Available online at www.sciencedirect.com
ScienceDirect
XXXXXX期刊 00（2023年）1–17
XXXXXX期刊

arXiv:2302.00962v1 [cs.AI] 2 Feb 2023

FV-MgNet：用于可解释时间序列预测的全连接 V 循环 MgNet

朱建清[a], 何俊才[b], 张莲[c], 徐进超[b,d,*]

[a]北京工业大学理学院，中国北京市朝阳区，邮编 *100124*，中华人民共和国[b]阿卜杜拉国王科技大学（*KAUST*）计算机、电气与数学科学及工程部，沙特阿拉伯图瓦勒，邮编 *23955-6900*[c]深圳国际工业与应用数学中心，深圳大数据研究院，龙岗区龙翔大道 *2001* 号，广东省深圳市，邮编 *518172*，中华人民共和国[d]宾夕法尼亚州立大学数学系，美国宾夕法尼亚州大学公园，*PA 16802*

摘要

通过研究约束线性模型的迭代方法，我们提出了一类新的全连接 V 循环 MgNet，用于长期时间序列预测——这是预测任务中最困难的任务之一。MgNet 是一种卷积神经网络模型，最初基于用于求解离散偏微分方程（PDEs）的多重网格（MG）方法，为图像分类而提出。我们将现有 MgNet 中的卷积运算替换为全连接运算，并将其应用于预测问题。受 MG 中 V循环结构的启发，我们进一步提出 FV-MgNet，即全连接 MgNet 的 V 循环版本，以分层提取特征。通过在常用数据集上评估 FV-MgNet 的性能，并将其与最先进模型进行比较，我们发现 FV-MgNet 取得了更好的结果，同时内存使用量更少、推理速度更快。此外，我们还进行了消融实验，以证明在众多变体中，FV-MgNet 的结构是最佳选择。

关键词： 时间序列预测，神经网络，MgNet，可解释性

1. 引言

长期时间序列预测在零售（[1, 2]）、医疗保健（[3, 4]）和工程（[5, 6]）等众多领域的各类应用中发挥着重要作用。人们已开发出多种用于时间序列预测的深度学习模型，其中循环神经网络（RNN）可能是研究最为广泛的模型（[7, 8]）。近期，Transformer 模型在自然语言处理（NLP）和计算机视觉（CV）研究（[9, 10, 11, 12]）中取得成功后，已成为时间序列预测领域最热门的研究方向之一（[13, 14, 15]）。高效自注意力机制应运而生，人们也提出了许多结构来处理时间序列预测任务，例如分解序列（[15]）、设计特殊的线性注意力结构（[13, 14]），以及将傅里叶变换（[15]）或小波变换（[16]）应用于自注意力结构。然而，Transformer 模型的计算成本较高，许多实验发现，随着输入序列长度增加，预测结果会变差（[17]）。

*通讯作者电子邮箱： jqzhu@emails.bjut.edu.cn (朱建清), juncai.he@kaust.edu.sa (何俊才), zhanglian@sribd.cn (张莲), xu@multigrid.com (徐进超)网址： https://juncaihe.github.io (何俊才), http://www.personal.psu.edu/jxx1/(徐进超)

通过研究约束线性模型的迭代方法，我们提出了一类用于长期时间序列预测的新型全连接 V 循环 MgNet（FV-MgNet）。这是 MgNet 类模型首次成功应用于这一领域。FV-MgNet 的一项主要技术，是将现有 MgNet 中的卷积运算替换为全连接运算（[18, 19, 20]）。现有 MgNet 是一种用于图像分类的卷积神经网络（CNN）模型，其设计基于求解离散偏微分方程（PDEs）的多重网格（MG）方法（[21, 22, 23]）。我们提出的另一项技术，是采用全连接 MgNet 的 V 循环版本来分层提取特征；这一设计受到多重网格中 V 循环结构的启发，在该结构中，不同频率信息可通过网格之间的尺寸变换加以捕获。通过在常用数据集上评估 FV-MgNet 的性能，并将其与最先进模型进行比较，我们发现 FV-MgNet 取得了更好的结果，同时内存使用量大幅降低、推理速度更快（如图1和图2所示）。我们还证明，随着输入序列长度增加，数值结果不会变差（如表3所示）。

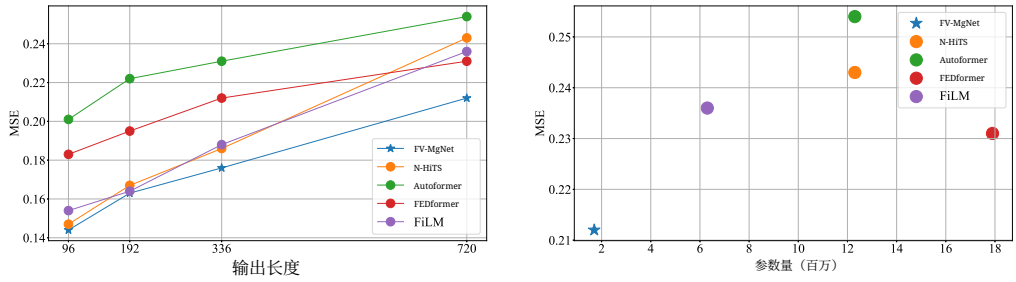


图1：Electricity 多变量数据集上的均方误差。左：均方误差随输出长度的变化。右：输出长度固定为 720 时，均方误差随参数量的变化。均方误差越低，预测效果越好。

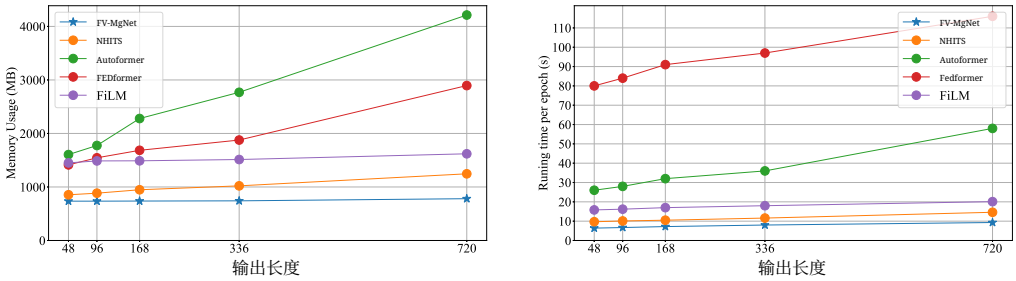


图2：Electricity 多变量数据集上的计算效率基准测试，批量大小为 16。时间开销以每轮训练时间衡量。左：输入长度固定为 96 时，内存使用量随输出长度的变化。右：输入长度固定为 96 时，每轮时间开销随输出长度的变化。

本文结构安排如下。第 2 节回顾相关工作。第 3 节首先将时间序列预测问题视为一个受约束的线性模型，并使用多重网格法（一种迭代法）求解；随后介绍用于预测问题的 FV-MgNet 模型。第 4 节和第 5 节分别给出数值实验和结论。

2. 相关工作

循环神经网络广泛应用于时间序列预测问题。LSTM（[24]）和 GRU（[25]）可通过门控机制缓解梯度消失和梯度爆炸等固有问题。许多基于 LSTM 和 GRU 的研究提出了大量时间序列预测模型（例如（[26, 27, 28, 29]）。然而，单步预测方法无法准确捕捉长期时间序列特征，且难以并行化。近年来，卷积神经网络也被应用于这项任务，并提出了时间卷积

网络（TCN）（[30]）。尽管这种方法效率较高，但在长期时间序列预测问题中仍存在预测瓶颈。随着 Transformer 在自然语言处理和计算机视觉领域的创新（[9, 10, 11]），基于 Transformer 的模型也被提出，并在时间序列预测中表现出色（[13, 14, 15]）。基于 Transformer 的模型擅长为序列数据建模长期依赖关系。此外，为克服原始 Transformer 的二次计算复杂度且不降低性能，研究者设计了许多适用于时间序列预测的高效 Transformer。LogTrans（[13]）采用对数稀疏注意力，实现了$N\log^2 N$复杂度。Reformer（[31]）引入局部敏感哈希，将复杂度降至$N\log N$。Informer（[14]）使用基于 KL 散度的方法从注意力矩阵中选取 top-k，复杂度为$N\log N$。最近，Autoformer（[15]）引入自相关模块取代标准注意力，以执行子序列级注意力；借助快速傅里叶变换（FFT）以及在自相关矩阵中进行 top-k 选择，其复杂度达到$N\log N$。Fedformer（[16]）通过应用傅里叶变换或小波变换来设计频域注意力。ETSformer（[32]）提出指数平滑注意力（ESA）和频率注意力（FA），从而提升准确率和效率。

近年来，许多非 Transformer 模型在预测问题上也取得了良好效果。FiLM ([33]) 在时间序列预测中应用勒让德多项式投影、傅里叶投影和低秩近似。NBEATS ([34]) 模型提出了一种由全连接层构成的双残差结构，在单变量预测中取得了良好效果。N-HiTS ([35]) 在 NBEATS 中加入多尺度结构，使其能够更好地捕捉序列中不同时期的信息，并在多变量模型中取得了更好的效果。在 NBEATS 和 N-HiTS 中，我们观察到全连接层；多尺度结构对于提升模型性能至关重要。MgNet ([18, 20]) 是一种用于图像分类的卷积神经网络模型，其设计基于多重网格；多重网格已成为求解偏微分方程的一种成功方法。MgNet 通过运用多重网格思想求解约束线性模型 ([19]) 来阐释，因此具有精心设计的多尺度结构。我们在 MgNet框架下研究全连接层和多尺度结构，以解决预测问题。

3. 方法

本节介绍我们提出的用于时间序列预测的模型 FV-MgNet，该模型采用约束线性模型和多重网格结构；FV-MgNet 的详细描述见算法1。将数据-标签对视为（f，y），其中$f \in \mathbb{R}^I$是输入时间序列，$y \in \mathbb{R}^O$是预测的时间序列，我们提出以下特征提取和插值方案，作为一种可解释模型，用于构建从f到y的映射：

$$\text{Input: } f \xrightarrow[\text{constrained linear model: } Au=f]{\text{feature extraction}} \text{Feature: } u \xrightarrow[\text{neural network approximation}]{\text{feature interpolation: } \mathcal{I}} \text{Output: } y. \tag{1}$$

总而言之，我们有

$$y = \mathcal{I}(u(f)). \tag{2}$$

这里，我们将$u \in \mathbb{R}^{I'}$记作特征向量。一般而言，特征中所包含的信息（本征维度）应小于或等于数据中的信息（嵌入维度），例如流形学习（[36, 37, 38]）；具体而言，$I' \le I$。为提取更多内在特征，同时尽可能保留信息，我们取$I' = I$，并采用 V循环结构（类似于 U-Net架构）进行特征提取。首先，我们提出应用约束线性模型（[19]），该模型最初是为视觉任务中的特征提取而设计的。然后，我们提出使用全连接算子作为底层线性模型，取代图像处理中使用的卷积算子。最后，我们引入多重网格结构，以多分辨率构建特征并降低计算复杂度。得到特征向量u后，我们提出使用单隐藏层 ReLU神经网络来近似 $\mathcal{I}(u)$，其表达式为

$$\mathcal{I}(u) \approx W^2\sigma\left(W^1 u\right). \tag{3}$$

所提出的模型以多重网格法为主要框架，对不同分辨率的输入进行残差校正，通过延拓算子整合其他网格上的信息，随后采用单隐层 ReLU神经网络进行插值近似，从而实现长时域精确预测。

此外，与基于 Transformer架构的模型相比，我们的模型采用全连接层实现，结构更加简洁优雅，计算效率更高，内存开销更低。

3.1. 约束线性模型：用于特征提取的可解释模型

约束线性模型（[19]）最初是为视觉任务提出的；在本研究中，我们证明它也可作为一种可解释模型，用于时间序列问题中的特征提取。对于数据空间中的任意输入时间序列$f \in \mathbb{R}^I$，约束线性模型定义为以下数据—特征映射

$$Au = f, \tag{4}$$

其中，$u \in \mathbb{R}^{I'}$ 是底层特征向量，满足

$$[u]_i \geq 0, \quad i = 1 : I'. \tag{5}$$

这里，我们假设$A: \mathbb{R}^I \mapsto \mathbb{R}^{I'}$ 是从特征空间到数据空间的线性（仿射）算子。求解带有非线性约束（5）的线性方程（4）至关重要。通常，迭代法是求解代数系统的重要组成部分。例如，给定代数系统$Au = f$，可通过残差校正方法获得其解

$$u^{i+1} = u^i + B^i(f^i - Au^i), \quad i = 1 : \nu, \tag{6}$$

其中 $B^i : \mathbb{R}^{I'} \mapsto \mathbb{R}^I$ 是数据-特征映射 A 的逆算子的近似。回顾激活函数 $\sigma(x) = \text{ReLU}(x) := \max\{0,x\}$ 的定义，上述迭代过程可以自然地修改，以保持约束 (5)：

$$u^i = u^{i-1} + \sigma \circ B^i\sigma \circ (f - Au^{i-1}), \quad i = 1 : \nu. \tag{7}$$

这里，$\sigma \circ(\cdot)$ 表示对非线性激活函数 σ 逐元素应用。我们指出，$\sigma(x)$ =ReLU(x) 并非保持约束 (5) 的唯一选择。例如，帽函数 [20] 以及 $\text{ReLU}^k(x)$ [39] 的组合也可用于此目的。如 ([19]) 所示，如果选择 $r = f - Au$ 作为特征，上述迭代方案可退化为 ResNet ([40]) 的基本块。然而，([19]) 中的理论论证和数值结果表明，u 作为特征可能是更合适的选择。在本研究中，我们也将 (7) 作为模型中特征提取的基本块。

3.2. 底层系统 A：稠密还是稀疏？

在视觉任务中，A和 B^i 通常被选作可学习的多通道卷积算子。例如，用于图像分类的基本块（MgNet 中的 [18]）为：

$$u^i = u^{i-1} + \sigma \circ B^i * \sigma \circ \left(f - A * u^{i-1}\right),$$

其中 A 和 B^i 都是卷积算子。通常，时间序列中也会使用卷积算子从频率角度进行信号分解处理，尤其是在小波等经典方法中（[41, 42]）。然而，在长期时间序列预测问题中，例如我们将在§4.1中使用的数据集，时间序列可能非常复杂。例如，图3展示了来自三个不同数据集的三个典型示例，它们的频率和振幅可能存在显著差异。因此，卷积算子无法处理

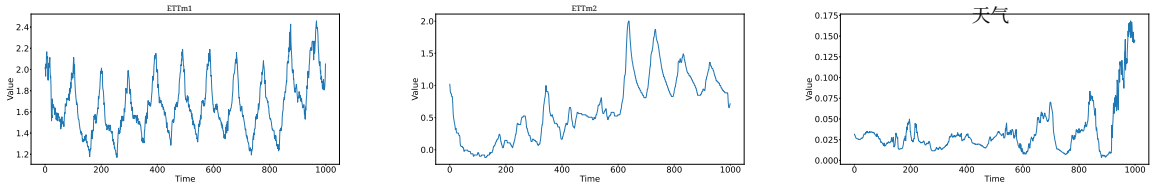


图3：来自不同数据集的一些典型长期时间序列。

以统一的方式处理这些不同的数据集。对于底层系统A，更合适的做法是将其视为一般的

线性映射，而不赋予其任何特殊结构。因此，我们使用全连接层来学习系统A。相应地，我们也选择将B^i 作为全连接层。因此，我们有

$$u^{\ell,i} = u^{\ell,i-1} + \sigma \circ B^{\ell,i} \circ \left(f^\ell - A^\ell u^{\ell,i-1}\right), \quad B^{\ell,i} \in \mathbb{R}^{I_\ell \times I_\ell}, A^\ell \in \mathbb{R}^{I_\ell \times I_\ell}. \tag{8}$$

为简单起见，我们分别用$A^\ell \in \mathbb{R}^{I_\ell \times I^\ell}$ 和$B^{\ell,i} \in \mathbb{R}^{I_\ell \times I^\ell}$ 表示带偏置的全连接算子。表5中的数值结果表明，A 与B^i 的最佳组合是全连接算子。

3.3. 多　网格结构：一种分层特征提取　方案

分层特征提取方案通常是经典图像处理方法和基于CNN的方法中用于时间序列处理的标准方案。对于时间序列信号，多层结构（多分辨率）有助于捕捉信号中具有不同频率的不同特征。例如，N-HiTS（[35]）将层次结构引入N-BEATS（[34]）的预测模块，以更少的参数实现了更高的准确率。

这里，我们考虑多重网格结构，这是一种较为成熟且适用于求解数值偏微分方程所产生的线性系统的多层方法（[21, 22, 23]）。在多重网格法中，我们首先为特征和数据空间设置一系列网格（分辨率）。更具体地说，令$I_\ell = I_{2^{\ell-1}}$ 表示第 ℓ 层网格的网格大小（分辨率）。相应地，我们在不同网格上建立受约束的线性模型，

$$A^\ell u^\ell = f^\ell, \tag{9}$$

其中u^ℓ、f^ℓ和A^ℓ 分别表示第 ℓ 层的特征、数据和底层系统。应用式（7）中的数值格式后，我们得到第 ℓ 层的基本块

$$u^{\ell,i} = u^{\ell,i-1} + \sigma \circ B^{\ell,i}\sigma(f^\ell - A^\ell u^{\ell,i-1}), \quad i = 1 : \nu_\ell. \tag{10}$$

这里，ν_ℓ 表示每一层的迭代次数，是模型的一个超参数。值得注意的是，第 ℓ 层只有一个A^ℓ，而我们需要学习 ν_ℓ 个不同的$B^{\ell,\,i}$。这是约束线性模型的一项基本假设；具体而言，每一层只能有一个系统算子A^ℓ，但可以有不同的求解算子$B^{\ell,\,i}$ 来求解模型$A^\ell u^\ell = f^\ell$。

为了整合不同网格，必须引入限制算子和延拓算子，例如典型卷积神经网络模型中的不同池化操作。近期，MgNet（[18, 19, 20]）已成功将多重网格结构应用于图像分类问题，并使用带步幅 $= 2$ 的卷积作为限制算子，以实现多尺度变换。特征u 和数据f 的尺度变换定义如下：

$$\begin{aligned} u^{\ell+1,0} &= \Pi_\ell^{\ell+1} *_2 u^{\ell,\nu_\ell} \\ f^{\ell+1} &= R_\ell^{\ell+1} *_2 (f^\ell - A^\ell * u^{\ell,\nu_\ell}) + A^{\ell+1} * u^{\ell+1,0}, \end{aligned} \tag{11}$$

其中，$\Pi_\ell^{\ell+1}*_2$ 和$R_\ell^{\ell+1}*_2$ 是步幅为 $= 2$的卷积算子。在图像分类的延拓过程中，MgNet 使用步幅为 $= 2$的反卷积算子 $\Pi_{\ell+1}^{\ell}*^2$，将粗网格上的信息插值到细网格。

在我们的模型中，考虑到全连接结构，我们将限制（下采样）和延拓（插值）视为 $\mathbb{R}^{I_\ell}$ 与 $\mathbb{R}^{I_{\ell+1}}$之间的线性映射。具体而言，我们将$\Pi_\ell^{\ell+1}$ 、$R_\ell^{\ell+1} \in \mathbb{R}^{I_{\ell+1} \times I^\ell}$ 和 $\Pi_{\ell+1}^{\ell} \in \mathbb{R}^{I_\ell \times I^{\ell+1}}$ 作为用于恢复和延拓算子的全连接层（带偏置）。

引入多重网格结构的另一个实际原因是进一步降低所提出模型的计算复杂度。表 4 中的数值结果展示了引入多重网格结构所带来的效率提升和性能改进。

这里，$u = u^{1,\nu_1}$ 是通过采用多重网格方案的全连接约束线性模型提取的特征。此外，我们将$W_1 \in \mathbb{R}^{I \times O}$ 和$W_2 \in \mathbb{R}^{O \times O}$ 作为全连接层（带偏置），以构建用于特征插值的单隐层近似 $\mathcal{I}$（u）。

3.4. FV-MgNet：用于时间序列预测的全连接 *V* 循环 *MgNet* 模型

结合上述讨论的特征提取过程与式（3）中的特征插值近似，我们得到用于时间序列预测的 FV-MgNet 算法。

算法1 y = FV-MgNet(f, J, ν_ℓ, I, O)

1: **输入**：序列$f \in \mathbb{R}^I$，网格数 J，平滑迭代次数 ν_ℓ for $\ell = 1 : J$，输入长度 I，输出长度 O。 2: 初始化：$f^1 = Wf$，$W \in \mathbb{R}^{I\times I}$，$u^{1,\,0} = 0$，$I_\ell = I2^{\ell-1}$3: **对于** $\ell = 1 : J$ **执行**4: **对于**$i = 1 :$ ν_ℓ **执行**5: 特征提取（平滑）：

$$u^{\ell,i} = u^{\ell,i-1} + \sigma \circ B^{\ell,i} \circ \left(f^\ell - A^\ell u^{\ell,i-1}\right), \quad B^{\ell,i} \in \mathbb{R}^{I_\ell \times I_\ell}, A^\ell \in \mathbb{R}^{I_\ell \times I_\ell}$$

6: **结束循环**7: 注：
$u^\ell = u^{\ell,\nu_\ell}$8: **如果**
$\ell < J$ **则**9: 限制：

$$u^{\ell+1,0} = \Pi_\ell^{\ell+1} u^\ell, \quad \Pi_\ell^{\ell+1} \in \mathbb{R}^{I_{\ell+1}\times I_\ell}$$
$$f^{\ell+1} = R_\ell^{\ell+1}(f^\ell - A^\ell u^\ell) + A^{\ell+1}u^{\ell+1,0}, \quad R_\ell^{\ell+1} \in \mathbb{R}^{I_{\ell+1}\times I_\ell}$$

10: **结束 if**11: **结束**
循环12: **对于** $\ell = J$
$-1 : 1$ **执行**13: 延拓

$$u^{\ell,0} = u^\ell + \Pi_{\ell+1}^\ell (u^{\ell+1} - u^{\ell+1,0}), \quad \Pi_{\ell+1}^\ell \in \mathbb{R}^{I_\ell \times I_{\ell+1}}$$

14: **对于**$i = 1 :$ ν_ℓ **执行**15: 特征提取
（平滑）：

$$u^{\ell,i} = u^{\ell,i-1} + \sigma \circ \overline{B}^{\ell,i} \circ \left(f^\ell - \overline{A}^\ell u^{\ell,i-1}\right), \quad \overline{B}^{\ell,i} \in \mathbb{R}^{I_\ell \times I_\ell}, \overline{A}^\ell \in \mathbb{R}^{I_\ell \times I_\ell}$$

16: **结束循环**
17: **结束循环**
特征插值：

$$y = W^2 \sigma\left(W^1 u^{1,\nu_1}\right), \quad u^{1,\nu_1} \in \mathbb{R}^I,\ W_1 \in \mathbb{R}^{I\times O},\ W_2 \in \mathbb{R}^{O\times O}$$

4. 实验

为验证 FV-MgNet 的性能，我们在六个广受欢迎的真实世界数据集上进行了实验，涵盖电力、经济、交通、天气和疾病领域。我们比较的模型包括 Transformer 模型，如 FEDformer（[16]）、ETSformer（[32]）、Autofomer（[15]）、基于全连接层的模型 N-HiTS（[35]），以及频率增强模型 FiLM（[33]）。

4.1. 数据集

我们考察了来自不同领域的六个基准数据集。这些数据集的详细信息如下。1) ETT[1] ([14])数据集包含两个子数据集（ETT1 和 ETT2），采集自两台电力变压器；每个数据集都有两个不同分辨率（15 分钟和 1 小时）的版本，其中包含多条负荷序列和一条油温序列。2) Electricity[2] 数据集包含客户的用电量，每列对应一位客户。3) Exchange[3] ([43])数据集包含八个国家的汇率。

[1]https://github.com/zhouhaoyi/ETDataset [2]
https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014
3https://github.com/laiguokun/multivariate-time-series-data

4) Traffic[4] 数据集包含加利福尼亚州（美国）高速公路系统的占用率。5) Weather[5] 数据集包含德国一年期间的 21 项气象指标。6) Illness[6] 数据集包含美国的流感样疾病患者数据，数据集的详细信息见表1。

表1：八个数据集的特征详情汇总。

数据集	LEN	DIM	FREQ
ETTm2	69680	7	15分钟
Electricity	26304	321	1H
Exchange	7588	8	1 天
Traffic	17544	862	1H
天气	52696	21	10 分钟
ILI	966	8	7 天

4.2. 实验设置

我们采用均方误差（MSE）和平均绝对误差（MAE）评估模型在预测长度 H 下的性能，这两项指标也是以往研究中最常用的指标，

$$MSE = \frac{1}{H}\sum_{\tau=t}^{t+H}(y_\tau - \hat{y}_\tau)^2 \qquad MAE = \frac{1}{H}\sum_{\tau=t}^{t+H}|y_\tau - \hat{y}_\tau| \tag{12}$$

所有数据集均按时间顺序划分为训练集、验证集和测试集，比例为 70/10/20；ETTm2 则遵循惯例，按 60/20/20 的比例划分。数据预处理采用基于训练集统计量的标准化方法。

在实验设置中，ILI 数据集的输出长度设为 24、36、48 和 60，其他数据集则设为 96、192、336 和 720。所有 Transformer 模型的输入长度均固定为 36（ILI 数据集）或 96（其他数据集）。N-HiTS 和 FiLM 通过调整参数并考察其在不同数据集上的表现来确定输入长度。为公平比较这些模型的性能，我们采用两种方式选择输入长度：1）将输入长度固定为 96，记作 FV-MgNet (96)，其设置与 Transformer 相同；2）对每个数据集，输入长度固定为 96、192、336 或 720 之一，记作 FV-MgNet。后一种方法能更好地展现 FV-MgNet 的优势。

4.3. 实验结果

图1总结了 FV-MgNet 在多变量电力数据集上的整体表现，并将其与不同模型进行了比较。对于不同的输出长度，FV-MgNet 的均方误差均显著优于其他基准模型，尤其是在输出长度较大时。此外，如图1右侧图表所示，FV-MgNet 的参数总量仅为 Transformer类模型和 N-HiTS 的 15%，以及 FiLM 的 25%。接下来，我们将从不同角度考察 FV-MgNet 的性能。

预测结果如表2所示，与其他模型相比，FV-MgNet 在六个数据集上均取得了最佳性能，但在 Exchange 和 Traffic 数据集上不及 FiLM。与在 Transformer 类型模型中达到最先进水平的 ETSformer 相比，FV-MgNet 的均方误差降低了 18% 以上；在 Traffic 和流感样疾病数据集上的改进尤为明显，均方误差分别降低了 31% 和 29%。与采用全连接层作为基本块的 N-HiTS 相比，FV-MgNet 的均方误差降低了 8%。尽管 FV-MgNet 和 FiLM 在 Exchange 和 Traffic 数据集上的性能相当，但在其他数据集上，FV-MgNet 的均方误差和平均绝对误差分别比 FiLM 低 5% 和 6%。值得注意的是，FV-MgNet 在所有数据集中最长的时间序列预测任务中取得了最佳结果，

[4] http://pems.dot.ca.gov [5] https://www.bgc-jena.mpg.de/wetter/ [6] https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

（流感样疾病数据集的输出长度为 60，其他数据集的输出长度为 720），这表明与其他模型相比，我们的模型更适合长距离预测。

表2：多变量 不同预测长度下的结果 $O \in \{96, 192, 336, 720\}$。均方误差或平均绝对误差越低，表示预测效果越好。最佳 t
结果以粗体突出显示 以粗体显示。

方法		FV-MgNet		FiLM		N-HiTS		ETSformer		FEDformer		Autoformer	
指标		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.173	**0.253**	**0.165**	0.256	0.176	0.255	0.189	0.280	0.203	0.287	0.255	0.339
	192	0.230	**0.296**	**0.222**	**0.296**	0.245	0.305	0.253	0.319	0.269	0.328	0.281	0.340
	336	0.279	**0.329**	**0.277**	0.333	0.295	0.346	0.314	0.357	0.325	0.366	0.339	0.372
	720	**0.367**	**0.385**	0.371	0.389	0.401	0.426	0.414	0.413	0.421	0.415	0.422	0.419
Electricity	96	**0.144**	**0.250**	0.154	0.267	0.147	0.249	0.187	0.304	0.183	0.297	0.201	0.317
	192	**0.163**	0.262	0.164	**0.258**	0.167	0.269	0.199	0.315	0.195	0.308	0.222	0.334
	336	**0.176**	**0.276**	0.188	0.283	0.186	0.290	0.212	0.329	0.212	0.313	0.231	0.338
	720	**0.212**	**0.308**	0.236	0.338	0.243	0.340	0.233	0.345	0.231	0.343	0.254	0.361
Exchange	96	0.082	0.206	**0.079**	**0.204**	0.092	0.211	0.085	0.204	0.139	0.276	0.197	0.323
	192	0.184	0.314	**0.159**	**0.292**	0.208	0.322	0.182	0.303	0.256	0.369	0.300	0.369
	336	0.307	0.416	**0.270**	**0.398**	0.371	0.443	0.348	0.428	0.426	0.464	0.509	0.524
	720	**0.554**	**0.582**	0.830	0.721	0.888	0.723	1.025	0.774	1.090	0.800	1.447	0.941
Traffic	96	**0.396**	**0.285**	0.416	0.294	0.402	0.282	0.607	0.392	0.562	0.349	0.613	0.388
	192	0.417	0.295	**0.408**	**0.288**	0.420	0.297	0.621	0.399	0.562	0.346	0.616	0.382
	336	0.436	0.302	**0.425**	**0.298**	0.448	0.313	0.622	0.396	0.570	0.323	0.622	0.337
	720	**0.468**	**0.315**	0.520	0.353	0.539	0.353	0.632	0.396	0.596	0.368	0.660	0.408
Weather	96	**0.155**	**0.196**	0.199	0.262	0.158	0.195	0.197	0.281	0.217	0.296	0.266	0.336
	192	**0.201**	**0.239**	0.228	0.288	0.211	0.247	0.237	0.312	0.276	0.336	0.307	0.367
	336	**0.244**	**0.279**	0.267	0.323	0.274	0.300	0.298	0.353	0.339	0.380	0.359	0.395
	720	**0.313**	**0.329**	0.319	0.361	0.351	0.353	0.352	0.388	0.403	0.428	0.419	0.428
ILI	24	**1.647**	**0.764**	1.970	0.875	1.862	0.869	2.527	1.020	2.203	0.963	3.483	1.287
	36	**1.841**	**0.839**	1.982	0.859	2.071	0.969	2.615	1.007	2.272	0.976	3.103	1.148
	48	**1.831**	**0.853**	1.868	0.896	2.346	1.042	2.359	0.972	2.209	0.981	2.669	1.085
	60	**1.765**	**0.814**	2.057	0.929	2.560	1.073	2.487	1.016	2.545	1.061	2.770	1.125

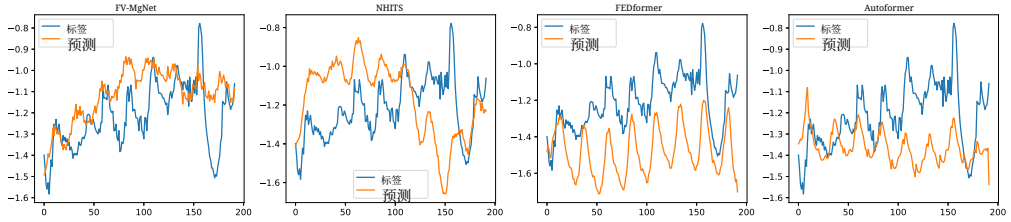
预测结果可视化图4展示了四种模型（FV-MgNet、N-HiTS、FEDformer 和 Autoformer）在三个具有不同时间模式的数据集上的预测结果：ETTh1（测试集序列 650，变量 7）、ETTh2（测试集序列 150，变量 7）和 Weather（测试集序列 185，变量 20）。在 ETTh1 上，FV-MgNet 能够准确预测季节性信息和趋势信息，N-HiTS 能够部分捕捉趋势信息，而 FedFormer 和 Autoformer 只能捕捉季节性信息。在 ETTh2 上，四种方法都能准确捕捉季节信息，且 FV-MgNet 捕捉趋势信息的准确度高于其他方法。在 Weather 上，FV-MgNet 能够准确预测季节信息和趋势信息，N-HiTS 能够部分捕捉趋势信息，而 FEDformer 和 Autoformer 均无法预测季节信息或趋势信息。

计算效率我们记录了训练过程中每个训练轮次的内存使用量和运行时间，如图2所示，其中输入长度设为 96，并改变了不同的输出长度。FV-MgNet 的内存使用量和运行时间均为最低。FV-MgNet 的运行速度是 Autoformer 的十倍、FiLM 的两倍。FV-MgNet 的运行速度是 Autoformer 的十倍、FiLM 的两倍，且比 N-HiTS 快 10%。FV-MgNet 的内存使用量是 Autoformer 的五分之一、FiLM 的二分之一、N-HiTS 的三分之二。此外，随着输出长度增加，FV-MgNet 的内存使用量和运行时间几乎保持不变。

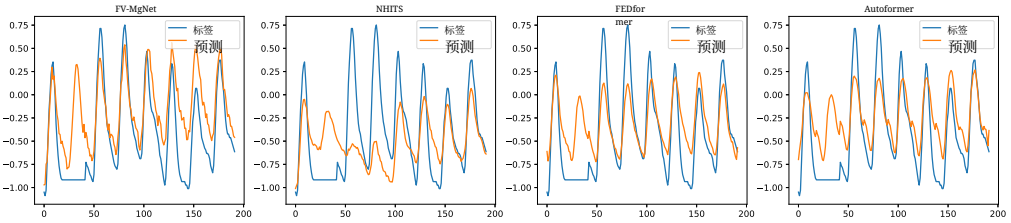
鲁棒性 直观而言，当输出长度固定时，输入长度越长，结果应该越好。我们将输入长度设为 96，并将其从 96 增加到 720。如表3所示，随着输入长度增加，Transformer 类型模型的预测结果有所下降。然而，FV-MgNet 在 ETT数据集上保持了稳定的预测结果，并且在天气数据集上取得了显著提升。结果表明，FV-MgNet 能够有效捕捉长输入序列中的特征；在长输入序列的情况下，FV-MgNet 的鲁棒性显著优于 Transformer模型。

4.4. 消融实验结果

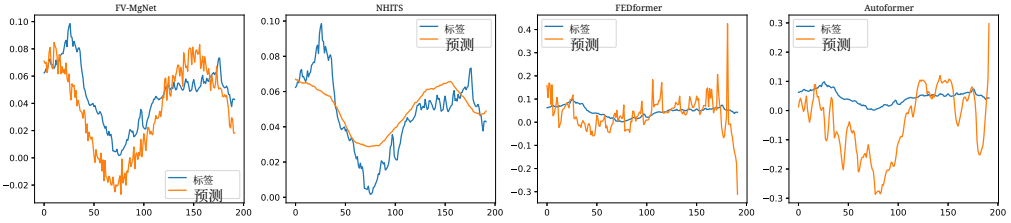
本小节中，我们研究了 MgNet 在预测问题中的结构。首先，我们在 V 循环结构的 MgNet 中，测试了用于 A 和 B 的两种不同算子：卷积算子和全连接算子。



(a) *ETTh*1



(b) *ETTh*2



(c) 天气

图4：分别展示了四种模型（FV-MgNet、N-HiTS、FEDformer 和 Autoformer）在 ETTh1、ETTh2 和天气数据集上的预测输出可视化结果，其中输出长度为$O = 192$ （X 轴），预测值为 Y 轴。

表3：在延长输入长度的情况下，对 ETTm2 和天气数据集进行 96 步预测的鲁棒性实验均方误差比较。

方法		FV-MgNet	FEDformer	Autoformer	Informer	Reformer	LogFormer	Transformer
指标		MSE	MSE	MSE	MSE	MSE	MSE	MSE
ETTm2	96	0.176	0.203	0.239	0.428	0.615	0.667	0.557
	192	0.173	0.275	0.265	0.385	0.686	0.697	0.710
	336	0.173	0.347	0.375	1.078	1.359	0.937	1.078
	720	0.180	0.392	0.315	1.057	1.443	2.153	1.691
Weather	96	0.185	0.217	0.266	0.300	0.689	0.458	0.604
	192	0.167	0.253	0.276	0.325	0.724	0.537	0.674
	336	0.155	0.278	0.336	0.346	0.736	0.742	0.692
	720	0.153	0.276	0.547	0.421	0.701	0.884	0.740

表5中的结果表明，采用全连接算子的 A 和 B 均显著优于其他选项。这验证了在 MgNet 框架下，全连接层比卷积层更适用于时间序列预测任务。接下来，我们探讨了全连接 MgNet 的结构，并设计了三种模型：(1) 残差网络，在不同网格中对 A 和 B 使用相同的维度（算法见算法2）；(2) \-MgNet，仅采用单侧层次结构（算法见算法3）；(3) FV-MgNet，如算法1所示。(3) FV-MgNet，如算法1所示。残差网络得到的结果较差，但加入单侧层次结构后，可改进为 a \-MgNet。然而，数值结果表明，采用 V 循环（双侧层次结构）的 MgNet，即 FV-MgNet，是最佳选择。

表4：消融实验，对比 FV-MgNet、\-MgNet 和残差模型。数值结果表明，多重网格结构用于特征提取可以确实提高方法的准确率。

方法		FV-MgNet		\-MgNet		残差	
指标		MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	**0.173**	**0.253**	0.183	0.270	0.212	0.318
	192	**0.230**	**0.296**	0.247	0.311	0.275	0.340
	336	**0.279**	**0.329**	0.321	0.356	0.354	0.385
	720	**0.367**	**0.385**	0.434	0.422	0.482	0.462
Electricity	96	**0.144**	**0.250**	0.179	0.287	0.213	0.312
	192	**0.163**	**0.262**	0.204	0.305	0.262	0.349
	336	**0.176**	**0.276**	0.216	0.317	0.272	0.370
	720	**0.212**	**0.308**	0.280	0.369	0.341	0.436
Exchange	96	**0.082**	**0.206**	0.093	0.223	0.104	0.283
	192	**0.184**	**0.314**	0.242	0.361	0.298	0.394
	336	**0.307**	**0.416**	0.358	0.455	0.443	0.552
	720	**0.554**	**0.582**	1.258	0.834	1.430	0.968
Traffic	96	**0.396**	**0.285**	0.717	0.398	0.589	0.348
	192	**0.417**	**0.295**	0.686	0.401	0.567	0.345
	336	**0.436**	**0.302**	0.661	0.388	0.553	0.342
	720	**0.468**	**0.315**	0.666	0.384	0.612	0.351
Weather	96	**0.155**	**0.196**	0.181	0.222	0.217	0.263
	192	**0.201**	**0.239**	0.218	0.259	0.255	0.308
	336	**0.244**	**0.279**	0.267	0.303	0.302	0.331
	720	**0.313**	**0.329**	0.335	0.345	0.389	0.403
ILI	24	**1.647**	**0.764**	3.123	1.130	3.261	1.280
	36	**1.841**	**0.839**	3.247	1.182	3.538	1.461
	48	**1.831**	**0.853**	2.890	1.063	3.295	1.198
	60	**1.765**	**0.814**	3.218	1.153	3.418	1.229

4.5. 其他研究

其他实验研究的详情如下：（1）单变量预测实验；（2）相同设置下的 Transformer 模型实验；（3）网格数和迭代次数的参数敏感性实验；（4）不同输入长度下的实验；以及（5）ETT 完整基准测试实验。

4.5.1. 单变量预测结果

我们在单变量设置下对这两个数据集进行了测试，结果如表6所示。与大量基线方法相比，FV-MgNet取得了有竞争力的结果，尤其是在 ETTm2 上进行长期预测时。FV-MgNet 可将均方误差降低 12%。

算法2 $y = $残差$(f, \nu, I, O)$

输入：序列$f \in \mathbb{R}^I$，平滑迭代次数 ν，输入长度 I，输出长度 O。

初始化：$f^1 = Wf$，$W \in \mathbb{R}^{I\times I}$，$u^0 = 0$

对于$i = 1:\nu$ **执行**

特征提取（平滑）：

$$u^i = u^{i-1} + \sigma \circ B^i\left(f - Au^{i-1}\right), \quad B^i \in \mathbb{R}^{I\times I}, A \in \mathbb{R}^{I\times I}$$

</style}结束循环

特征插值：

$$y = W^2\sigma\left(W^1(u^\nu) + b^1\right), \quad u^\nu \in \mathbb{R}^I,\ W_1 \in \mathbb{R}^{I\times O},\ W_2 \in \mathbb{R}^{O\times O}$$

算法3 $y = $ \–MgNet(f, J, ν_ℓ, I, O)

输入：序列$f \in \mathbb{R}^I$，网格数 J，平滑迭代次数 ν_ℓ for $\ell = 1:J$，输入长度 I，输出长度 O。

初始化：$f^1 = Wf$、$W \in \mathbb{R}^{I\times I}$、$u^{1,0} = 0$、$I_\ell = \frac{I}{2^{\ell-1}}$

对 $\ell = 1:J$ **执行**

对$i = 1:\nu_\ell$ **执行**

特征提取（平滑）：

$$u^{\ell,i} = u^{\ell,i-1} + \sigma \circ B^{\ell,i}\left(f^\ell - A^\ell u^{\ell,i-1}\right), \quad B^{\ell,i} \in \mathbb{R}^{I_\ell\times I_\ell},\ A^\ell \in \mathbb{R}^{I_\ell\times I_\ell}$$

结束循环

注：$u^\ell = u^{\ell,\nu_\ell}$

如果 $\ell < J$ **则**

限制：

$$u^{\ell+1,0} = \Pi_\ell^{\ell+1}u^\ell, \quad \Pi_\ell^{\ell+1} \in \mathbb{R}^{I_{\ell+1}\times I_\ell}$$

$$f^{\ell+1} = R_\ell^{\ell+1}(f^\ell - A^\ell u^\ell) + A^{\ell+1}u^{\ell+1,0}, \quad R_\ell^{\ell+1} \in \mathbb{R}^{I_{\ell+1}\times I_\ell}$$

结束条件判断

结束循环

特征插值：

$$y = W^2\sigma\left(W^1u^{J,\nu_J}\right), \quad u^{1,\nu_1} \in \mathbb{R}^{I_J},\ W_1 \in \mathbb{R}^{I_J\times O},\ W_2 \in \mathbb{R}^{0\times O}$$

表5：A和B的不同编码方法，标记为(A,B)，第一个元素表示A编码方法，第二个元素表示B编码方法。第一行表示所选的编码方法，其中Conv表示卷积层，FC表示全连接层。

(A,B)		(FC,FC)		(Conv,Conv)		(Conv,FC)		(FC，Conv)	
指标		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	**0.173**	**0.253**	0.464	0.383	0.304	0.379	0.355	0.351
	192	**0.230**	**0.296**	0.608	0.610	0.438	0.540	0.448	0.366
	336	**0.279**	**0.329**	0.893	1.276	0.481	0.624	0.588	0.487
	720	**0.367**	**0.385**	1.151	2.350	0.824	1.319	0.737	0.628
Electricity	96	**0.144**	**0.250**	0.407	0.336	0.340	0.418	0.349	0.423
	192	**0.163**	**0.262**	0.410	0.336	0.335	0.420	0.447	0.498
	336	**0.176**	**0.276**	0.393	0.316	0.460	0.507	0.423	0.493
	720	**0.212**	**0.308**	0.409	0.335	0.384	0.457	0.379	0.447
Exchange	96	**0.082**	**0.206**	0.798	0.875	0.492	0.348	0.528	0.366
	192	**0.184**	**0.314**	0.977	1.364	0.615	0.447	0.428	0.474
	336	**0.307**	**0.416**	0.903	1.180	0.676	0.482	0.486	0.493
	720	**0.554**	**0.582**	1.873	2.469	1.506	1.092	1.629	1.086
Weather	96	**0.155**	**0.196**	0.256	0.174	0.174	0.258	0.168	0.251
	192	**0.201**	**0.239**	0.308	0.243	0.226	0.309	0.225	0.305
	336	**0.244**	**0.279**	0.351	0.289	0.321	0.362	0.296	0.365
	720	**0.313**	**0.329**	0.401	0.366	0.585	0.490	0.482	0.454

表6：不同预测长度下的单变量预测结果$O \in \{96, 192, 336, 720\}$。均方误差或平均绝对误差越低，预测效果越好。最佳结果以粗体突出显示。

方法		FV-MgNet		FiLM		N-HiTS		ETSformer		FEDformer		Autoformer	
指标		均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差
ETTm2	96	0.077	0.213	0.065	0.189	0.066	**0.185**	0.080	0.212	**0.063**	0.189	0.065	0.189
	192	0.101	0.247	0.094	0.233	**0.087**	**0.223**	0.150	0.302	0.102	0.245	0.118	0.256
	336	0.121	0.272	0.121	0.274	**0.106**	**0.251**	0.175	0.334	0.130	0.279	0.154	0.305
	720	**0.153**	**0.310**	0.173	0.323	0.157	0.312	0.224	0.379	0.178	0.325	0.182	0.335
Exchange	96	0.123	0.245	0.110	0.259	**0.093**	**0.223**	0.099	0.230	0.131	0.284	0.241	0.387
	192	0.295	0.402	**0.207**	0.352	0.230	**0.313**	0.223	0.353	0.277	0.420	0.300	0.369
	336	0.337	0.486	**0.327**	**0.461**	0.370	0.486	0.421	0.497	0.426	0.511	0.509	0.524
	720	**0.641**	0.663	0.811	0.708	0.728	**0.569**	1.114	0.807	1.162	0.832	1.260	0.867

4.5.2. 相同设置下与 *Transformer*模型的比较

Transformer类模型将 ILI 的输入长度 I 设置为 36，将其他数据集的输入长度设置为 96。我们还在相同设置下对 FV-MgNet 进行了实验，并将该模型称为 FV-MgNet(96)；Transformer模型的基线包括 FEDformer（[16]）、ETSformer（[32]）、Autofomer（[15]）、Informer（ [14]）、LogTransfomer（[13]）和 Reformer（[31]），结果如表7所示。与 ETSformer 相比，FV-MgNet 将均方误差降低了 8%以上，将平均绝对误差降低了 11%。在 Exchange、Traffic 和 ILI 数据集上，改进尤为明显，均方误差分别降低了 15%、10%和 25%。实验结果表明，即使在较短的输入长度下，FV-MgNet 也能很好地捕获时间序列中的信息。

表7：不同预测长度下的多变量结果$O \in \{96, 192, 336, 720\}$。均方误差或平均绝对误差越低，预测效果越好。ILI 的输入长度I设为 36，其他数据集设为 96。最佳结果以粗体标出。

方法		FV-MgNet(96)		ETSformer		FEDformer		Autoformer		Informer		LogTrans		Reformer	
指标		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	**0.176**	**0.252**	0.189	0.280	0.203	0.287	0.255	0.339	0.365	0.453	0.768	0.642	0.658	0.619
	192	**0.241**	**0.296**	0.253	0.319	0.269	0.328	0.281	0.340	0.533	0.563	0.989	0.757	1.078	0.827
	336	**0.303**	**0.335**	0.314	0.357	0.325	0.366	0.339	0.372	1.363	0.887	1.334	0.872	1.549	0.972
	720	**0.401**	**0.399**	0.414	0.413	0.421	0.415	0.422	0.419	3.379	1.388	3.048	1.328	2.631	1.242
Electricity	96	0.204	**0.287**	0.187	0.304	**0.183**	0.297	0.201	0.317	0.274	0.368	0.258	0.357	0.312	0.402
	192	0.208	**0.291**	0.199	0.315	**0.195**	0.308	0.222	0.334	0.296	0.386	0.266	0.368	0.348	0.433
	336	0.222	**0.306**	**0.212**	0.329	**0.212**	0.313	0.231	0.338	0.300	0.394	0.280	0.380	0.350	0.433
	720	0.258	**0.332**	0.233	0.345	**0.231**	0.343	0.254	0.361	0.373	0.439	0.283	0.376	0.340	0.420
Exchange	96	**0.082**	0.206	0.085	**0.204**	0.139	0.276	0.197	0.323	0.847	0.752	0.968	0.812	1.065	0.829
	192	0.184	0.314	**0.182**	**0.303**	0.256	0.369	0.300	0.369	1.204	0.895	1.040	0.851	1.188	0.906
	336	**0.307**	**0.416**	0.348	0.428	0.426	0.464	0.509	0.524	1.672	1.036	1.659	1.081	1.357	0.976
	720	**0.554**	**0.582**	1.025	0.774	1.090	0.800	1.447	0.941	2.478	1.310	1.941	1.127	1.510	1.016
Traffic	96	0.564	0.350	0.607	0.392	**0.562**	**0.349**	0.613	0.388	0.719	0.391	0.684	0.384	0.732	0.423
	192	**0.546**	**0.345**	0.621	0.399	0.562	0.346	0.616	0.382	0.696	0.379	0.685	0.390	0.733	0.420
	336	**0.546**	0.340	0.622	0.396	0.570	**0.323**	0.622	0.337	0.777	0.420	0.733	0.408	0.742	0.420
	720	**0.565**	**0.347**	0.632	0.396	0.596	0.368	0.660	0.408	0.864	0.472	0.717	0.396	0.755	0.423
Weather	96	**0.185**	**0.219**	0.197	0.281	0.217	0.296	0.266	0.336	0.300	0.384	0.458	0.490	0.689	0.596
	192	**0.231**	**0.256**	0.237	0.312	0.276	0.336	0.307	0.367	0.598	0.544	0.658	0.589	0.752	0.638
	336	**0.283**	**0.256**	0.298	0.353	0.339	0.380	0.359	0.395	0.578	0.523	0.797	0.652	0.639	0.596
	720	0.358	**0.343**	**0.352**	0.388	0.403	0.428	0.419	0.428	1.059	0.741	0.869	0.675	1.130	0.792
ILI	24	**1.648**	**0.805**	2.527	1.020	2.203	0.963	3.483	1.287	5.764	1.677	4.480	1.444	4.400	1.382
	36	**1.894**	**0.837**	2.615	1.007	2.272	0.976	3.103	1.148	4.755	1.467	4.799	1.467	4.783	1.448
	48	**1.962**	**0.856**	2.359.	0.972	2.209	0.981	2.669	1.085	4.763	1.469	4.800	1.468	4.832	1.465
	60	**1.970**	**0.876**	2.487	1.016	2.545	1.061	2.770	1.125	5.264	1.564	5.278	1.560	4.882	1.483

4.5.3. 参数敏感性

在这里，我们测试网格数J 和 ν_ℓ 平滑迭代次数对预测准确率的影响。表8展示了迭代次数为2时网格数的影响。我们将网格数设为2、3、4和5，并观察到，在预测准确率与计算效率之间进行权衡时，网格数为3是最佳选择。随后，我们在此条件下进一步探究网格内的迭代次数，将每个网格的迭代次数设为1、2、3、4和5。表9中的结果表明，不同数据集的最佳迭代次数各不相同：迭代次数为3最适合ETTm2、Exchange和天气数据集，而迭代次数为4最适合Electricity和Traffic数据集。

表8：每次网格迭代的网格数均为2；均方误差或平均绝对误差越低，预测效果越好。最佳结果以粗体标出。

网格数		2		3		4		5	
指标		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.244	0.317	0.183	0.268	0.176	0.266	**0.175**	**0.262**
	192	0.263	0.335	0.238	0.313	0.245	0.315	**0.231**	**0.298**
	336	0.336	0.376	**0.283**	**0.334**	0.298	0.350	0.306	0.360
	720	0.465	0.452	**0.378**	**0.397**	0.390	0.400	0.400	0.406
Electricity	96	0.243	0.376	0.154	0.262	**0.149**	**0.253**	0.152	0.256
	192	0.253	0.367	**0.170**	0.274	0.172	0.275	0.166	**0.267**
	336	0.263	0.366	0.182	0.287	**0.178**	**0.280**	0.201	0.301
	720	0.364	0.446	0.224	0.321	**0.223**	**0.318**	0.226	0.322
Exchange	96	0.099	0.230	0.105	0.234	**0.086**	**0.216**	0.105	0.236
	192	0.312	0.423	0.273	0.384	**0.228**	**0.353**	0.196	0.326
	336	0.336	0.446	0.337	0.434	**0.319**	**0.428**	0.332	0.435
	720	1.405	0.924	**0.530**	**0.562**	1.026	0.764	0.980	0.751
Traffic	96	0.524	0.399	**0.419**	**0.293**	0.421	0.297	0.419	0.296
	192	0.545	0.309	**0.431**	**0.303**	0.441	0.306	0.432	0.299
	336	0.550	0.414	**0.446**	**0.305**	0.448	0.307	0.447	0.306
	720	0.578	0.424	**0.470**	**0.311**	0.477	0.318	0.471	0.322
Weather	96	0.180	0.265	**0.158**	0.221	0.159	0.209	0.160	0.209
	192	0.259	0.345	0.205	0.251	**0.200**	**0.246**	0.203	0.251
	336	0.329	0.326	**0.242**	**0.280**	0.246	0.281	0.247	0.292
	720	0.374	0.401	**0.310**	**0.333**	0.315	0.335	0.317	0.342

表9：网格数为3时的迭代次数；均方误差或平均绝对误差越低，预测效果越好。最佳结果以粗体标出。

迭代次数		1		2		3		4		5	
指标		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.176	0.268	0.183	0.268	**0.173**	**0.253**	0.174	0.261	0.175	0.265
	192	0.239	0.313	0.238	0.313	**0.230**	**0.296**	0.230	0.300	0.233	0.301
	336	0.281	0.334	0.283	0.334	**0.279**	**0.329**	0.280	0.329	0.282	0.330
	720	0.386	0.415	0.378	0.415	**0.367**	**0.385**	0.377	0.393	0.379	0.395
Electricity	96	0.145	**0.248**	0.154	0.262	0.161	0.266	**0.144**	0.250	0.202	0.301
	192	0.163	0.265	0.170	0.274	0.166	0.267	**0.163**	**0.262**	0.178	0.279
	336	0.180	0.281	0.182	0.287	0.182	0.284	**0.176**	**0.276**	0.186	0.287
	720	0.221	0.318	0.224	0.321	0.217	0.314	**0.212**	**0.308**	0.216	0.314
Exchange	96	0.084	0.211	0.105	0.234	**0.082**	**0.206**	0.087	0.216	0.083	0.208
	192	0.161	0.295	0.273	0.384	**0.184**	**0.314**	0.208	0.338	0.186	0.315
	336	0.336	0.433	0.337	0.434	**0.307**	**0.416**	0.314	0.427	0.327	0.431
	720	0.996	0.757	**0.530**	**0.562**	0.554	0.582	0.632	0.613	0.916	0.719
Traffic	96	0.413	0.289	0.419	0.293	0.423	0.296	**0.396**	**0.285**	0.429	0.299
	192	0.434	0.301	0.436	0.303	0.435	0.301	**0.417**	**0.295**	0.440	0.304
	336	0.448	0.303	0.446	0.305	0.445	0.304	**0.436**	**0.302**	0.448	0.305
	720	0.472	0.314	0.472	0.318	0.470	**0.313**	**0.468**	0.315	0.472	0.315
Weather	96	0.156	0.206	0.165	0.221	**0.155**	**0.196**	0.160	0.217	0.174	0.265
	192	**0.199**	0.246	0.205	0.251	0.201	**0.239**	0.200	0.251	0.232	0.300
	336	0.247	0.283	0.248	0.284	**0.244**	**0.279**	0.244	0.284	0.280	0.329
	720	0.321	0.351	0.310	0.333	0.313	**0.329**	**0.309**	0.332	0.367	0.385

4.5.4. 输入长度

我们通过为流感样疾病数据集设置输入长度$I \in \{24, 36, 48, 60\}$，为其他数据集设置输入长度$I \in \{96, 192, 336, 720\}$，考察了不同输入长度对预测误差的影响。结果如表10所示。除Exchange数据集外，随着输入长度增加，预测误差均显著降低。这表明，随着输入长度增加，FV-MgNET能够更好地提取时间序列中的相关特征，并做出更准确的预测。

4.5.5. ETT 完整基准测试

我们在四个 ETT 多变量数据集上测试了模型，结果如表11所示。ETTh 和 ETTm 的记录频率不同；ETTh1 和 ETTh2 数据集按小时记录，而 ETTm1 和

表10：不同输入长度下的多变量长序列时间序列预测结果。流感样疾病输入长度$I \in 24$、36、48、60及其他$I \in 96$、192、336、720

输入长度		96		192		336		720	
指标		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.176	0.252	0.173	0.251	**0.173**	**0.253**	0.180	0.260
	192	0.241	0.296	0.234	**0.291**	**0.230**	0.296	0.249	0.303
	336	0.303	0.335	0.295	0.330	**0.279**	**0.329**	0.304	0.341
	720	0.401	0.393	0.386	0.387	**0.367**	**0.385**	0.398	0.397
Electricity	96	0.204	0.287	0.162	0.259	**0.144**	**0.250**	0.151	0.252
	192	0.208	0.291	0.179	0.274	**0.163**	**0.262**	0.167	0.267
	336	0.222	0.306	0.194	0.289	**0.176**	**0.276**	0.186	0.286
	720	0.258	0.332	0.231	0.316	**0.212**	**0.308**	0.229	0.319
Exchange	96	**0.082**	**0.206**	0.101	0.225	0.119	0.242	0.110	0.240
	192	**0.184**	**0.314**	0.253	0.353	0.251	0.352	0.245	0.366
	336	**0.307**	**0.416**	0.482	0.495	0.435	0.473	0.482	0.503
	720	**0.554**	**0.582**	0.948	0.726	1.219	0.778	1.370	0.830
Traffic	96	0.564	0.350	0.446	0.302	0.414	0.290	**0.396**	**0.285**
	192	0.546	0.345	0.463	0.312	0.435	0.300	**0.417**	**0.295**
	336	0.546	0.340	0.476	0.314	0.449	0.304	**0.436**	**0.302**
	720	0.565	0.347	0.494	0.323	0.473	0.315	**0.468**	**0.315**
Weather	96	0.185	0.219	0.167	0.203	**0.155**	**0.196**	0.153	0.198
	192	0.231	0.256	0.213	0.244	**0.201**	**0.239**	0.203	0.245
	336	0.283	0.296	0.263	0.284	**0.244**	**0.279**	0.256	0.285
	720	0.358	0.343	0.340	0.337	**0.313**	**0.329**	0.330	0.335
输入长度		24		36		48		60	
ILI	24	3.487	1.067	1.648	0.805	1.787	0.792	**1.647**	**0.764**
	36	3.785	1.135	1.894	0.837	1.990	0.873	**1.841**	**0.839**
	48	2.487	0.987	1.962	0.856	1.845	0.835	**1.831**	**0.853**
	60	2.387	0.971	1.970	0.876	1.812	0.824	**1.765**	**0.814**

ETTm2 每15分钟记录一次。与多种基线模型相比，我们取得了相当的结果；尤其是在 ETTh1 和 ETTh2 数据集的平均绝对误差方面，FV-MgNet 达到了最先进水平。

5. 结论

数值偏微分方程中的迭代法和多重网格法已有成熟的数学理论。我们从约束模型的角度考虑预测问题，并利用多重网格法和 MgNet，提出了用于长期时间序列预测的 FV-MgNet 模型。我们研究了 FV-MgNet 的各种变体，发现采用 V 循环的全连接 MgNet 效果最佳。与最先进模型相比，我们展示了 FV-MgNet 在不同数据集上的良好性能：它以更少的参数和更快的推理速度取得了更好的结果。此外，随着输入序列长度的增加，FV-MgNet 的预测结果也显著改善。

参考文献

[1] J.-H. B¨ose、V. Flunkert、J. Gasthaus、T. Januschowski、D. Lange、D. Salinas、S. Schelter、M. Seeger、Y. Wang，《大规模概率需求预测》，VLDB Endowment 论文集 10 (12) (2017) 1694–1705。[2] P. Courty、H. Li，《季节性销售的时机》，《商业杂志》72 (4) (1999) 545–572。[3] B. Lim、A. Alaa、M. van der Schaar，《利用循环边际结构网络预测治疗反应随时间的变化》，NeurIPS 18 (2018) 7483–7493。[4] J. Zhang、K. Nawata，《基于调整型长短期记忆网络的流感暴发多步预测》，《流行病学与感染》146 (7) (2018) 809–816。[5] C.Zhang、D.Song、Y. Chen、X. Feng、C. Lumezanu、W. Cheng、J. Ni、B. Zong、H. Chen、N. V. Chawla，《用于多变量时间序列数据无监督异常检测与诊断的深度神经网络》，载于：《人工智能促进协会人工智能会议论文集》，第33卷，2019，1409 1416页。– 。[6] A. Gonzalez-Vidal、F. Jimenez、A. F. Gomez-Skarmeta，《一种基于特征选择的智能建筑能源多变量时间序列预测方法》，《能源与建筑》196 (2019) 71–82。[7] J. T. Connor、R. D. Martin、L. E. Atlas，《循环神经网络与鲁棒时间序列预测》，《IEEE 神经网络汇刊》5 (2) (1994) 240–254。

表11：ETT 完整基准测试上的多变量长序列时间序列预测结果。最佳结果以粗体标出。

方法		FV-MgNet		FiLM		FEDformer		Autoformer		Informer		LogTrans		Reformer	
指标		均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差
ETTh1	96	**0.364**	**0.382**	0.371	0.394	0.376	0.419	0.449	0.459	0.865	0.713	0.878	0.740	0.837	0.728
	192	**0.409**	**0.415**	0.414	0.423	0.420	0.448	0.500	0.482	1.008	0.792	1.037	0.824	0.923	0.766
	336	0.452	**0.443**	**0.442**	0.445	0.459	0.465	0.521	0.496	1.107	0.809	1.238	0.932	1.097	0.835
	720	0.473	0.482	**0.465**	**0.472**	0.506	0.507	0.514	0.512	1.181	0.865	1.135	0.852	1.257	0.889
ETTh2	96	**0.278**	**0.382**	0.284	0.397	0.346	0.388	0.358	0.397	3.755	1.525	2.116	1.197	2.626	1.317
	192	0.360	**0.436**	**0.357**	0.452	0.429	0.439	0.456	0.452	5.602	1.931	4.315	1.635	11.12	2.979
	336	**0.365**	**0.457**	0.377	0.486	0.482	0.480	0.482	0.486	4.721	1.835	1.124	1.604	9.323	2.769
	720	**0.418**	**0.432**	0.439	0.456	0.463	0.474	0.515	0.511	3.647	1.625	3.188	1.540	3.874	1.697
ETTm1	96	0.320	0.352	**0.302**	**0.345**	0.378	0.418	0.505	0.475	0.672	0.571	0.600	0.546	0.538	0.528
	192	0.363	0.380	**0.338**	**0.368**	0.426	0.441	0.553	0.496	0.795	0.669	0.837	0.700	0.658	0.592
	336	0.400	0.407	**0.373**	**0.388**	0.445	0.459	0.621	0.537	1.212	0.871	1.124	0.832	0.898	0.721
	720	0.456	0.441	**0.420**	**0.420**	0.543	0.490	0.671	0.561	1.166	0.823	1.153	0.820	1.102	0.841
ETTm2	96	0.173	0.253	**0.165**	**0.256**	0.203	0.287	0.255	0.339	0.365	0.453	0.768	0.642	0.658	0.619
	192	0.240	**0.296**	**0.222**	**0.296**	0.269	0.328	0.281	0.340	0.533	0.563	0.989	0.757	1.078	0.827
	336	0.296	**0.333**	**0.277**	**0.333**	0.325	0.366	0.339	0.372	1.363	0.887	1.334	0.872	1.549	0.972
	720	0.378	**0.385**	**0.371**	0.389	0.421	0.415	0.422	0.419	1.338	3.379	3.048	1.328	2.631	1.242

[8] H. Hewamalage、C. Bergmeir、K. Bandara，《用于时间序列预测的循环神经网络：现状与未来方向》，《国际预测期刊》37 (1) (2021) 388–427。[9] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、Ł. Kaiser、I. Polosukhin，《注意力就是你所需要的一切》，神经信息处理系统进展（NeurIPS）30。[10] J. Devlin、M. Chang、K. Lee、K. Toutanova，《BERT：用于语言理解的深度双向变换器预训练》，载于：2019年计算语言学协会北美分会人类语言技术会议（NAACL-HLT）论文集，美国明尼阿波利斯，2019年6月2日至7日，2019，第4171–4186页。[11] A. Dosovitskiy、L. Beyer、A. Kolesnikov、D. Weissenborn、X. Zhai、T. Unterthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly、J. Uszkoreit、N. Houlsby，《一张图像胜过16×16个词：用于大规模图像识别的Transformer》，载于：第9届国际学习表征会议（ICLR），奥地利，2021年5月3日至7日，2021。[12] Y. Rao、W. Zhao、Z. Zhu、J. Lu、J. Zhou，《用于图像分类的全局滤波网络》，神经信息处理系统进展（NeurIPS）34。[13] S. Li、X. Jin、Y. Xuan、X. Zhou、W. Chen、Y.-X. Wang、X. Yan，《增强局部性并突破Transformer在时间序列预测中的内存瓶颈》，载于：神经信息处理系统进展（NeurIPS），第32卷，2019。[14] H. Zhou、S. Zhang、J. Peng、S. Zhang、J. Li、H. Xiong、W. Zhang，《Informer：超越高效Transformer，实现长序列时间序列预测》，载于：第三十五届人工智能促进协会人工智能会议（AAAI），第35卷，AAAI出版社，2021，第11106–11115页。[15] H. Wu、J. Xu、J. Wang、M. Long，《Autoformer：利用自相关进行长期序列预测的分解变换器》，载于：神经信息处理系统进展（NeurIPS），2021，第101–112页。[16] T. Zhou、Z. Ma、Q. Wen、X. Wang、L. Sun、R. Jin，《FEDformer：用于长期序列预测的频率增强分解变换器》，arXiv预印本 arXiv:2201.12740。[17] Q. Wen、T. Zhou、C. Zhang、W. Chen、Z. Ma、J. Yan、L. Sun，《时间序列中的Transformer：综述》，arXiv预印本 arXiv:2202.07125。[18] J. He、J. Xu，《MgNet：多重网格与卷积神经网络的统一框架》，《中国科学：数学》(2019) 1–24。[19] J. He、J. Xu、L. Zhang、J. Zhu，《ResNet和MgNet的可解释约束线性模型》，arXiv预印本 arXiv:2112.07441。[20] J. Wang、J. Xu、J. Zhu，《具有紧致激活函数的卷积神经网络》，载于：国际计算科学会议，施普林格，2022，第319–327页。[21] J. Xu，《基于空间分解和子空间校正的迭代方法》，《SIAM评论》34 (4) (1992) 581–613。[22] J. Xu、L. Zikatanov，《希尔伯特空间中的交替投影法与子空间校正法》，《美国数学会杂志》15 (3) (2002) 573–597。[23] J. Xu、L. Zikatanov，《代数多重网格方法》，《数值学报》26 (2017) 591–721。[24] A. Graves，《长短期记忆》，《利用循环神经网络进行监督式序列标注》(2012) 37–45。[25] J. Chung、C. Gulcehre、K. Cho、Y. Bengio，《门控循环神经网络在序列建模上的实证评估》，arXiv预印本 arXiv:1412.3555。[26] R. Wen、K. Torkkola、B. Narayanaswamy、D. Madeka，《多时域分位数循环预测器》，arXiv预印本 arXiv:1711.11053。[27] D. Salinas、V. Flunkert、J. Gasthaus、T. Januschowski，《DeepAR：基于自回归循环网络的概率预测》，《国际预测期刊》36 (3) (2020) 1181–1191。[28] T. Guo、T. Lin、N. Antulov-Fantulin，《探索面向多变量数据的可解释LSTM神经网络》，载于：国际机器学习会议，PMLR，2019，第2494–2504页。[29] C. Fan、Y. Zhang、Y. Pan、X. Li、C. Zhang、R. Yuan、D. Wu、W. Wang、J. Pei、H. Huang，《利用时间注意力学习进行多时域时间序列预测》，载于：第25届ACM SIGKDD知识发现与数据挖掘国际会议论文集，2019，第2527–2535页。[30] R. Sen、H.-F. Yu、I. S. Dhillon，《立足全局，着眼局部：一种高维时间序列预测的深度神经网络方法》，神经信息处理系统进展32。[31] N. Kitaev、L. Kaiser、A. Levskaya，《Reformer：高效的Transformer》，载于：第8届国际学习表征会议（ICLR），

亚的斯亚贝巴，埃塞俄比亚，2020年4月26日至30日，2020年。[32] G. Woo、C. Liu、D. Sahoo、A. Kumar、S. Hoi，ETSformer：用于时间序列预测的指数平滑变换器，arXiv预印本 arXiv:2202.01381。[33] T. Zhou、Z. Ma、Q. Wen、L. Sun、T. Yao、R. Jin 等，FiLM：用于长期时间序列预测的频率改进勒让德记忆模型，arXiv预印本 arXiv:2205.08897。[34] B. N. Oreshkin、D. Carpov、N. Chapados、Y. Bengio，N-BEATS：用于可解释时间序列预测的神经基展开分析，arXiv预印本 arXiv:1905.10437。[35] C. Challu、K. G. Olivares、B. N. Oreshkin、F. Garza、M. Mergenthaler、A. Dubrawski，N-HiTS：用于时间序列预测的神经分层插值，arXiv预印本 arXiv:2201.12886。[36] J. B. Tenenbaum、V. De Silva、J. C. Langford，用于非线性降维的全局几何框架，《科学》290（5500）（2000）2319–2323。[37] S. T. Roweis、L. K. Saul，通过局部线性嵌入进行非线性降维，《科学》290（5500）（2000）2323–2326。[38] D. L. Donoho、C. Grimes，海森特征映射：用于高维数据的局部线性嵌入技术，美国国家科学院院刊100（10）（2003）5591–5596。[39] Q. Chen、W. Hao、J. He，幂级数展开神经网络，《计算科学杂志》59（2022）101552。[40] K. He、X. Zhang、S. Ren、J. Sun，用于图像识别的深度残差学习，载于：电气电子工程师学会计算机视觉与模式识别会议论文集，2016年，第770–778页。[41] D. B. Percival、A. T. Walden，时间序列分析的小波方法，第4卷，剑桥大学出版社，2000年。[42] T. W. Joo、S. B. Kim，基于小波滤波的时间序列预测，《专家系统与应用》42（8）（2015）3868–3874。[43] G. Lai、W.-C. Chang、Y. Yang、H. Liu，利用深度神经网络建模长期和短期时间模式，载于：第41届ACM信息检索研究与发展国际会议，2018年，第95–104页。