2. 北京市气象信息中心, 北京 100089)

(3. 中国气象局发展研究中心, 北京 100081)

**[摘要]** 为了提高 PM<sub>2.5</sub> 浓度短期预报的准确率, 解决现有 PM<sub>2.5</sub> 浓度短期预报准确率不高的问题, 提出了一种基于卷积神经网络和长短时记忆的深度学习预报模型。首先, 综合考虑气温、相对湿度、降水量、风力、能见度等多种气象要素, 综合分析气象要素与 PM<sub>2.5</sub> 浓度相关性。其次, 利用 PM<sub>2.5</sub> 浓度数据、气象站点观测数据和气象要素网格实况分析数据进行融合处理, 生成用于训练和测试的时空序列数据, 并使用卷积神经网络和长短时记忆网络获取时空特征。通过大量实验确定模型中关键参数, 然后利用最优参数建立预测模型。最后, 使用模型对 PM<sub>2.5</sub> 未来 24 h 浓度进行预测, 并与支持向量机、业务中的预报模型进行对比。实验结果表明, 相比其他机器学习方法和预报方法, 卷积神经网络和长短时记忆相结合的预测方法能有效提高 PM<sub>2.5</sub> 浓度未来 24 h 预测精度, 并具有较高的泛化能力。

**[关键词]** PM<sub>2.5</sub> 浓度预测, 机器学习, 长短时记忆, 深度学习, 卷积神经网络

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2019)03-0032-10

## Forecasting Model of Short-Term PM<sub>2.5</sub> Concentration Based on Deep Learning

Zhao Wenfang<sup>1,2</sup>, Lin Runsheng<sup>2</sup>, Tang Wei<sup>3</sup>, Zhou Yong<sup>3</sup>

(1. Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China)

(2. Beijing Meteorological Information Center, Beijing Meteorological Bureau, Beijing 100089, China)

(3. Development and Research Center, China Meteorological Administration, Beijing 100081, China)

**Abstract:** In order to improve the accuracy of PM<sub>2.5</sub> concentration forecast in Beijing Meteorological Bureau, a deep learning prediction model based on convolutional neural network (CNN) and long short term memory neural network (LSTM) was proposed. Firstly, the feature vectors extraction was carried out by using the correlation analysis technique from meteorological data such as temperature, wind, relative humidity, precipitation, visibility and atmospheric pressure. Secondly, taking into account the fact that PM<sub>2.5</sub> concentration was significantly affected by surrounding meteorological impact factors, meteorological grid analysis data were novel involved into the model, as well as the historical PM<sub>2.5</sub> concentration data and meteorological observation data of the present station. Spatiotemporal sequence data were generated from these data after integrated processing. Highlevel spatiotemporal features were extracted through the combination of the CNN and LSTM. Finally, future 24-hour prediction of PM<sub>2.5</sub> concentration was made by the model. The comparison among the accuracy of this optimized model, support vector machine (SVM) and existing PM<sub>2.5</sub> forecast system is performed to evaluate their performance. The results show that the proposed CNN-LSTM model performs better than SVM and current operational models in Beijing Meteorological Bureau, which has effectively improved the prediction accuracy of PM<sub>2.5</sub> concentration for different time predictions scales in the next 24 hours.

**Key words:** forecast for PM<sub>2.5</sub> concentration, machine learning, long short term memory, deep learning, convolutional neural network

收稿日期: 2019-07-05.

基金项目: 国家自然科学基金(41575156)、中国气象局软科学研究重点课题(2019ZDIANXM19).

通讯联系人: 唐伟, 博士, 高级工程师, 研究方向: 气象数据分析、模式识别、深度学习. E-mail: weitang@cma.gov.cn

PM2.5(fine particulate matter)是指一种悬浮于大气中的空气动力学直径小于等于  $2.5\ \mu\text{m}$  的细颗粒物<sup>[1-3]</sup>,其浓度的高低直接影响空气质量和气候环境,过高浓度甚至危害人类健康.近年来,由PM2.5引起的空气污染受到了各级政府、部门及社会各界的关注.从2005至2014年,我国某些省的PM2.5排放量的增长率超过了200%<sup>[4]</sup>;2015年我国城市区域的PM2.5浓度水平居全球前列<sup>[5]</sup>;2015年12月的中国74个城市空气条件月报指出,平均空气污染天数占全年44.4%,其中有39个城市受污染天数超过全年50%<sup>[6]</sup>;还有一些研究表明,PM2.5与一些疾病的就诊率、住院率和死亡率存在显著的正相关关系<sup>[7]</sup>.因此,PM2.5浓度的准确预测对加强大气污染防治、政府决策和保护人们日常健康都具有重要意义.

目前PM2.5的预报主要有数值模式法和统计学预报法.数值模式预报法主要基于空气动力学理论和物理化学过程,使用数学方法建立大气污染浓度的稀释和扩散模型动态预测空气质量和主要污染物的浓度变化.气象领域常用的数值模式包括美国环保局研发的通用多尺度空气质量模式(the community multi-scale air quality)<sup>[8-11]</sup>,气象-化学在线完全耦合的区域空气质量模式WRF-Chem<sup>[12-13]</sup>,中国气象局研发的雾霾数值预报模式<sup>[14]</sup>,以及北京市气象局自主研发的区域环境气象数值预报模式(Beijing regional environmental meteorology pre-diction system, BREMPS)<sup>[15]</sup>.总体上来说,这些模式较为全面地考虑了物理和化学过程,但由于模式中有关污染物排放、传输和沉降过程中参数存在较大不确定性,导致预报结果也存在不确定性<sup>[16]</sup>.统计学预报法是利用统计数学方法开展天气预报.常用的方法有多元线性回归、支持向量机、人工神经网络、遗传算法、小波分析等.已有大量学者利用空气质量观测数据(例如PM2.5浓度、SO<sub>2</sub>浓度、CO浓度、PM10浓度、O<sub>3</sub>浓度等)、气象观测数据和数值模式预报数据,应用一种或多种统计学方法建立预测模型,对PM2.5浓度和其他污染物浓度进行预报<sup>[17-24]</sup>.然而,除了气象条件,污染物浓度还受排放量、交通条件、人口密度等因素的共同影响,使用单一统计方法很难建立准确度高的预报模型.

深度学习是人工智能领域中一种新颖的机器学习方法,可以对大量输入数据的特征表示进行有效学习,为气象时间序列的预测提供了新的研究思路和方法.深度学习的主要神经网络模型主要有卷积神经网络(convolutional neural networks, CNN)、递归神经网络(recurrent neural network, RNN)、长短时记忆(long short term memory, LSTM)、对抗神经网络等,已有一些学者利用这些模型开展气象预测的研究. Zheng Y等<sup>[23]</sup>使用历史空气质量数据和气象数据提出了一个FC-LSTM预测模型来预测未来24 h PM2.5污染浓度; Vidushi C<sup>[25]</sup>等提出了一个多层的LSTM模型预测未来空气污染物的浓度; Weber S A等<sup>[26]</sup>构建了多层次LSTM和编码器-解码器(Encoder-Decoder)两个预测模型并进行比较,实验结果表明LSTM模型优于Encoder-Decoder框架,该LSTM模型对南京PM2.5的预测结果与监测数据非常接近. Chiou J H等<sup>[27]</sup>开发一种深层神经网络模型,以历史小时降水、风速风向和PM2.5浓度数据作为输入,先进行多次一维卷积处理,再将结果输入LSTM来预测PM2.5浓度.此外,还有一些学者将LSTM模型与特征空间相关性相结合应用于PM2.5浓度预报,如: Cong W等<sup>[28]</sup>提出了一种新的时空卷积长短期记忆神经网络,将当前站点和近邻站点PM2.5浓度数据经过一维卷积运算后输入到模型中; PING S W等<sup>[29]</sup>提出LSTM预测模型,使用当前站点和近邻站点的PM2.5浓度数据、气象数据和地形数据,其中地形数据用以提取地形对空气质量的影响.然而,大城市的空气质量监测站点数量十分有限,而且分布可能不均匀,两个最近邻的站点可能分布在不同区域,将这样的临近站点加入到模型中反而有可能导致预测不准确.此外,大部分的研究使用的都是离散的站点数据,使用网格化的气象要素实况数据相对甚少,使用与PM2.5浓度有很强相关性的能见度数据也甚少<sup>[30-32]</sup>.事实上,网格化的气象要素实况数据综合考虑了复杂地形和动力降尺度因素,比站点数据能更客观、更准确地反映气象要素的空间特征,过去一段时间的气象要素与未来时刻的气象要素也存在较强的相关性.3D-CNN模型可以从空间和时间的维度提取特征,然后进行3D卷积,以获取更多的可用信息.

综上所述,综合考虑PM2.5浓度的时空相关性、网格化的气象要素实况数据特点和3D-CNN优势,本文提出了一种基于3D-CNN和LSTM的预测模型.将每个观测时刻的网格化的气象要素实况数据看成一幅图像,过去一段时间的数据看成序列化的图像,使用3D-CNN提取多种气象要素的空间特征. LSTM用于从站点序列数据提取时间特征,将时间特征与空间特征接入全连接层,得到最终的预测结果.

1 数据分析

1.1 北京 PM2.5 浓度的时间序列分析

以 2008 年-2018 年北京地区 20 个国家级气象站的逐日观测数据为基础,综合分析逐年霾出现的平均日数(以下简称霾日数)、一个霾天气过程中两个临近站点 PM2.5 浓度的变化规律和单个站点的空气污染物浓度随时间变化规律. 根据站点的地理位置,可将 20 个国家级气象站分为山区站和平原站. 图 1(a)反映了北京地区 2008 年-2018 年 20 个国家级气象站逐年霾日数的变化趋势,横坐标为年份,纵坐标为霾日数. 可以看出,山区站点的霾日数低于所有站的均值水平,平原站则高于均值水平;2014 年霾日数最多,其中位于山区的延庆站的霾日数到达 182 天;而海淀站却在 2016 年出现霾日数峰值. 图 1(b)反映了 2018 年 12 月 2 日-12 月 3 日霾天气过程中,海淀区 2 个最临近自动站的 PM2.5 浓度随时间变化趋势. 这两个站的 PM2.5 浓度峰值差异很大,变化趋势也不近相同,时空相关性并不显著. 由此可见,PM2.5 浓度存在明显的空间差异,对于空气污染物监测站点稀少而且分布不均匀的大城市而言,从网格化的数据获取空间特征相关性会比从离散站点中获取更可靠一些. 图 1(c)反映了该次霾天气过程中海淀站点 PM2.5、PM10、SO<sub>2</sub> 和 O<sub>3</sub> 浓度随时间的变化趋势,PM10 浓度峰值明显高于 PM2.5,但是 O<sub>3</sub> 和 SO<sub>2</sub> 的浓度偏低,而且变化不明显,和 PM2.5 变化趋势也不相同.

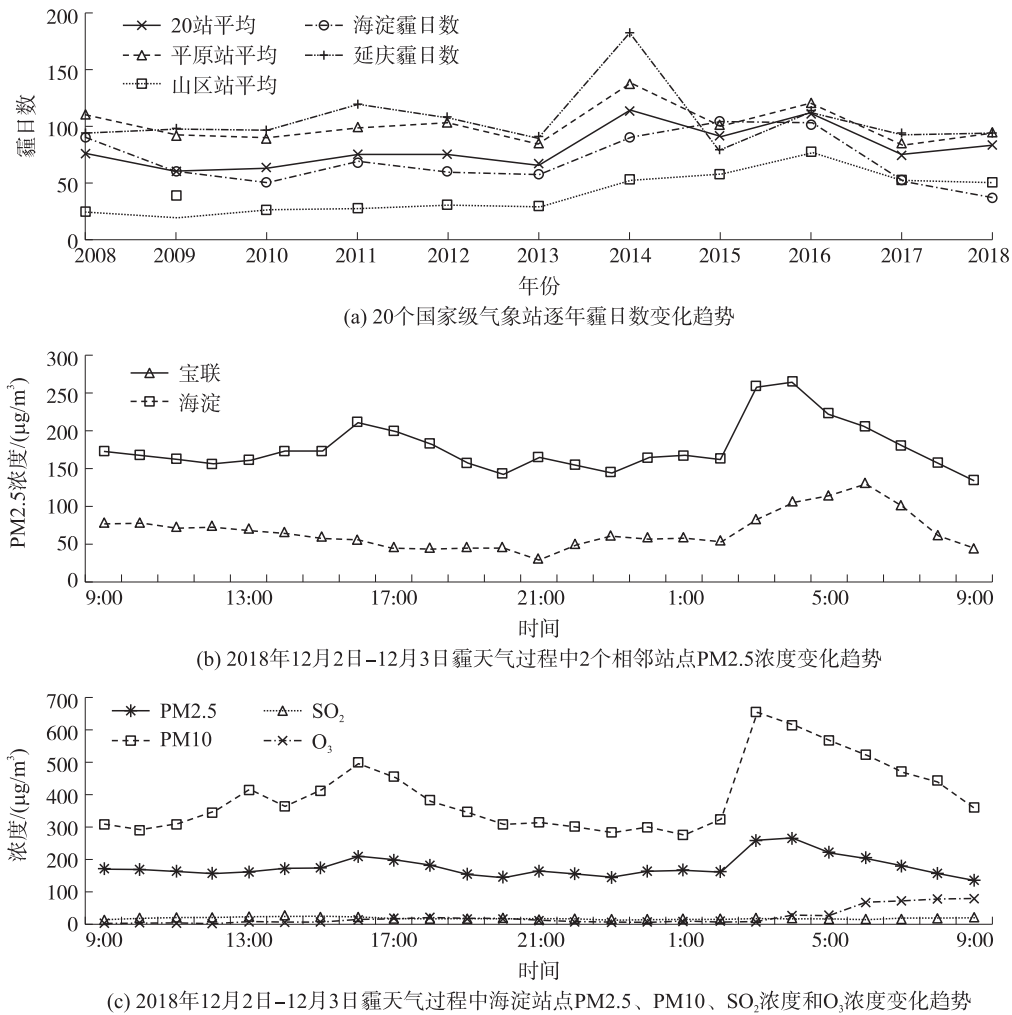


图 1 北京地区不同尺度的 PM2.5 浓度时间序列分析

Fig. 1 Time series analysis of PM2.5 concentrations at different scales in Beijing area

1.2 PM2.5 浓度与气象要素相关性分析

气象因素对 PM2.5 浓度的影响十分复杂,往往是不同气象因素相互影响的结果. 如果分别考虑各个因素,则不能很好体现多个因素相互作用对 PM2.5 浓度产生的耦合效应,从而影响预报模型的准确性. 通

过收集整理,建立了10个观测站2014年-2017年的逐小时空气污染物浓度和11个气象要素的数据样本集.利用Spark并行计算框架的机器学习库,编写程序并在大数据环境下运行,快速分析出PM2.5浓度与各气象要素相关性,如图2所示.从图2可以看出:温度、最高温度、平均相对湿度、能见度与PM2.5浓度呈现正相关;2 m风、10 m风、极大风速、小时降水量与PM2.5浓度呈负相关;PM10的浓度与PM2.5浓度相关性较强,而O<sub>3</sub>、SO<sub>2</sub>浓度与PM2.5浓度几乎不相关,与文献32中阐述的结论相符合.

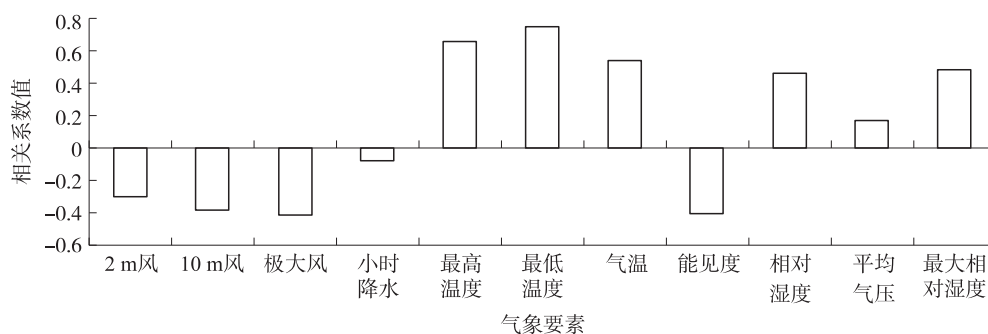


图2 PM2.5浓度与气象要素相关性比较

Fig. 2 Correlation coefficients between PM2.5 concentration and meteorological factors

## 2 PM2.5 预测模型

### 2.1 问题的描述

PM2.5浓度预报的目标是使用过去固定时长(例如24 h, 48 h, 72 h)的观测值,预测目标站点固定时长(例如未来24 h, 36 h或48 h)的PM2.5浓度.假设有 $N$ 个不同的观测站点,对于给定的时刻 $t$ ,获取的数据组成集合 $ST = \{ST_1, ST_2, \dots, ST_n\}$ ,站点 $n$ 在过去 $L$  h内的数据表示为 $ST_n = \{X_{t-L}, X_{t-L+1}, \dots, X_t\}$ (共 $L$ 个时间步),序列中的每个观测值 $X_t$ 是一个 $d$ 维向量,由污染物浓度和一些气象要素观测值构成.用 $M \times N$ 网格表示北京的空间区域大小,网格中的每个单元格内部有 $P$ 个随着时间而变化的测量值, $t$ 时刻北京气象观测数据可以用张量 $R_t^{P \times M \times N}$ 表示,过去 $L$  h的数据可以形成序列张量 $RT = \{R_{t-L}^{P \times M \times N}, R_{t-L+1}^{P \times M \times N}, \dots, R_t^{P \times M \times N}\}$ .从 $RT$ 张量找到目标站点 $n$ 的空间特征,结合使用数据集 $ST$ 预测它在时刻 $(t+1, t+2, \dots, t+K)$ 的对应值.从机器学习的角度来看,这个问题可以看作是一个时空序列预测问题.

### 2.2 LSTM 模型

LSTM是一种用于解决梯度消亡现象的改进型RNN,广泛用于自然语言处理、图片标注、机器翻译和时间序列预测等领域.使用sigmoid神经网络和按位做乘法的两个操作合在一起构成了LSTM的“门”结构,让信息有选择性地影响循环神经网络中每个时刻的状态.LSTM拥有3个这样的“门”结构,“遗忘门”、“输入门”和“输出门”.“遗忘门”会根据当前的输入 $X_t$ 、上一个时刻状态 $C_{t-1}$ 和上一时刻输出 $h_{t-1}$ 共同决定哪一部分记忆需要被遗忘.“输入门”会根据 $X_t$ 、 $C_{t-1}$ 和 $h_{t-1}$ 决定哪些部分进入当前时刻的状态 $C_t$ .在LSTM结构计算得到最新的状态 $C_t$ 后,输出门会根据 $C_t$ 、 $X_t$ 、 $h_{t-1}$ 决定该时刻的输出 $h_t$ .

### 2.3 PM2.5 浓度预测模型

PM2.5浓度预测模型包含2个主要模块:获取空间特征的3D-CNN与获取时间特征的LSTM,使用的数据包括北京市气象局10个观测站逐小时的空气污染物浓度观测资料、逐小时气象要素观测资料和北京地区网格化三维气象要素客观分析资料.为了获取与站点更加相关更加精准的空间特征,以站点为中心,从整个三维气象要素客观分析网格中取出 $112 \times 112$ 的子网格作为输入,使用3D-CNN分别获取温度、相对湿度、风要素的空间特征.获取时间特征的LSTM模型包含2层LSTM结构,每层有128个节点,气象要素和空气污染浓度数据经过预处理后作为站点的特征向量输入模型计算时间特征.时间特征与多个气象要素的空间特征输入全连接层,得到预测结果.整个模型的结构如图3所示.

为了评估模型的性能,使用均方根误差(Root Mean Square Error, RMSE)、平均绝对误差(Mean Absolute Error, MAE)和平均绝对百分比错误(MAPE)作为评价指标. RMSE和MAE用于评估绝对误差,而MAPE用于测量相对误差. RMSE和MAE反映了预测的极值效应和误差范围值, MAPE反映了平均值特异性预测值.

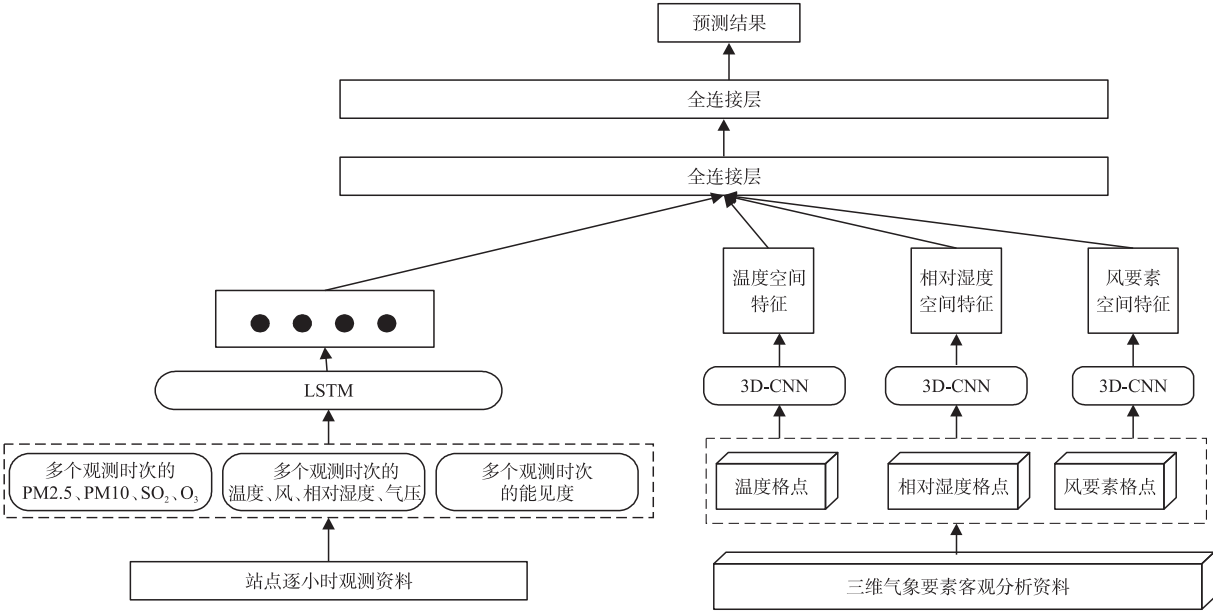


图 3 PM2.5 浓度预测模型架构

Fig. 3 Network framework of predict model for PM2.5 concentration

2.3.1 获取空间特征的 3D-CNN

气象要素客观分析资料经过卷积、采样、激活函数和扁平化处理,得到一维列向量作为空间特征的输出。3D-CNN 需要学习的是与站点高度相关的短期时空特征,不需要很深,包含 4 个卷积层和 3 个池化层<sup>[33-35]</sup>,结构如图 4 所示。每个卷积层的内核大小为 3×3×3,步长为 1;第一层池化层内核大小为 1×2×2,其余池化层内核大小为 2×2×2,步长为 2,所有卷积层的卷积核个数分别为 64、128、256 和 256。输入的图片大小是 112×112,输出的图像大小为 28×28,经过全连接层后得到 4096×1 的特征向量。

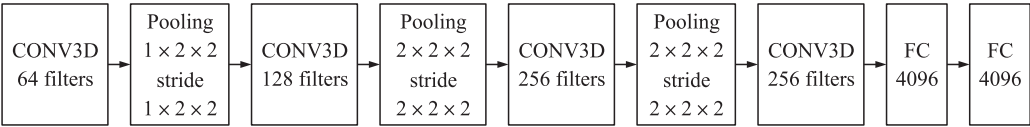


图 4 获取空间特征的 3D-CNN 模型

Fig. 4 The 3D-CNN component

2.3.2 获取时间特征的 LSTM

与单层 LSTM 相比,堆叠 LSTM 的表示能力更强大,因此本文使用 2 层 LSTM 结构,每层 128 个节点。站点的空气污染物浓度、能见度、气象要素观测资料、时间信息等作为该站某个时刻的特征向量,输入 LSTM 模型。为了防止过拟合,使用比例为 0.15 的 Dropout,即含有 Dropout 的网络层在训练过程中,会有 15% 的节点(及与之相连的权重连接)会被隐去。

2.3.3 模型算法流程

PM2.5 预测模型算法的具体步骤如下:

- (1)对站点的空气污染物浓度和气象要素观测数据进行预处理。首先,计算每类数据的均值和方差,根据均值和方差相互关系的基本规律,将均值在缩小而偏差在增大的数据看作错误数据,并从样本集中剔除。其次,所有缺测值使用均值替代,再进行归一化处理。最后,这些数据与站点基本信息、时间信息一并作为 LSTM 模型的输入特征向量。
- (2)对气象要素客观分析资料预处理。原始的气象要素客观分析资料是京津冀范围,以站点为中心,将原始数据处理为 112×112 的子网格,按温度、相对湿度、风要素网格分别存储。对数据进行归一化处理。
- (3)数据融合与格式化,设置滑动时间窗口,生成用于训练和测试模型的时间序列数据,完成原始数据到张量的转变。
- (4)使用时间序列数据训练 LSTM 和 3D-CNN 深度神经网络,并使用特定的评价指标 RMSE、MAE 和



MAPE 来评估网络的预测效果.

3 实验及分析

3.1 实验环境及数据

本文使用的实验数据来源于北京市气象局,分为 2 类:(1) 10 个站点的逐小时观测资料,包括 PM2.5 浓度、PM10 浓度、O<sub>3</sub> 浓度、SO<sub>2</sub> 浓度、能见度和气象要素,数据每 1 h 更新一次,具体见表 1;(2) 北京地区的网格化三维气象要素客观分析资料,空间分辨率为 1 km,每 1 h 更新一次,主要包括温度、风和相对湿度要素.

表 1 输入 LSTM 模型的站点特征向量  
Table 1 Input feature vectors of LSTM

字段名称	描述	字段名称	描述
ID	站号	tem_min	最低气温
PM2.5	PM2.5 浓度	win_2m	2 m 风速
PM10	PM10 浓度	win_10m	10 m 风速
SO <sub>2</sub>	二氧化硫浓度	win_max	极大风速
O <sub>3</sub>	臭氧逐小时浓度	pre	1 h 降水
vis	能见度	rh	相对湿度
tem	平均气温	rh_max	最大相对湿度
tem_max	最高气温	prs	平均气压

使用 python 和机器学习库 scikit-learn 完成数据的预处理和基于支持向量机回归 (Support Vector Regression,简称 SVR) 的 PM2.5 浓度预测模型建立,该模型主要用来与深度神经网络模型进行比较. 使用 python、深度学习库 tensorflow 和 Keras,搭建并训练了上述的 LSTM 和 3D-CNN 深度神经网络结构.

3.2 LSTM 模型关键超参数选择

PM2.5 浓度预测模型包括 3D-CNN 和 LSTM 2 个子模型,3D-CNN 子模型的层次和参数已经在图 4 中阐明,这里主要考虑 LSTM 子模型的参数选择. 隐藏层数量、隐藏神经元个数和训练步长是 LSTM 模型中几个关键的参数,根据 Stuart 等<sup>[36-37]</sup> 的研究,改变隐藏节点的数量可能会减少过度拟合和增加模型泛化;多层 LSTM 的结构可以增加模型的学习能力,但是网络参数也会随之增加,这对模型泛化能力和训练时间都有直接影响. 为了找到隐藏层数量、隐藏神经元个数和训练步长参数的最优值,进行了大量实验.

使用验证集数据测试了 5 种不同结构的 LSTM 模型,每种结构的层数、隐藏神经元个数、网络参数和误差见表 2. 可以看出,单纯的增加层数或者节点数并不能减少模型的误差;当每层节点超过 400 时误差明显增大,超过 4 层 LSTM 堆叠时 RMSE 误差增加了 3.16 倍;具有 2 层结构每层 128 个结点的模型误差最小,呈现最佳性能. 因此,选用这个结构的 LSTM 模型用于后续实验.

表 2 不同结构的 LSTM 模型误差比较  
Table 2 Error comparison among five LSTM structure

LSTM 结构	层数	每层隐藏节点	MAPE	RMSE	MAE
LSTM-1	1	64	27.14	17.85	9.91
LSTM-1	1	128	24.49	19.12	11.69
LSTM-2	2	128	<b>23.8</b>	<b>15.57</b>	<b>9.72</b>
LSTM-3	3	384	46.53	64.79	9.29
LSTM-4	4	512	48.62	64.06	27.45

过大的训练步长会引起模型对训练数据过度拟合,也会消耗更多的时间. 图 5 显示了训练集和验证集数据的 MAPE 随训练步长的变化. 训练集和验证集上,MAPE 误差都随迭代次数增加逐步减少. 迭代次数超过 3 000 左右时,该模型似乎过度适应,不仅泛化能力没有改善,而且还出现微弱波动. 因此,设置迭代步长为 3 000. 此外,在参考大量文献的基础上,本文中对模型中的部分关键参数做出了合理的设定:学习率设置为 10<sup>-3</sup>,衰减率设置为 0.95,参数初始化范围设置为 [-0.08,0.08],模型优化算法选用 Adam 算法.

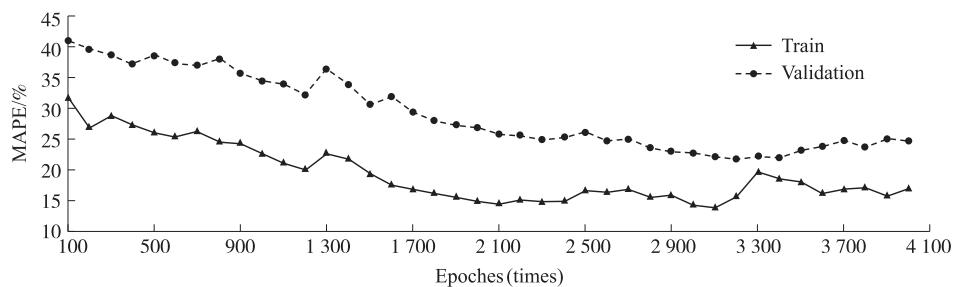


图 5 MAPE 随步长变化趋势  
Fig. 5 The trend of MAPE over epochs

3.3 实验结果与分析

使用过去 72 h 的序列数据预测各站点未来 24 h 的 PM2.5 浓度,每个样本包含 73 个时间序列(输入 72,输出 1)数据,而每个时间序列数据则由表 1 中的 16 个特征向量组成,批量训练一次输入 64 个样本.训练数据集的时间跨度为 2014 年-2017 年,其中 80%的数据用来训练模型,20%的数据用来验证模型,测试集的时间跨度为 2018 年 10 月-12 月.

为了方便模型之间的对比,除了 3DCNN-LSTM,本文还设计了基于机器学习的 SVR 模型<sup>[38]</sup>,Just-LSTM 模型和 Local-LSTM 模型. Just-LSTM 和 Local-LSTM 结构与 3DCNN-LSTM 的 LSTM 部分完全相同,使用同样的超参数,输入序列长度也都相同,只是序列数据中包含的特征向量不同. Just-LSTM 模型的输入仅包含 3DCNN-LSTM 模型中用于生成时间特征的特征向量,而 local-LSTM 模型的输入包含 3DCNN-LSTM 模型中用于生成时间特征的特征向量和站点周边距离最近 6 个站点的空气质量数据.

(1)几个模型预测结果的误差分析

使用验证集进行模型预测结果的误差分析.表 3 列出了 4 个模型对 10 个站点预测得到的整体精度.可以看出,3DCNN-LSTM 模型预测效果最佳,在两个指标中均优于所有其他模型.相比仅使用本站观测数据的 Local-LSTM 模型,对于未来 24 h 的 PM2.5 浓度预测,3DCNN-LSTM 对 MAPE 和 RMSE 的改进分别至少提高 28.57%和 43.91%.图 6(a)、(b)显示了 4 个模型在 10 个站上的 RMSE 和 MAE 误差分布情况.3DCNN-LSTM 预测模型对 10 个站点的 RMSE 在[12,18]范围区间,MAE 在[3,16]范围区间,其中有 9 个站的 MAE 小于 15;ID 为 3 号的站点是所有站里预测效果最好的.总体上,预测结果与观测数据趋势一致,10 个站点之间的 RMSE 和 MAE 差别不大,绝对误差均在可接受范围内.10 个站点有 4 个分布在郊区,其中 1 个站点位于山区,6 个在城区,站点地理位置差异较大,但是模型的预测结果没有出现与观测偏差特别大的情况,这说明模型学习空间特征效果比较好. Just-LSTM 对 10 个站点的 RMSE 在[17,38]区间,MAE 在[15,25]范围区间;Local-LSTM 模型的 RMSE 和 MAE 分布与 Just-LSTM 类似,峰值比 Just-Seq2Seq 略高;SVR 模型的预测误差明显高于其他 3 个深度学习模型.

表 3 不同模型预测结果的误差比较

Table 3 Error comparison among different models

模型	MAPE	RMSE
SVR	54.79	49.90
Just-LSTM	35.91	25.46
Local-LSTM	31.53	20.47
3DCNN-LSTM	25.65	14.28

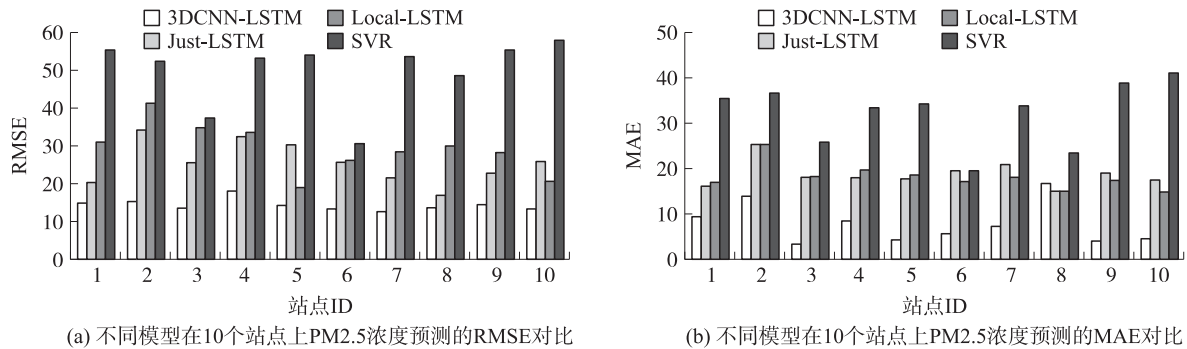


图 6 不同模型在 10 个站点上 PM2.5 浓度预测的误差对比

Fig. 6 Error comparison of PM2.5 concentration prediction from different models over ten stations

## (2) 不同深度学习模型的预测结果比较分析

从SVR、Just-LSTM、local-LSTM和3DCNN-LSTM模型预测的误差分布可看出,3个深度学习模型的误差均比SVR模型小,因此,在测试集上,仅对3个深度模型的预测准确性进行对比.选择2018年10月14日雾霾天气过程为例,3DCNN-LSTM模型、Just-LSTM和Local-LSTM模型下的ID为1的站点PM2.5逐小时浓度预测曲线随时间的变化如图7所示.由图7可知,3种模型的预测曲线与实际观测曲线的趋势都基本保持一致.3DCNN-LSTM模型的预测曲线与实际观测曲线最为接近,尤其是在浓度波峰附近,表明该模型能较好地对PM2.5 1 h浓度峰值进行预测,这对PM2.5浓度的预测有着十分积极的作用,因为PM2.5浓度峰值直接关系到空气污染防御和政府的服务决策. Just-LSTM模型的预报曲线在PM2.5浓度波谷部分拟合有波动,对未来18 h后PM2.5浓度的预测偏离较大,预测效果不理想. local-LSTM模型的预报曲线起伏最大,对PM2.5浓度峰值的预测值比实际观测值偏大很多,对PM2.5浓度波谷的预测值仅有2个时次接近实际观测值,总之和实际观测曲线偏离较大.相比之下,3DCNN-LSTM模型预测效果最好.

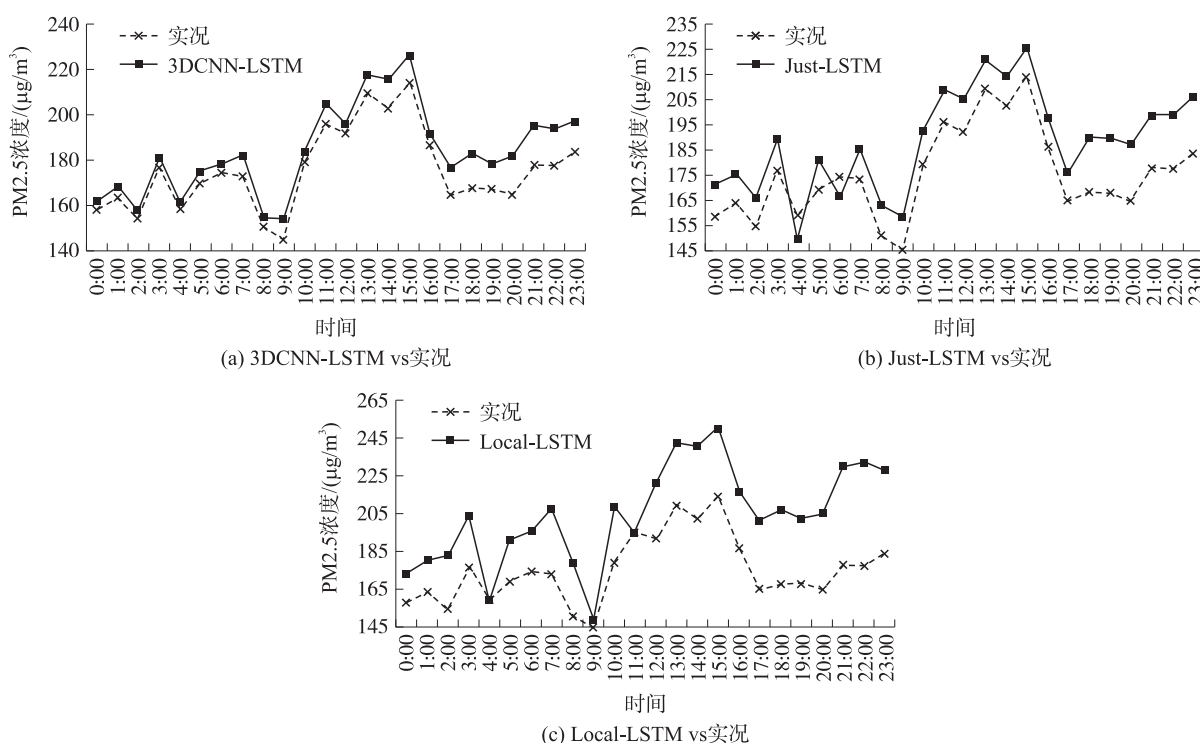


图7 不同模型预测逐小时PM2.5浓度的误差对比

Fig. 7 Comparison of forecast accuracy of different models

由此可见,对未来24 h PM2.5浓度预测而言,考虑时空特征的深度学习模型准确度最高,仅考虑站点及周边邻近站点的时间序列模型的准确度和仅考虑站点自身的时序模型的预测效果都存在一定程度偏差.这些恰恰说明空间相关性对PM2.5浓度预测很重要,在预测模型中考虑站点周围气象要素是十分合理,同时也是非常必要的.3个深度学习模型的预测准确率均高于机器学习模型,可以看出,深度神经网络方法比无监督的机器学习可以自主学习到更多的信息,从而能提高预测准确性.

## 4 结论

气象因素一直是影响PM2.5预测的主要因素,多个气象因素产生的耦合效果对它的影响十分复杂.本文在分析气象要素与PM2.5相关性基础上,提出了一种使用深度学习进行预测的方法,利用PM2.5观测实况数据、气象实况数据和气象要素网格实况资料,建立3DCNN-LSTM预测模型.通过几个深度学习模型预测结果和误差的对比表明,3DCNN-LSTM预测模型能有效地获取时空特征,适合解决时空序列数据的预测问题.在后续的研究中,可考虑将气象数值模式预报、气象格点预报等产品加入到模型中,与现有的预报系统进行全面的对比分析,进一步提高模型的效率和精度.



## [参考文献]

- [1] KAUFMAN Y J, DIDIER T, OLIVIER B. A satellite view of aerosols in the climate system[J]. *Nature*, 2002, 419: 215–233.
- [2] CHAN C, YAO X. Air pollution in mega cities in China[J]. *Atmospheric environment*, 2008, 42(1): 1–12.
- [3] 张西雅, 扈海波. 基于多源数据的北京地区 PM2.5 暴露风险评估[J]. *北京大学学报(自然科学版)*, 2018, 54(5): 1103–1113.
- [4] JIN Q, FANG Y, WEN B, et al. Spatio-temporal variations of pm2.5 emission in china from 2005 to 2014[J]. *Chemosphere*, 2014, 183: 429–436.
- [5] Van DONKELAAR A, MARTIN R V, BRAUER M, et al. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate[J]. *Environmental health perspectives*, 2015, 123(2): 135–143.
- [6] FANG C L, WANG Z B, XU G. Spatial-temporal characteristics of pm2.5 in china: A city level perspective analysis[J]. *Journal of geographical sciences*, 2016, 26: 1519–1532.
- [7] LIM S S, VOS T, FLAXMAN A D, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010[J]. *The lancet*, 2012, 380: 2224–2260.
- [8] CHEN R, WANG X, MENG X, et al. Communicating air pollution-related health risks to the public: an application of the air quality health index in Shanghai, China[J]. *Environment international*, 2013, 1(5): 168–173.
- [9] CHEN J, LU J, AVISE J C, et al. Seasonal modeling of PM 2.5 in California's San Joaquin Valley[J]. *Atmospheric environment*, 2014, 92: 182–190.
- [10] WU Q Z, SHI A, LI Y, et al. Air quality forecast of PM10 in Beijing with community multi-scale air quality modeling(CMAQ) system: emission and improvement[J]. *Geoscientific model development*, 2014, 7(5): 2243–2259.
- [11] 邓涛, 吴兑, 邓雪娇, 等. 珠三角空气质量暨光化学烟雾数值预报系统[J]. *环境科学与技术*, 2013(4): 62–68.
- [12] 周广强, 谢英, 吴剑斌, 等. 基于 WRF-Chem 模式的华东区域 PM2.5 预报及偏差原因[J]. *中国环境科学*, 2016, 36(8): 2251–2259.
- [13] 高怡, 张美根. 2013 年 1 月华北地区重雾霾过程及其成因的模拟分析[J]. *气候与环境研究*, 2014, 19(2): 140–152.
- [14] 张恒德, 咸云浩, 谢永华, 等. 基于时间序列分析和卡尔曼滤波的霾预报技术[J]. *计算机应用*, 2017, 37(11): 279–284.
- [15] 赵秀娟, 徐敬, 张自银, 等. 北京区域环境气象数值预报系统及 PM2.5 预报检验[J]. *应用气象学报*, 2016, 27(2): 160–172.
- [16] WANG T J, JIANG F, DENG J J, et al. Urban air quality and regional haze weather forecast for Yangtze River Delta[J]. *Atmospheric environment*, 2012, 58(15): 70–83.
- [17] 刘杰, 杨鹏, 吕文生, 等. 模糊时序与支持向量机建模相结合的 PM2.5 质量浓度预测[J]. *北京科技大学学报*, 2014, 36(12): 1694–1703.
- [18] 李龙, 马磊, 贺建峰, 等. 基于特征向量的最小二乘支持向量机 PM2.5 浓度预测模型[J]. *计算机应用*, 2014, 34(8): 2212–2216.
- [19] VOUKANTISIS D, KARATZAS K, KUKKONEN J, et al. Inter-comparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki[J]. *Science of the total environment*, 2011, 409(7): 1266–1276.
- [20] MISHRA D, GOLYAL P, UPADHYAY A. Artificial intelligence based approach to fore-cast PM2.5 during haze episodes: a case study of Delhi, India[J]. *Atmospheric environment*, 2015, 120: 239–248.
- [21] LI J D, CHANG J Z, LEI M M. Dynamic forecasting model of short-term PM2.5 concentration based on machine learning[J]. *Journal of computer applications*, 2017, 37(11): 3057–3063.
- [22] JIA C Z, FANG D, YE Y C, et al. Long short-term memory-fully connected(LSTM-FC)neural network for PM2.5 concentration prediction[J]. *Chemosphere*, 2019, 220: 486–492.
- [23] ZHENG Y, YI X, LI M, et al. Forecasting fine-grained air quality based on big data[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, 2015.
- [24] YI X, ZHENG Y, ZHANG J, et al. ST-MVL: filling missing values in geo-sensory time series data[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016.
- [25] VIDUSHI C, ANAND D, VIJAYANAND K, et al. Time Series Based LSTM Model to Predict Air Pollutant's Concentration for Prominent Cities in India[C]//Proceedings of the first International Workshop on Utility-Driven Mining. London, 2018.

- [26] WEBER S A,INSAF T Z,HALL E S,et al. Assessing the impact of fine particulate matter(PM2.5)on respiratory-cardiovascular chronic diseases in the New York city metropolitan area using hierarchical bayesian model estimates[J]. Environment research, 2016,151:399-409.
- [27] CHIOU J H,PING H K. A Deep CNN-LSTM Model for Particulate Matter(PM2.5)Forecasting in Smart Cities[J]. Sensors, 2018,18(7):2220-2241.
- [28] CONG W,SHU L,XIAO J Y,et al. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction[J]. Science of the total environment,2019,654:1091-1099.
- [29] PING W S,JIA W C,JEN W H. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations[J]. IEEE access 6,neurocomputing,2018(6):38186-38200.
- [30] CHU D A,KAUFMAN Y J,ZIBORDI G,et al. Global monitoring of air pollution over land from the Earth observing system-terra moderate resolution imaging spectroradiometer( MODIS) [J]. Geophysis research,2003,108( D21 ):4661-4667.
- [31] KOELEMEEIJER R,HOMAN C,MATTHIJSEN J. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe[J]. Atmosphere environment,2013,40(27):5304-5315.
- [32] SAIDE P E,CARMICHAEL G R,SPAK S N,et al. Forecasting urban PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model[J]. Atmosphere environment,2011,45(16):2769-2780.
- [33] SU X,GOUGH W,SHEN Q. Correlation of pm 2.5 and meteorological variables in Ontario cities:statistical downscaling method coupled with artificial neural network [C]//Proceedings of the 24th International Conference on Modeling, Monitoring and Management of Air Pollution(AIR 2016). Crete,2016.
- [34] LIANG Z,GUANG Z,PEI Y S. Learning Spatiotemporal Features using 3DCNN and Convolutional LSTM for Gesture Recognition[C]//Proceedings of the International Conference on Computer Vision. Venice,2017.
- [35] DU T,BOURDEV L,FERGUS R,et al. Learning spatiotemporal features with 3D convolutional networks[C]//Proceedings of the International Conference on Computer Vision. Santiago,2015.
- [36] STUART G,ELIE B,RENE D. Neural networks and the bias/variance dilemma[J]. Neural computer,1992,41(1):1-58.
- [37] WIKRAM R,PAVAN Y,SHRESTHA M. Deep air: forecasting air pollution in Beijing, China. [C]//Proceedings of IEEE International Conference on Computer Vision. New York:IEEE Press,2017:1-9.
- [38] 赵文芳,王京丽,尚敏,等. 基于粒子群优化和支持向量机的花粉浓度预测模型[J]. 计算机应用,2019,39(1):98-104.

[责任编辑:顾晓天]

(上接第31页)

- [8] Zhang Z,Yang L,Zheng Y. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR),IEEE,Salt Lake City, UT,2018:9242-9251.
- [9] Ronneberger O,Fischer P,Brox T. U-Net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI),Springer,Cham,Munich,2015:234-241.
- [10] Hong S,Oh J,Han B,et al. Learning transferrable knowledge for semantic segmentation with deep convolutional neural network[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR),IEEE,Las Vegas,NV,2016:3204-3212.
- [11] Kingma D P,Ba J. Adam:a method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980,2014.
- [12] <http://www.itksnap.org/pmwiki/pmwiki.php>.
- [13] Shelhamer E,Long J,Darrell T. Fully convolutional networks for semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR),IEEE,Boston,MA,2015:3431-3440.

[责任编辑:陆炳新]