
时间序列作为图像: Vision Transformer

不规则采样时间序列

Zekun Li, Shiyang Li, Xifeng Yan University
of California, Santa Barbara {zekunli,
shiyangli, xyan}@cs.ucsb.edu

抽象的

不规则采样的时间序列越来越普遍,尤其是在医学领域。虽然已经开发出各种专门的方法来处理这些不规则性,但有效地建模它们的复杂动态和明显的稀疏性仍然是一个挑战。本文引入了一种新颖的视角,即将不规则采样的时间序列转换为线图图像,然后利用强大的预训练视觉转换器以与图像分类相同的方式进行时间序列分类。这种方法不仅大大简化了专门的算法设计,而且还有可能成为时间序列建模的通用框架。值得注意的是,尽管方法很简单,但我们的方法在几个流行的医疗保健和人类活动数据集上的表现优于最先进的专门算法。

尤其是在严格的传感器排除设置中,即在测试过程中省略一部分变量,我们的方法对不同程度的缺失观测值表现出很强的稳健性,即使一半的变量被屏蔽,绝对 F1 得分点也比领先的专业基线提高了 42.8%。代码和数据可在<https://github.com/Leezekun/ViTST> 上找到。

1 简介

时间序列数据在医疗保健、金融、交通和气候科学等众多领域中无处不在。随着 LSTM [13]、时间卷积网络 (TCN) [18] 和 Transformer [38] 等深度学习架构的进步,已经开发出许多用于时间序列建模的算法。然而,这些方法通常假设以规则间隔完全观察到数据,并假设数值输入大小固定。因此,这些方法在面对不规则采样的时间序列时会遇到困难,这些时间序列由一系列在观察时间之间具有不规则间隔的样本组成。为了应对这一挑战,已经开发出高度专业化的模型,这需要大量的先验知识和在模型架构选择和算法设计方面的努力 [24, 20, 3, 16, 49, 34, 48]。

与此同时,基于 Transformer 的预训练视觉模型 (尤其是 Vision Transformer)¹ 也已出现,并在图像分类和物体检测等各种视觉任务中表现出强大的能力,几乎接近人类水平的表现。受人类通过可视化分析复杂数值时间序列数据的灵活有效方式的启发,我们提出了一个问题:这些强大的预训练视觉 Transformer 能否像人类一样捕捉可视化时间序列数据中的时间模式?

为了研究这个问题,我们提出了一种称为 ViTST (视觉时间序列) 的极简方法 (视觉 Transformer),其涉及将不规则采样的多元时间序列转换为线图,将其组织成标准 RGB 图像格式,并对预先训练的视觉进行微调

¹在本文中,我们使用术语“视觉变换器”来表示一类基于预训练变换器的视觉模型,例如 ViT [9]、Swin Transformer [22]、DeiT [36] 等

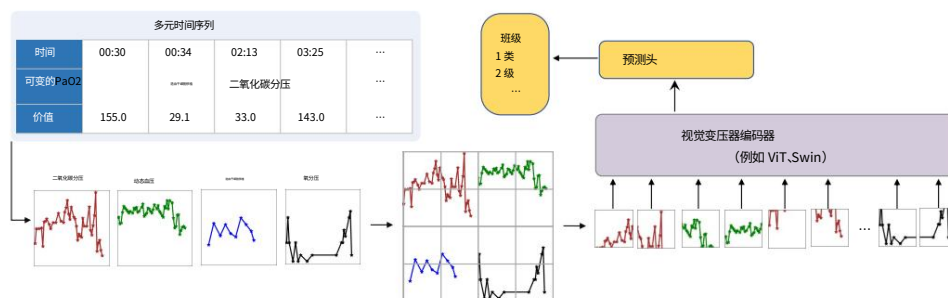


图 1:我们的方法 ViTST 的说明。示例来自医疗保健数据集 P12 [12], 它提供了 36 个患者变量的不规则抽样观测值 (我们只显示了 4 个为简单起见, 此处不再赘述)。表中的每一列都是对一个变量的观察, 观察时间和值。我们为每个变量绘制单独的线图, 并将它们排列成单张图像, 然后将其输入到视觉转换器中进行分类。

transformer 使用生成的图像作为输入进行分类。我们的方法可以这样说明如图 1 所示。

折线图是一种有效且高效的时间序列数据可视化技术, 无论不规则性、结构和尺度。它们可以捕捉关键模式, 例如时间各个线图中表示的动态以及整个过程中变量之间的相互关系单独的图表。这种可视化技术有利于我们的方法, 因为它既简单又人类可以直观地理解, 从而可以直接做出时间序列到图像的决策转换。以这种方式利用视觉模型进行时间序列建模反映了提示工程的概念, 其中个人可以直观地理解和制作提示从而潜在地提高语言模型的处理效率。

我们对所提出的方法 ViTST 进行了全面的调查和验证, 该方法已经证明了其优于专门设计的最先进 (SoTA) 方法的性能对于不规则采样的时间序列。具体来说, ViTST 比之前的 SoTA 分别高出 2.2% 和 0.7% 医疗保健数据集的绝对 AUROC 点数分别高出 1.3% 和 2.9% P19 [29] 和 P12 [12]。对于人类活动数据集 PAM [28], 我们观察到准确率提高 7.3%, 精确率提高 6.3%, 召回率提高 6.2%, F1 得分提高 6.7% (绝对点) 优于现有的 SoTA 方法。我们的方法还表现出对缺失值的强大鲁棒性观察结果, 绝对 F1 得分点明显超过之前领先的解决方案 42.8% 在不使用传感器的情况下, 测试集中一半的变量在测试。此外, 在对常规时间序列数据 (包括具有不同数字的数据) 进行测试时, 变量和扩展序列长度, ViTST 仍然取得了与专为常规时间序列建模而设计的专用算法。这强调了我们的方法, 因为传统的针对定期采样的时间序列的方法经常会遇到不规则采样的数据, 反之亦然。

综上所述, 本研究的贡献有三点: (1) 我们提出了一个简单但高度多元不规则采样时间序列分类的有效方法。尽管方法简单, 但我们的方法在与高度专业化的 SoTA 方法的竞争中取得了很好的效果。(2) 我们的方法在不规则和规则的时间序列数据上都表现出了优异的效果, 展示了其多功能性和作为时间序列建模通用框架的潜力。它提供了能够处理具有不同特征的各种时间序列数据集的强大解决方案。(3) 我们的这项工作证明了经过预先训练的视觉转换器的知识成功转移自然图像到合成可视化时间序列线图图像。我们预计这将促进快速发展和充分研究的计算机视觉技术的利用系列领域, 如更好的模型架构[23]、数据增强[32]、可解释性[4], 以及自我监督的预训练[15]。

2 相关工作

不规则采样时间序列。不规则采样时间序列是一系列观测值时间间隔各不相同。在多变量设置中,

同一时间序列可能不一致。这些特征对标准时间序列建模方法提出了重大挑战,因为标准时间序列建模方法通常假设完全观察到并定期采样数据点。处理不规则采样的常用方法是将连续时间观测转换为固定时间间隔[24, 20]。为了解释观测之间的动态,已经提出了几种模型,例如 GRU-D [3],它基于门控循环单元 (GRU) [6] 衰减隐藏状态。同样, [46]提出了一种基于多向 RNN 的方法,它可以捕捉序列间和序列内模式。除了基于递归和微分方程的模型架构外,最近的研究还探索了基于注意机制的模型。Transformer [38]自然能够处理任意的观测序列。ATTAIN [50]将注意力机制与 LSTM 结合起来对时间不规则性进行建模。SeFT [16]将不规则时间序列映射到基于可微集函数的一组观测值,并利用注意力机制进行分类。mTAND [34]学习连续时间嵌入,并结合多时间注意机制来处理连续时间输入。UTDE [48]将来自 mTAND 的嵌入和经典插补时间序列与可学习门相结合,以处理复杂的时间模式。

Raindrop [49]将不规则采样的时间序列建模为图形,并利用图形神经网络来建模变量之间的关系。虽然这些方法专门用于不规则时间序列,但我们的工作探索了一种简单且通用的基于视觉变换器的不规则采样时间序列建模方法。

使用 Transformer 进行数值时间序列建模。Transformer在时间序列建模中获得了极大的关注,因为它们具有出色的捕捉序列数据中远距离依赖关系的能力。大量基于 Transformer 的方法已被提出并成功应用于各种时间序列建模任务,例如预测[19, 51, 41, 52]、分类[47]和异常检测[45]。然而,这些方法通常是常规时间序列设置而设计的,其中同一时间戳的多变量数值被视为一个单位,并且明确建模不同单位之间的时间交互。最近的一项研究[25]建议将每个单变量时间序列划分为一系列子序列,并独立建模它们的相互作用。

这些方法都基于数值进行操作,并假设输入完全可观察,而我们提出的方法则以视觉方式处理时间序列数据。通过将时间序列转换为可视化的线图,我们可以有效地处理不规则采样的时间序列,并利用预训练视觉转换器强大的视觉表征学习能力。

成像时间序列。先前的研究已经探索了将时间序列数据转换为不同类型的图像,例如格拉姆场[39]、循环图[14, 37] 和马尔可夫转换场[40]。

这些方法通常采用卷积神经网络 (CNN)进行分类任务。然而,它们通常需要领域专业知识来设计专门的成像技术,并且并非普遍适用于各个领域。另一种方法[35]涉及利用卷积自动编码器完成从时间序列转换而来的图像,专门用于预测目的。

类似地, [31]利用 CNN 对从时间序列转换而来的图像进行编码,并使用回归器进行数值预测。然而,这些方法仍然需要大量专门的设计和修改才能适应时间序列建模。此外,它们仍然落后于当前领先的数值技术。相比之下,我们提出的方法利用预先训练的视觉变换器的强大能力,通过将时间序列转换为线图图像来实现卓越的结果,从而避免了对先验知识和特定修改和设计的需求。

3 方法

如图 1 所示,ViTST 包含两个主要步骤:(1) 将多变量时间序列转换为串联线图图像,以及 (2) 利用预先训练的视觉变换器作为分类任务的图像分类器。首先,我们介绍一些基本符号和问题表述。

符号:设 $D = \{(S_i, y_i) | i = 1, \dots, N\}$ 表示包含 N 个样本的时间序列数据集。
每个数据样本都与一个标签 $y_i \in \{1, \dots, C\}$ 相关联,其中 C 是类别的数量。
每个多变量时间序列 S_i 最多由 D 个变量的观测值组成 (有些可能没有观测值)。每个变量 d 的观测值由具有观测 v_d 时间和值的元组序列给出 $[(t_1), (t_2)], \dots, (t_{nd})$, 其中 nd 是变量的观测值数量。
 d . 如果观测时间间隔 $[t_i, t_{i+1}]$ 为不规则采样时间序列。 否则,它为 $[t_1, t_2], \dots, [t_{nd-1}, t_{nd}]$ 在不同变量或样本之间有所不同,规则时间序列。

问题表述给定包含 N 个多元时间序列的数据集 $D = \{(S_i, y_i) | i = 1, \dots, N\}$, 我们的目标是预测每个时间序列 S_i 的标签 $y_i \in \{1, \dots, C\}$ 。我们的框架主要有两个组成部分: (1) 将时间序列 S_i 转换为图像 x_i 的函数; (2) 以线图图像 x_i 作为输入并预测标签 y_i 的图像分类器。

3.1 时间序列到图像的转换

时间序列折线图。折线图是一种常用的可视化时间数据点的方法。在这种表示中,每个点都表示一个以时间和值标记的观察结果:横轴表示时间戳,纵轴表示值。观察结果按时间顺序用直线连接,任何缺失值都无缝插入。这种绘图方法允许用户灵活地将时间序列绘制为图像,直观地适合视觉转换器的处理效率。这种做法反映了使用语言模型时的提示工程,用户可以理解和调整自然语言提示以提高模型性能。

在我们的实践中,我们使用标记符号 “ ” 来表示折线图中观察到的数据点。

由于不同变量的尺度可能存在很大差异,我们将每个变量的观测值绘制在单独的线图中,如图 1 所示。每个线图 g_i 的尺度在不同的时间序列 S_i 中保持不变。我们为每个线图使用不同的颜色以进行区分。我们的初步实验表明,刻度标签和其他图形组件是多余的,因为观测值的位置本身就表明了其相对时间和值的大小。我们在第 4.3 节中研究了时间序列到图像转换的不同选择的影响。

图像创建。给定时间序列 S_i 的一组时间序列线图 $G_i = g_1, g_2, \dots, g_D$, 我们使用预定义的网格布局将它们排列在单个图像 x_i 中。我们默认采用方形网格,遵循[10]。具体而言,如果 $l \times (l - 1) < D \leq l \times l$, 我们将 D 个时间序列线图排列在大小为 $l \times l$ 的网格中;如果 $l \times l < D \leq l \times (l + 1)$, 我们将 D 个时间序列线图排列在大小为 $l \times (l + 1)$ 的网格中。例如, P19、P12 和 PAM 数据集分别包含 34、36 和 17 个变量,相应的默认网格布局为 6×6 、 6×6 和 4×5 。任何未被线图占用的网格空间都保持为空。图 6 展示了生成的图像的示例。至于变量的顺序,我们根据不规则采样的时间序列的缺失率对它们进行排序。我们在第 4.3 节探讨了不同网格布局和变量顺序的影响。

3.2 用于时间序列建模的 Vision Transformers

给定从时间序列 S_i 转换而来的图像 x_i , 我们利用图像分类器来感知图像并执行分类任务。线图图像中的时间序列模式涉及局部 (即线图中单个变量的时间动态) 和全局 (不同线图中变量之间的相关性) 上下文。为了更好地捕捉这些模式,我们选择了最近开发的视觉变换器。与主流 CNN 不同,视觉变换器被证明擅长保持空间信息,并且具有更强的捕捉局部和全局依赖关系的能力 [9, 22]。

初步。Vision Transformer (ViT) [9] 最初改编自 NLP。输入图像被分割成固定大小的块,每个块都线性嵌入并用位置嵌入进行增强。然后将得到的向量序列输入到由多头注意模块 (MSA) 和 MLP 堆栈组成的标准 Transformer 编码器中,以获得块表示。将额外的分类标记添加到序列中以执行分类或其他任务。ViT 对所有块对之间的全局单元间交互进行建模,这对于高分辨率图像来说可能在计算上很昂贵。另一方面, Swin Transformer 采用具有多级特征图的分层架构,并在非重叠窗口内局部执行自注意,从而降低计算复杂度并提高性能。除非另有说明,否则我们使用 Swin Transformer 作为默认的主干视觉模型,但也可以在此框架内使用任何其他视觉模型。

Swin Transformer 通过构建分层表示来捕获局部和全局信息,从较早层中的小块开始,并逐渐合并较深层中的相邻块。具体而言,在 W -MSA 块中,在每个不重叠的窗口内计算自注意力,从而可以捕获变量 d 的单个线图的局部变量内相互作用和时间动态。然后,移位窗口块 SW -MSA 可以实现

跨越不同线图的不同窗口之间的连接,以捕捉全局相互作用。图 2 说明了这个过程。从数学上讲,连续的 Swin Transformer 块计算如下:

$$\begin{aligned} z^{l+1} &= \text{W-MSA LN } z^l + z^{l-1}, \\ \text{和 } z^{l+1} &= \text{MLP LN } z^l + z^l, \\ z^{l+1} &= \text{SW-MSA LN } z^{l+1} + z^l, \\ \text{和 } z^{l+1} &= \text{MLP LN } z^{l+1} + z^{l+1}, \end{aligned} \tag{1}$$

其中 z^l 和 z^{l-1} 表示输出特征 (S)W-MSA 模块和用于块的 MLP 模块 l 分别为;LN 代表层正则化[1]。经过多阶段的块处理后,全局所有线图之间的相互作用可以捕捉到这些变量之间的相关性,从而能够对不同变量之间的相关性进行建模。我们还探索了使用额外的位置嵌入,包括局部位置嵌入来指示位置每个图块在其对应的线图的位置,以及全局位置嵌入来表示索引整个图像中相关线图的改善。然而,我们没有观察到持续的改善已经极具竞争力的表现,可能意味着原始的预训练位置嵌入已经能够捕获有关局部和全局补丁位置的信息。

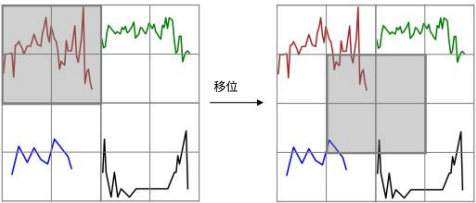


图 2:移位窗口的图示 Swin Transformer 中的方法。自我注意力在每个窗口 (灰色框)内进行计算。当窗口包含在单个线图,它捕捉局部相互作用。之后转移,窗口包括来自不同的折线图,可以对整体跨变量相互作用进行建模。

推理。我们使用视觉转换器来预测时间序列的标签,方式与图像分类。最后阶段的 Swin Transformer 块的输出用作补丁表示,在其上应用具有线性头的扁平层来获得预测 y_i 。

4 实验

4.1 实验装置

表 1:我们实验中使用的非规则采样时间序列数据集的统计数据。

数据集	#样本	#变量	#平均值	#类别	人口统计信息	不平衡	缺失率
P19	38,803	34	401	2	真的	真的	94.9%
P12	11,988	36	233	2	真的	真的	88.4%
美国糖尿病	5,333	17	4,048	8	错误的	错误的	60.0%

数据集和指标。我们使用三种广泛使用的医疗保健和人类活动数据集,如表 1 所示。P19 数据集[29]包含来自 38,803 患者,有 34 个传感器变量和一个表示败血症的二进制标签。P12 数据集[12]包括来自 11,988 名患者的数据,包括 36 个传感器变量和一个表示生存的二进制标签住院期间。最后,PAM 数据集[28]包括来自 8 个不同人类活动,每个样本提供 17 个传感器变量。我们使用提供的处理后的数据 Raindrop [49]为确保一致性,我们在所有比较样本中采用相同的数据分割基线,并使用 ROC 曲线下面积等标准指标评估性能 (AUROC)和不平衡 P12 和 P19 数据集的精确度-召回率曲线下面积 (AUPRC) 。对于更加平衡的 PAM 数据集,我们报告了准确度、精确度、召回率和 F1 分数。实施。我们利用 Matplotlib 包绘制线图并将其保存为标准 RGB 图像。对于 P19、P12 和 PAM 数据集,我们实现了 $6 \times 6 \times 6$ 和 4×5 。为了公平比较,我们为每个网格单元分配了 64×64 的固定大小 (线图),得到的图像大小分别为 384×384 、 384×384 和 256×320 。重要的是需要注意的是,图像大小也可以直接设置为任意大小,与网格单元尺寸无关。我们根据训练集上的值尺度绘制图像。我们采用了 Swin 的检查点

表 2:与不规则采样时间序列分类任务的基线方法的比较。
粗体表示表现最佳,下划线表示表现第二好。

方法	P19		P12		医疗数据集			
	意大利疟疾	意大利疟疾-仅静态特征	意大利疟疾	意大利疟疾-仅静态特征	准确性	精确	记起	F1 分数
变压器 80.7 ± 3.8	83.7 ± 1.8	GRU-D 42.7 ± 7.7	83.3 ± 0.7	47.9 ± 3.6	83.5 ± 1.5	46.3 ± 4.0	84.8 ± 1.5	85.0 ± 1.3
83.9 ± 1.7	SeFT 81.2 ± 2.3	mTAND 45.8 ± 3.2	82.6 ± 2.0	83.7 ± 2.3	46.1 ± 4.7	83.3 ± 1.6	84.9 ± 2.6	85.1 ± 2.4
± 1.3	IP-Net 84.6 ± 1.3	DGM2 -O 46.9 ± 2.1	81.9 ± 2.1	31.1 ± 4.1	67.1 ± 2.2	48.2 ± 3.4	84.6 ± 1.2	85.2 ± 1.6
MTGNN Raindrop		41.9 ± 3.1	73.9 ± 2.5	74.6 ± 4.3	47.6 ± 3.1	74.3 ± 3.8	70.0 ± 2.4	68.2 ± 1.5
		50.6 ± 2.0	84.2 ± 0.8				74.3 ± 4.0	79.5 ± 2.8
		38.1 ± 3.7	82.6 ± 1.4				75.6 ± 2.1	77.9 ± 2.2
	86.7 ± 3.4	44.7 ± 11.7	84.4 ± 1.6	39.9 ± 8.9	47.3 ± 3.6	82.4 ± 2.3	35.5 ± 6.0	85.2 ± 1.2
	81.9 ± 6.2	74.4 ± 6.7	51.8 ± 5.5	82.8 ± 1.7	83.4 ± 1.9	44.0 ± 3.0	88.5 ± 1.5	85.2 ± 1.7
	87.0 ± 2.3						89.9 ± 1.5	89.9 ± 0.6
维特斯特	89.2 ± 2.0	53.1 ± 3.4	85.1 ± 0.8	51.1 ± 4.1	95.8 ± 1.3	96.2 ± 1.3	96.1 ± 1.1	96.5 ± 1.2

在 ImageNet-21K 数据集上预训练的 Transformer2 默认的补丁大小和窗口大小都是4
分别为和 7。

训练。鉴于 P12 和 P19 数据集的高度不平衡性,我们采用了上采样
少数类的规模与多数类的规模相匹配。我们对 Swin Transformer 进行了微调
在上采样的 P19 和 P12 数据集上分别进行 2 次和 4 次迭代,在
PAM 数据集。P19 和 P12 训练使用的批大小为 48,PAM 训练使用的批大小为 72,而
学习率设置为2e-5。使用具有48G内存的A6000 GPU训练模型。

结合静态特征。在实际应用中,特别是在医疗保健领域,
不规则的时间序列数据通常伴随其他信息,例如分类或文本
特征。在 P12 和 P19 数据集中,每个患者的人口统计信息,包括体重、
高度和 ICU 类型。此静态信息随时间保持不变,并且可以
使用自然语言表达。为了将这些信息纳入我们的框架,我们采用了
将其转换为自然语言句子的模板,然后对结果文本进行编码
使用 RoBERTa-base [21]文本编码器。将得到的文本嵌入与
从视觉转换器获得的图像嵌入进行分类。静态特征
也适用于所有比较基线。

4.2 主要结果

与最新技术的比较。我们将我们的方法与几种最先进的方法进行了比较,
专门针对不规则采样的时间序列而设计,包括 Transformer [38]、 Trans-mean
(使用插补方法将缺失值替换为平均观测值
变量值)、GRU-D [3]、SeFT [16]、 mTAND [34]、 IP-Net [33] 和Raindrop [49]。
此外,我们还将我们的方法与最初为预测任务设计的两种方法进行了比较,
即 DGM2 -O [42]和 MTGNN [43]。这些模型的实现和超参数设置
基线与 Raindrop [49] 中使用的基线保持一致。具体来说,批量大小为 128
我们使用 100 个样本,所有比较模型都进行了 20 次训练。为了确保评估的公平性,
我们对 5 个数据分割中每种方法的性能进行了平均,这些分割在所有数据分割中保持一致
比较了方法。

如表 2 所示,我们提出的方法在专门的
在所有三个数据集上都采用了最先进的算法。具体来说,在 P19 和 P12 数据集上,ViTST
绝对 AUROC 点数分别提高了 2.2% 和 0.7%,绝对
AUPRC 分别指出了最先进的结果。对于 PAM 数据集,改进
更为显著,准确率提高了 7.3%,精确率提高了 6.3%,召回率提高了 6.2%,
绝对 F1 得分点为 6.7%。

排除传感器设置。我们进行了额外的评估,以评估我们的
在更具挑战性的场景下建模,其中对一组传感器 (变量)的观测
在测试期间被屏蔽。此设置模拟了现实世界中某些传感器发生故障或
变得无法到达。按照[49]中采用的方法,我们尝试了两种设置
使用 PAM 数据集: (1)保留固定传感器,即在所有数据集中删除一组固定的传感器

²<https://huggingface.co/microsoft/swin-base-patch4-window7-224-in22k>

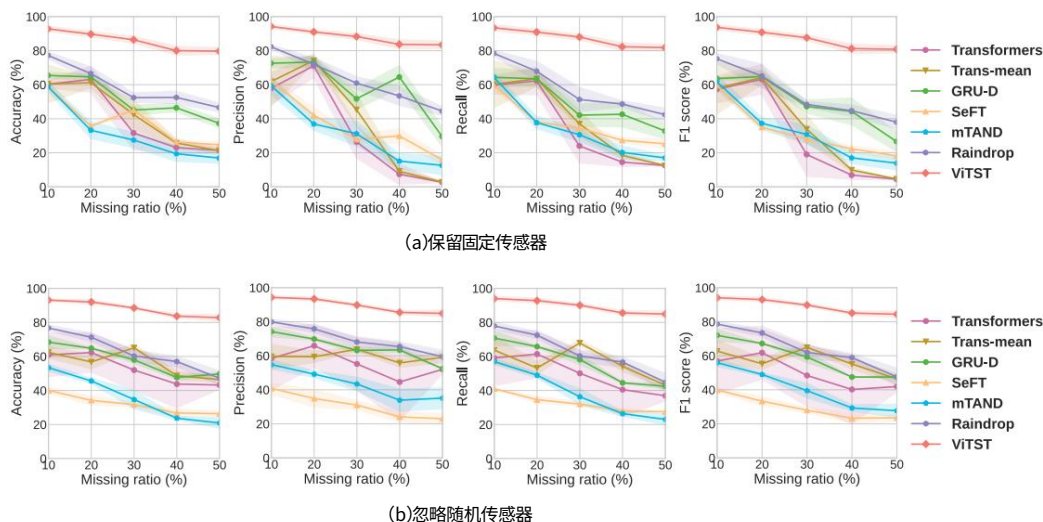


图 3: PAM 数据集上保留固定传感器和保留随机传感器设置的性能。x 轴是“缺失率”,表示屏蔽变量的比例。

样本和比较方法,以及 (2)随机丢弃传感器,即随机丢弃传感器。

值得注意的是,我们只删除了验证集和测试集中的观测值,而训练集保持不变。为了确保公平比较,我们在保留固定传感器的设置中删除了与 [49] 中相同的一组传感器。

结果如图 3 所示,从中我们可以看出,我们的方法始终保持最佳性能,并且远远超过所有专门的基线。随着缺失率从 10% 增加到 50%,我们的方法保持强劲的性能,保持在 80% 以上。相比之下,比较基线则出现明显下降。ViTST 相对于比较基线的优势变得越来越明显。即使删除了一半的变量,我们的方法仍然能够实现 80 以上的可接受性能,在保留固定传感器设置的情况下,准确率、精确度、召回率和 F1 得分分别比表现最佳的基线 Raindrop 高出 33.1%、40.9%、39.4% 和 42.8% (全部以绝对点数表示)。我们还注意到,与基线相比,我们的结果方差明显较低。这些结果表明,我们的方法对时间序列中不同程度的缺失观测具有高度的鲁棒性。

4.3 附加分析

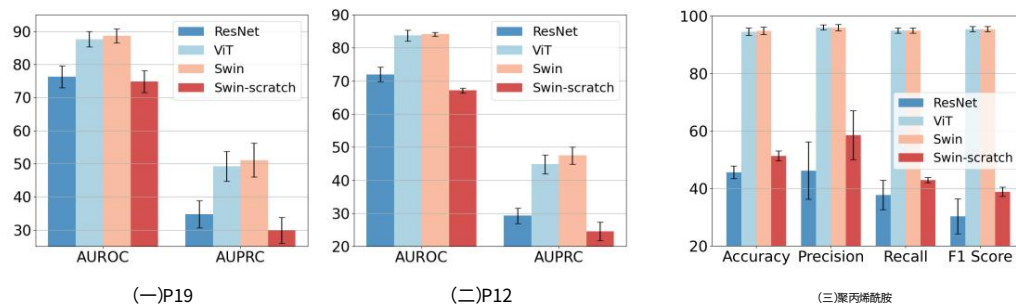


图 4:不同骨干视觉模型在 P19、P12 和 PAM 数据集上的表现。我们在此方法中不使用静态特征以排除它们的影响。

性能从何而来?我们的方法将时间序列转换为线图图像,从而允许使用视觉转换器进行时间序列建模。我们假设视觉转换器可以利用从自然图像 (例如 ImageNet [30]) 的大规模预训练中获得通用图像识别能力来捕获线图图像中的信息模式。为了验证这一点,我们比较了预训练的 Swin

表 3:时间序列到图像转换的不同策略的消融研究。

方法	P19		P12		融合策略			
	意大利语数据集	意大利语数据集-插值	意大利语数据集	意大利语数据集-插值	准确性	精确	记忆	F1 分数
默认	89.2 ± 2.0	53.1 ± 3.4 85.1 ± 1.0	51.1 ± 4.1 95.8 ± 1.3	96.2 ± 1.1	96.2 ± 1.3	96.5 ± 1.2		
无插值 89.6 ± 2.1	无颜色 无顺序	52.9 ± 3.4 85.7 ± 1.0	51.9 ± 3.4 95.6 ± 1.1	96.6 ± 0.9	95.9 ± 1.0	96.2 ± 1.0		
	89.0 ± 2.1 51.7 ± 2.5	85.3 ± 0.9 50.3 ± 3.2	95.8 ± 1.1 96.9 ± 0.7	96.0 ± 1.0	96.4 ± 0.9			
	88.8 ± 1.8 51.4 ± 4.1	84.4 ± 0.7 47.0 ± 2.9	95.0 ± 1.0 96.2 ± 0.7	95.3 ± 1.0	95.7 ± 0.9			
	89.3 ± 2.3 52.7 ± 4.5	84.0 ± 1.8 47.8 ± 4.6			-	-	-	-

使用从头开始训练的 Swin Transformer 进行测试,如图 4 所示。

在没有预先训练的情况下,性能下降提供了 Swin transformer 可以迁移的证据

将通过对自然图像进行预训练获得的知识运用到我们的合成时间序列线图中

图像,取得了令人印象深刻的表现。然而,底层机制仍需要进一步研究

在未来的研究中进行探索和探究。

不同的视觉模型表现如何?我们对几种骨干视觉模型进行了基准测试

在我们的框架内。具体来说,我们尝试了另一种流行的预训练视觉变换器 ViT³

以及基于 CNN 的预训练模型 ResNet⁴。结果如图 4 所示。预训练的

基于 Transformer 的 ViT 和 Swin Transformer 表现出了相当的性能,都优于之前最先进的方法 Raindrop。相比之下,基于 CNN 的预训练 ResNet

远远落后于视觉变换模型。这种性能差距与

在 ImageNet 等数据集上的图像分类任务中,视觉转换器

与传统的 CNN 模型相比,它已被证明在保留空间信息方面表现出色。这

优势使视觉转换器能够有效地捕捉每个块内的位置

线图子图像和整个图像,并有助于复杂动态的建模和

变量之间的关系。

如何创建时间序列线图图像?使用线图来可视化时间序列可以为我们带来

一种直观的方式来解释数据并调整可视化策略以增强清晰度和

潜在地增强了性能。为了对有效的时间序列到图像转换提供见解,我们分析了实践中几个关键元素的影响:

(1)默认的线性插值到

在折线图上连接部分观察到的数据点;(2)标记以指示观察到的数据点;

(3)使用特定于变量的颜色来区分代表不同变量的线图;(4)

在单个图像中组织多个线图时由缺失率确定的顺序。

结果如表 3 所示。鉴于 PAM 数据集中的缺失率均衡,我们

排除没有排序的结果。有趣的是,只绘制没有排序的观察到的数据点

线性插值在 P19 和 P12 数据集上取得了更好的结果。这可能归因于

插值可能引入不准确性,模糊了观测值和

插值点。此外,省略标记会使模型的辨别任务变得复杂

观测数据点与插值数据点不同,降低了其性能。缺乏独特的

颜色导致最明显的性能下降,强调使用不同色调的必要性

单独的线图可以帮助模型区分它们。虽然特定的排序顺序可能不能

确保所有数据集都能获得最佳结果,但它确实在多个数据集上提供了相对稳定的结果

数据集和评估指标。对于 PAM 数据集,这些细微差别似乎影响很小,

表明我们的方法在某些情况下对这些变化具有很强的稳健性。

网格布局和图像大小的影响。我们探讨了网格布局和图像大小的影响

尺寸对我们方法的有效性的影响。为了在网格布局之间进行公平比较,我们固定了尺寸

将每个网格单元的大小设为 64 × 64 ,并改变网格布局。如图 5 所示,我们观察到

方法对网格布局变化的鲁棒性,方形布局始终能产生良好的

不同数据集和指标的结果,这在 P12 数据集中尤为明显。

至于图片尺寸,当我们保持网格布局,但减小整体图片尺寸时,

在 P12 和 PAM 数据集上观察到明显的性能下降,这符合

我们的直觉。

对各种绘图参数的稳健性。为了衡量我们的方法对

不同的绘图参数,我们评估了线条样式/宽度和标记样式/大小等方面,

³<https://huggingface.co/google/vit-base-patch16-224-21k>

⁴<https://huggingface.co/microsoft/resnet-50>

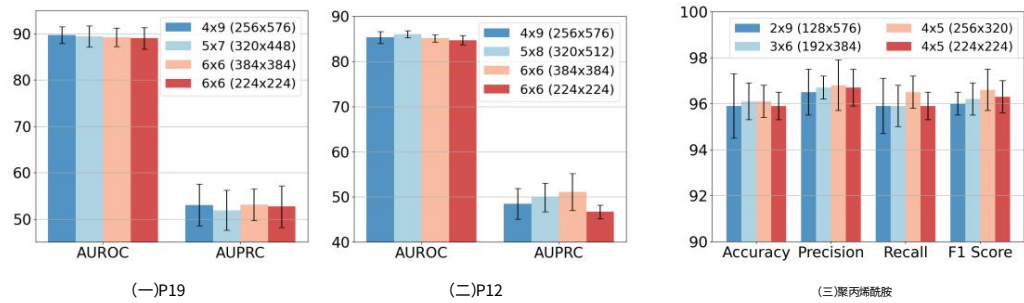


图 5:网格布局和图像大小影响的消融研究。例如,4x9 (256x576) 表示 4×9 的网格布局,图像大小为 256×576 像素。

主要使用 P19 数据集。如表 4 所示,我们的方法证明了对这些参数的变化,在不同的绘图配置中保持强劲的性能。

ViTST 捕获了哪些内容?要获得见解
融入 ViTST 当时捕捉到的模式
系列线图图像,我们分析了ViTST 模型的平均注意力图,其中 ViT
如图 6 所示。注意力图显示,该模型始终

关注信息部分,即线图
图像中的轮廓。此外,我们观察到模型适当地关注

观测数据点和斜率区域
线的变化。相反,平线图
缺乏动态模式的似乎较少
注意。这表明 ViTST 可能
能够辨别线图图像中的信息性特征和非信息性特征,

使其能够提取有意义的模式。

表 4:款式和尺寸的稳健性
行和标记。在括号中,第一个元素
第一个表示款式,第二个表示尺寸。

线	标记物	AUROC	
(实线,1)	(*, 2)	89.2 ±2.0	53.1 ±3.4
(虚线,1) (点	(*, 2)	89.2 ±2.1	53.7 ±4.1
线,1) (实	(*, 2)	89.2 ±2.1	52.8 ±4.0
线,0.5) (实	(*, 2)	88.6 ±1.7	53.0 ±3.6
线,1) (实	(*, 2)	89.2 ±2.0	53.1 ±3.4
线,2) (实	(*, 2)	88.5 ±2.3	53.6 ±3.1
线,1) (实	(*, 2)	89.2 ±2.0	53.1 ±3.4
线,1) (实	(^, 2)	89.3 ±1.9	52.6 ±4.0
线,1) (实	(°, 2)	89.1 ±1.9	51.3 ±4.2
线,1) (实	(*, 1)	88.2 ±1.4	52.1 ±4.5
线,1) (实	(*, 2)	89.2 ±2.0	53.1 ±3.4
线,1) (实线,1)	(*, 3)	88.9 ±1.9	52.8 ±3.2

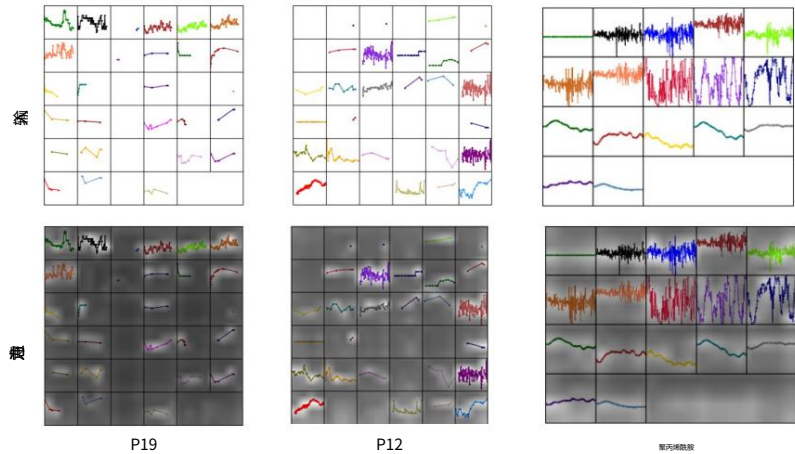


图 6:ViTST 平均注意力图的说明。

4.4 常规时间序列分类

我们的方法的一个优点是它能够对不同形状和尺度的时间序列进行建模,无论它们是规则的还是不规则的。为了评估我们的方法在规则时间序列数据上的性能,

表 5:常规多元时间序列数据集上的性能比较。粗体表示最佳表现最佳,下划线代表第二佳。

数据集	EC	UW	SCP1	SCP2	JV	SAD	HB	FD	PS	EW	平均值	
数据集统计												
#变量	3			6	7	12	13	896	1,152	29	93	模特表演
长度	3	1,751	315						405	62	17984	
DTWD	0.323	0.903	0.775	0.539	0.949	0.963	0.717	0.529	0.711	0.618	LSTM	0.323
0.399	XGBoost	0.437	0.759	0.846	0.489	0.865	0.696	0.732	0.633	0.983	Rocket	0.452
0.756	0.647	0.751	TST	0.326	0.913	0.922	0.604	0.997	0.998	0.776	0.689	0.896
0.561	0.946	0.985	0.766	0.632	0.913	0.878					维特斯特	0.456
												0.862
												0.898
												-
												0.717
												0.523
												0.727
												0.741
												0.791
												0.780

我们对来自 UEA Time 的 10 个代表性多元时间序列数据集进行了实验
系列分类档案[2]。这些数据集表现出不同的特征,总结如下

表 5. 值得注意的是,我们评估的 PS 数据集包含异常高数量的
变量(963),而 EW 数据集具有极长的时间序列(17984)。我们特别
选择这两个数据集来评估我们的方法在处理大量
变量和长时间序列。我们遵循[47]使用这些基线进行比较: DTWD
代表与扩张 CNN [11]、LSTM [13]、XGBoost [5] 相结合的维度相关 DTW,
Rocket [7] 和一个基于 Transformer 的 TST [47],它对完全观察到的数值时间序列进行操作。

我们的方法在常规时间序列数据集上的表现始终很好,如表 5 所示。平均准确率排名第二,与

表现最佳的基线方法 TST,我们的方法展示了其竞争能力。值得注意的是,
它在处理具有大量变量和观测长度的具有挑战性的数据集 PS 和 EW 时表现出色。这些
使用与其他数据集相同的图像分辨率 (384×384)获得结果,表明
我们的方法的有效性和效率。我们的方法能够处理不规则
而有规律的时间序列数据进一步强调了它的多功能性和广泛的适用性。

5 结论

在本文中,我们介绍了一种对不规则采样时间序列进行建模的新视角。通过
将时间序列数据转换为线图图像,我们可以有效地利用优势
预先训练的视觉转换器。这种方法简单但有效且用途广泛,
能够对各种特征的时间序列进行建模,无论其无规律性、不同
结构和尺度。通过大量实验,我们证明了我们的方法超越了
为不规则时间序列设计的最先进的方法,并且对变化保持强大的鲁棒性
缺失观测值的程度。此外,我们的方法在常规
时间序列数据。我们设想它作为各种时间序列的通用框架的潜力
任务。我们的研究结果强调了适应快速发展的计算机视觉技术的潜力
时间序列建模。

6 局限性和未来工作

在这项工作中,我们利用一种简单的方法来对多变量时间序列进行成像,方法是将它们转换为
使用 matplotlib 将其转换为线图图像,然后将其保存为 RGB 图像。虽然我们的结果是
前景光明,并且表现出对时间序列到图像转换过程中变化的鲁棒性,
可能还有其他方法可以可视化数据。这包括可能更可控和
精确的绘图方法或线图以外的不同图像表示。我们的研究结果还
强调了预训练视觉变换器对时间序列分类的有效性,建议
这些模型可能会利用从自然图像预训练中获得的知识。然而,
他们取得显著成功的深层原因仍需要更深入的探索和研究。
该研究是该领域的一个有希望的起点,提出了各种潜在的方向。
我们将这些进一步的探索和调查留待未来的工作。

参考

- [1] Ba, J., Kiros, JR 和 Hinton, GE. 层归一化。arXiv, abs/1607.06450, 2016 年。
- [2] Bagnall, A., Dau, HA, Lines, J., Flynn, M., Large, J., Bostrom, A., Southam, P. 和 Keogh, E. ueda 多元时间序列分类档案, 2018 年。arXiv 预印本 arXiv:1811.00075, 2018 年。
- [3] Che, Z., Purushotham, S., Cho, K., Sontag, D. 和 Liu, Y. 具有缺失值的多元时间序列的循环神经网络。《科学报告》, 8(1):1–12, 2018 年。
- [4] Chefer, H., Gur, S. 和 Wolf, L. 超越注意力可视化的 Transformer 可解释性。《IEEE/CVF 计算机视觉与模式识别会议论文集》, 第 782–791 页, 2021 年。
- [5] Chen, T. 和 Guestrin, C. Xgboost: 可扩展的树提升系统。第 22 届 acm sigkdd 国际知识发现和数据挖掘会议论文集, 第 785–794 页, 2016 年。
- [6] Chung, J., Gulcehre, C., Cho, K. 和 Bengio, Y. 门控循环神经网络的实证评估
序列建模网络。arXiv 预印本 arXiv:1412.3555, 2014 年。
- [7] Dempster, A., Petitjean, F. 和 Webb, 《GI Rocket: 使用随机卷积核实现异常快速准确的时间序列分类》。《数据挖掘与知识发现》, 34 (5):1454–1495, 2020 年。
- [8] DeVries, T. 和 Taylor, GW 改进了卷积神经网络的正则化
剪切。arXiv 预印本 arXiv:1708.04552, 2017 年。
- [9] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. 等人。一张图片胜过 16x16 个单词: 用于大规模图像识别的 Transformers。国际学习表征会议, 2020 年。
- [10] Fan, Q., Chen, C.-F. 和 Panda, R. 图像分类器能够满足动作识别的需求吗?
2021 年国际学习表征会议。
- [11] Franceschi, J.-Y., Dieuleveut, A. 和 Jaggi, M. 多元时间序列的无监督可扩展表示学习。神经信息处理系统进展, 32, 2019 年。
- [12] Goldberger, AL, Amaral, LA, Glass, L., Hausdorff, JM, Ivanov, PC, Mark, RG, Mietus, JE, Moody, GB, Peng, C.-K. 和 Stanley, HE Physiobank, physiotoolkit 和 physionet: 复杂生理信号新研究资源的组成部分。circuitry, 101(23): e215–e220, 2000。
- [13] Graves, A. 长短期记忆。使用循环神经网络进行监督序列标记, 第 37–45 页, 2012 年。
- [14] Hatami, N., Gavet, Y. 和 Debayle, J. 使用深度卷积神经网络对时间序列图像进行分类。第十届国际机器视觉会议 (ICMV 2017), 第 10696 卷, 第 242–249 页。SPIE, 2018 年。
- [15] He, K., Chen, X., Xie, S., Li, Y., Dollár, P. 和 Girshick, R. 蒙版自动编码器是可扩展的视觉学习器。IEEE/CVF 计算机视觉与模式识别会议 (CVPR) 论文集, 第 16000–16009 页, 2022 年 6 月。
- [16] Horn, M., Moor, M., Bock, C., Rieck, B. 和 Borgwardt, K. 时间序列的集合函数。国际机器学习会议, 第 4353–4363 页。PMLR, 2020 年。
- [17] Langley, P. 撰写机器学习论文。载于 Langley, P. (ed.), 第 17 届国际机器学习会议 (ICML 2000) 论文集, 第 1207–1216 页, 斯坦福, 加州, 2000 年。Morgan Kaufmann。
- [18] Lea, C., Flynn, MD, Vidal, R., Reiter, A. 和 Hager, GD, 《用于动作分割和检测的时间卷积网络》。《IEEE 计算机视觉和模式识别会议论文集》, 第 156–165 页, 2017 年。

- [19] Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X. 和 Yan, X. 增强Transformer 在时间序列预测中的局部性并突破内存瓶颈。神经信息处理系统进展,32,2019 年。
- [20] Lipton, ZC.Kale, D. 和 Wetzel, R. 使用 rnns 直接对序列中的缺失数据进行建模:改进临床时间序列的分类。在医疗保健机器学习会议上,第 253-270 页。PMLR,2016 年。
- [21] Liu, Y.,Ott, M.,Goyal, N.,Du, J.,Joshi, M.,Chen, D.,Levy, O.,Lewis, M.,Zettlemoyer, L. 和Stoyanov, V. Roberta:一种稳健优化的 bert 预训练方法。arXiv preprint arXiv:1907.11692,2019 年。
- [22] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. 和 Guo, B. Swin transformer:使用移位窗口的分层视觉转换器。IEEE/CVF国际计算机视觉会议论文集,第 10012-10022 页,2021 年。
- [23] 刘志、胡红、林勇、姚志、谢志、魏勇、宁建、曹勇、张志、董、L.,等人。
Swin transformer v2:扩大容量和分辨率。在 IEEE/CVF计算机视觉和模式识别会议论文集,第 12009-12019 页,2022 年。
- [24] Marlin, BM.Kale, DC.Khemani, RG 和 Wetzel, RC,《使用概率聚类模型在电子医疗数据中进行无监督模式发现》。《第二届ACM SIGHT 国际健康信息学研讨会论文集》,第 389-398 页,2012 年。
- [25] Nie, Y.,Nguyen, NH.Sinthong, P. 和 Kalagnanam, J. 一个时间序列值得 64 个字:使用 transformers 进行长期预测。arXiv 预印本 arXiv:2211.14730,2022 年。
- [26] OpenAI.Gpt-4 技术报告,2023 年。
- [27] Radford, A.,Kim, JW.Hallacy, C.,Ramesh, A.,Goh, G.,Agarwal, S.,Sastry, G.,Askell, A.,Mishkin, P.,Clark, J. 等人。
从自然语言监督中学习可迁移的视觉模型。国际机器学习会议,第 8748-8763 页。PMLR,2021 年。
- [28] Reiss, A. 和 Stricker, D. 引入用于活动监控的新基准数据集。2012年第 16 届可穿戴计算机国际研讨会,第 108-109 页。IEEE,2012 年。
- [29] Reyna, MA.Josef, C.,Seyedi, S.,Jeter, R.,Shashikumar, SP.Westover, MB.Sharma, A.,Nemati, S. 和 Clifford, GD,
根据临床数据对脓毒症进行早期预测:2019 年心脏病学计算挑战赛。2019 年心脏病学计算 (CinC),第 15 页。
第 1 页。IEEE,2019 年。
- [30] Russakovsky, O.,Deng, J.,Su, H.,Krause, J.,Satheesh, S.,Ma, S.,Huang, Z.,Karpthy, A.,Khosla, A.,Bernstein, M. 等。Imagenet 大规模视觉识别挑战赛。国际计算机视觉杂志,115:211-252,2015 年。
- [31] Semenoglou, A.-A.,Spiliotis, E. 和 Assimakopoulos, V. 基于图像的时间序列预测:一种深度卷积神经网络方法。《神经网络》,157:39-53,2023 年。
- [32] Shorten, C.和Khoshgoftaar, TM. 一项关于深度学习的图像数据增强的调查。
大数据杂志,6(1):1-48,2019。
- [33] Shukla, SN 和 Marlin, B. 不规则采样时间序列的插值预测网络。在 2018 年国际学习表征会议上。
- [34] Shukla, SN 和 Marlin, B.针对不规则采样时间序列的多时间注意力网络。
在 2020 年国际学习表征会议上。
- [35] Sood, S.,Zeng, Z.,Cohen, N.,Balch, T. 和 Veloso, M. 视觉时间序列预测:一种图像驱动的方法。第二届 ACM 金融人工智能国际会议论文集,第 1-9 页,2021 年。
- [36] Touvron, H.,Cord, M.,Douze, M.,Massa, F.,Sablayrolles, A. 和 Jégou, H. 训练数据 通过注意力实现高效的图像变换和提炼。国际机器学习会议,第 10347-10357 页。PMLR,2021 年。

- [37] Tripathy, R. 和 Acharya, UR,利用 rr 时间序列和 eeg 信号的特征在深度神经网络框架中自动分类睡眠阶段。生物控制论和生物医学工程,38(4):890-902,2018 年。
- [38] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AN, Kaiser, . 和 Polosukhin, I. 注意力就是你所需要的一切。神经信息处理系统进展, 2017 年 30 日。
- [39] Wang, Z. 和 Oates, T. 通过对时间序列进行成像来改进分类和归纳。第二十四届国际人工智能联合会议,2015 年。
- [40] Wang, Z. 和 Oates, T. 使用卷积神经网络对时间相关性进行空间编码以对时间数据进行分类。arXiv 预印本 arXiv:1509.07481,2015 年。
- [41] Wu, H., Xu, J., Wang, J. 和 Long, M. Autoformer :用于长期序列预测的具有自相关的分解变压器。神经信息处理系统进展,34:22419-22430,2021 年。
- [42] 吴英、倪建、程伟、宗斌、宋丹、陈哲、刘燕、张鑫、陈华和 SB Davidson,基于动态高斯混合的深度生成模型,用于稀疏多元时间序列的稳健预测。《AAAI 人工智能会议论文集》,第 35 卷,第 651-659 页,2021 年。
- [43] 吴哲、潘胜、龙国、江建、张晓玲和张晨。连接点 :使用图神经网络进行多元时间序列预测。第 26 届 ACM SIGKDD 国际知识发现与数据挖掘会议论文集,第 753-763 页,2020 年。
- [44] Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q. 和 Hu, H. Simmim :一种用于蒙版图像建模的简单框架。在 IEEE/CVF 计算机视觉与模式识别会议论文集上,第 9653-9663 页,2022 年。
- [45] Xu, J., Wu, H., Wang, J. 和 Long, M. Anomaly Transformer :具有关联差异的时间序列异常检测。arXiv preprint arXiv:2110.02642,2021 年。
- [46] Yoon, J., Zame, WR 和 van der Schaar, M. 多向循环神经网络 :一种估算缺失数据的新方法。2017 年国际机器学习会议时间序列研讨会。
- [47] Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A. 和 Eickhoff, C. 基于 Transformer 的多元时间序列表示学习框架。第 27 届 ACM SIGKDD 知识发现与数据挖掘会议论文集,第 2114-2124 页,2021 年。
- [48] Zhang, X., Li, S., Chen, Z., Yan, X. 和 Petzold, L. 通过不规则多模式电子健康记录建模改善医疗预测。ArXiv,abs/2210.12156,2022 年。
- [49] Zhang, X., Zeman, M., Tsiligkaridis, T. 和 Zitnik, M. 不规则采样多元时间序列的图引导网络。在国际学习表征会议上, 2022 年。
- [50] Zhang, Y. Attain :基于注意力机制的时间感知 lstm 网络用于疾病进展建模。在第 28 届国际人工智能联合会议 (IJCAI- 2019)的论文集上,第 4369-4375 页,中国澳门,2019 年。
- [51] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., 和 Zhang, W. Informer: Beyond efficient Transformer for long serial time-series prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pp. 11106-11115, 2021 年。
- [52] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., 和 Jin, R. Fedformer :用于长期序列预测的频率增强分解变压器。arXiv preprint arXiv:2201.12740, 2022 年。

关于时间序列线图图像创建的更多细节

实现时间序列到图像的转换可以使用 Matplotlib 来实现 package5包含以下几行代码。

```
1 def TS2Image (t, v,D,颜色,图像高度,图像宽度,网格高度,网格宽度) :
2     导入matplotlib.pyplot作为plt
3     plt.figure(figsize=(图像高度/100,图像宽度/100), dpi=100)
4     for d in range(D): # 枚举多个变量
5         plt.subplot(grid_height, grid_width, d+1) # 网格中的位置
6         # 绘制变量 d 的线图
7         plt.plot(t[d], v[d], color=colors[d], linestyle= - , marker= * )
```

折线图的轴限值轴限值决定了折线图的绘图区域和显示时间戳和值的范围。默认情况下,我们将 x 轴和 y 轴的限制设置为所有观察到的时间戳和值的范围。然而,我们发现一些极端的观察某些变量的值会大大扩展 y 轴的范围,导致大多数绘制点观测结果会聚集在一个小区域,从而导致平线图。常见的规范化和标准化方法无法解决这个问题,因为相对量级在创建的图像。因此,我们尝试了以下策略来消除极端值并缩小范围我们初步实验中的 y 轴:

- 四分位距 (IQR):IQR 是用于异常值分析的最广泛使用的方法之一检测和去除。四分位距是根据第一和第三数据集中每个变量的所有观测值的四分位数,然后用于计算上限和下限。
- 标准差 (SD):上边界和下边界通过取 3 个标准差计算得出数据集中每个变量观测值与平均值的偏差。这种方法通常假设数据呈正态分布。
- 修正的 Z 分数 (MZ):Z 分数衡量某个值偏离标准差的多少该方法与检测异常值的标准差方法类似。然而,z 分数可能受到极端值的影响,而修改后的 z 分数可以更好地处理这种影响。我们将上限和下限设置为修改后的 z 分数为 3.5 和 -3.5 的值。

我们在图 7 中展示了使用这些策略创建的图像的示例。

表 6:对确定线图极限的不同策略的初步实验。默认策略是直接将轴限值设置为数据集上所有观测值的范围。“IQR”, “SD”和 “MZS”表示三种去除极值的策略,即四分位距、标准差偏差和修正的 Z 分数。报告的数字是 5 个数据分割的平均值。

策略	P19		P12		多元模型			
	澳大利亚奥罗克	意大利奥罗克-德蒙特中心	澳大利亚奥罗克	意大利奥罗克-德蒙特中心	准确性	精确	记起	F1 分数
默认	89.4 ±1.9 52.8 ±3.8 85.6 ±1.1 49.8 ±2.5 96.1 ±0.7					96.8 ±1.1	96.5 ±0.7	96.6 ±0.9
四分位距	88.2 ±0.8	49.6 ±1.7 84.5 ±1.1 51.2 ±3.6	48.9 ±2.6 95.9 ±0.7 96.4 ±0.7		96.8 ±0.7 96.1 ±0.7 47.1 ±2.9			
	87.4 ±1.6	84.6 ±1.7 50.8 ±3.7 84.3 ±1.4	96.6 ±0.9 97.1 ±0.8 97.0 ±0.6 97.0 ±0.7					
	87.3 ±1.0		47.1 ±2.1 96.0 ±1.1 96.6 ±0.9		96.8 ±0.99 96.4 ±0.9			

显示了使用不同策略创建的图像训练的模型的性能比较在表 6 中。我们观察到,去除极端值的方法会损害性能,除非在 PAM 数据集上进行 SD 测试。虽然这些方法缩小了值范围并突出了线图的动态模式,它们会丢弃本身可能具有参考价值的极端值。因此,我们坚持默认方式,将轴限制设置为所有观察值的范围。

计算成本我们列出了不同方法在测试集上的推理时间(以秒为单位)。表 7 中的三个数据集。所有推断均在单个 Nvidia A6000 GPU 中进行。观察到我们基于视觉的方法比非视觉基线消耗更多的推理时间。然而,我们相信,在当今的机器学习实践中,这一成本仍在可接受的范围内,并且医疗应用中,考虑每个样本的推理仅花费大约 0.01 秒。

⁵<https://matplotlib.org/>

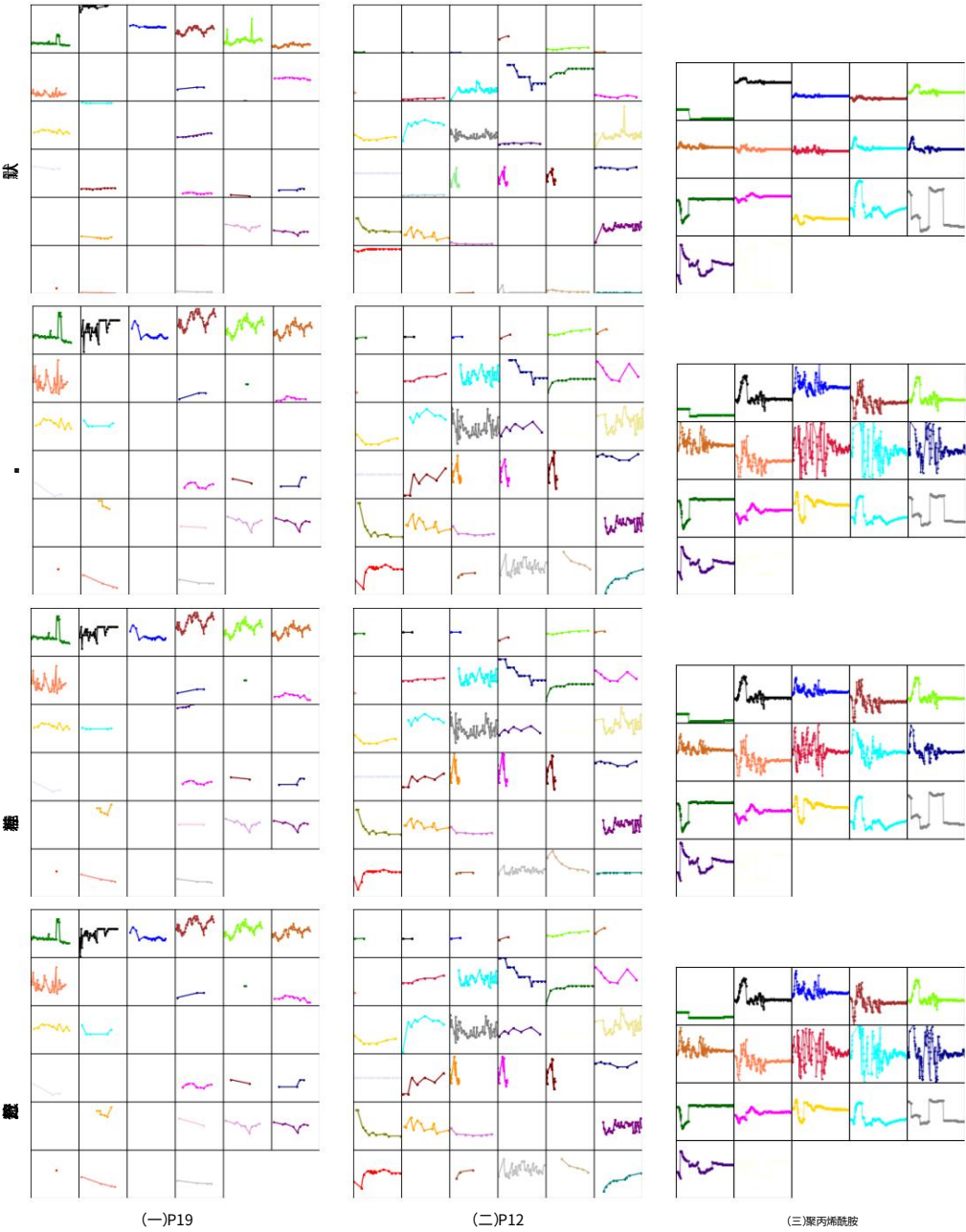


图 7:使用不同策略对 P19、P12 和 PAM 的三个样本创建的图像数据集（样本“p000019”代表 P19，“132548”代表 P12，“0”代表 PAM）。

表 7:不同方法在测试集上的推理时间（以秒为单位）。

数据集	变形金刚	mTAND	SeFT	雨滴	MTGNN	DGM2-O	GRU-D	ViTST
P19	0.21	0.52	2.72	3.05	3.62	2.47	31.04	44.51
P12	0.12	0.44	0.97	1.27	1.46	2.80	10.13	12.14
聚丙烯酰胺	0.06	0.23	0.89	0.67	1.16	2.98	4.55	5.30

B 更多实验细节

B.1 数据集

我们使用了[49]处理的数据集,其详细信息如下。

P19:2019 年 PhysioNet 败血症早期预测挑战赛。⁶ P19 数据集[29]包括临床研究人员对 38,803 名患者的数据进行了分析,旨在预测未来 6 小时内是否会发生败血症。数据集包括 34 个不规则采样的传感器,每个传感器有 8 个生命体征和 26 个实验室值患者,以及 6 个人口统计特征。为了处理静态特征,我们使用了以下模板在表 8 中,并利用预先训练的 Roberta 模型来提取文本特征。这些文本然后将特征与从视觉转换器获得的视觉特征相结合,执行二分类。数据集高度不平衡,只有 4% 的样本为正样本,并且缺失率为94.9%。

P12:2012 年 PhysioNet 死亡率预测挑战赛。⁷ P12数据集[12]包括临床数据来自 11,988 名 ICU 患者,其中 36 个不规则采样传感器观测值和 6 个静态人口统计为每个患者提供特征。目标是预测患者死亡率,这是一个二元分类任务。数据集高度不平衡,约 86% 的样本为负数。该数据集的缺失率为88.4%。

PAM:PAMAP2 体力活动监测。⁸ PAM 数据集最初包含 18 9 名受试者佩戴 3 个惯性测量装置进行体力活动。然而,为了使其适合对于不规则时间序列分类, [49]排除了第九个受试者,因为它的传感器长度较短读数,并且样本少于 500 个的 18 个活动中有 10 个也被排除在外。作为结果,该任务是一个 8 路分类,有 5,333 个样本,每个样本有 600 个连续观测。为了模拟不规则的时间序列设置,60% 的观测值被随机删除。没有静态特征,8个类别大致平衡。缺失率为60.0%。

表 8:将静态特征转换为自然语言句子的模板。

数据集静态特征		模板A 患者	例子
P19	年龄、性别、单位 1 (内科 ICU)、单位 2 (外科 ICU)、住院时间; ICULOS (ICU 住院时间)	年龄为 {Age} 岁, {性别},前往{Unit1&Unit2} 入院后 {HospAdmTime} 小时,住院时间已达 {ICU- LOS} 小时。	患者今年65岁, 女性,10 小时后被送往重症监护室 医院承认,已经住院 在那里呆了20个小时。
	记录ID、年龄、性别、 身高 (厘米)、ICU类型、 重量 (公斤)	患者年龄为 {Age} 岁,{Gender},{Height} cm, {Weight} kg, 住在{ICUType}。	患者今年48岁, 男性,171厘米,78公斤, 住在外科ICU。

表 9:针对静态特征编码的不同方法的消融研究。

方法	P19		P12	
	澳大利亚奥罗克	澳大利亚阿德莱德-詹姆斯研究所	澳大利亚奥罗克	澳大利亚阿德莱德-詹姆斯研究所
雨滴	87.0 ± 2.3	51.8 ± 5.5	82.8 ± 1.7	44.0 ± 3.0
斯温	89.4 ± 1.8	50.2 ± 3.0	84.3 ± 0.6	49.3 ± 3.7
斯温-MLP	88.6 ± 1.3	51.4 ± 3.7	84.6 ± 0.9	48.7 ± 3.2
斯温-罗伯塔	89.4 ± 1.9	52.8 ± 3.8	85.6 ± 1.1	49.8 ± 2.5

B.2 静态特征实验

时间序列数据通常与其他模态的信息相关,例如文本临床医疗保健领域的电子健康记录 (EHR) 中的注释。我们的方法自然适合因为我们将时间序列数据转换为图像,所以需要整合这些信息,因此各种可以利用视觉语言和多模式技术来整合视觉 (时间序列) 信息和其他模态的信息。例如,CLIP [27]学习一个共享的

⁶<https://phyonet.org/content/challenge-2019/1.0.0/>
⁷<https://phyonet.org/content/challenge-2012/1.0.0/>
⁸<https://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring>

隐藏特征空间,其中配对的图像和文本保持接近。在我们的框架下,这种共享还可以学习配对的视觉时间序列图像和文本临床记录的空间,这是我们未来的研究方向。它也为 GPT- 4 [26]等多模态模型的应用铺平了道路,可以同时处理可视化的时间序列数据和临床记录。在我们目前的实验中,我们使用了文本编码器 Roberta-base,对文本人口统计信息进行编码。P19 和 P12 数据集。我们还尝试对原始分类特征进行归一化,并使用 MLP 对其进行编码,就像以前的工作一样,并与强基线 Raindrop 进行比较。结果如表 9 所示。我们观察到,即使不使用静态特征,我们的方法也具有已经超越了 Raindrop。此外,利用 Roberta 编码并整合文本特征比在分类特征上应用 MLP 更有效。

表 10:在两种设置上融合不同变量的线图图像的初步实验。默认设置是先将所有线图图像排列成一张图像,然后学习分类的表示。ViT -subimage表示我们首先学习每个线图子图像分别进行,然后将它们的表示连接起来以进行分类。

策略	P19		P12		默认设置		
	澳大利亚奥罗克	澳大利亚奥罗克公司	澳大利亚奥罗克公司	准确性	精确	记起	F1 分数
维特	87.9 ± 2.5 51.6 ± 3.7 84.8 ± 1.3 48.1 ± 3.8 93.4 ± 0.7	94.7 ± 0.9 94.1 ± 0.7 94.3 ± 0.7					
ViT 子图像	85.1 ± 1.5 47.9 ± 3.4 77.6 ± 3.2 35.5 ± 6.2 90.4 ± 1.3 92.9 ± 1.0 91.1 ± 0.9 91.9 ± 1.0						

B.3 图像融合实验

在我们的初步实验中,我们研究了两种不同的方法来融合每个多元时间序列数据中的线图子图像。首先,我们独立处理每个子图像以学习补丁/图像表示,然后将它们各自的补丁表示连接起来输入到最终的预测层。相比之下,我们的默认方法将所有子图像聚合成单个图像来学习补丁嵌入。这些策略之间的关键区别在于补丁可以关注来自其他子图像的补丁,以及位置嵌入在表示学习阶段。除此之外,两种方法的其他参数都是一致的。我们在 ViT 上进行了测试,表 10 显示了这两种设置的性能比较。有证据表明,关注其他子图像的补丁具有优势。这可能允许在自注意力层中以粒度级别捕获跨变量相关性,而不是仅在最后的线性预测层。

表 11:评估的常规多元时间序列数据集的统计数据和超参数设置。

数据集变量类别长度训练规模网格布局图像大小学习率时期								
电子商务		4	1,751	261	2×2	256×256	1e-4	20
你的		8	315	120	2×2	256×256	1e-4	100
SCP1	3	2	896	268	2×3	256×384	1e-4	100
SCP2	3	2	1,152	200	3×3	384×384	5e-5	100
合演	6	9	29	270	4×4	384×384	1e-4	100
伤心	7	10	93	6599	4×4	384×384	1e-5	20
乙肝	12	2	405	204	4×4	384×384	1e-4	100
FD	13	2	62	5890	12×12	384×384	5e-4	100
附言	61	7	144	267	32×32	384×384	5e-4	100
那个	144 963 6	5	17984	128	2×3	256×384	2e-5	100

B.4 规则时间序列实验

我们从 UEA 时间序列分类档案[2]中选择了 10 个具有代表性的多变量时间序列数据集,这些数据集具有不同的特征,包括类别、变量和时间的数量系列长度。我们选择的数据集是 EthanolConcentration (EC)、Handwriting (HW)、UWaveGestureLibrary (UW)、SelfRegulationSCP1 (SCP1)、SelfRegulationSCP2 (SCP2)、JapaneseVowels (JV)、SpokenArabicDigits (SAD)、Heartbeat (HB)、FaceDetection (FD)、PEMS-SF (PS) 和 EigenWorms (EW)。值得注意的是,PS 数据集的变量数量异常多(963),而 EW 数据集具有极长的时间序列(17984)。这两个数据集使我们能够评估有效性在处理大量变量和长时间序列时,我们的方法非常有用。我们应用了根据这些数据集的网格布局,设置不同的图像大小。超参数设置如下:

如表 11 所示,我们将 cutout [8]数据增强方法应用于 SCP1,SCP2 和 JV 由于训练集规模较小,因此数据集。

B.5 自监督学习

我们对时间序列线上的蒙版图像建模自监督预训练进行了初步探索 图形图像。我们在每个线图上随机屏蔽宽度为 32 的块列,每个块位于 网格单元。掩蔽率设置为 50%。我们对 Swin Transformer 模型进行了 10 个 epoch 的微调 批处理大小为 48。学习率为 2e-5。根据[44],我们使用线性层来重构 像素值并对掩蔽像素采用l1损失:

L = \frac{1}{\Omega(p_M - p_M)} \sum_{p \in \Omega} |p - p_M|, \tag{2}

其中p_M和p_M分别为掩蔽像素和重建像素; \Omega(\cdot)表示 元素。通过自监督蒙版图像建模,AUPRC 性能提升 1.0 点 (百分比)从 52.8 (\pm 3.8)降至 53.8 (\pm 3.2)。AUROC 点 (百分比)略有下降 从89.4 (\pm 1.9)下降到88.9 (\pm 2.1)。

B.6 完整实验结果

我们在表 12 中展示了不考虑传感器的设置下的完整实验结果。

表 12:PAM 数据集上保留传感器设置的完整结果。“缺失率”表示 掩蔽变量的比例。

丢失的 比率	方法	PAM (保留固定传感器)				PAM (不考虑随机传感器)			
		准确性	精确	记起	F1 分数	准确性	精确	记起	F1 分数
10%	变压器	60.3 \pm 2.4	57.8 \pm 9.3	59.8 \pm 5.4	57.2 \pm 8.0	60.9 \pm 12.8	58.4 \pm 18.4	59.1 \pm 16.2	56.9 \pm 18.9
	均值	60.4 \pm 11.2	61.8 \pm 14.9	60.2 \pm 13.8	58.0 \pm 15.2	62.4 \pm 3.5	62.7 \pm 6.4		
	胃肠道疾病	65.4 \pm 1.7	72.6 \pm 2.6	64.3 \pm 5.3	63.6 \pm 0.4	68.4 \pm 3.7	72.0 \pm 3.7	59.6 \pm 7.2	63.7 \pm 8.1
	58.9 \pm 2.3	62.5 \pm 1.8	59.6 \pm 2.6	59.6 \pm 2.6	40.0 \pm 1.9	39.9 \pm 1.5		74.2 \pm 3.0	70.8 \pm 4.2
	平均 TAND	58.8 \pm 2.7	59.5 \pm 5.3	64.4 \pm 2.9	61.8 \pm 4.1	53.4 \pm 2.0	55.9 \pm 2.2	40.8 \pm 3.2	41.0 \pm 0.7
	雨滴	77.2 \pm 2.1	82.3 \pm 1.1	78.4 \pm 1.9	75.2 \pm 3.1	76.7 \pm 1.8	78.6 \pm 1.8	54.8 \pm 2.7	57.0 \pm 1.9
20%	维特斯特	92.8 \pm 1.6	94.2 \pm 1.3	93.4 \pm 1.8	93.7 \pm 1.6	93.1 \pm 0.9	94.3 \pm 0.9	94.0 \pm 1.2	94.1 \pm 1.1
	变压器	63.1 \pm 7.6	71.1 \pm 7.1	62.2 \pm 8.2	均值	61.2 \pm 3.0	74.2 \pm 1.8	63.5 \pm 6.3	2 \pm 8.7
	GRU-D	64.6 \pm 1.8	73.3 \pm 3.6	63.5 \pm 4.6	SeFT	35.7 \pm 0.5	42.1 \pm 4.8	64.1 \pm 4.1	56.8 \pm 4.1
	38.1 \pm 1.3	mTAND	33.2 \pm 5.0	36.9 \pm 3.7	37.7 \pm 3.7	72.0 \pm 3.9	67.9 \pm 5.8	64.8 \pm 3.6	64.8 \pm 0.4
	66.5 \pm 4.0	89.7 \pm 1.7	91.0 \pm 1.4	90.9 \pm 1.9				69.8 \pm 0.8	65.8 \pm 0.5
	雨滴							67.2 \pm 0.0	
30%	维特斯特							35.0 \pm 2.2	34.2 \pm 2.8
	变压器	31.6 \pm 10.0	26.4 \pm 9.7	均值	42.5 \pm 8.6	45.3 \pm 24.0	10.0 \pm 12.8	52.0 \pm 11.9	55.2 \pm 15.3
	9.6	64.9 \pm 1.7						50.1 \pm 13.3	48.4 \pm 18.2
	格雷特迪瓦	45.1 \pm 2.9	51.7 \pm 6.2	59.3 \pm 3.5				65.1 \pm 6.1	59.4 \pm 3.4
	SeFT	32.7 \pm 2.3	27.9 \pm 2.4	28.0 \pm 1.6				53.2 \pm 3.9	55.3 \pm 3.5
	平均输出速率	27.5 \pm 4.5	31.2 \pm 7.3	39.5 \pm 4.4				64.8 \pm 0.4	69.8 \pm 0.8
40%	雨滴		52.4 \pm 2.8	60.9 \pm 3.8	61.9 \pm 3.9			65.8 \pm 0.5	67.2 \pm 0.0
	维特斯特		86.4 \pm 2.1	88.3 \pm 1.8	88.0 \pm 1.7	87.6 \pm 1.7	88.5 \pm 0.7	89.8 \pm 0.9	90.1 \pm 1.0
	变压器	23.0 \pm 3.5	均值	25.7 \pm 2.5	7.4 \pm 6.0	14.5 \pm 2.6	6.9 \pm 2.6	43.8 \pm 14.0	44.6 \pm 23.0
	GRU-D	46.4 \pm 2.5	SeFT	26.3 \pm 9.1	\pm 2.3	18.5 \pm 1.4	9.9 \pm 1.1	40.5 \pm 15.9	40.2 \pm 20.1
	0.9	mTAND	19.4 \pm 4.5	雨滴	52.5	64.5 \pm 6.8	42.6 \pm 7.4	55.8 \pm 2.6	54.2 \pm 3.0
	\pm 3.7	ViTST	80.0 \pm 2.6	83.7 \pm 29.9	\pm 4.5	27.3 \pm 1.6	22.3 \pm 1.9	63.4 \pm 1.6	44.5 \pm 0.5
50%	2.7	82.3 \pm 2.4	81.2 \pm 2.7	83.7 \pm 15.1	\pm 4.4	20.2 \pm 3.8	17.0 \pm 3.4	24.1 \pm 3.4	28.0 \pm 1.2
	1.3	85.5 \pm 1.1	85.6 \pm 1.4	85.1 \pm 53.4	\pm 5.6	48.6 \pm 1.9	44.7 \pm 3.4	33.9 \pm 6.5	26.4 \pm 1.6
	变压器	21.4 \pm 1.8	均值	21.3 \pm 1.6	2.7 \pm 0.2	12.5 \pm 0.4	4.4 \pm 0.3	65.4 \pm 2.7	56.7 \pm 3.1
	GRU-D	37.3 \pm 2.7	SeFT	24.7 \pm 2.8	\pm 0.4	12.5 \pm 0.7	4.6 \pm 0.2	52.0 \pm 2.5	36.9 \pm 3.1
	1.7	mTAND	16.9 \pm 3.1	雨滴	46.6	32.8 \pm 4.6	26.6 \pm 5.9	59.1 \pm 3.2	43.1 \pm 2.2
	\pm 2.6	ViTST	79.7 \pm 2.1	83.4 \pm 15.9	\pm 2.7	25.3 \pm 2.6	18.2 \pm 2.4	52.4 \pm 0.3	42.5 \pm 1.7