

基于深度时空融合网络的雷达回波外推模型

方巍^{1,2,3,4}, 庞林^{1,4}, 易伟楠¹

- (1. 南京信息工程大学计算机学院数字取证教育部工程研究中心, 江苏南京 210044;
2. 气象灾害国家重点实验室, 北京 100081; 3. 南京信息工程大学江苏省大气环境与装备技术协同创新中心, 江苏南京 210044;
4. 中国气象局气象干部培训学院, 北京 100081)

摘要: 基于深度学习的雷达回波外推是实现短临降水预报的重要方法, 由于雷达回波数据具有显著的非刚性的运动特征, 使得数据的统计特性随时间在不断变化, 意味着其具有高阶非平稳性, 而现有深度学习方法难以捕捉回波序列的非刚性运动特征, 且难以建模雷达数据的高阶非平稳性. 为此, 本文针对雷达数据特征提出了一种新的时空融合网络 STUNNER (Spatio-Temporal Fusion Neural Network). STUNNER 设计了一种两路时空融合架构, 通过交叉连接时间差分网络和时空轨迹网络实现高效的雷达回波外推. 时间差分网络通过引入差分的思想提取高阶非平稳数据中平稳性特征来学习雷达回波的长期趋势, 时空轨迹网络利用动态卷积将卷积循环神经网络中普通卷积固定的参数采样位置优化为随时间变化的动态位置来提取雷达回波的瞬时变化, 同时采用两路融合策略将长期趋势与瞬时变化融合, 实现长短时关联记忆. 所提模型与其他四个模型在两个公开数据集上进行了实验对比. 在雷达回波外推任务中当雷达反射率阈值为 45 dBZ 时, STUNNER 在 POD (Probability Of Detection)、CSI (Critical Success Index)、FAR (False Alarm Rate) 上相比 MIM (Memory In Memory) 分别优化了 0.020, 0.023, 0.043. 实验结果表明新模型在处理雷达回波外推任务上具有更高的准确率.

关键词: 雷达回波外推; 短临降水预报; 高阶非平稳性; 动态卷积; 两路时空融合

基金项目: 国家自然科学基金 (No. 42075007); 灾害天气国家重点实验室开放课题 (No. 2021LASWB19)

中图分类号: TP399

文献标识码: A

文章编号: 0372-2112(2023)09-2526-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211551

Radar Echo Extrapolation Model Based on Deep Spatio-Temporal Fusion Neural Network

FANG Wei^{1,2,3,4}, PANG Lin^{1,4}, YI Wei-nan¹

(1. School of Computer, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science & Technology, Nanjing, Jiangsu 210044, China;

2. State Key Laboratory of Severe Weather, Beijing 100081, China;

3. Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China;

4. Training College of Meteorological Officers of China Meteorological Administration, Beijing 100081, China)

Abstract: Radar echo extrapolation based on deep learning is an important method for precipitation nowcasting. Since radar echo data has significant non-rigid motion characteristics, the statistical characteristics of the data are constantly changing with time, which means that it has high-order non-stationarity. However, it is difficult for existing deep learning methods to capture the non-rigid motion characteristics of echo sequences and model the high-order non-stationarity of radar data. In this paper, we propose a new spatio-temporal fusion neural network (STUNNER) for radar data features, which designs a two-stream spatio-temporal fusion architecture to achieve efficient radar echo extrapolation by cross-connecting the temporal differencing network and the spatio-temporal trajectory network. The temporal differencing network learns the long-term trend of the radar echo by introducing the idea of difference to extract the stationary features in the high-order non-stationary data. The spatio-temporal trajectory network uses dynamic convolution to optimize the fixed parameter sampling position of conventional convolution in the convolutional recurrent neural network to a dynamic position that changes

with time to extract the transient variation of radar echo. And the long-term trend and the transient variation are fused by a two-stream fusion strategy to realize long- and short-term association memory. The proposed model is experimentally compared with four other models on two public datasets. In the radar echo extrapolation task, when the radar reflectivity threshold is 45 dBZ, compared with MIM (Memory in Memory), STUNNER has a 0.020 higher POD (Probability of Detection), 0.023 higher CSI (Critical Success Index) and 0.043 lower FAR (False Alarm Rate). The experimental results show that the innovative model has higher accuracy in dealing with the extrapolation of radar echoes.

Key words: radar echo extrapolation; precipitation nowcasting; high-order non-stationarity; dynamic convolution; two stream spatio-temporal fusion

Foundation Item(s): National Natural Science Foundation of China (No.42075007); Open Grants of the State Key Laboratory of Severe Weather (No.2021LASWB19)

1 引言

近年来,强对流天气愈发频繁,对人们的生命财产安全造成了严重的威胁,及时准确的预报强对流天气的发生及演变过程具有重要意义.短临降水预报的目的为预测未来两小时内的降水强度与降水落区,目前主要采用基于雷达回波外推的方法实现.雷达回波外推的任务为根据已有的雷达回波图像序列去预测未来雷达回波运动趋势.多普勒雷达图像的时间分辨率为6~10 min,雷达回波图像序列为局部区域随时间变化的四维格点数据.该数据在空间维度上用 $M \times N$ 个格点表示固定大小的空间区域,每一个格点包含 C 个通道,在任意时间点对该区域的观测可以用一个矩阵 $\mathbf{X} \in \mathbb{R}^{1 \times M \times N}$ 表示,其中 \mathbb{R} 指观测特征的域.随时间变化的 T 个时间步的观测形成了一组动态序列,用矩阵序列 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T$ 表示.雷达回波外推问题可以建模为在给出之前的 J 帧数据的条件下外推出未来最有可能的 K 帧数据,如式(1)所示.

$$\tilde{\mathbf{X}}_{t+1}, \dots, \tilde{\mathbf{X}}_{t+K} = \underset{\mathbf{X}_{t+1}, \dots, \mathbf{X}_{t+K}}{\operatorname{argmax}} p(\mathbf{X}_{t+1}, \dots, \mathbf{X}_{t+K} | \hat{\mathbf{X}}_{t-J+1}, \dots, \hat{\mathbf{X}}_t) \quad (1)$$

相对于传统的雷达回波外推方法,基于深度学习的方法能够充分的学习海量雷达回波数据中潜在的天气演变规律,并已经在准确率上超越了传统方法.由于雷达序列数据属于高维时空数据,所以高效提取长短期时段中空间分布演变特征、充分学习时空关联性信息即时空融合是提高雷达回波外推准确率的关键性要素.现有研究时空融合可以分为两大类:第一种是单一模型融合,例如使用3D卷积神经网络(3D Convolutional Neural Networks, 3DCNN)^[1-4]或卷积长短期记忆网络(Convolutional Long Short-Term Memory, ConvLSTM)^[5,6]进行时空相关性学习;另一种是两路(two-stream)形式的融合^[7-9].本文采用了后一种方式,在一个网络架构中融合了时间差分网络和时空轨迹网络,以减轻模型学习大量时空信息的负担并提高模型学习效率.

时间差分网络运用了差分思想,学习雷达回波运动

的确定性长期趋势.云层的运动除了整体的迁移外还会伴随着其自身的旋转、放缩、生成和消散等运动形式,具有显著的非刚性运动特征,如图1所示.这种非刚性运动特征进一步造成了雷达图像序列的统计特性随时间在不断变化,即始终会有“新息”(innovation)的引入.新息是当前信号值与预测信号值的差,而预测信号是根据过去信号的统计特性推测出来的,也就是说新息是不可预测的部分.新息的引入表现为相同时间的不同局部位置 and 不同时间的同一局部位置的回波分布存在很大的不同,如图2(a)所示.文献[10]中将这种现象称为高阶非平稳性.由Cramer分解定理,任何一个时间序列都可以分解成为确定项、时间变量多项式与零均值随机项之和,时间差分网络中通过对底层循环网络提取到的时空序列特征进行差分,来提取非平稳序列中确定性的部分,如图2(b)为差分前的信息分布,图2(c)为差分后的信息分布,并利用多层循环结构逐层抽象差分特征,学习长期时段中确定性趋势信息,使得雷达回波的非平稳过程可预测.

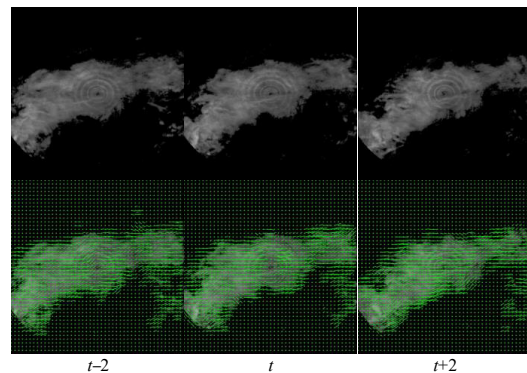


图1 间隔两个时刻的三张雷达图像的运动趋势

时空轨迹网络运用了动态卷积的思想,学习雷达回波运动短期的瞬时变化.该网络优化MIM(Memory In Memory)^[10]中各时间步普通卷积固定的参数采样位置得到随时间变化的动态采样位置.该网络一方面改进了MIM中普通卷积局部不变性参数结构在应对局部变化

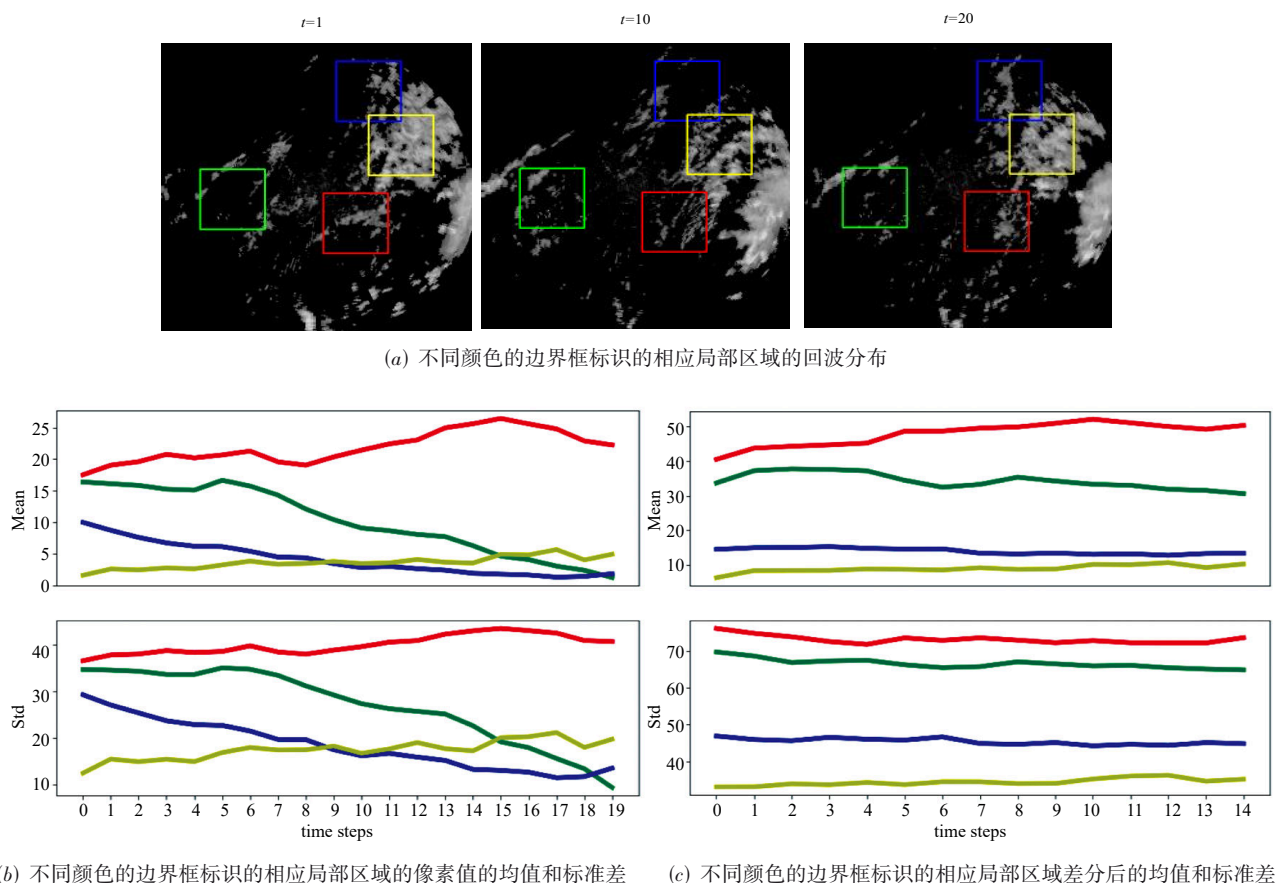


图2 20张连续雷达图差分前后回波分布

的雷达图像时的局限性,构建了一种全时间域的空间动态连接结构,实现对短期时段中瞬时变化的捕获;另一方面继承了MIM复杂化遗忘门的优势来融合时间差分网络提取的长期趋势和动态连接结构学习的瞬时变化,使得该两路融合模型能够对雷达回波序列的高阶非平稳性信息进行建模同时高效学习长短期时空关联信息.

2 相关工作

2.1 雷达回波外推

气象部门主要通过分析雷达图像对临近的天气过程进行预测,采用的方法为雷达回波外推技术.传统的雷达回波外推技术如单体质心法、交叉相关法、光流法等由于算法本身的局限性使得预测的准确率并不高.近年来,基于深度学习的方法取得了一定的进展,并在预报准确率上已经超越了传统的雷达回波外推方法.2015年,Shi等人^[5]将卷积结构整合到长短期记忆单元(Long Short-Term Memory, LSTM)中,提出卷积长短期记忆网络ConvLSTM. ConvLSTM继承了LSTM时序学习优势并整合了卷积结构空间特征提取能力,实现

了时空特征融合,提高了时空序列预测的准确率.此后,基于ConvLSTM的雷达回波外推方法受到广泛应用^[10-12].为了改进ConvLSTM中不适用于雷达回波图像的局部不变的卷积结构,Shi等人进一步提出了轨迹门控循环单元(Trajectory Gated Recurrent Unit, TrajGRU)^[13].该模型针对雷达回波序列的非刚性运动特征而提出了一种卷积采样位置随循环时间步变化的动态连接结构,并设计了一种权重损失函数缓解降水不均问题给模型造成的影响.2019年,Song等人^[14]组合了U-Net, ResNet, Squeeze-and-Excitation等结构为北京气象局构建了实时短临降水预报系统,该系统对未来一帧的雷达图像进行预测,并取得了优于TrajGRU的预测准确率.同年,Tian等人^[15]提出了基于对抗性卷积编码解码网络的GA-ConvGRU(Generative Adversarial Gated Recurrent Unit)模型,该模型以改进的编码解码ConvGRU模型做生成器,卷积神经网络做判别器,博弈学习生成更加清晰的图片,同时克服MSE(Mean-Square Error)倾向于建模单峰分布的问题.2020年,Google research^[16]提出了MetNet模型,该模型根据雷达(MRMS)和卫星(GOES-16)等数据输入,依靠一个时间

信息通道预测未来某个时刻降水的概率分布,模型结构上包括空间下采样器、时序编码器、空间聚合器等模块,是首个在预测未来8小时的时间规模上超越了基于物理模型的数值算法HRRR(High Resolution Rapid Refresh)和光流法的机器学习模型。同年,Jing等人^[17]提出了一种HPRNN(Hierarchical Prediction Recurrent Neural Network)模型,该模型由分层堆叠的RNN模块和一个细化模块组成,目的是为了减轻预测误差随时间的积累,有助于进行长期外推。2021年,DeepMind团队^[18]设计了一种基于条件生成对抗网络的深度生成模型DGMR(Deep Generative Models of Radar)。该模型采用基于ConvGRU的生成器,和两个空间时间上的判别器进行博弈训练,输出未来90min的降水概率。该模型能够更好地捕捉环流强度和结构,并能够更准确地预测回波运动和降水量,同时生成清晰的预测。通过50多位气象专家的系统评估,DGMR对降水的位置、范围、运动和强度等方面的预测与其他广泛使用的临近预报方法相比,准确性和实用性均排名第一。

2.2 动态卷积

Jaderberg等人^[19]提出STN(Spatial Transformer Network)模型,该模型通过学习一种仿射变换来对目标进行矫正,并将单个变换应用于整个特征图。Patraucean等人^[20]通过自编码器优化STN使得每个位置都对应一个变换。Klein等人^[21]学习一个将输入映射成卷积核的函数,实现卷积核参数的动态变化。Brabandere等人^[22]采用了同样的思想,根据输入的不同动态生成特定样本和特定位置的卷积核。Shi等人^[13]设计了TrajGRU,采用动态连接结构来捕捉像素点的运动轨迹。Dai等人^[23]提出的DCN(Deformable Convolutional Networks)模型能够依据输入学习偏移量来动态调整空间采样位置,学习图像高层语义特征,与STN异曲同工。

2.3 现存问题和主要贡献

现有的雷达回波外推方法虽然已经在准确率上超越了传统方法HRRR和光流法,但仍然存在一定的缺陷。TrajGRU结合雷达图像本身的非刚性运动的特点设计一种轨迹预测结构,但却没有学习长期特征,导致生成的雷达回波容易消散。HPRNN模型和MIM分别设计了一种堆叠结构和差分逐层平稳化的方式来对长期的运动模式进行建模,但缺少对瞬时空间变化的学习。针对之前模型所存在的问题,本文对现有模型进行了改进,主要贡献总结如下:

(1) 本文提出一种新的时空融合网络模型STUNNER(Spatio-Temporal Fusion Neural Network),它采用一种两路融合的时空堆叠结构,通过将时间差分网络和时空轨迹网络交叉融合的方式,先分别提取时间上长期趋势和空间上的瞬时变化,再进行二者的时

空融合。一方面减轻了模型学习长时空序列的负担,另一方面使两部分更加专注于长期或瞬时信息的学习。

(2) 本文在STUNNER的时间差分网络中利用了差分操作来提取雷达特征图中的差分特征,建模雷达序列的高阶非平稳性,通过逐时间步差分 and 差分特征,逐层抽象学习序列中确定性长期趋势。

(3) 本文通过实验证明差分虽能够在一定程度上将非平稳信息平稳化,但过度的差分会造成有效信息的削减。MIM^[10]中提出的四层差分结构只保留第一层的差分效果更好(实验结果见第4节)。所以在时间差分网络中只在第一层应用了差分。

(4) 本文通过分析雷达回波图像固有的非刚性运动特征和传统卷积局部不变的局限性,将动态卷积应用到时空融合网络STUNNER的时空轨迹网络中。这种动态卷积操作摒弃了传统卷积固定的参数几何结构和固定的采样位置,利用了空间轨迹上的时间连续性特点来捕获序列中的瞬时变化。并通过将长期趋势与瞬时变化融合达到有效外推,从而提高雷达外推准确率。

3 时空融合网络

在这部分中,主要对时空融合网络STUNNER的网络结构进行详细的描述。本文设计了一种新的时空融合网络结构,网络的整体架构如图3所示。该网络通过交叉连接时间差分网络(Temporal Differencing Network, TDN)和时空轨迹网络(Spatio-Temporal trajectory Network, STN),实现长期趋势和瞬时变化的纵向与横向传递。网络的底层采用了ST-LSTM(Spatio-Temporal LSTM)单元,该循环单元负责抽象特征的提取,从而将时空序列编码进网络。其他层为TDiff-LSTM(Temporal Differencing LSTM)和STraj-LSTM(Spatio-Temporal Trajectory LSTM)堆叠而成的STUNNER-block。网络共四层,底层为ST-LSTM提取输入特征,后三层的第一个时间步为STraj-LSTM单元,输入为底层ST-LSTM的隐状态。剩余时间步为由TDiff-LSTM和STraj-LSTM整合而成的单元模块STUNNER-block,如灰色虚线框内所示。其中第二层TDiff-LSTM的输入为底层ST-LSTM隐状态的差分,剩下两层TDiff-LSTM的输入为前一层TDiff-LSTM的隐状态;第二层STraj-LSTM的输入为底层ST-LSTM的隐状态与同层TDiff-LSTM的隐状态,剩下两层STraj-LSTM的输入为前一层STraj-LSTM的隐状态与同层TDiff-LSTM的隐状态。其中,TDiff-LSTM和STraj-LSTM分别为TDN和STN的循环单元。TDN将ST-LSTM输出状态的差分作为输入,逐层抽象差分特征以学习时空序列中的确定性的长期趋势。在STN中,每层循环单元STraj-LSTM学习一种动态参数采样位置以捕获非刚

性运动的瞬时变化,然后将瞬时变化与从TDiff-LSTM输入的长期趋势进行时空融合实现预测.为方便理解,下面将分别对TDN和STN两个网络的结构进行详细阐述.

3.1 时空长短期记忆网络

时空长短期记忆网络ST-LSTM由Wang等人在文献[11]中提出,文中指出空间相关性和时间动态性是时空预测中的两个重要的方面,并指出之前的LSTM模型倾向于学习时间上的变化,却忽视了空间表现对于外推的重要性.ConvLSTM中输入信息的空间维度被逐层编码,而时间维度彼此独立,导致最底层循环单元会忽略之前的时间步中的最高层的时间信息.ST-LSTM将前一时间步的顶层信息以之字形传递到当前时间步底层,使得同一个记忆单元中同时学习空间表现和时间变化,克服了ConvLSTM中记忆状态受限于同层LSTM内的缺陷,如式(2)~(11)所示.

$$g_t = \tanh(W_{xg} * X_t + W_{hg} * H_{t-1}^l + b_g) \quad (2)$$

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1}^l + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1}^l + b_f) \quad (4)$$

$$C_t^l = f_t \otimes C_{t-1}^l + i_t \otimes g_t \quad (5)$$

$$g'_t = \tanh(W_{xg}' * X_t + W_{mg}' * M_{t-1}^{l-1} + b'_g) \quad (6)$$

$$i'_t = \sigma(W_{xi}' * X_t + W_{mi}' * M_{t-1}^{l-1} + b'_i) \quad (7)$$

$$f'_t = \sigma(W_{xf}' * X_t + W_{mf}' * M_{t-1}^{l-1} + b'_f) \quad (8)$$

$$M_t^l = f'_t \otimes M_{t-1}^{l-1} + i'_t \otimes g'_t \quad (9)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1}^l + W_{co} * C_t^l + W_{mo} * M_t^l + b_o) \quad (10)$$

$$H_t^l = o_t \otimes \tanh(W_{1 \times 1} * [C_t^l, M_t^l]) \quad (11)$$

其中,*和 \otimes 分别表示卷积和Hadamard操作; σ 是sigmoid激活函数;输入序列 X_1, \dots, X_t 、时间记忆状态 C_1, \dots, C_t 、时空记忆状态 M_1, \dots, M_t 、隐状态 H_1, \dots, H_t 和门 $g_t, i_t, f_t, g'_t, i'_t, f'_t, o_t \in \mathbb{R}^{1 \times M \times N}$,其构成都是3维矩阵.ST-LSTM相比与ConvLSTM网络增加了时空记忆状态 M ,如图3红色连接线所示.ST-LSTM中时空记忆状态 M 不同于 C 的水平传递,它可以进行信息的垂直传递,考虑到了高层的抽象信息与低层细节信息的交互.式(6)~(9)中 M 的更新过程与式(2)~(5)中 C 相同.最后,时空记忆状态 M_t^l 与时间记忆状态 C_t^l 融合得到最后的输出 H_t^l .

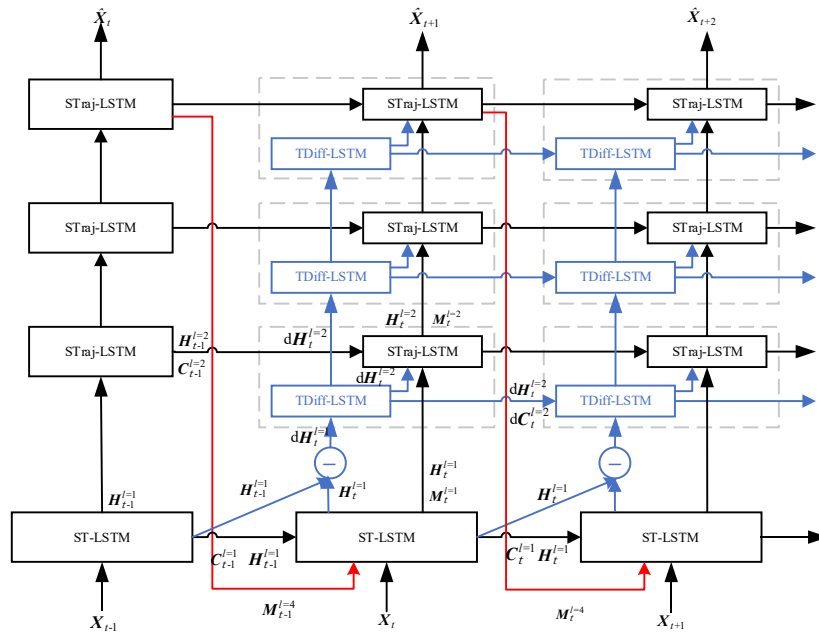


图3 时空融合网络STUNNER.

3.2 时间差分网络

时间差分网络TDN将差分操作引入到时空序列预测中使确定性的部分可预测.网络结构图如图4所示.

该网络共四层,首先底层ST-LSTM对输入序列进行编码:

$$H_t^{l=1}, C_t^{l=1}, M_t^{l=1} = \text{ST-LSTM}(X_t, H_{t-1}^{l=1}, C_{t-1}^{l=1}, M_{t-1}^{l=4}) \quad (12)$$

其中 $M_t^{l=1} = M_{t-1}^{l=4}$,将前一时间步顶层的时空记忆状态 M 垂直传递到底层.TDN的第2层TDiff-LSTM将差分 $dH_t^{l=1} = H_t^{l=1} - H_{t-1}^{l=1}$ 作为输入.根据Cramer分解定理,差分可以将非平稳信息转化为可预测的平稳性信息,但过度差分会在一定程度上造成有用信息的丢失(实验结果见第4节),所以在时间差分网络中仅在第一层运用了差分.TDN采用了堆叠的TDiff-LSTM逐层抽象差分特征:

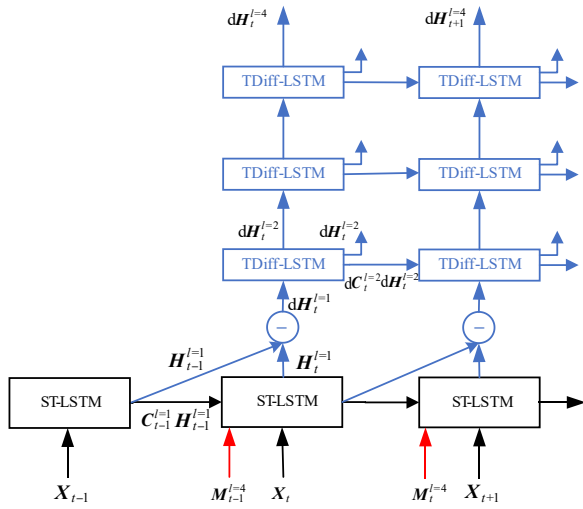


图4 时间差分网络TDN

$$dH_t^l, dC_t^l = \text{TDiff-LSTM}(dH_{t-1}^{l-1}, dH_{t-1}^l, dC_{t-1}^l) \quad (13)$$

其中 $l \in 2, 3, 4$. TDiff-LSTM 的结构如图 5 所示, 与 ConvLSTM 最大的区别在于输入为差分, TDiff-LSTM 的公式表示如 (14)~(21) 所示.

$$dH_t^{l=1} = H_t^{l=1} - H_{t-1}^{l=1} \quad (14)$$

$$H_t^{l=1} = \text{concat}(dH_{t-1}^{l=1} - dC_{t-1}^{l=1}) \quad (15)$$

$$g_t = \tanh(W_{dg} * dH_t^{l=1} + W_{hg} * H_{t-1}^l + b_g) \quad (16)$$

$$i_t = \sigma(W_{di} * dH_t^{l=1} + W_{hi} * H_{t-1}^l + b_i) \quad (17)$$

$$f_t = \sigma(W_{df} * dH_t^{l=1} + W_{hf} * H_{t-1}^l + b_f) \quad (18)$$

$$dC_t^l = f_t \otimes dC_{t-1}^l + i_t \otimes g_t \quad (19)$$

$$o_t = \sigma(W_{do} * dH_t^{l=1} + W_{ho} * H_{t-1}^l + b_o) \quad (20)$$

$$dH_t = o_t \otimes \tanh(W_{1 \times 1} * dC_t^l) \quad (21)$$

其中, TDN 的第二层 TDiff-LSTM 中的所有门 g_t, i_t, f_t, o_t 通过底层两个连续隐状态的差分 $H_t^{l=1} - H_{t-1}^{l=1}$ 和同层差分隐状态 $dH_{t-1}^{l=2}, dC_{t-1}^{l=2}$ 进行更新. 剩下两层 TDiff-LSTM 的所有门通过前一层 TDiff-LSTM 提取到的差分隐状态 $dH_t^{l=1}$ 和同层差分隐状态 dH_{t-1}^l, dC_{t-1}^l 进行更新. TDN 利用时间连续性提取长时差分特征, 并逐层抽象来学习时空序列中长期趋势. 从图 5 可以看出, TDN 的差分细胞状态 dC 是独立的在 TDN 内部的, 只对长期趋势的学习负责. 且每层 STUNNER-block 中都将 TDiff-LSTM 输出的差分隐状态 dH 输入到 STraj-LSTM 中进行时间上的长期趋势和瞬时变化的融合. TDN 不但保证了时间模块学习差分信息的专注性, 还在一定程度上分担了模型学习时空序列中高维时空信息的负担.

3.3 时空轨迹网络

时空轨迹网络 STN 中的网络结构图如图 6 所示.

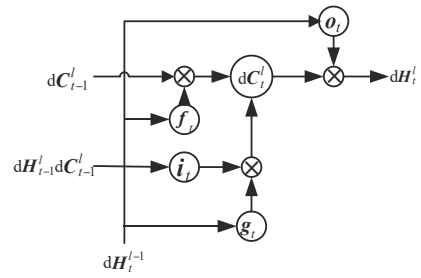


图5 TDiff-LSTM结构

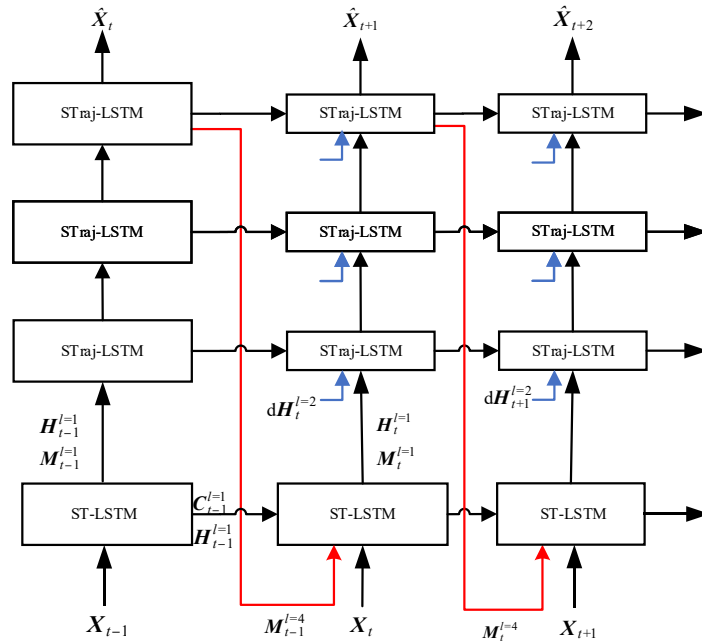


图6 时空轨迹网络STN

网络的底层与TDN共用同一个ST-LSTM进行时空抽象信息的提取. 第二层STraj-LSTM的输入为底层ST-LSTM输出的状态信息 $\mathbf{H}_t^{l-1}, \mathbf{M}_t^{l-1}$. 在STraj-LSTM中, 首先利用动态卷积进行雷达回波非刚性运动特征的提取, 然后与输入的差分隐状态 $d\mathbf{H}_t^l$ 进行时空融合, 其中 $l \in 2, 3, 4$.

STraj-LSTM的结构如图7(a)所示. 文献[10]中指出神经网络中的遗忘门在大多数情况下都是饱和的, 也就是网络的预测是一个线性推理的过程, 即遗忘门饱和后将不再参考后续的输入更新记忆. 所以STraj-LSTM将ST-LSTM中的遗忘门式(4)复杂化, 网络的遗忘门由Differencing Memory(Diff-M)模块取代如图7(a)所示, 用于融合TDiff-LSTM输入的确定性趋势(蓝色箭头). Diff-M模块的结构如图7(b)所示, 公式表示如式(23)~(28).

$$\mathbf{g}_t = \tanh(\mathbf{W}_{dg} * d\mathbf{H}_t^{l-1} + \mathbf{W}_{cg} * \mathbf{C}_{t-1}^l + b_g) \quad (23)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_{di} * d\mathbf{H}_t^{l-1} + \mathbf{W}_{ci} * \mathbf{C}_{t-1}^l + b_i) \quad (24)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{df} * d\mathbf{H}_t^{l-1} + \mathbf{W}_{cf} * \mathbf{C}_{t-1}^l + b_f) \quad (25)$$

$$d\mathbf{C}_t^l = \mathbf{f}_t \otimes \mathbf{S}_{t-1}^l + \mathbf{i}_t \otimes \mathbf{g}_t \quad (26)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{do} * d\mathbf{H}_t^{l-1} + \mathbf{W}_{co} * \mathbf{C}_{t-1}^l + \mathbf{W}_{so} * \mathbf{S}_t^l + b_o) \quad (27)$$

$$\mathbf{T}_t^l = \mathbf{o}_t \otimes \tanh(\mathbf{S}_t^l) \quad (28)$$

其中, \mathbf{S} 表示Diff-M模块中水平传输的内存单元, \mathbf{T} 是通过复杂化遗忘门的输出, $*$ 表示卷积. 为了实现非刚性运动空间运动轨迹的捕获, 本文对ST-LSTM的式(2)、式(3)中隐状态 \mathbf{H} 的局部不变的普通卷积进行了优化, 利用时间连续性使其形成一种动态连接的结构, 捕获瞬时变化. ST-LSTM中的普通卷积对输入特征图 \mathbf{X} 的矩形邻域 $\mathcal{R} = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$ 进行内积计算, 如式(29)定义了一个 3×3 的卷积核, 假设输入全为0且门全为1. $y(p_0)$ 为在点 $p_0(0, 0)$ 处卷积的输出值, p_n 为 \mathcal{R} 的枚举.

$$y(p_0) = \sum_{p_n \in \mathcal{R}} \mathbf{W}_h(p_n) \cdot \mathbf{H}_{t-1}(p_0 + p_n) \quad (29)$$

在STraj-LSTM中, 为了实现动态连接, 对输入的状态 \mathbf{H}_{t-1} 进行warp操作, 如式(30)~(32)所示. 即先将 \mathbf{H}_{t-1} 要求内积的坐标偏移符合几何语义的位置, 如式(30)所示. 偏移量 Δp 是通过一个卷积神经网络学到的, 是一个动态的输出结果, 如式(31)所示. 该卷积神经网络由三层 3×3 的卷积操作构成, 三层卷积分别以 $\mathbf{X}_t, \mathbf{H}_{t-1}$ 前两层输出的拼接作为输入. 再利用双线性插值计算得到偏置后 $p_0 + p_n + \Delta p$ 坐标上的 $\mathbf{H}_{t-1}(p_0 + p_n + \Delta p)$ 值, 如式(32)所示. 最后对 $\mathbf{H}_{t-1}(p_0 + p_n + \Delta p)$ 在循环结构中应用普通卷积对不规则采样位置进行卷积, 如式(33)、式(34)所示, 利用时间连续性和每个时间步动态采样位置的变化形成一种动态连接的结构, 实现雷达回波瞬时变化的捕获.

$$y(p_0) = \sum_{p_n \in \mathcal{R}} \mathbf{W}_h(p_n) \cdot \mathbf{H}_{t-1}(p_0 + p_n + \Delta p) \quad (30)$$

$$\Delta p = \text{CNN}(\mathbf{X}_t, \mathbf{H}_{t-1}) \quad (31)$$

$$\mathbf{H}_{t-1}(p_0 + p_n + \Delta p) = \mathbf{H}_{t-1}(q + p)$$

$$= \sum_q \max(0, 1 - |q_x - p_x|) \max(0, 1 - |q_y - p_y|) \cdot \mathbf{H}_{t-1}(q) \quad (32)$$

STraj-LSTM的公式如式(33)~(42)所示. 该模型的输入 $\mathbf{H}_{t-1}(p_0 + p_n + \Delta p)$ 为warp操作后得到的根据雷达图像特征动态变化的采样位置, 形成时间上的动态连接, 对该采样区域运用普通卷积 $\mathbf{W} * \mathbf{H}_{t-1}(p_0 + p_n + \Delta p)$ 来实现动态卷积. 该动态卷积只应用在前一时间步输入的状态 \mathbf{H}_{t-1}^l 上, 这样既可以提取每一时间步输入 \mathbf{H}_{t-1}^l 的抽象特征信息, 又可以通过隐状态 \mathbf{H}_{t-1}^l 在时间上的流动实现雷达图非刚性运动瞬时变化的捕获. 复杂化的遗忘门Diff-M中记忆输入的长时趋势 $d\mathbf{H}_t^{l-1}$ 信息, 然后将长时趋势信息与瞬时变化融合得到记忆 \mathbf{C}_t^l . 同时 $\mathbf{g}_t', \mathbf{i}_t', \mathbf{f}_t'$ 三个门通过时空记忆状态 \mathbf{M}_t^{l-1} 更新. 并且两种记忆类型 \mathbf{C}, \mathbf{M} 使用共享的输出门实现无缝记忆融合, 可以有效地模拟时空序列中的形状变形和运动轨迹.

$$\mathbf{g}_t = \tanh(\mathbf{W}_{xg} * \mathbf{H}_t^{l-1} + \mathbf{W}_{hg} * \mathbf{H}_{t-1}^l(p_0 + p_n + \Delta p) + b_g) \quad (33)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_{xi} * \mathbf{H}_t^{l-1} + \mathbf{W}_{hi} * \mathbf{H}_{t-1}^l(p_0 + p_n + \Delta p) + b_i) \quad (34)$$

$$\mathbf{T}_t^l = \text{Diff-M}(d\mathbf{H}_t^{l-1}, \mathbf{C}_{t-1}^l, \mathbf{S}_{t-1}^l) \quad (35)$$

$$\mathbf{C}_t^l = \mathbf{T}_t^l + \mathbf{i}_t \otimes \mathbf{g}_t \quad (36)$$

$$\mathbf{g}_t' = \tanh(\mathbf{W}_{xg}' * \mathbf{H}_t^{l-1} + \mathbf{W}_{mg}' * \mathbf{M}_t^{l-1} + b_g') \quad (37)$$

$$\mathbf{i}_t' = \sigma(\mathbf{W}_{xi}' * \mathbf{H}_t^{l-1} + \mathbf{W}_{mi}' * \mathbf{M}_t^{l-1} + b_i') \quad (38)$$

$$\mathbf{f}_t' = \sigma(\mathbf{W}_{xj}' * \mathbf{H}_t^{l-1} + \mathbf{W}_{mj}' * \mathbf{M}_t^{l-1} + b_f') \quad (39)$$

$$\mathbf{M}_t^l = \mathbf{f}_t' \otimes \mathbf{M}_t^{l-1} + \mathbf{i}_t' \otimes \mathbf{g}_t' \quad (40)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{xo} * \mathbf{H}_t^{l-1} + \mathbf{W}_{ho} * \mathbf{H}_{t-1}^l + \mathbf{W}_{co} * \mathbf{C}_t^l + \mathbf{W}_{mo} * \mathbf{M}_t^l + b_o) \quad (41)$$

$$\mathbf{H}_t^l = \mathbf{o}_t \otimes \tanh(\mathbf{W}_{1 \times 1} * [\mathbf{C}_t^l, \mathbf{M}_t^l]) \quad (42)$$

4 实验结果与分析

这部分将在人工合成数据集和真实世界数据集上评估STUNNER模型的表现. 所有实验中模型都为四层, 网络结构如图3所示, 每一层的特征通道数都为64. 实验所设计代码均是基于3.6版本的Python语言开发, 深度学习网络模型主要通过PyTorch开源神经网络库实现, 并且使用初始学习率为 10^{-3} 的ADAM优化器优化模型. 损失函数定义为L1+L2损失来同时增强生成图片的锐度和平滑度. 文中实验的batch_size大小为10, 且为了减少协变量偏移问题, 对模型采用了层归一化. 模型训练采用了提前停止策略, 当超过5次迭代验证集损失没有下降则训练停止. 文中所有其他方法的结果

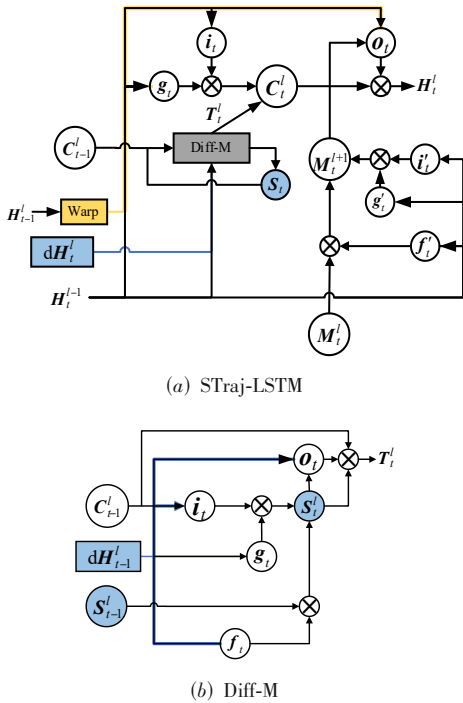


图7 STraj-LSTM与STraj-LSTM中复杂化遗忘门模块Diff-M的结构

均来源于本文的实验. 硬件上, 结合 GPU 加速深度学习实验的经验, 采用 NVIDIA GeForce GTX 1080Ti 进行 GPU 并行计算加速.

4.1 模拟数据集上的实验结果与分析

该数据集为人工合成数据集, 使用文献[24]中描述的方法生成数据集. 该数据集每一个样本中包含 20 帧连续的图片, 预测任务是根据前 10 帧预测后 10

帧图像. 每一帧图片又包含两个或三个手写数字, 数字在 64×64 格点的图像中做无规律的弹跳运动. 这些手写数字都是从 MNIST 数据集 (<http://yann.lecun.com/exdb/mnist/>) 中随机的选取的. Moving MNIST 数据集的生成首先随机初始化数字的位置, 然后以指定的速度和随机的方向运动. 当数字触碰到边界或另一个数字时会发生反弹^[5]. 这些生成过程中的随机因素给模型预测增加了难度, 模型没有建模这种瞬时变化的能力很难做到准确的预测. 该实验采用这种方法生成无限大小的训练集, 验证集和测试集为固定的 5 000 个序列. 数据维度为: $(B, 20, 64, 64, 1)$, B 为 batch_size, 时间维度 20 个图像帧中, 前 10 张为输入, 后 10 张为目标标签. 为了加速训练, 训练时将图像分割为 4 份补丁并在通道维度上拼接为 $(B, 20, 16, 16, 16)$ 的数据维度. 文中的模型用单帧图像包含两个数字的数据集训练, 为了验证模型的泛化和迁移能力, 本文实验在另外的每帧图像包含三个数字的数据集上测试.

MIM^[10]中将每一层的差分作下一层的输入, 但过多的差分在一定程度上会削减重要信息, 所以本文在 MIM 的基础上进行改进提出 MIM_r, 只在第一层使用了差分. 此外, 本实验还对模型的层数进行了实验, 实验结果表明 4 层的架构能够在模型参数量和准确度上实现平衡. 本文中对特征通道数为 64 和 128 分别进行了实验, 实验结果如表 1 所示. 特征通道数为 128 时模型大小增加为特征通道数为 64 时的 4 倍, 但模型表现相较于特征通道数均为 64 时并无显著增加, 所以本文中每一层的特征通道数设为 64.

表 1 在 Moving MNIST-2 测试集上的特征通道数实验 (较低 MSE 表示更好的预测准确性)

模型	MNIST-2(SSIM/frame)	MNIST-2(MSE/frame)	参数量	FLOPS/序列
STUNNER-64	0.863	39.2	10.694 M	54.126 G
STUNNER-128	0.879	38.1	41.869 M	211.854 G

为了进一步证明网络中各个模块的必要性, 通过分别拆卸 TDN 网络、STN 中的 Diff-M 和 warp 模块来对 STUNNER 进行消融实验, 实验结果如表 2 所示. 没有 TDN 网络的表现比没有 Diff-M 和 warp 模块的 STUNNER 稍好一些, 没有 warp 模块表现最差. 三者中的任何一个在 MSE 和 SSIM 上的相对 MIM 都有显著的提升, 说明了将他们嵌入在 STUNNER 中的必要性. 同时具备 TDN 网络,

warp 模块和 Diff-M 模块的 STUNNER 模型达到最佳性能.

如表 3 所示, STUNNER 将每帧的 MSE 损失降低到了 39.2, 相比其他模型得到了最低 MSE 损失. 图 8 可视化了预测一个 Moving MNIST 测试样本的逐帧对比结果. 该序列在预测阶段出现了两个数字重合的现象, 使得预测更具有挑战性. 可以看出 MIM 的预测和 STUNNER 最为清晰, 在一定程度上解决了外推模糊的

表 2 在 Moving MNIST-2 测试集上的消融实验 (较低 MSE 表示更好的预测准确性)

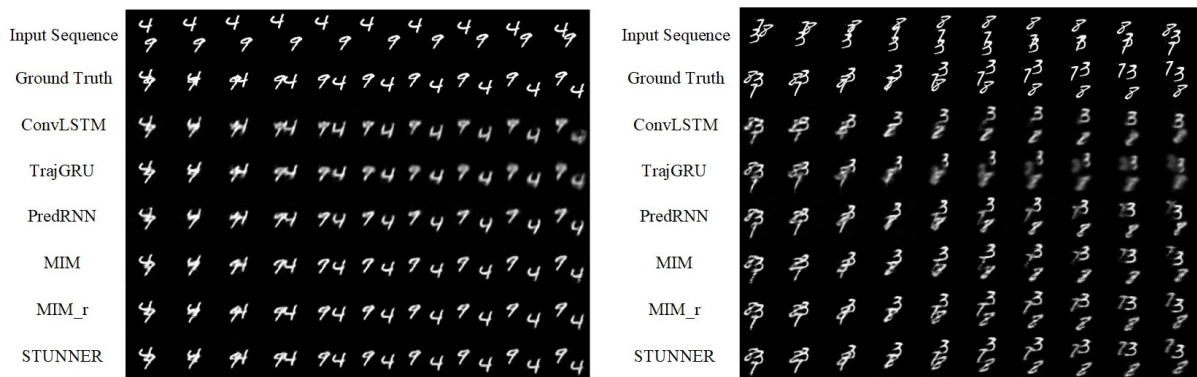
模型	MNIST-2(SSIM/frame)	MNIST-2(MSE/frame)	参数量	FLOPS/序列
STUNNER w/o Diff-M	0.846	44.2	8.235 M	41.536 G
STUNNER w/o TDN	0.855	41.9	8.235 M	42.165 G
STUNNER w/o warp	0.834	47.1	9.256 M	46.761 G
STUNNER	0.863	39.2	10.694 M	54.126 G

问题. 但 MIM 的预测逐渐将正确的数字“9”预测偏离为错误的数字“7”, 而只在第一层进行差分的 MIM_r 却得到了正确的预测结果, 说明过多的差分确实会导致信息的损失. 从图 9 可以看出, TrajGRU 刚开始表现的比较好, 但出现了十分严重的退化现象, 这是因为基于光流的方法使用局部变化状态转换来捕捉短期的时间依赖性, 但缺乏对长期运动趋势的表征.

STUNNER 得到了最佳的表现是因为它能够通过时间差分网络 TDN 中的差分操作建模序列的时空非平稳性, 学习长期的趋势; 同时又能够通过时空轨迹网络 STN 学习时空关联性, 时间空间分离减轻模型学习的负担并记住更多的时空信息帮助外推. 从图 8(b) 和图 9(b) 中可以看出 STUNNER 同样具有很好的泛化表现.

表 3 在 Moving MNIST 测试集上生成序列平均每帧的 MSE (较低 MSE 表示更好的预测准确性)

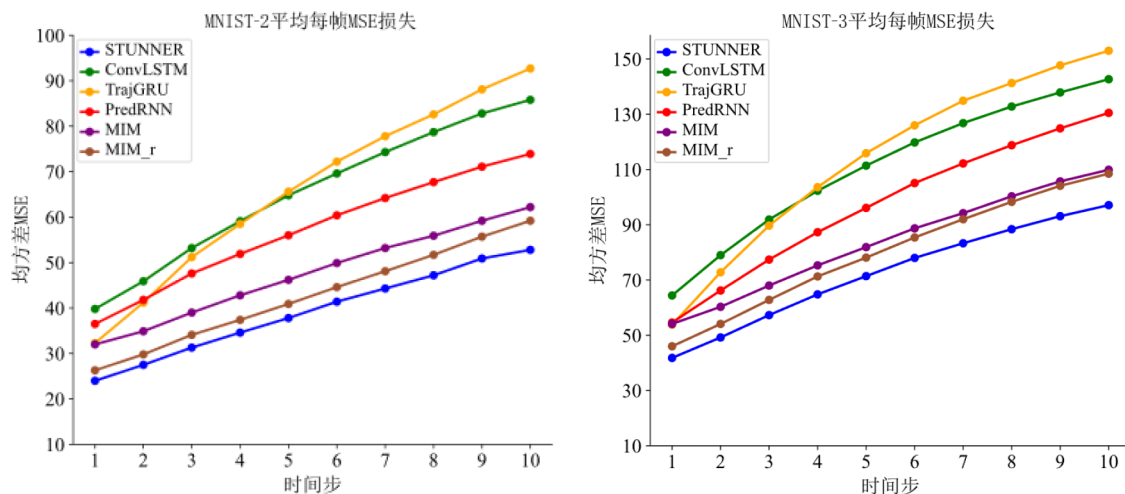
模型	MNIST-2(SSIM/frame)	MNIST-2(MSE/frame)	MNIST-3(MSE/frame)	参数量	FLOPS/序列
ConvLSTM ^[5]	0.743	65.4	110.9	4.619 M	11.824 G
TrajGRU ^[13]	0.756	66.2	113.9	2.301 M	10.890 G
PredRNN ^[11]	0.788	57.1	97.3	6.054 M	30.996 G
MIM ^[10]	0.830	47.5	83.8	3.866 M	19.792 G
MIM_r	0.850	42.8	80.1	3.866 M	19.792 G
STUNNER	0.863	39.2	72.4	10.694 M	54.126 G



(a) 测试集在 MNIST-2 上的表现

(b) 测试集在 MNIST-3 上的泛化表现

图 8 在 Moving MNIST-2 和 MNIST-3 测试集上的测试实例



(a) 测试集在 MNIST-2 上的表现

(b) 测试集在 MNIST-3 上的表现

图 9 不同模型在 Moving MNIST 测试集上的逐帧 MSE 比较

4.2 雷达回波数据集上的实验结果及分析

本文中使用的雷达回波数据集来自上海市气象学会

和上海进博会气象台联合举办的全国人工智能天气预报创新大赛. 雷达数据集覆盖上海市约 1 km×0.866 km 的

区域,水平分辨率为 0.01° ,网格数量为 460×460 .每个雷达数据序列样本覆盖时长为3 h,输入数据时长为1 h,目标数据时长为2 h.输入数据的时间分辨率为6 min,目标数据的时间分辨率为12 min.一个样本序列的雷达图像共20个时次,实验需求为根据前10个时次的雷达图像外推后10个时次的雷达图像,从而实现根据前一个小时的雷达回波外推后两小时的回波分布.雷达数据的反射率数据为组合反射率.雷达回波通过相关系数(CC)质量控制,数据范围为15~70 dBZ,大于70 dBZ的值设为70,小于15 dBZ的值和缺测值都设为0,这是由于降水回波的反射率一般在15 dBZ以上.大赛提供的雷达数据以灰度图PNG格式存储,图片灰度值和组合反射率的关系为: $G = \text{round}(255 \times \frac{\text{dBZ}}{70})$.对于数据处理,首先将空间尺度由 460×460 缩小到 100×100 ;然后进行最大值归一化;接着读取灰度图像为Numpy数组并将样本内数据在时间维度上拼接;最后将所有样本合并成为一个五维的数组,其中训练集包含40 000个样本,数据维度为:(40 000, 20, 100, 100, 1),时间维度的20个图像帧中,前10个为输入,后10个为目标标签.由于强对流天气雷达图像值的长尾分布,导致雷达回波外推相比于Moving MNIST更具有挑战性.雷达回波数据集中模型参数与Moving Mnist数据集中一致.同样在训练时将图像分割为4份补丁并在通道维度上拼接

为(40 000, 20, 25, 25, 16)的数据维度.

图10展示了这5个模型的一个雷达回波外推测试实例.从图中可以看出,雷达图像的演化是一个具有非刚性运动特征的高阶非平稳性过程.这种时间和空间上的对未来的高度不确定性给雷达回波外推带来了很大的挑战.在这个实例中,左上角的雷达回波在生成,只有MIM和STUNNER能够正确的捕获到这种雷达回波的生成趋势,这是因为二者都具备建模高阶非平稳性的能力,使得雷达图像生成过程更加可预测.此外,相比MIM,STUNNER对雷达图像中部区域的强对流聚集趋势的预测表现更好,这是因为相比于MIM,STUNNER具备独立学习长期趋势的时间差分网络,同时两路融合的结构使得模型更加专注于从长期和短期中把握全局的动态规律.除此之外,ConvLSTM相比PredRNN缺少层间信息交互的能力,且都不具备建模高阶非平稳信息的能力,所以表现趋于模糊.TrajGRU在刚开始表现较好,是因为TraGRU能够利用提取到的光流动态特征捕捉短期的运动趋势,但由于缺少学习长期趋势的能力使得TrajGRU表现迅速下降.STUNNER的时空轨迹网络中不仅集成了TrajGRU动态学习轨迹特征的优势,做到及时抓取空间上的轨迹变化,同时具备MIM建模高阶非平稳信息的能力,掌握时间上长期确定性趋势.STUNNER两路融合结构进行时空融合,高效学习长短期信息从而提升模型预测准确率.

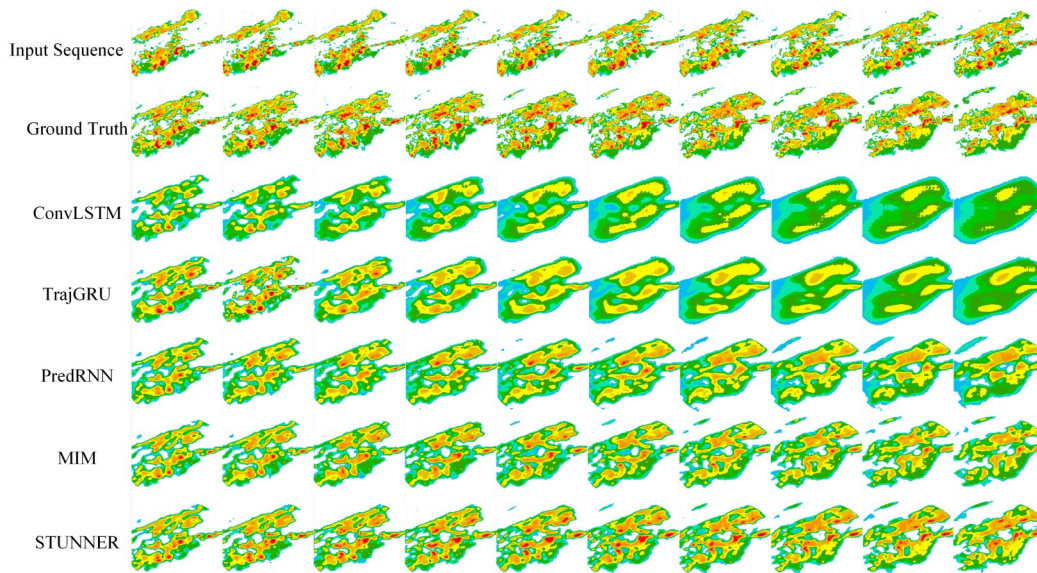


图10 五个模型在测试实例上可视化外推的结果

STUNNER通过利用两路时空融合的方式提高模型学习长期和短期信息的效率,相比MIM性能得到进一步提升.STUNNER不仅可以预测回波的长期运动趋势而且随着外推时间步的增加依然能够保留雷达图像的细节.图11展示了具有明显强对流的生消演

变过程的雷达回波外推个例.图11的左侧和图12的下方有局部强对流的生成,可以看出STUNNER对生成过程具有一定的预测能力.图12中心的强对流有明显的消散趋势,STUNNER对消散过程也能做到较为准确的预测.

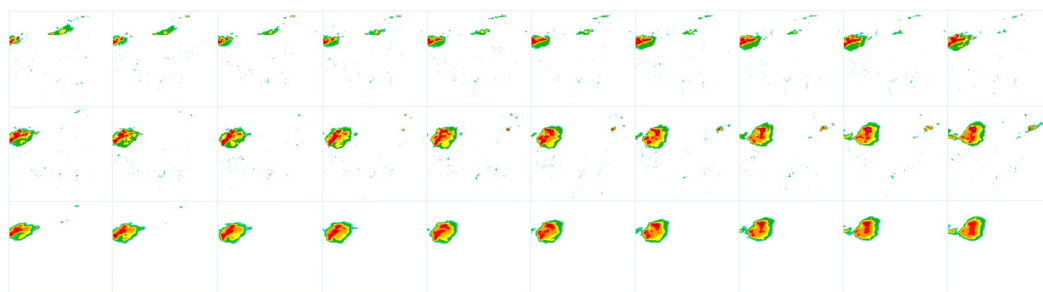


图 11 第一行为输入序列,第二行为目标序列,第三行为 STUNNER 的预测

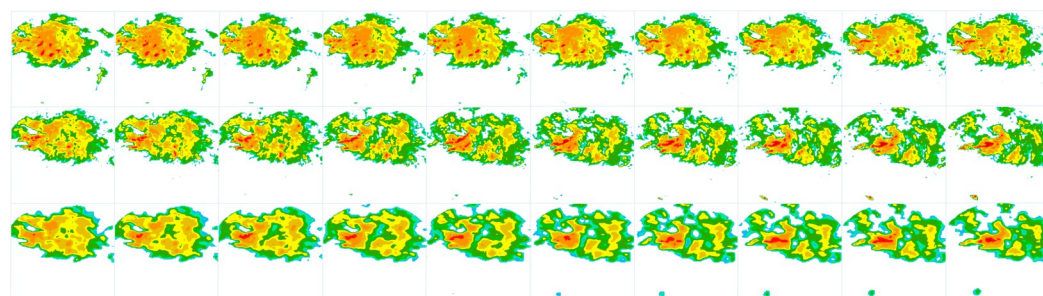


图 12 第一行为输入序列,第二行为目标序列,第三行为 STUNNER 的预测

本文采用降水短时预报指标来评估实验的结果,这些指标分别为:临界成功指数 CSI (Critical Success Index), 命中率 POD (Probability Of Detection) 和误报率 FAR (False Alarm Rate). 在计算 CSI、POD、FAR 时,首先将像素值转换为以 dBZ 为单位的雷达回波强度,然后以某个固定阈值为界将数据转换为 0/1 二值化数据.接着计算匹配数 hits (prediction=1, truth=1),漏报数 misses (prediction=0, truth=1) 和错报数 falsealarms (prediction=1, truth=0) 的值.三个评分指标的计算公式为 $CSI = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{falsealarms}}$, $POD = \frac{\text{hits}}{\text{hits} + \text{misses}}$, $FAR = \frac{\text{falsealarms}}{\text{hits} + \text{falsealarms}}$, 其中,更高的 CSI 和 POD 表示更好的预测效果,而 FAR 越高则表明模型的误报越多.

表 4~6 为 4 个对比模型 ConvLSTM、TrajGRU、PredRNN 和 MIM 与 STUNNER 在 MSE 和三个评价指标 POD、CSI、FAR 上的表现,评估的是测试集中外推雷达图像第 5 帧和最后一帧的输出,也就是模型一个小时和两个小时后的外推效果.四个对比模型 ConvLSTM、TrajGRU、PredRNN 和 MIM 同样被提出用于处理雷达回波外推问题.由于雷达回波着重关注的是 35 dBZ 以上的强对流情况,所以本文实验选取了反射率 35 dBZ、40 dBZ 和 45 dBZ 作为对应阈值来进行模型评估.雷达反射率阈值越大代表发生恶劣天气条件的概率越高.由于计算指标时,反射率阈值越高过滤掉的信息更多,所以当阈值越大,效果自然会下降.从数据中同样可以看出 TrajGRU 由于只专注于

学习瞬时变化导致衰减的速度最快.而 STUNNER 借鉴了 MIM 和 TrajGRU 的优势,在长期和短期的预测表现上最为平稳.从 3 个表中可以看到,同样是预测未来两个小时的雷达回波,5 个模型在 3 个反射率检验阈值的 CSI、POD、FAR 的表现 STUNNER 均优于其他 4 种模型.当雷达反射率阈值为 35 dBZ 时,STUNNER 在 3 个评价指标 POD、CSI、FAR 上比 MIM 分别优化了 0.016, 0.014, 0.005; 当雷达反射率阈值为 40 dBZ 时,STUNNER 在 POD、CSI、FAR 上比 MIM 分别优化了 0.021, 0.024, 0.032; 当雷达反射率阈值为 45 dBZ 时,STUNNER 在 POD、CSI、FAR 上比 MIM 分别优化了 0.020, 0.023, 0.043. 实验结果表明 STUNNER 模型能够在雷达回波外推方法中进一步提高降水预报的质量.

表 4 反射率 ≥ 35 dBZ,外推 10 个未来回波图像帧时,5 个模型在 4 个评价指标上的评估结果

外推时间	模型	POD	CSI	FAR	MSE
60 min	ConvLSTM ^[5]	0.584	0.523	0.166	138.824
	TrajGRU ^[13]	0.580	0.516	0.176	140.916
	PredRNN ^[11]	0.615	0.547	0.168	82.051
	MIM ^[10]	0.639	0.575	0.148	67.075
	STUNNER	0.653	0.592	0.136	64.113
120 min	ConvLSTM ^[5]	0.406	0.358	0.248	194.485
	TrajGRU ^[13]	0.398	0.349	0.261	200.327
	PredRNN ^[11]	0.497	0.429	0.242	82.686
	MIM ^[10]	0.527	0.461	0.214	84.360
	STUNNER	0.543	0.475	0.209	79.570

表 5 反射率 ≥ 40 dBZ,外推 10 个未来回波图像帧时,5 个模型在 4 个评价指标上的评估结果

外推时间	模型	POD	CSI	FAR	MSE
60 min	ConvLSTM ^[5]	0.446	0.378	0.287	138.824
	TrajGRU ^[13]	0.440	0.371	0.297	140.916
	PredRNN ^[11]	0.519	0.452	0.222	82.051
	MIM ^[10]	0.538	0.474	0.201	67.075
	STUNNER	0.556	0.493	0.187	64.113
120 min	ConvLSTM ^[5]	0.267	0.216	0.469	194.485
	TrajGRU ^[13]	0.256	0.204	0.498	200.327
	PredRNN ^[11]	0.401	0.333	0.337	82.686
	MIM ^[10]	0.424	0.359	0.299	84.360
	STUNNER	0.445	0.383	0.267	79.570

表 6 反射率 ≥ 45 dBZ,外推 10 个未来回波图像帧时,5 个模型在 4 个评价指标上的评估结果

外推时间	模型	POD	CSI	FAR	MSE
60 min	ConvLSTM ^[5]	0.367	0.298	0.387	138.824
	TrajGRU ^[13]	0.353	0.285	0.403	140.916
	PredRNN ^[11]	0.417	0.349	0.318	82.051
	MIM ^[10]	0.443	0.379	0.276	67.075
	STUNNER	0.461	0.399	0.252	64.113
120 min	ConvLSTM ^[5]	0.187	0.142	0.629	194.485
	TrajGRU ^[13]	0.168	0.125	0.672	200.327
	PredRNN ^[11]	0.296	0.231	0.487	82.686
	MIM ^[10]	0.328	0.263	0.427	84.360
	STUNNER	0.348	0.286	0.384	79.570

5 结论

本文提出了一种时空融合网络 STUNNER,该网络由时间差分网络和时空轨迹网络两部分交叉融合而成.时间差分网络中引入了时间序列的差分操作学习序列中的长期运动趋势.时空轨迹网络中利用了动态卷积捕获瞬时的变化,同时以两路融合的方式将瞬时变化信息与长期运动趋势融合,从而对时空序列的高阶非平稳性建模,实现更加稳定精确的外推.模型在两个数据集上进行了对比实验, Moving MNIST 和雷达回波数据上的实验结果表明,本文所提出的时空融合网络 STUNNER 相比其他四个模型 ConvLSTM、TrajGRU、PredRNN 和 MIM 在准确率上得到了进一步的提升.当外推未来两小时雷达图像且雷达反射率阈值为 45 dBZ 时, STUNNER 在 POD、CSI、FAR 上三个评价指标上相较于 MIM 模型分别优化了 0.020, 0.023, 0.043.

参考文献

[1] JI S W, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013,

35(1): 221-231.

[2] ZHOU Y Z, SUN X Y, LUO C, et al. Spatiotemporal fusion in 3D CNNs: A probabilistic view[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9826-9835.

[3] WANG L M, LI W, LI W, et al. Appearance-and-relation networks for video classification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1430-1439.

[4] ZHOU Y Z, SUN X Y, ZHA Z J, et al. MiCT: Mixed 3D/2D convolutional tube for human action recognition[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 449-458.

[5] SHI X J, CHEN Z R, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. Montreal: MIT Press, 2015: 802-810.

[6] DÜZÇEKER A, GALLIANI S, VOGEL C, et al. DeepVideoMVS: Multi-view stereo on video with recurrent spatiotemporal fusion[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 15319-15328.

[7] TULYAKOV S, LIU M Y, YANG X D, et al. MoCoGAN: Decomposing motion and content for video generation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1526-1535.

[8] WANG L M, XIONG Y J, WANG Z, et al. Temporal segment networks: Towards good practices for deep action recognition[C]//European Conference on Computer Vision. Cham: Springer, 2016: 20-36.

[9] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1. Montreal: MIT Press, 2014: 568-576.

[10] WANG Y B, ZHANG J J, ZHU H Y, et al. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9146-9154.

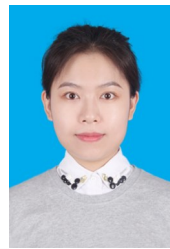
[11] WANG Y B, LONG M S, WANG J M, et al. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs[C]//Proceedings of the 31st International Conference on Neural Information Processing

- Systems. Long Beach: Curran Associates Inc, 2017: 879-888.
- [12] LIN Z H, LI M M, ZHENG Z B, et al. Self-attention ConvLSTM for spatiotemporal prediction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 11531-11538.
- [13] SHI X J, GAO Z H, LAUSEN L, et al. Deep learning for precipitation nowcasting: A benchmark and a new model [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc, 2017: 5617-5627.
- [14] SONG K, YANG G W, WANG Q X, et al. Deep learning prediction of incoming rainfalls: An operational service for the city of Beijing China[C]//2019 International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE, 2020: 180-185.
- [15] TIAN L, LI X T, YE Y M, et al. A generative adversarial gated recurrent unit model for precipitation nowcasting [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(4): 601-605.
- [16] SØNDERBY C K, ESPEHOLT L, HEEK J, et al. MetNet: A neural weather model for precipitation forecasting [EB/OL]. (2020-03-24) [2021-09-10]. <https://arxiv.org/abs/2003.12140>.
- [17] JING J R, LI Q, PENG X, et al. HPRNN: A hierarchical sequence prediction model for long-term weather radar echo extrapolation[C]//ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2020: 4142-4146.
- [18] RAVURI S, LENC K, WILLSON M, et al. Skilful precipitation nowcasting using deep generative models of radar [J]. Nature, 2021, 597(7878): 672-677.
- [19] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2. Montreal: MIT Press, 2015: 2017-2025.
- [20] PATRAUCEAN V, HANDA A, CIPOLLA R. Spatio-temporal video autoencoder with differentiable memory [EB/OL]. (2015-11-25) [2021-09-10]. <https://arxiv.org/abs/1511.06309>.
- [21] KLEIN B, WOLF L, AFEK Y. A dynamic convolutional layer for short rangeweather prediction[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 4840-4848.
- [22] BRABANDERE D B, JIA X, TUYTELAARS T, et al. Dynamic filter networks[M]//Advances in Neural Information Processing Systems. Montreal: MIT Press, 2016: 667-675.
- [23] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 764-773.
- [24] SRIVASTAVA N, MANSIMOV E, SALAKHUTDINOV R. Unsupervised learning of video representations using LSTMs[C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37. New York: ACM, 2015: 843-852. .

作者简介



方巍 男,1975年7月出生于安徽省黄山市. 硕博毕业于苏州大学计算机专业. 现为南京信息工程大学计算机学院、软件学院、网络空间安全学院教授、硕士生导师. 研究方向为人工智能、气象大数据分析、深度学习和云计算.
E-mail: hsfangwei@sina.com



庞林 女,1996年9月出生于山东省枣庄市. 硕士毕业于南京信息工程大学计算机学院、软件学院、网络空间安全学院. 研究方向为深度学习、雷达回波外推.
E-mail: panglin0912@126.com



易伟楠 男,1996年10月出生于江苏省南通市. 硕士毕业于南京信息工程大学计算机学院、软件学院、网络空间安全学院. 研究方向为深度学习、行人重识别.
E-mail: lxywn96@aliyun.com