

耦合编码解码卷积长短期记忆网络求解时变偏微分方程的参数识别问题

郭 琦¹, 贾宏恩^{1,2*}, 王鸿斌³

¹太原理工大学数学学院, 山西 太原

²密码关键技术创新与融合应用实训实验室, 山西 太原

³山西工学院通识教育学院, 山西 朔州

收稿日期: 2025年1月24日; 录用日期: 2025年2月17日; 发布日期: 2025年2月25日

摘要

文章基于编码、解码结构提出了一种同时反演参数和求解偏微分方程的耦合编码解码卷积长短期记忆网络(ED-ConvLSTM)。编、解码结构中分别使用卷积层和转置卷积层, 用于提取和恢复物理约束的空间信息。为了有效逼近时变偏微方程中的时间演化规律, ED-ConvLSTM引入编解码模块。通过几个数值试验证明所提出的耦合ED-ConvLSTM网络框架能够有效准确地预测场变量并且反演未知参数, 特别是对于高噪声水平的输入序列。

关键词

参数识别, 编码, 解码, ConvLSTM, 时变偏微分方程

Coupled Encoder-Decoder ConvLSTM Structure for the Parameter Identification Problems of Time-Dependent Partial Differential Equations

Qi Guo¹, Hong'en Jia^{1,2*}, Hongbin Wang³

¹School of Mathematics, Taiyuan University of Technology, Taiyuan Shanxi

²Key Cryptographic Technology Innovation and Integrated Application Training Laboratory, Taiyuan Shanxi

³College of General Education, Shanxi College of Technology, Shuzhou Shanxi

Received: Jan. 24th, 2025; accepted: Feb. 17th, 2025; published: Feb. 25th, 2025

*通讯作者。

文章引用: 郭琦, 贾宏恩, 王鸿斌. 耦合编码解码卷积长短期记忆网络求解时变偏微分方程的参数识别问题[J]. 应用数学进展, 2025, 14(2): 341-355. DOI: [10.12677/aam.2025.142075](https://doi.org/10.12677/aam.2025.142075)

Abstract

This paper proposes a coupled encoder-decoder ConvLSTM network (ED-ConvLSTM) based on an encoder and decoder to identify parameters and solve partial differential equations simultaneously. Convolutional and transposed convolutional layers are utilized in the encoder and decoder modules to extract and retrieve spatiotemporal information in physical laws. ConvLSTM modules are embedded into the encoder and decoder modules to accurately approximate the time evolution in time-dependent PDEs. Through intensive numerical examples, the coupled neural network can predict field variables and unknown parameters accurately and efficiently, especially for the input with large noise.

Keywords

Parameter Identification, Encoder, Decoder, ConvLSTM, Time-Dependent PDE

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

卷积神经网络(CNN)在诸多领域已经获得了许多显著的成就,如图像识别[1]、视频分类[2]、自然语言识别[3]等等。随着研究的不断深入,残差网络、稠密连接卷积网络、VGG、AlexNetCNN等卷积神经网络变体不断涌现,研究者逐渐将目光转移至偏微分方程(PDEs)求解领域。由于卷积网络能够从局部空间中自动提取特征且具有平移不变性,这有效地减少模型训练参数和缩短模型训练时间。此外,Zhou等人[4]通过数值试验说明深度卷积神经网络对复杂函数的建模能力,并证实其具有普遍逼近性。Hamzehloo等人[5]利用CNN将原始二维泊松方程转化为一个齐次问题和四个非齐次问题,然后通过求解五个子问题得到原方程的解。Long等人[6]受到小波理论的启发提出了PDE-Net,通过训练卷积核逼近微分算子并在训练前对所有滤波器进行初始化约束,使得该模型实现对动力系统相对长时间的演化。Mo等人[7]利用深度卷积结构搭建编解码网络模型,将其应用于多相流动态预测问题,并通过数值算例证实该方法能够准确预测2,500维磁导率场变化趋势。Han等人[8]提出物理信息集合自适应卷积神经网络(PhyGeoNet),利用椭圆坐标映射实现规则参考域卷积和不规则参考域卷积间的相互转化,并通过求解稳态方程证实该模型的有效性和准确性。Li等人[9]提出了一种新的神经算子,它通过在傅里叶空间参数化积分核学习无穷维函数空间之间的映射,从而提高偏微分方程求解的精度。

为了有效地预测时变偏微分方程的潜在解,除了对空间相关性的考察,同时也需要考虑时间信息对PDEs解的影响。递归神经网络(RNN)是一种经典的时序数列处理方法,长短期记忆网络(LSTM)作为一种特殊的RNN,通过引入门单元结构实现有效控制信息流的目的[10]。相较于标准RNN,LSTM能更好地捕捉序列中的长期依赖关系,并在长短期时间序列中均表现优异。Hu等人[11]提出Bi-LSTM网络用于模拟多维度偏微分方程,文章证实该模型能在20个训练循环内有效学习PDEs并生成高精度解。这证实了LSTM能够根据先验周期信息对PDEs进行长期的更新和演化。

对于时空偏微分方程,时间和空间之间的相互作用会显著影响解的特征。因此,在分析求解时空偏微分方程过程中整合这两方面信息至关重要。通过组合几个深度学习模型得到的混合网络模型,不仅能

减少独立模型的局限性, 还可以利用模型各自的特性提高混合模型的整体性能。Shi 等人[12]首次将卷积结构与 LSTM 相结合提出的卷积长短期记忆网络(ConvLSTM)模型, 用来预测天气预报。文章指出 ConvLSTM 在输入和状态以及状态和状态之间引入卷积结构分析空间相关性, 同时利用 LSTM 结构解析时间特性, 综合分析时空序列内部关联性。Jalalifar 等人[13]利用 ConvLSTM 的优势, 捕获电力负荷数据中的时空模式, 有效地对负荷需求中的时间依赖性和空间相关性进行建模, 短期预测精度得到提高。Mehrdad 等人[14]将注意力机制与 ConvLSTM 集成起来, 通过加强对重要特征的关注, 同时利用 ConvLSTM 的时空处理能力改进从有限数据集中提取模态频率的能力。随着 ConvLSTM 及其变体的发展, 它逐渐被用于求解时间依赖的 PDEs。Raissi 等人[15]利用 PINN 求解逆标量参数问题, 通过数值算例证实该方法能够从不完整的模型和数据中预测未知量。Rao 等人[16]提出了 PeRCNN 来解决反应扩散方程中的标量参数识别问题, PeRCNN 模型通过对偏微分方程物理规律施加硬约束提高了预测的准确性和鲁棒性。Kakka 等人[17]提出 AE-ConvLSTM, 该模型使用深度卷积编码器、解码器网络结构对原始数据进行自动编码和预测, 并通过对无噪声数据和噪声数据的动力学系统预测证实其长期演化能力。

因此, 本文提出一种耦合编解码卷积长短期记忆网络(ED-ConvLSTM)用于反演 PDEs 未知参数, 同时求解该方程。该网络引入深度卷积编码器和解码器结构有效捕获空间信息, 并利用 ConvLSTM 模块解析时间信息, 从而提高模型对动态系统长期演化的预测准确度。此外, 在编码器和解码器对应模块之间建立跳跃连接, 为解码器部分梯度流提供了一条捷径, 从而缓解了梯度消失问题, 并加快模型收敛速度。

2. 问题描述

时变偏微分方程的一般形式如下:

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} = F(\mathbf{x}, r, u, \nabla u, \Delta u, \nabla \cdot (r \nabla u)), \quad (1)$$

其中, r 是以坐标为自变量的物理参数, $u(\mathbf{x}, t)$ 是方程在 $\mathbf{x} := (x, y) \in \Omega$ 位置、时刻的潜在解, ∇ 、 $\nabla \cdot$ 、 Δ 分别是梯度算子、散度算子和拉普拉斯算子, 是一个线性或者非线性函数。此外, 方程初始条件和边界条件如下所示:

$$\begin{aligned} u(x, y, 0) &= f(x, y), (x, y) \in \Omega, \\ u(x, y, t) &= g(x, y, t), t \in [0, T], (x, y) \in \partial\Omega \end{aligned} \quad (2)$$

通过时间离散和空间剖分, 对于某个给定时刻问题域内的函数可以看作一张二维图像, 连续时间内的三维数据能被离散为多张二维空间图像, 因此可将时变偏微分方程转化为一个时空序列预测问题。针对该问题, 本文提出一种编码解码卷积长短期记忆网络来预测的方法, 其网络结构和损失函数如下所示:

$$\mathbf{U}_{i+k, l} := (\mathcal{U}_{i+k+1}, \mathcal{U}_{i+k+2}, \dots, \mathcal{U}_{i+k+l}) \quad (3)$$

$$(\mathbf{U}_{i+k, l}, r(x, y)) = \mathcal{N}(\mathcal{U}_{i+1}^*, \mathcal{U}_{i+2}^*, \dots, \mathcal{U}_{i+k}^*; \theta) \quad (4)$$

$$L_u(\theta) = \frac{1}{N_{bd}} \sum_{i=1}^{N_{bd}} \left\| \mathcal{N}(\mathbf{U}_{i, k}^*; \theta) - \mathbf{U}_{i+k, l}^* \right\|_2^2 + \lambda_u \cdot \frac{1}{N_{int}} \sum_{i=1}^{N_{int}} \left\| \mathcal{N}(\mathbf{U}_{i, k}^*; \theta) - \mathbf{U}_{i+k, l}^* \right\|_2^2 \quad (5)$$

$$L_r(\theta) = \frac{1}{N_{bd}} \sum_{i=1}^{N_{bd}} \left\| r(x, y) - r^*(x, y) \right\|_2^2 + \lambda_r \cdot \frac{1}{N_{int}} \sum_{i=1}^{N_{int}} \left\| r(x, y) - r^*(x, y) \right\|_2^2 \quad (6)$$

其中, \mathcal{U}_i^* 是 i 时刻的精确解, $\mathbf{U}_{i+k, l}$ 是在 $i+k$ 时刻之后连续 l 个时刻的网络预测解, $r(x, y)$ 是随空间变化的未知参数, θ 是网络参数, $L(\theta)$ 是均方误差损失函数。从式(4)可以看出, ED-ConvLSTM 模型就是将 k 个连续时间步的精确解作为输入数据, 通过网络模拟给出输入时刻之后连续 l 个时刻的偏微分方程预测

解 $\mathbf{U}_{i+k,l}$ 和其中的未知参数 $r(x, y)$ 。

3. 网络架构

3.1. 卷积长短期记忆网络(ConvLSTM)

作为一种特殊的 RNN, LSTM 在循环网络原有框架下引入细胞状态和门控结构, 使其能够避免 RNN 中固有的梯度消失和爆炸问题, 有效解决长期序列预测问题。ConvLSTM 作为 LSTM 的一个变体, 在保持上述特性的同时将其中部分矩阵运算替换为卷积运算, 这有助于 ConvLSTM 在捕获时间相关性的同时减轻空间数据的冗余。其公式表述如下[12]:

$$\begin{aligned} f_t &= \sigma(W_{xf} * \mathcal{U}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f), \\ i_t &= \sigma(W_{xi} * \mathcal{U}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i), \\ \tilde{\mathcal{C}}_t &= \tanh(W_{xc} * \mathcal{U}_t + W_{hc} * \mathcal{H}_{t-1} + b_c), \\ o_t &= \sigma(W_{xo} * \mathcal{U}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o), \\ \mathcal{C}_t &= f_t \circ \mathcal{C}_{t-1} + i_t \circ \tilde{\mathcal{C}}_t, \\ \mathcal{H}_t &= o_t \circ \tanh(\mathcal{C}_t), \end{aligned} \quad (7)$$

其中 ‘.’、‘*’ 分别表示逐元素乘积和卷积运算, i_t 、 f_t 、 o_t 分别表示输入门、遗忘门、输出门; \mathcal{U}_t 、 \mathcal{H}_t 、 \mathcal{C}_t 分别表示 t 时刻输入、隐藏特征和细胞状态。所有的量都是 $B \times H \times W$ 的三维张量, 第一个维度表示时间步长, 后两个维度表示问题域大小。其结构图见图 1, 其中红色线条表示细胞状态 \mathcal{C}_t 在门控结构间的流动。

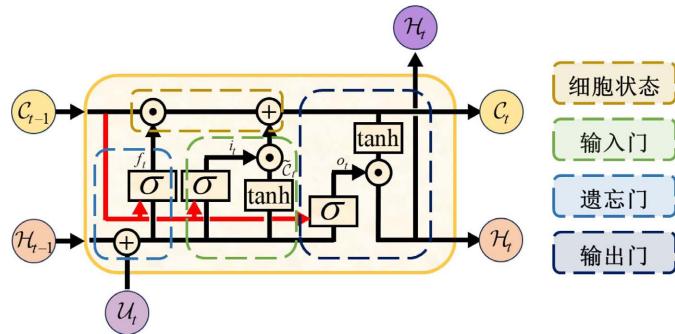


Figure 1. The structure of ConvLSTM
图 1. ConvLSTM 结构示意图

3.2. 卷积块和转置卷积块

如图 2 所示, 蓝色矩形和绿色矩形分别表示卷积块(Conv)和对应的转置卷积块(ConvT), 每个卷积块包含两个卷积层和两个 ReLU 激活函数, 按照卷积 - 激活 - 卷积 - 激活的顺序依次连接。与此类似, ConvT 由转置卷积和激活函数按顺序拼接而成。当特征通过这两个模块时, 其维度不断变化将原始场信息尽可能提取为一系列独立特征并最终还原为原始特征维度。简单来说, 该过程就是对原始数据的降维和升维。不同尺寸的卷积核具有不同的卷积效果, 较大的卷积核通过处理大范围的空间信息来捕获全局特征, 而较小的卷积核则通过处理小范围邻域信息提取局部特征。卷积核尺寸的灵活选择使模型能够有效捕捉特征数据中的多种空间尺度和复杂性, 促进了模型对数据中潜在模式的深入理解, 从而增强其学习和表征能力。

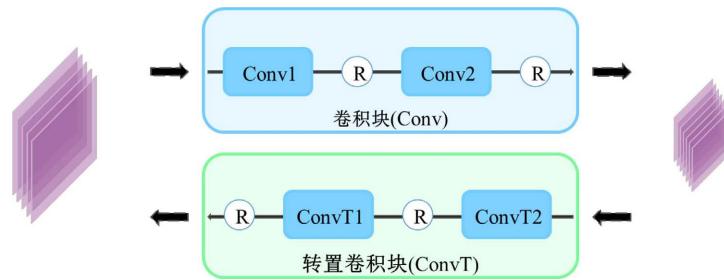


Figure 2. The structure of convolutional block and responding transposed convolutional block

图 2. 卷积和转置卷积结构示意图

3.3. ED-ConvLSTM

本文提出的 ED-ConvLSTM 由编码(Encoder-ConvLSTM)和解码(Decoder-ConvLSTM)两大模块组成。二者都是由多个卷积块和 ConvLSTM 单元构成。编码模块利用卷积单元从邻近区域内提取独立特征, 将原始场信息尽可能转化为不相关的特征。该模块通过降低信息维度降低模型复杂度、减少模型计算量、节约计算成本。解码模块中, 利用转置卷积单元将低维特征恢复到原始维度, 从而有效地解决了编码阶段所导致的细节信息损失。这一过程不仅有助于重建高维数据, 还确保了在特征恢复过程中尽可能保留原始信息。

ED-ConvLSTM 结构如图 3 所示, 其中紫色矩形表示由不同时刻场值组成的 3D 张量。在编码阶段, 卷积单元(由蓝色矩形表示)通过不同尺寸卷积核从输入数据中捕获高低维特征, 该过程中 3D 张量不断变长变扁。在解码阶段, 转置卷积单元(由绿色矩形表示)利用与 Conv 相对应的转置卷积将解码模块得到的特征不断变宽变短, 直到恢复数据原始维度。每个卷积单元和转置卷积单元都配备一个 ConvLSTM 单元(由橙色矩形表示)用于处理时间信息。具体来说, 该单元通过捕获不同时刻数据间的依赖关系来模拟学习动力系统的演化过程。因为每个输入数据都是由多个连续时刻的场值构成的序列, 所以模型能更好地从先验数据中保存重要时空特征, 或者特征随时间演变的规律, 进而增强模型对动力系统的演化能力。此外, 卷积单元和转置卷积单元间的特征映射、编解码模块中相应 ConvLSTM 之间的状态映射实现了信息的重复利用, 增强了模型保留细节的能力, 并且提高了上采样质量。

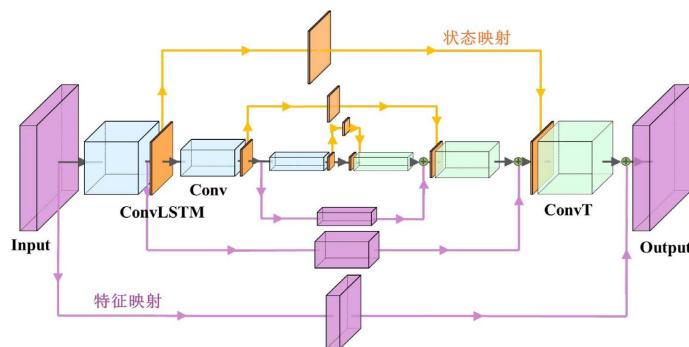


Figure 3. The structure of ED-ConvLSTM

图 3. 编码解码卷积长短期网络结构示意图

3.4. 耦合 ED-ConvLSTM 模型

耦合 ED-ConvLSTM 由两个 ED-ConvLSTM 网络组成, 一个是前向网络(\mathcal{N}_u)用于求解偏微分方程,

另一个是反演网络(\mathcal{N}_r)用于求解偏微分方程的未知参数。耦合 ED-ConvLSTM 中编码模块的基本原理可由如下公式表述:

$$\begin{aligned}\mathbf{U}^m &\triangleq [\mathcal{U}_i^m, \mathcal{U}_{i+1}^m, \mathcal{U}_{i+2}^m, \dots, \mathcal{U}_{i+k}^m], \mathbf{H}^m \triangleq [\mathcal{H}_i^m, \mathcal{H}_{i+1}^m, \dots, \mathcal{H}_{i+k}^m], \\ (\mathbf{U}_i^1; \mathcal{C}^1, \mathcal{H}^1) &= \text{ConvLSTM}_1(\mathbf{U}_i^0 * W_1; \mathcal{C}_i^1, \mathcal{H}_i^1), \\ (\mathbf{U}_i^2; \mathcal{C}^2, \mathcal{H}^2) &= \text{ConvLSTM}_2(\mathbf{U}_i^1 * W_2; \mathcal{C}_i^2, \mathcal{H}_i^2), \\ (\mathbf{U}_i^3; \mathcal{C}^3, \mathcal{H}^3) &= \text{ConvLSTM}_3(\mathbf{U}_i^2 * W_3; \mathcal{C}_i^3, \mathcal{H}_i^3),\end{aligned}\quad (8)$$

其中, $W_n (n=1,2,3)$ 表示卷积块的权重参数, \mathbf{U}_i^m 表示第 i 时刻的输入信息, \mathcal{H}^m 、 \mathcal{C}^m 分别表示相应 ConvLSTM 单元的隐藏状态、细胞状态信息。耦合 ED-ConvLSTM 中解码模块如下所述:

$$\begin{aligned}\mathbf{U}^l &\triangleq [\mathcal{U}_{i+k+1}^l, \mathcal{U}_{i+k+2}^l, \dots, \mathcal{U}_{i+k+l}^l], \\ (\mathbf{U}_i^4) &= \text{ReLU}(\text{ConvLSTM}_4(\mathbf{U}_i^3; \mathcal{C}^3, \mathcal{H}^3) * W_1^T + \mathbf{U}_i^3), \\ (\mathbf{U}_i^5) &= \text{ReLU}(\text{ConvLSTM}_5(\mathbf{U}_i^4; \mathcal{C}^2, \mathcal{H}^2) * W_2^T + \mathbf{U}_i^4), \\ (\mathbf{U}_i^6) &= \text{ReLU}(\text{ConvLSTM}_6(\mathbf{U}_i^5; \mathcal{C}^1, \mathcal{H}^1) * W_3^T + \mathbf{U}_i^5),\end{aligned}\quad (9)$$

其中, $W_n^T (n=1,2,3)$ 表示转置卷积块的权重参数, \mathbf{U}_i^m 表示相应 ConvLSTM 单元在第 l 时刻的预测输出, \mathcal{H}^m 、 \mathcal{C}^m 、 \mathbf{U}_i^m 均保留式(8)中对应参数的值, 通过复制数值实现状态映射、特征映射, 最终达到跳跃连接的目标, 缩短模型预测时间, 提高模型预测精度。

如图 4 所示, 本文采用 10 个连续时刻的场值作为前向网络的一个输入样本, 通过前向网络模拟未来 10 个连续时刻的偏微分方程解, 并将这组值作为前向网络的下一个输入样本和反演网络的输入。为降低预测解变化幅度对反演网络预测精度的影响, 先对前向网络预测值进行无量纲化处理后再将其输入反演网络, 得到 PDEs 待反演参数。重复上述过程九次, 并保存各阶段预测值, 将其拼接得到未来 90 个连续时刻的预测值。最后, 计算预测值和待反演参数的 MSE 损失。为了更细致地说明, 表 1 给出了耦合 ED-ConvLSTM 模型训练过程的伪代码。本文对模型共设置 200 个循环, 每次得到的 MSE 损失都通过 Adam 优化器对模型参数进行新一轮的更新, 并且每 100 次循环对优化器学习率减半直至循环结束。

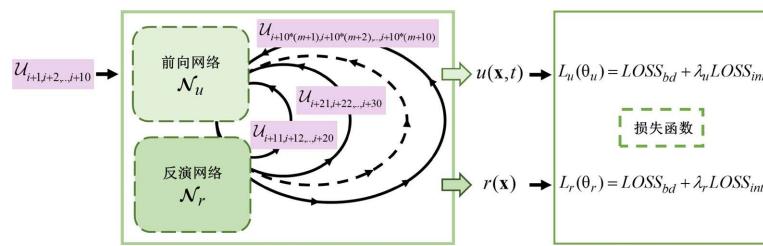


Figure 4. The data flow in the coupled ED-ConvLSTM
图 4. 耦合 ED-ConvLSTM 模型中的数据流动示意图

4. 数值实验与结果

4.1. 模型及其设置

在本节中, 将本文提出的耦合 ED-ConvLSTM 网络应用到线性热方程和非线性扩散方程的求解。另外, 为了更好地说明网络的性能及预测精度, 在相同数据集上将该网络与 AE-ConvLSTM 网络的预测结果进行对比。这两个网络结构如下所述:

Table 1. The training procedure of the coupled ED-ConvLSTM model
表 1. 耦合 ED-ConvLSTM 模型的训练过程

```

输入: 训练样本  $\mathbf{U}_{i,k}^*(x, y, t)$ , 前向网络  $\mathcal{N}_u$ , 反演网络  $\mathcal{N}_r$ 。
输出: 潜在解  $\mathbf{U}_{i,l}^*$ , 待反演参数  $r^*(x, y)$ 。
初始化: 初始化网络参数  $\theta$ , Adam 优化器的学习率  $lr$ 。
for epoch in range (200) do
    epoch = i,  $\mathbf{I} = \mathbf{U}_{i,k}^*$ ,
    for l in range (m) do
         $\mathbf{U}_{i,l} = \mathcal{N}_u(\mathbf{I}; \theta_u)$ ,
         $L_u(\theta) = \frac{1}{N_{bd} N_{bd}} \sum \|\mathbf{U}_{i,l} - \mathbf{U}_{i,l}^*\|_2^2 + \lambda_u \cdot \frac{1}{N_{int} N_{int}} \sum \|\mathbf{U}_{i,l} - \mathbf{U}_{i,l}^*\|_2^2$ ,
         $r(x, y) = \mathcal{N}_r(\mathbf{U}_{i,l}^T; \theta_r)$ ,
         $L_r(\theta) = \frac{1}{N_{bd} N_{bd}} \sum \|r(x_i, y_i) - r^*(x_i, y_i)\|_2^2 + \lambda_r \cdot \frac{1}{N_{int} N_{int}} \sum \|r(x_i, y_i) - r^*(x_i, y_i)\|_2^2$ ,
         $LOSS_u += L_u(\theta), LOSS_r += L_r(\theta)$ ,
         $\mathbf{I} = \mathbf{U}_{i,l}$ ,
    end for
    利用损失函数  $LOSS_u, LOSS_r$  更新网络参数  $\theta$ 。
    when epoch % 100 == 0 do  $lr /= 2$  end when
end for

```

AE-ConvLSTM[17]: 包含一个编码模块和一个解码模块的自动编码卷积长短期网络。编码模块由三个 Conv-ConvLSTM 组成, 解码模块由两个 ConvT-ConvLSTM 和 Conv-ConvLSTM 一个组成。

耦合 ED-ConvLSTM: 由一个 Encoder-ConvLSTM 模块和一个 Decoder-ConvLSTM 模块组成的网络。两模块分别包含三个 Conv-ConvLSTM 单元和三个 ConvT-ConvLSTM 单元, 并且相应子单元之间通过跳跃连接相连。

网络的评价指标选取相对 L2 误差 e_r 和 L1 绝对误差 e , 其定义如下:

$$e_r = \frac{\|\mathbf{U} - \mathbf{U}^*\|_2^2}{\|\mathbf{U}^*\|_2^2}, \quad (10)$$

$$e = \|\mathbf{U} - \mathbf{U}^*\|_1,$$

其中, \mathbf{U}^* 、 \mathbf{U} 分别表示 PDEs 的精确解和网络预测解, $\|\cdot\|_1$ 、 $\|\cdot\|_2$ 分别表示 L1 范数和 L2 范数。为了对比的公平性, 两个网络均采用 Adam 优化器自适应调整网络参数, 且统一将学习率初始化为 10^{-3} 。此外, 两个网络采用相同的宽度和深度, 并对内部的超参数都进行随机正态分布初始化。

4.2. 热传导方程

热方程用于描述一个区域内温度随时间的变化规律, 广泛应用于金融数学、量子力学、影响分析等领域, 是一类重要的偏微分方程。本文对具有温度边界条件的线性热方程进行研究, 其定义如下:

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} - r(\mathbf{x}) \Delta u(\mathbf{x}, t) = 0, \mathbf{x} \in \Omega = [0, 1]^2, t \in [0, T], \quad (11)$$

$$u(\mathbf{x}, 0) = g(\mathbf{x}), \mathbf{x} \in \Omega,$$

其中, $u(\mathbf{x}, t)$ 是温度值, $r(\mathbf{x})$ 是方程待反演参数, 它们的精确解如下:

$$\begin{aligned} r(x, y) &= (x^2 + y^2)/4, \\ u^*(\mathbf{x}, t) &= e^{Ct}(x^2 + y^2), \end{aligned} \quad (12)$$

式(12)中的 C 是一个与时间步长有关的常数, 用于将温度值约束在较小范围内。从温度场的解析式中可以得到初始时刻的温度分布如下:

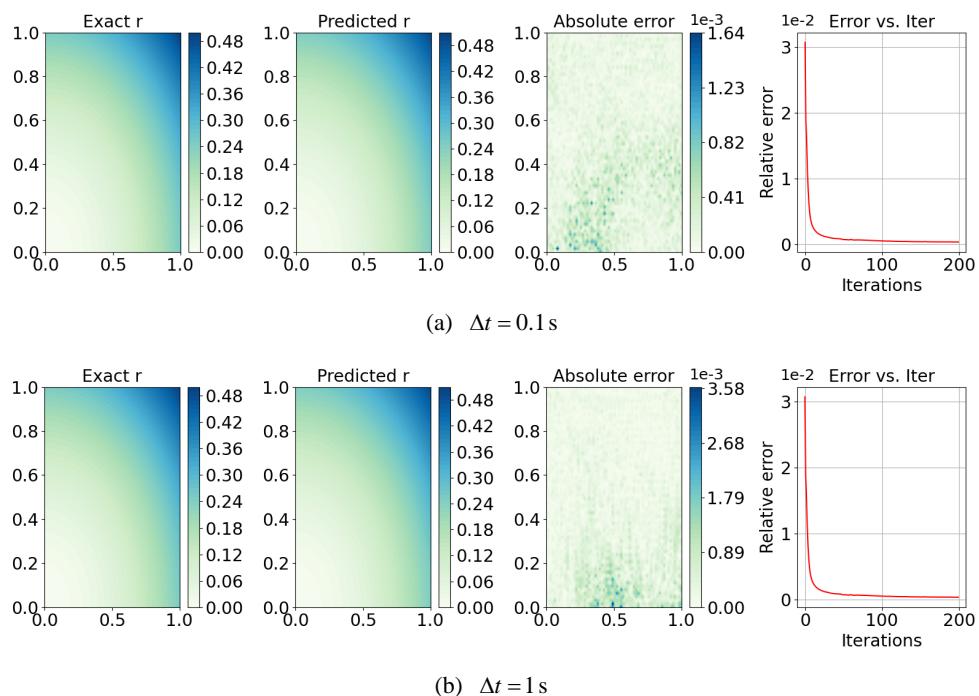
$$u(\mathbf{x}, 0) = x^2 + y^2, (x, y) \in [0, 1] \times [0, 1] \quad (13)$$

输入前向网络时, 将原始连续问题域 $[0, 1] \times [0, 1]$ 均匀离散为 64×64 的网格, 即某一时刻场温度值的尺寸为 64×64 。当采用不同时间步长 $\Delta t = 0.1, 1, 10$ s 时, 对应式(12)中的常数 C 分别取 $1e-4, 1e-6, 1e-8$ 。

另外, 时间域 $[0, T]$ 被分割成 10,000 小份, 其中前 9000 个时刻作为训练样本, 剩余 1000 个时刻作为测试集。在前向网络中, 将 10 个连续时刻的温度值作为输入, 网络能够输出未来连续 90 个时刻的温度场预测值。接着将这些值经过无量纲处理后输入反演网络, 最终得到方程待反演参数。

本文首先利用耦合 ED-ConvLSTM 反演方程未知参数 $r(\mathbf{x})$, 如图 5 所示。从结果可以看出, 在不同时间步长下, 随着迭代次数的增加, 待反演参数预测值的损失也在逐步下降, 这说明在迭代过程中模型逐渐学习到温度场的变化规律, 并且在不同时间步长下都能获得较为准确的预测结果。其相对 L2 误差具体数值见表 2, 可以看出在采用大时间步长 $\Delta t = 10$ s 情况下, 待反演参数的预测精度也能保持在 $1e-3$ 。所以耦合 ED-ConvLSTM 能有效求解线性热方程的待反演参数, 即使在大时间步长情况下, 该模型也能取得较为准确的预测结果。这表明该模型可用于解决大时间步长的参数识别问题, 并具有良好的鲁棒性。

基于深度学习模型求解 PDEs 待反演参数的同时, 一般也会同时求解 PDEs。为了验证该模型求解方程的准确性, 耦合 ED-ConvLSTM 模型在预测方程待反演参数的同时, 也给出了偏微分方程解的预测值 $u(\mathbf{x}, t)$, 其结果见表 2, 可以看到在不同时间步长的情况下, 方程预测解的相对 L2 误差均保持在 $1e-4$, 这说明模型对于温度场的预测可以达到较高的精度。



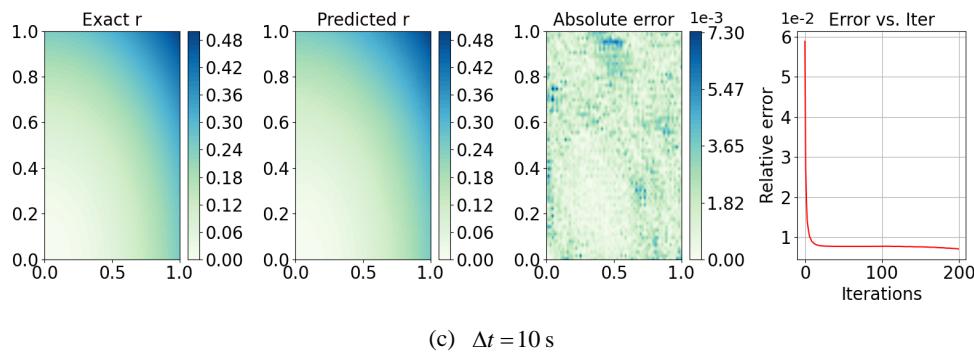
(c) $\Delta t = 10 \text{ s}$

Figure 5. Linear heat equation: The exact parameter, predicted parameter, point-wise absolute error and relative L2 error of the inverse parameter $r(\mathbf{x})$ with different time steps $\Delta t = 0.1, 1, 10 \text{ s}$ obtained by the coupled ED-ConvLSTM model after 200 iterations

图 5. 线性热方程：经过 200 次迭代后耦合 ED-ConvLSTM 模型在不同时间步长 $\Delta t = 0.1, 1, 10 \text{ s}$ 下得到的待反演参数 $r(\mathbf{x})$ 的精确解、预测解、逐点绝对误差、相对 L2 误差图

为了进一步证实本文所提出模型的有效性，将其与同等参数设置的 AE-ConvLSTM 模型进行比较，图 6 给出了两个模型在不同时间步下，经过 200 次迭代得到的参数预测值的绝对误差。图 7 展示了二者在第 100 个外延时刻温度场预测值的相对 L2 误差值。不难看出，无论是温度场预测值还是待反演参数，本文提出的模型都优于 AE-ConvLSTM，并且在不同时间步长下温度场预测值的精度都保持稳定。这充分说明该模型在不同时间步长下都具有显著的鲁棒性，无论时间步长如何变化，耦合 ED-ConvLSTM 模型都能保持其良好的预测性能。

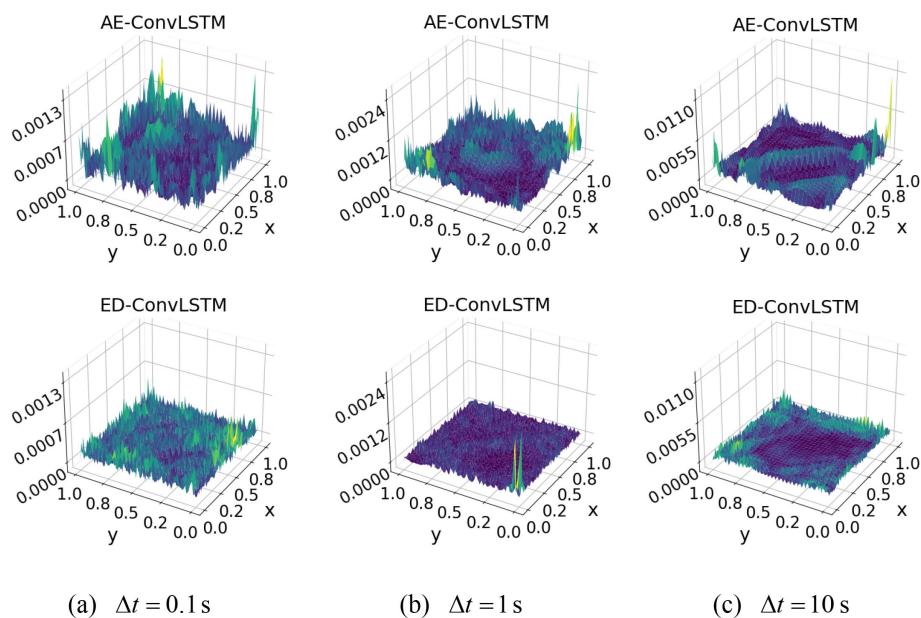
(a) $\Delta t = 0.1 \text{ s}$ (b) $\Delta t = 1 \text{ s}$ (c) $\Delta t = 10 \text{ s}$

Figure 6. Linear heat equation: The absolute point-wise error of the inverse parameter $r(\mathbf{x})$ with different time steps $\Delta t = 0.1, 1, 10 \text{ s}$ obtained by AE-ConvLSTM and ED-ConvLSTM after 200 iterations

图 6. 线性热方程：经过 200 次迭代后 AE-ConvLSTM、耦合 ED-ConvLSTM 模型在不同时间步长 $\Delta t = 0.1, 1, 10 \text{ s}$ 下得到的待反演参数 $r(\mathbf{x})$ 的逐点绝对误差图

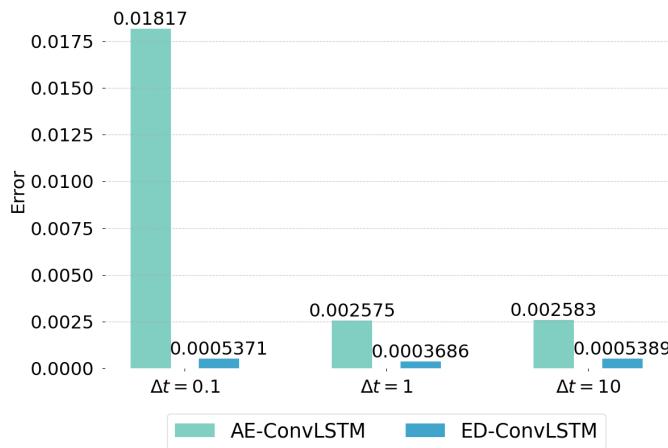


Figure 7. Linear heat equation: The L2 relative error of PDE solution $u(\mathbf{x}, t)$ at the 100th extrapolated time with different timesteps solved by AE-ConvLSTM and ED-ConvLSTM

图 7. 线性热方程: AE-ConvLSTM、耦合 ED-ConvLSTM 模型在不同时间步长下得到的第 100 个外延时刻温度场预测值 $u(\mathbf{x}, t)$ 的相对 L2 误差图

最后, 本文试验模型在不同噪声水平下的表现, 分析其对噪声的鲁棒性, 并评估该模型处理病态反问题的能力。对于温度场和参数精确解的边界区域分别增加如下噪声:

$$u_b = (1 + \zeta \varepsilon) u^*, r_b = (1 + \zeta \varepsilon) r^* \quad (14)$$

其中, ε 服从正态分布, ζ 表示噪声水平, 取值为 1%、5%、10%。通过模型训练得到的参数预测值和温度场预测值的相对于 L2 误差被展示在表 2 中。可以观察到, 随着噪声水平的增加预测值的相对 L2 误差增加相对缓慢, 这说明本文提出的模型对噪声具有良好的鲁棒性, 能够有效解决反问题中常见的不适定问题。

Table 2. Linear heat equation: L2 relative error of the inverse parameter $r(\mathbf{x})$ and predicted solution $u(\mathbf{x}, t)$ by the coupled ED-ConvLSTM model with different timesteps and noise levels

表 2. 线性热方程: 耦合 ED-ConvLSTM 模型在不同时间步长和噪声水平下待反演参数 $r(\mathbf{x})$ 和场温度 $u(\mathbf{x}, t)$ 预测值的相对 L2 误差

Time step	L2 relative error			
	Noise level = 0%	Noise level = 1%	Noise level = 5%	Noise level = 10%
$\Delta t = 0.1$ s	$e_r(u)$	5.371e-4	6.224e-4	1.803e-3
	$e_r(r)$	4.443e-4	2.861e-3	1.650e-2
$\Delta t = 1$ s	$e_r(u)$	3.682e-4	7.735e-4	1.773e-3
	$e_r(r)$	3.828e-4	3.587e-3	1.057e-2
$\Delta t = 10$ s	$e_r(u)$	5.388e-4	4.402e-4	1.597e-3
	$e_r(r)$	1.514e-3	4.167e-3	1.201e-2

4.3. 非线性扩散方程

非线性偏微分方程是 PDEs 研究的重要分支, 被广泛应用于实际工程领域中, 如流体力学、流行病

学、化工循环系统等。本节对具有狄利克雷边界条件的非线性热方程进行研究, 其定义如下:

$$\begin{aligned} \frac{\partial u(\mathbf{x}, t)}{\partial t} - \nabla \cdot (r(\mathbf{x}) \nabla u(\mathbf{x}, t)) &= f(\mathbf{x}, t), \mathbf{x} \in \Omega := [0, 1]^2, t \in [0, T], \\ u(\mathbf{x}, 0) &= g(\mathbf{x}), \mathbf{x} \in \Omega, \end{aligned} \quad (15)$$

其中, $r(\mathbf{x})$ 是待反演参数, $f(\mathbf{x}, t)$ 是源项, 定义如下:

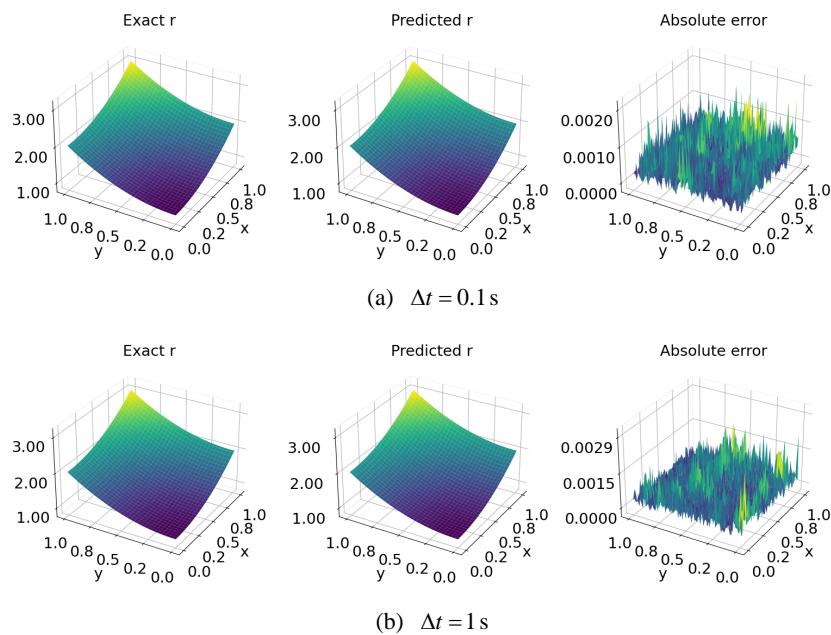
$$r(\mathbf{x}) = 1 + x^2 + y^2, \quad (16)$$

$$\begin{aligned} f(\mathbf{x}, t) &= \sin\left(\frac{\pi}{2}x\right) \cos\left(\frac{\pi}{2}y\right) \left[1 + \frac{t(1+|\mathbf{x}|^2)}{2} \right] \\ &\quad - \pi x t \cos\left(\frac{\pi}{2}x\right) \cos\left(\frac{\pi}{2}y\right) + \pi y t \sin\left(\frac{\pi}{2}x\right) \sin\left(\frac{\pi}{2}y\right). \end{aligned} \quad (17)$$

定义在式(15)中的偏微分方程的解析解为:

$$u(\mathbf{x}, t) = t \sin\left(\frac{\pi}{2}x\right) \cos\left(\frac{\pi}{2}y\right), \quad (18)$$

在本例中, 首先展示模型在不同时间步长下得到的待反演参数预测精度。图 8 给出耦合 ED-ConvLSTM 模型得到的参数预测值与真实值的绝对误差 3D 云图。可以明显看出, 在不同的时间步长下, 该模型得到的预测解和真实解有很好的一致性, 绝对误差最大值都保持在 $1e-3$ 的量级。其具体相对 L2 误差见表 3, 可以看到随着时间步长从 0.1 s 增加到 10 s, 待反演参数的相对 L2 误差始终保持在 $1e-4$ 量级。这说明模型可以较为准确地预测非线性扩散方程的待反演参数。此外, 该模型能够在求解待反演参数的同时预测非线性扩散方程的解。图 9 给出了在时间步长 $\Delta t = 1$ s 的情况下, 耦合 ED-ConvLSTM 得到的在外延时间段 9910 s~10,000 s 的温度预测值, 值得注意的是, 该时间段并不包含在训练数据集中。该时间段内温度预测值的相对 L2 误差在 10,000 s 取得最大值, 为 $6.435e-6$; 在 9920 s 处取得最小值, 为 $2.385e-6$ 。这表明该模型具有很好的外延性, 能够根据原始数据推演温度场的未来变化情况。



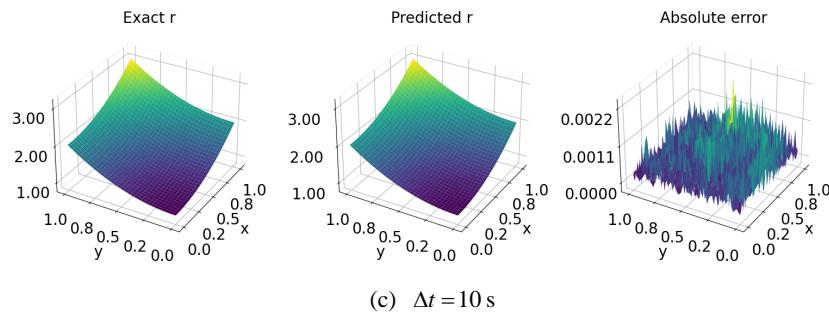


Figure 8. Nonlinear diffusion equation: The exact parameter, predicted parameter and point-wise absolute error of the inverse parameter $r(\mathbf{x})$ with different timesteps $\Delta t = 0.1, 1, 10$ s obtained by ED-ConvLSTM model after 200 iterations

图 8. 非线性扩散方程: 经过 200 次迭代后耦合 ED-ConvLSTM 模型在不同时间步长 $\Delta t = 0.1, 1, 10$ s 下得到的待反演参数 $r(\mathbf{x})$ 的精确解、预测解、逐点绝对误差图

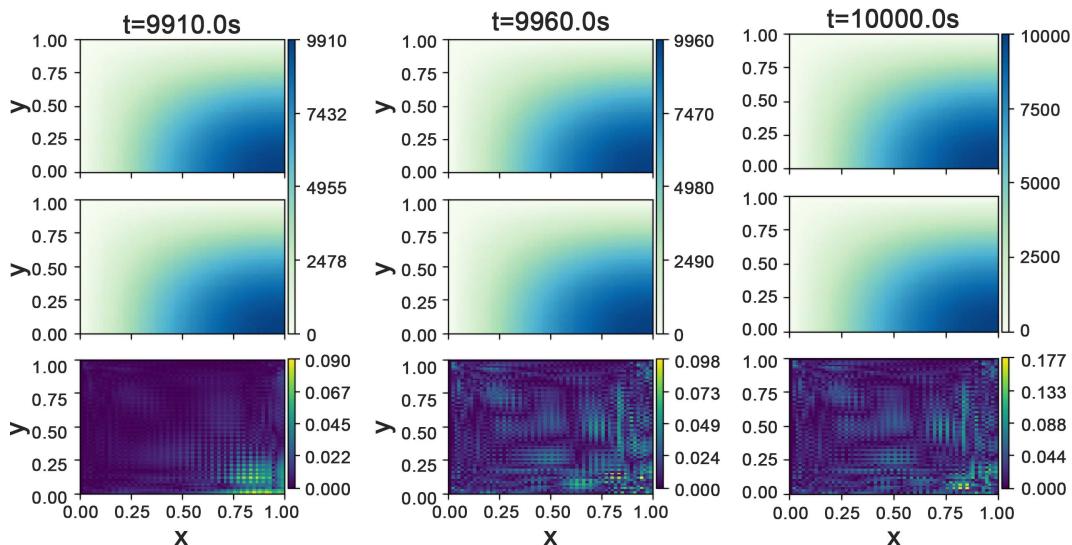


Figure 9. Nonlinear diffusion equation: The exact solution u^* , predicted solution u and point-wise absolute error (top to bottom) for the forward diffusion problem at different extrapolated times obtained by the coupled ED-ConvLSTM model with timestep $\Delta t = 1$ s

图 9. 非线性扩散方程: 在时间步长 $\Delta t = 1$ s 下经过 200 次迭代后耦合 ED-ConvLSTM 模型得到的在外部延时间段上温度场的精确解 u^* 、预测解 u 、逐点绝对误差图(按从上到下的顺序)

Table 3. Nonlinear diffusion equation: Relative L2 error of the inverse parameter $r(\mathbf{x})$ and predicted solution $u(\mathbf{x}, t)$ by the ED-ConvLSTM model with different timesteps

表 3. 非线性扩散方程: 耦合 ED-ConvLSTM 模型在不同时间步长下待反演参数 $r(\mathbf{x})$ 和场温度 $u(\mathbf{x}, t)$ 预测值的相对 L2 误差

Relative L2 error	Time step		
	$\Delta t = 0.1$ s	$\Delta t = 1$ s	$\Delta t = 10$ s
$u(\mathbf{x}, t)$	1.627e-5	6.435e-6	7.609e-6
$r(\mathbf{x})$	2.923e-4	3.476e-4	3.058e-4

接着, 将本文提出的模型与同参数设置的 AE-ConvLSTM 模型进行比较。图 10 展示了 200 次迭代后得到的待反演参数的逐点绝对误差云图。图 11 给出在第 100 个外延时刻的温度场预测值的相对 L2 误差图。可以看出, 不论是待反演参数还是温度场的预测值, 耦合 ED-ConvLSTM 的精度都高于 AE-ConvLSTM 模型, 这再次证实本文提出的网络在预测精度上的优越性。综上, 耦合 ED-ConvLSTM 在实现高精度参数识别和解预测方面具备有效性和内在优势。

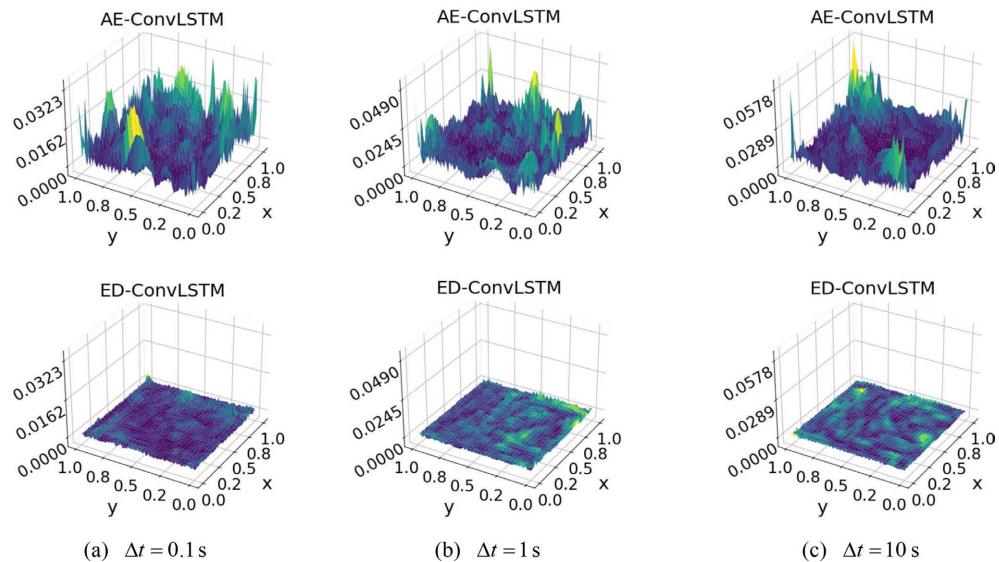


Figure 10. Nonlinear diffusion equation: The point-wise absolute error of the inverse parameter $r(\mathbf{x})$ with different timesteps $\Delta t = 0.1, 1, 10 \text{ s}$ obtained by AE-ConvLSTM and ED-ConvLSTM after 200 iteration.

图 10. 非线性扩散方程: 经过 200 次迭代后 AE-ConvLSTM、耦合 ED-ConvLSTM 模型在不同时间步长 $\Delta t = 0.1, 1, 10 \text{ s}$ 下得到的待反演参数 $r(\mathbf{x})$ 的逐点绝对误差图

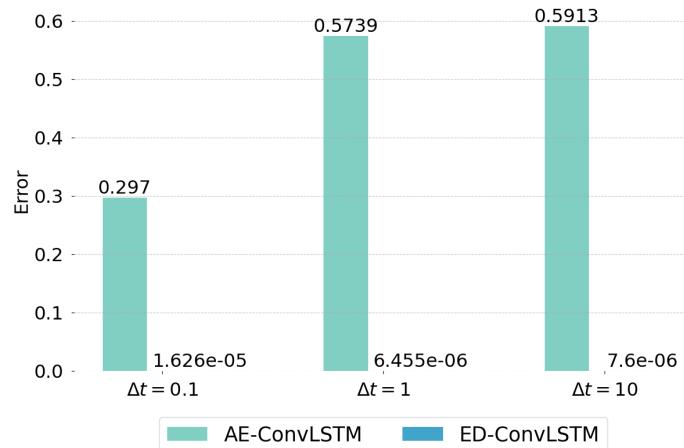


Figure 11. Nonlinear diffusion equation: The relative L2 error of PDE solution $u(\mathbf{x}, t)$ at the 100th extrapolated time with different timesteps solved by AE-ConvLSTM and coupled ED-ConvLSTM

图 11. 非线性扩散方程: AE-ConvLSTM、耦合 ED-ConvLSTM 模型在不同时间步长下得到的第 100 个外延时刻温度场预测值 $u(\mathbf{x}, t)$ 的相对 L2 误差图

最后, 测试该模型对噪声的鲁棒性, 进一步评估模型在生产实际中的可用性和潜力。模型在不同噪声水平下得到的待反演参数见表 4。可以看到随着噪声水平的增加, 模型的待反演参数预测精度略有下降, 最高达到 $1\text{e-}2$ 量级。同时温度场预测值的相对 L2 误差保持在 $1\text{e-}6\text{--}1\text{e-}3$ 范围内。虽然模型的性能随着噪声水平增加而下降, 但是其预测能力仍保持在令人满意的水平。耦合 ED-ConvLSTM 在不同时间步长下的一致性, 进一步凸显了该模型在处理参数识别问题、PDEs 正问题时的鲁棒性和可靠性。

Table 4. Nonlinear diffusion equation: Relative L2 error of the inverse parameter $r(\mathbf{x})$ and predicted solution $u(\mathbf{x}, t)$ at the 100th extrapolated time with different timesteps and noise levels solved by the coupled ED-ConvLSTM model

表 4. 非线性扩散方程: 耦合 ED-ConvLSTM 模型在不同时间步长和噪声水平下得到的第 100 个外延时刻待反演参数 $r(\mathbf{x})$ 、温度场 $u(\mathbf{x}, t)$ 预测值的相对 L2 误差

Time step	Relative L2 error			
	Noise level = 0%	Noise level = 1%	Noise level = 5%	Noise level = 10%
$\Delta t = 0.1\text{ s}$	$e_r(u)$	1.627e-5	2.524e-4	1.145e-3
	$e_r(r)$	2.923e-4	1.796e-3	1.035e-2
$\Delta t = 1\text{ s}$	$e_r(u)$	6.435e-6	2.692e-4	1.254e-3
	$e_r(r)$	3.476e-4	1.871e-3	1.049e-2
$\Delta t = 10\text{ s}$	$e_r(u)$	7.609e-6	2.244e-4	5.077e-4
	$e_r(r)$	3.058e-4	1.927e-3	1.045e-2

5. 结论

本文基于卷积和 LSTM 提出了一种具有跳跃连接的耦合编解码卷积长短期记忆网络, 用于求解时变偏微分方程和参数识别问题。一方面, 该模型能够有效求解 PDEs 中的待反演参数; 另一方面, 它能够预测动力系统的演变情况, 并且其预测精度优于 AE-ConvLSTM 模型。在测试阶段, 仅仅输入 10 个连续时刻的真实解, 网络就能对未来 100 个时刻的场值进行较为精确的预测。数值结果表明, 耦合 ED-ConvLSTM 可以有效求解二维热方程和扩散方程中的待反演参数、近似解。此外, 所提出的模型在不同噪声水平下, 对参数识别问题表现出充分的鲁棒性和可靠性。然而, 在实际工程应用中, 数据短缺使得很难获得足够的样本进行训练, 这一限制阻碍了全数据驱动的能力, 而这对于确保模型准确性、验证模型有效性至关重要, 作者将在未来的工作中解决这一难题。

参考文献

- [1] Archana, R. and Jeevaraj, P.S.E. (2024) Deep Learning Models for Digital Image Processing: A Review. *Artificial Intelligence Review*, **57**, Article No. 11. <https://doi.org/10.1007/s10462-023-10631-z>
- [2] Demir, F., Abdullah, D.A. and Sengur, A. (2020) A New Deep CNN Model for Environmental Sound Classification. *IEEE Access*, **8**, 66529-66537. <https://doi.org/10.1109/access.2020.2984903>
- [3] Abas, A.R., Elhenawy, I., Zidan, M. and Othman, M. (2022) BERT-CNN: A Deep Learning Model for Detecting Emotions from Text. *Computers, Materials & Continua*, **71**, 2943-2961. <https://doi.org/10.32604/cmc.2022.021671>
- [4] Zhou, D. (2020) Universality of Deep Convolutional Neural Networks. *Applied and Computational Harmonic Analysis*, **48**, 787-794. <https://doi.org/10.1016/j.acha.2019.06.004>
- [5] Özbay, A.G., Hamzehloo, A., Laizet, S., Tzirakis, P., Rizos, G. and Schuller, B. (2021) Poisson CNN: Convolutional Neural Networks for the Solution of the Poisson Equation on a Cartesian Mesh. *Data-Centric Engineering*, **2**, e6. <https://doi.org/10.1017/dce.2021.7>
- [6] Long, Z., Lu, Y. and Dong, B. (2019) PDE-Net 2.0: Learning PDEs from Data with a Numeric-Symbolic Hybrid Deep

Network. *Journal of Computational Physics*, **399**, Article ID: 108925. <https://doi.org/10.1016/j.jcp.2019.108925>

[7] Mo, S., Zhu, Y., Zabaras, N., Shi, X. and Wu, J. (2019) Deep Convolutional Encoder-Decoder Networks for Uncertainty Quantification of Dynamic Multiphase Flow in Heterogeneous Media. *Water Resources Research*, **55**, 703-728. <https://doi.org/10.1029/2018wr023528>

[8] Gao, H., Sun, L. and Wang, J. (2021) Phygeonet: Physics-Informed Geometry-Adaptive Convolutional Neural Networks for Solving Parameterized Steady-State PDEs on Irregular Domain. *Journal of Computational Physics*, **428**, Article ID: 110079. <https://doi.org/10.1016/j.jcp.2020.110079>

[9] Li, Z.Y., Kovachki, N.B., Azizzadenesheli, K., et al. (2020) Fourier Neural Operator for Parametric Partial Differential Equations. *International Conference on Learning Representations*. arXiv:2010.08895

[10] Gers, F.A., Schmidhuber, J. and Cummins, F. (2000) Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, **12**, 2451-2471. <https://doi.org/10.1162/089976600300015015>

[11] Hu, Y., Zhao, T., Xu, S., Lin, L. and Xu, Z. (2022) Neural-PDE: A RNN Based Neural Network for Solving Time Dependent PDEs. *Communications in Information and Systems*, **22**, 223-245. <https://doi.org/10.4310/cis.2022.v22.n2.a3>

[12] Shi, X., Chen, Z., Wang, H., et al. (2015) Convolutional LSTM Network: A Machine Learning Approach for Precipitation Now-Casting. *Advances in Neural Information Processing Systems*, **2015**, 802-810.

[13] Jalalifar, R., Delavar, M.R. and Ghaderi, S.F. (2024) SAC-ConvLSTM: A Novel Spatio-Temporal Deep Learning-Based Approach for a Short-Term Power Load Forecasting. *Expert Systems with Applications*, **237**, Article ID: 121487. <https://doi.org/10.1016/j.eswa.2023.121487>

[14] Dizaji, M.S., Mao, Z. and Haile, M. (2023) A Hybrid-Attention-ConvLSTM-Based Deep Learning Architecture to Extract Modal Frequencies from Limited Data Using Transfer Learning. *Mechanical Systems and Signal Processing*, **187**, Article ID: 109949. <https://doi.org/10.1016/j.ymssp.2022.109949>

[15] Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *Journal of Computational Physics*, **378**, 686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>

[16] Rao, C., Ren, P., Wang, Q., Buyukozturk, O., Sun, H. and Liu, Y. (2023) Encoding Physics to Learn Reaction-Diffusion Processes. *Nature Machine Intelligence*, **5**, 765-779. <https://doi.org/10.1038/s42256-023-00685-7>

[17] Kakka, P.R. (2022) Sequence to Sequence AE-ConvLSTM Network for Modelling the Dynamics of PDE Systems.