arXiv:2601.00446v1 [cs.LG] 1 Jan 2026

时间序列基础模型在异常检测中适应策略的比较研究

Miseon Park[1] 和 Kijung Yoon[1,2]

[1] 电子工程系，汉阳大学，首尔，韩国[2] 人工智能系，汉阳大学，首尔，韩国
{misun9631,kiyoon}@hanyang.ac.kr

摘要。 时间序列异常检测对于复杂系统的可靠运行至关重要，但现有大多数方法都需要大量的特定任务训练。我们探讨了在大规模异构数据上预训练的时间序列基础模型（TSFMs）能否作为异常检测的通用骨干网络。通过在多个基准测试上开展系统性实验，我们比较了零样本推理、全量模型适配和参数高效微调（PEFT）策略。结果表明，TSFMs 的表现优于特定任务基线，在 AUC-PR 和 VUS-PR 指标上取得了显著提升，尤其是在严重类别不平衡的情况下。此外，LoRA、OFT 和 HRA 等参数高效微调方法不仅降低了计算成本，在大多数情况下还达到了或超过了全量微调的效果，这表明即使时间序列基础模型最初是为预测而预训练的，也能够高效地适配于异常检测任务。这些发现表明，TSFMs 有望成为可扩展且高效的时间序列异常检测通用模型。

关键词： 时间序列分析 · 异常检测 · 基础模型 · 参数高效微调

1 引言

时间序列异常检测（TSAD）是数据驱动监控系统中的一个基础问题，在维护从工业过程控制到金融监测等现实应用的可靠性、安全性与保障性方面发挥着至关重要的作用 [3, 6]。TSAD 的目标是识别序列数据中显著偏离预期时间模式的异常行为。多年来，针对这一问题形成了两种主要的方法范式：基于预测的方法和基于重构的方法。基于预测的方法从历史观测中学习时间依赖关系，以预测未来值；当观测数据与这些预测值存在较大偏差时，即判定为异常 [8, 7, 36]。由于具有因果性，这些方法尤其适用于在线检测场景，因为在每个时间步只能获取过去的信息。典型实现依赖于循环神经网络 [9, 15], 、Trans-formers [37, 34], 或时序图神经网络 [40, 8], 等预测模型，并以预测误差

作为异常分数。相比之下，基于重构的方法旨在对输入时间序列进行编码和重构，并将较大的重构误差视为异常行为的指标 [24, 17, 37]。这些方法通常应用于回顾性场景，在此类场景中可以获得完整序列，其基本假设是：主要在正常数据上训练的模型无法忠实地重构罕见或异常模式。尽管两种范式都已展现出良好的效果，但它们通常都依赖大量的任务特定训练和细致调优，这限制了其在各种现实场景中的适应性。

近年来，时间序列基础模型（TSFM）已成为时间序列表示学习与预测的一种通用范式 [39, 16,2, 35, 33]。值得注意的是，TSFM 的发展并非单纯由异常检测驱动，而是服务于更广泛的目标：从大规模异构时间序列数据中学习可迁移的时间表示。受自然语言处理领域基础模型的启发，TSFM 通过预训练来捕捉通用时间结构，以支持广泛的下游任务。在这些任务中，基于预测的异常检测自然能够受益于 TSFM 强大的预测能力，因为可以通过预测结果与观测值之间的偏差识别异常。在实际应用中，TSFM 仅根据过去的观测值生成多步预测，因此可直接用于在线检测场景。此外，其强大的泛化能力支持零样本异常检测，无需针对特定任务重新训练即可标记异常。先前研究表明，TimeGPT [10], 、Chronos [2],、Moirai [35],和 Time-MoE [29]等模型在未见数据集上取得了具有竞争力的预测性能，凸显了它们作为广泛适用的预测骨干模型的潜力。

随着基础模型的兴起，参数高效微调（PEFT）被提出，作为一种通用策略，用于将大型预训练模型适配到下游任务。与时间序列基础模型类似，参数高效微调方法最初并非为时间序列预测或异常检测而开发，而是为了解决更广泛的挑战：如何在不更新所有参数的情况下，高效地使大规模模型适应特定任务。LoRA [14], IA^3 [19], OFT [25],和 HRA [38] 等技术仅需调整一小部分参数或插入轻量级模块，即可实现适配。将这些方法应用于时间序列基础模型时，它们提供了一种切实可行的手段，可调整预测行为以适应特定领域特征，进而提升异常检测性能。更重要的是，参数高效微调使实践者能够利用预训练时间序列基础模型的泛化能力，同时将计算和内存成本控制在合理范围内。

在本研究中，我们探讨如何将通用时间序列基础模型和参数高效微调技术重新用于 时间序列异常检测。我们的假设是，在大型且多样化的数据集上预训练的时间序列基础模型能够编码可迁移时间知识，为区分正常行为与异常行为提供强有力的先验。通过采用轻量级参数高效微调适配，无需付出全量微调的高昂成本，即可高效地将这一先验与特定异常检测场景对齐。尽管时间序列基础模型和参数高效微调已在各自领域得到广泛研究，但二者结合应用于时间序列异常检测仍未得到充分探索。

为弥补这一空白，我们在多个时间序列异常检测基准测试上开展了系统的实证评估，对比零样本推理、全量微调和基于参数高效微调的适应。我们还分析了评估指标和参数高效微调超参数对检测性能的影响。我们的贡献总结如下：

– 我们对时间序列基础模型在时间序列异常检测中的应用进行了系统的实证研究，比较了多种架构和基准测试下的零样本推理、全量微调和参数高效微调策略。– 我们证明，经过适配的时间序列基础模型始终优于特定任务基线；在类别严重不平衡的情况下，其在注重精确率的指标和基于 VUS 的指标上优势尤为明显，这凸显了评估方案和适配策略在时间序列异常检测中的重要性。– 我们还发现，LoRA、OFT 和 HRA 等参数高效微调方法通常能够达到或超越全量微调的效果，同时仅需更新模型的一小部分参数，因此无需从头训练模型，即可高效适配稠密型和混合专家时间序列基础模型。

2 相关工作

2.1 时间序列基础模型

时间序列基础模型近年来已成为一种通用范式，可用于学习具有迁移能力的时间表示，其主要预训练目标是预测。总体而言，现有时间序列基础模型可根据其如何利用大规模数据，以及最初为其他领域开发的架构进行分类。Moirai [35]等模型直接在海量且异构的时间序列语料库上训练，从而能够捕捉跨领域的通用时间模式。相比之下，Chronos [2]通过对数值进行词元化，并应用预训练的 Transformer 语言模型（如 T5 [27] 和 GPT-2 [26]），将大语言模型适配于时间序列预测。Time-MoE [29] 则介于两者之间，采用基于原始时间序列的自回归训练，并使用稀疏混合专家（MoE）架构，以提升可扩展性和模型容量。

除了训练范式之外，时间序列基础模型在架构设计和输入表示方面也存在显著差异。Moirai采用仅编码器架构和分块嵌入，有助于高效学习表示和进行长上下文建模，但可能会掩盖细粒度的时间变化。Chronos和Time-MoE则采用逐点嵌入和自回归生成，能够更精确地建模局部动态，但代价是对输入长度更加敏感，并且在长时域预测中容易出现误差累积。这些设计选择直接影响时间序列基础模型如何适配异常检测等下游任务，尤其是在预测行为至关重要的在线或零样本场景中。

2.2 参数高效微调

尽管时间序列基础模型展现出强大的泛化能力，但要将其适配到特定领域的任务，通常仍需进行某种形式的微调。参数高效微调已成为一种有效策略，既能让大型预训练模型适应特定任务，又可避免全量微调带来的计算和内存成本。代表性方法包括 IA^3 [19], ，它引入可训练的缩放向量来调节 Transformer 的关键激活；以及 LoRA [14],„ 它将低秩更新注入预训练权重矩阵。其他方法，如 OFT [25], ，将适配限制在正交变换范围内，以保留几何结构；而 HRA[38] 则通过 Householder反射的乘积 [13] 扩展了这一思路，以兼顾稳定性和表达能力。

参数高效微调技术最初是为大型语言模型和视觉模型开发的，对于时间序列基础模型尤其具有吸引力，因为这类模型的规模可能非常庞大，而从头开始重新训练通常并不现实。然而，尽管参数高效微调在预测和表征学习中的应用日益广泛，但其在时间序列异常检测中的作用仍鲜有研究，尤其是与全量微调和零样本推理相比。

2.3 时间序列基础模型与异常检测

近期研究开始探讨时间序列基础模型在异常相关任务中的适用性，但现有研究的范围仍然有限。Shyalika等 [30]区分了异常预测与异常检测，并指出，提前预测异常是一项更具挑战性的问题，需要对前兆或因果动态进行建模。因此，他们主要关注异常预测，并将其表述为监督分类或带标签预测任务。虽然作者还报告了使用少量时间序列基础模型进行零样本异常检测的结果，但其性能高度依赖于数据集，且相较于更简单的基线方法并未展现出稳定优势。此外，通过全量模型更新进行的微调实验仅带来了有限的改进，这表明，对时间序列基础模型进行简单适配可能不足以实现稳健的检测性能。

与本文更为相关的一项研究是 MOMENT框架 [11], ，该框架展示了单个预训练时间序列模型如何通过微调适用于多个下游任务，包括预测、插补、分类和异常检测。MOMENT是早期的重要探索，证明时间序列基础模型可以用于预测以外的任务；研究还报告了其在零样本设置下的出色表现，以及使用线性探测进行异常检测时具有竞争力的结果。不过，其评估仅涵盖规模相对较小的异常检测基准测试，且该研究未探讨线性层以外的参数高效微调策略。

与这些初步探索不同，我们对将时间序列基础模型用于时间序列异常检测进行了系统而全面的研究。我们明确聚焦于检测而非预测，并在多个基准测试上进行评估

，还详细比较了零样本推理、全量微调和多种参数高效微调方法。这项范围更广的实证分析阐明了时间序列基础模型何时以及如何能够成为有效的异常检测器，并凸显了参数高效微调在将基础模型适配到真实时间序列异常检测场景时的实际优势。

3 方法

3.1 数据集

在实证评估中，我们使用 TSB-AD-U 基准 [20], 。这是一个全面的单变量时间序列异常检测套件，旨在通过提升数据集完整性、多样性和评估可靠性，解决先前基准测试的局限性。TSB-AD-U 包含 23 个经过精心筛选的单变量数据集，涵盖不同领域、异常类型、序列长度和统计特征，其覆盖范围远超早期数据集 [23]。这种多样性使 TSB-AD-U 尤其适用于评估基础模型和参数高效微调策略，因为它能减少对特定异常模式的过拟合，更好地反映现实世界的异质性，从而为时间序列异常检测方法的基准测试提供稳健可靠的测试平台。遵循基准测试协议，TSB-AD-U 提供了预定义的训练集、调优集和评估划分。在本研究中，我们使用调优集进行性能评估。尽管该划分原本用于超参数选择，但由于其规模远小于原始评估集，我们将其用作评估集，以便更高效地开展实验。所有超参数均在评估前固定，且调优集仅用于报告最终结果，以确保评估公平且无偏。

3.2 时间序列基础模型骨干网络

我们评估了一系列多样化的时间序列基础模型骨干网络，它们在架构、规模和预训练策略方面各不相同。值得注意的是，我们在不同实验设置中对这些模型的使用方式有所不同。我们使用范围广泛的时间序列基础模型进行零样本异常检测，以评估其在未经任务特定适配情况下的泛化能力；而全量微调和参数高效微调仅应用于选定的代表性骨干模型子集。

零样本推理。 在零样本评估中，我们纳入了三个主要时间序列基础模型系列的代表性模型：Moirai [35],Chronos [2], 和 Time-MoE[29], ，如表 2 所示。在 Moirai 系列中，我们评估了稠密变体和混合专家变体，包括 Moirai 和 Moirai-MoE [21]。具体而言，Moirai-MoE-small 在每次前向传播中激活约 11M 个参数，相当于其总参数量的约 10%；更大型的变体则通过增加专家来提升整体模型容量，同时保持专家激活的稀疏性 [35]。在 Chronos 系列中，我们评估了基于 T5 的模型和

基于 Bolt 的骨干网络。Chronos-Bolt 变体包括 bolt-tiny（9M 个参数）、bolt-small（48M）和 bolt-base（205M），具体配置遵循官方 Chronos 模型卡 [2]。对于基于 T5 的 Chronos 模型，其参数量与底层 T5 架构一致：t5-small（46M 个参数）、t5-base（200M）和 t5-large（710M）；这些模型通过数值标记化和自回归解码适配时间序列预测 [27, 2]。此外，我们还纳入了 Time-MoE-base，它每次前向传播使用 50M 个参数，作为一种具有代表性的自回归混合专家时间序列基础模型，直接在大规模时间序列数据上进行预训练 [29]。这一广泛的模型选择有助于系统分析架构设计、模型容量和稀疏性如何影响零样本异常检测性能。

微调与参数高效适配。为研究特定任务适配的影响，我们将全量微调和参数高效微调（PEFT）实验聚焦于三种具有代表性的骨干模型：Moirai-base、Chronos-t5-base 和 Time-MoE-base。这些模型涵盖了不同的架构范式，包括仅编码器的基于分块建模（Moirai）、通过数值标记化实现的自回归 Transformer 预测（Chronos-t5），以及稀疏混合专家自回归建模（Time-MoE），同时在计算上仍足以支持反复进行适配实验。将微调和参数高效微调限制在这一子集，可在零样本推理、全量模型更新和轻量级适配策略之间进行受控比较，同时避免模型规模或训练成本差异过大所带来的混杂影响。

3.3 适应策略与训练方案

为将时间序列基础模型用于时间序列异常检测，我们考虑两种适配方案：全量微调和参数高效微调（PEFT）。全量微调使用任务专用训练数据更新预训练骨干模型的全部参数，作为适应灵活性方面的上界基线，但会带来高昂的计算和内存成本。对于参数高效微调，我们采用多种轻量级适配策略，仅修改一小部分参数，同时基本冻结预训练骨干模型。具体而言，我们评估 LoRA [14], OFT [25], 和 HRA [38],，这些方法会在 Transformer 架构中引入可训练的适配器模块。按照标准做法 [31, 12], ，这些适配器被插入注意力块的键、查询、值和输出投影层，而偏置项和丢弃层则不参与适配。对于这些方法，我们测试了秩为 4、8、16 和 32 的适配器，以考察参数效率与检测性能之间的权衡。此外，我们还评估 IA^3 [19], ，该方法通过学习乘法缩放向量而非加性低秩更新来适配预训练模型。对于 IA^3, ，可训练向量应用于键和值投影以及前馈网络层，从而以极低的参数开销轻量调节内部激活。

在所有微调实验中，模型均使用 150 个时间步的上下文窗口，固定训练 10 个训练轮次。每个模型均使用其原始时间序列基础模型（TSFM）设定中的

损失函数进行优化：Moirai 使用负对数似然，Chronos 使用交叉熵，Time-MoE 使用 Huber 损失。为考虑模型对优化超参数的敏感性，我们测试了学习率 10^{-2}、 10^{-3}、 10^{-4} 和 10^{-5}，并报告所有数据集上的平均表现最佳的配置所对应的结果。该方案使我们能够在一致的训练条件下，公平且系统地比较全量微调与不同的参数高效微调策略。

3.4 评估指标

为评估性能，我们采用四种无阈值指标：AUC-ROC、AUC-PR、VUS-ROC 和 VUS-PR [20, 4]。这些指标衡量模型在所有可能的决策阈值下的异常检测性能，从而无需手动调整或针对特定数据集调节阈值，即可对模型进行稳健比较。AUC-ROC 和 AUC-PR 分别量化真正率与假正率之间、精确率与召回率之间的权衡，但它们依赖于预测异常与真实标签之间的逐点对齐，因此即使检测结果仅有轻微的时间偏移，也会受到惩罚。

为缓解这一局限，我们还报告 VUS-ROC 和 VUS-PR。这两项指标通过在真实异常点周围引入可变缓冲区，对传统的基于 AUC 的指标进行了扩展。该设计使评估能够对发生在已标注异常位置附近、但未与其精确重合的检测结果给予部分分数，因此对基于预测的异常检测中常见的轻微时间错位更具容忍度。不过，这种灵活性提升的代价是计算成本更高：标准 AUC 指标只需遍历所有不重复的异常分数，将其作为候选阈值，即可高效计算；而基于 VUS 的指标则需要针对每个阈值，在多种缓冲区大小下重新计算检测结果。对于所有方法，我们均将每个时间步预测值与真实值之间的均方误差（MSE）作为异常分数，以此为指标计算提供一致的依据。

3.5 异常检测基线方法

为全面评估时间序列基础模型的性能，我们将其与多种广泛使用的时间序列异常检测基线方法进行比较，这些方法均不依赖基础模型预训练。这些基线涵盖基于预测、基于重构、统计方法以及传统机器学习方法，为不同建模范式提供全面的参考。具体而言，我们纳入 Anomaly Transformer [37], LSTMAD [22], 和 TimesNet [36] ，作为基于深度学习的预测或序列建模方法的代表，它们通过预测误差或学习到的时间异常来识别异常。基于重构的神经网络基线包括自编码器 [28] 和 OmniAnomaly [32], ，它们基于异常模式无法被忠实重构这一假设，通过重构误差检测异常。

表1。 TSB-AD-U基准上基线时间序列异常检测方法的性能比较。结果采用无阈值指标（AUC-PR、AUC-ROC、VUS-PR 和 VUS-ROC）进行报告，并在所有数据集上取平均值。基线方法包括传统机器学习方法、基于深度学习的预测方法和重构模型。最佳和次佳结果分别以粗体和下划线标出。

基线	模型规模	AUC-PR	AUC-ROC	VUS-PR	VUS-ROC
Sub-PCA	-	0.293	0.700	<u>0.332</u>	0.745
IForest	-	0.187	0.659	0.234	0.732
SAND	-	<u>0.299</u>	<u>0.714</u>	**0.353**	0.747
自编码器	$\sim 7.7K$	0.173	0.667	0.242	0.729
LSTMAD	$\sim 10K$	**0.332**	**0.724**	0.323	**0.820**
OmniAnomaly	$\sim 13K$	0.263	0.687	0.308	0.735
TimesNet	$\sim 73K$	0.184	0.637	0.259	<u>0.748</u>
Anomaly Transformer	$\sim 4.7M$	0.059	0.499	0.102	0.575

除神经网络时间序列异常检测基线外，我们还评估了时间序列异常检测基准测试中常用的传统和统计基线，包括 Sub-PCA [1], 孤立森林（IForest）[18], 和 SAND [5]。这些方法依赖降维、基于树的孤立或统计偏差建模；尽管方法简单，但在某些场景中仍具有竞争力。这些基线共同构成了一套全面且公正的比较方案，突显了相较于成熟异常检测技术，重新利用时间序列基础模型的优势与局限。

4 实验结果

我们在 TSB-AD-U基准上评估异常检测性能。该基准包含 23 个多样化的单变量数据集，覆盖多个领域和异常特征。总体而言，表 1–3 的结果表明，时间序列基础模型相比专门针对任务的异常检测基线具有显著优势。具体而言，时间序列基础模型展现出较强的零样本性能，并且通过全量微调和参数高效微调均可进一步提升，这证实了它们在时间序列异常检测方面的有效性和灵活性。

4.1 传统时间序列异常检测基线

表1总结了传统时间序列异常检测基线方法的性能。在经典方法和统计方法中，SAND 和 Sub-PCA 的整体表现最强，尤其是在基于 VUS 的指标上，表明它们能够稳健应对时间错位。在从头训练的神经网络模型中，LSTMAD 在大多数指标上表现最佳，在该组中取得了最高的 AUC-PR、AUC-ROC 和 VUS-ROC。相比之下，Anomaly Transformer 等更复杂的架构平均表现不佳，这表明它们在 TSB-AD-U 的异质数据集上缺乏稳健性。

表 2. 时间序列基础模型在 TSB-AD-U基准上的零样本异常检测性能。所有模型均未进行任务特定微调，并使用预测误差作为异常分数。该表比较了不同的时间序列基础模型系列（Moirai、Chronos 和 Time-MoE）、模型规模及架构变体，展示了模型设计和容量对零样本检测性能的影响。每项指标中的最佳和次佳结果分别以粗体和下划线标示。

TSFM	模型规模	AUC-PR	AUC-ROC	VUS-PR	VUS-ROC
Moirai	小型	0.335	**0.697**	<u>0.318</u>	**0.778**
	base	0.321	0.678	0.312	0.763
	大型	0.326	0.692	0.315	0.771
	moe-small	0.284	0.641	0.294	0.747
	moe-base	0.283	0.641	0.291	0.747
Chronos	t5-small	0.304	0.675	0.280	0.754
	t5-base	0.314	0.667	0.291	0.753
	t5-large	0.323	0.664	0.290	0.752
	bolt-tiny	<u>0.345</u>	<u>0.693</u>	0.318	<u>0.774</u>
	bolt-small	**0.346**	0.684	**0.319**	0.770
	bolt-base	0.340	0.678	0.317	0.765
Time-MoE	base	0.339	0.669	0.318	0.756

4.2 与时间序列基础模型零样本性能的比较

表 2 报告了时间序列基础模型在 TSB-AD-U基准上的零样本异常检测性能，评估依据是在未进行任何任务特定适配的情况下产生的预测误差。总体而言，零样本时间序列基础模型的表现优于表1所示的成熟基线方法，但在所有指标上并未持续超过表现最强的任务专用模型。具体而言，尽管若干时间序列基础模型变体的 AUC-PR 与传统和深度学习基线相当或略高，但 LSTMAD 在 AUC-ROC 和 VUS-ROC 方面仍是表现最佳的方法。

在时间序列基础模型中，Chronos 和 Moirai 系列的模型通常优于其他零样本配置。Chronos Bolt 变体在时间序列基础模型中取得了最高的 AUC-PR 和 VUS-PR 分数，而稠密 Moirai 模型则展现出较强的 AUC-ROC 和 VUS-ROC 性能。然而，不同时间序列基础模型架构及规模之间的性能差异相对较小，混合专家变体也未能始终在零样本结果上胜过其稠密对应模型。这些结果表明，尽管不经适配直接应用预训练时间序列基础模型，能够在多样化数据集上实现稳健且有竞争力的异常检测性能，但仅靠零样本推理仍不足以持续超越经过充分调优的特定任务基线。这一发现促使我们探索特定任务适配策略，并将在下一节对此展开讨论。

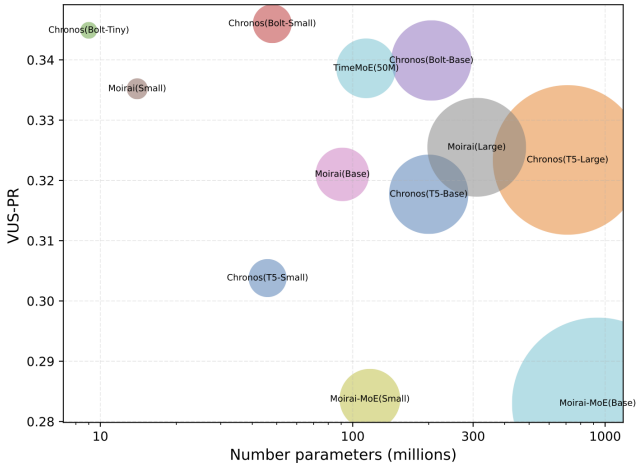


图 1. 模型规模与零样本异常检测性能之间的关系。图中展示了在 TSB-AD-U 基准上评估的不同时间序列基础模型变体的模型参数数量与平均 VUS-PR。气泡大小反映相对模型容量。

4.3 适应策略对性能的影响

表 3 中对各项指标的进一步分析表明，与零样本推理相比，对时间序列基础模型进行适配可在所有骨干网络上带来持续提升，其中 AUC-PR 和 VUS-PR 的提升最为显著。在异常检测所具有的严重类别不平衡情形下，这些指标尤为重要。稠密时间序列基础模型在 AUC-ROC 上的提升通常较为温和，这反映出零样本模型本身已具备合理的排序性能；而 Time-MoE 在适配后则取得了显著提升，说明混合专家架构从特定任务校准中获益更大。总体而言，适配后，以精确率为导向的指标和基于 VUS 的指标提升更明显，胜过仅依据排序的指标；这表明适配主要改善了对罕见异常的准确识别，而不只是优化整体排序表现。此外，适配后 VUS-ROC 和 VUS-PR 分数更高，说明模型对微小时间错位的容忍度有所提升；当实际应用中的异常边界不明确时，这一特性尤为理想。架构对比还表明，尽管稠密时间序列基础模型在零样本设置下表现强劲，Time-MoE 在适配后缩小了这一差距，并取得了有竞争力甚至更优的基于 VUS 的性能。这说明，经过恰当微调后，显式的时间与专家级专门化能够胜过稠密架构在稳健性方面的优势。

在这三种骨干模型中，全量微调相较于零样本推理始终能够提升性能。具体而言，Time-MoE 在基于 VUS 的指标上取得了最大的绝对增益：全量微调后，VUS-PR 从 0.318 提升至

表 3. 将时间序列基础模型应用于时间序列异常检测的适应策略比较。结果涵盖零样本推理、全量微调（全参数微调）和参数高效微调方法（包括 IA^3, LoRA、OFT 和 HRA），涉及三种有代表性的骨干模型。使用无阈值指标在 TSB-AD-U 基准上评估性能。该表展示了适应灵活性与检测性能之间的权衡。最佳和次佳结果分别以粗体和下划线标出。

TSFM	适应策略	AUC-PR	AUC-ROC	VUS-PR	VUS-ROC
Moirai（基础版）	零样本	0.321	0.678	0.312	0.763
	全参数微调	0.387	<u>0.757</u>	0.354	0.827
	IA^3	0.377	0.748	0.351	0.817
	LoRA	<u>0.388</u>	**0.763**	0.352	<u>0.829</u>
	OFT	**0.389**	0.756	**0.361**	**0.834**
	HRA	0.387	0.751	<u>0.357</u>	0.822
Chronos-t5（基础版）	零样本	0.318	0.675	0.290	0.752
	全参数微调	<u>0.361</u>	0.727	0.318	0.796
	IA^3	0.325	0.680	0.299	0.765
	LoRA	0.325	0.674	0.297	0.757
	OFT	<u>0.361</u>	<u>0.733</u>	**0.338**	<u>0.813</u>
	HRA	**0.367**	**0.738**	<u>0.336</u>	**0.816**
Time-MoE（基础版）	零样本	0.339	0.669	0.318	0.756
	全参数微调	**0.386**	0.750	**0.392**	**0.843**
	IA^3	0.347	0.712	0.341	0.787
	LoRA	<u>0.384</u>	<u>0.757</u>	<u>0.372</u>	0.830
	OFT	0.379	**0.762**	0.365	<u>0.840</u>
	HRA	0.351	0.736	0.351	0.813

0.392，VUS-ROC 则从 0.756 提升至 0.843。Moirai-base 的 VUS-ROC 同样从 0.763 提升至 0.827，而 Chronos-t5 在所有评估指标上均有稳定提升。这些结果证实，将预训练预测模型适配至目标数据分布，有助于捕捉数据集特定的时间模式。然而，不同架构的提升幅度存在显著差异，这表明适配效果不仅取决于数据对齐程度，也取决于专家路由或共享表示等架构组件与适配策略之间的相互作用。

参数高效微调方法通常只更新一小部分模型参数，其效果便可媲美甚至超越全量微调（见表4）。对于 Moirai-base，LoRA、OFT 和 HRA 等基于适配器的方法在大多数指标上始终优于全量微调，其中 OFT 的 VUS-PR 和 VUS-ROC 表现最佳（图2）。对于 Chronos-t5，OFT 和 HRA 的整体表现最为出色（图2），不仅超过零样本推理，也优于全量微调，尤其是在侧重精确率的指标和基于 VUS 的指标上。

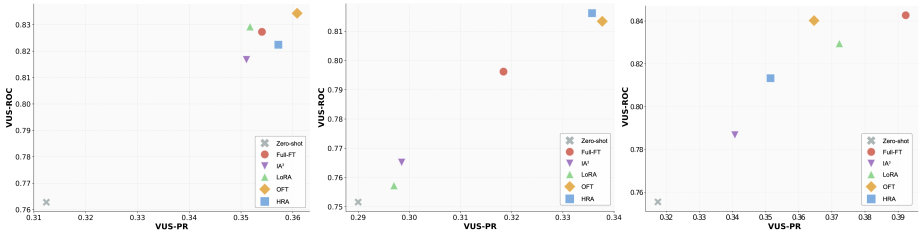


图 2. Moirai（左）、Chronos（中）和 Time-MoE（右）在零样本、全量微调和参数高效微调方法下的 VUS-PR 和 VUS-ROC 性能对比。

表 4. Moirai、Chronos-t5 和 Time-Moe 在参数高效微调中使用的可训练参数及其比例。

	Moirai（基础版）		Chronos-t5（基础版）		Time-MoE（基础版）	
参数高效微调	参数量	占比(%)	参数量	占比(%)	参数量	占比(%)
IA^3	$\sim 43K$	0.05	$\sim 0.1M$	0.06	$\sim 27K$	0.02
LoRA	$\sim 1.1M$	1.28	$\sim 3.5M$	1.73	$\sim 1.6M$	1.47
OFT	$\sim 0.8M$	0.94	$\sim 2.5M$	1.27	$\sim 4M$	3.48
HRA	$\sim 0.5M$	0.64	$\sim 1.7M$	0.87	$\sim 0.7M$	0.65

相比之下，对于 Time-MoE，全量微调在基于 VUS 的整体表现上最为出色，而 LoRA 和 OFT 只能部分弥补这一差距（图 2 ）。这表明，标准的参数高效微调方法可能不足以充分适配混合专家模型固有的专家选择、负载均衡或专家间协调机制。尽管参数高效微调仍能显著提升零样本推理的效果，但与全量微调之间仍存在差距，这意味着 Time-MoE 可能需要更专门的参数高效微调设计，例如具备专家感知能力的适配器或路由感知参数化，才能充分发挥其适配潜力。

最后，在比较参数高效微调变体时，基于适配器的方法（LoRA、OFT 和 HRA）在所有骨干网络和评估指标上始终优于 IA^3 。尽管 IA^3 在某些情况下提升了零样本性能，但其增益比基于适配器的方法更小，也不够稳定，尤其是在基于 VUS 的指标上。这一模式表明，仅靠乘法调制不足以重新校准复杂的时间动态和专家级动态，而加法适配器或正交适配器能够提供更丰富的适配空间。因此，对于将时间序列基础模型改用于时间序列异常检测，基于适配器的参数高效微调是一种更有效、更可靠的策略，尤其是在稠密架构这类专家交互有限的情况下。

5 结论

本研究展示了时间序列基础模型在异常检测中的有效性，并表明即使没有任务特定训练，时间序列基础模型在大量真实世界数据集上的表现也始终优于传统基线方法。借助从大规模、多样化时间序列语料库中获取的预训练时序知识，时间序列基础模型展现出强大的零样本性能和跨领域鲁棒异常检测能力。此外，包括 LoRA、OFT 和 HRA 在内的参数高效微调方法，为全量微调提供了颇具吸引力的替代方案；它们往往能达到甚至超越全量微调的性能，同时大幅降低计算和内存开销。

尽管取得了这些进展，仍有一些重要挑战有待解决，尤其是在架构复杂度、模型规模和计算效率之间取得平衡。虽然参数高效微调方法显著提升了密集型和混合专家型时间序列基础模型的性能，但仍需进一步研究，以更好地理解适应策略如何与不同的架构设计在大规模场景下相互作用。未来的工作可以探索更具可扩展性的参数高效微调变体、改进基于混合专家的时间序列基础模型中的专家专门化，并将这些模型扩展至更广泛的异常检测场景，例如多变量、流式或在线场景。总体而言，我们的研究结果凸显了时间序列基础模型结合参数高效微调的巨大潜力，这一范式可扩展且有效，适用于多种真实世界应用中的时间序列异常检测。

致谢 K.Y.部分得到韩国国家研究基金会（NRF）项目（编号：RS-2024-00337092）以及韩国政府（MSIT）资助的信息与通信技术规划评估院（IITP）项目的支持，包括编号为RS-2020-II201373的人工智能研究生院项目，以及编号为IITP-(2025)-RS-2023-00253914的人工智能半导体支持项目（汉阳大学）。

参考文献

1. Aggarwal, C.C., Sathe, S.：我应该使用哪种离群点检测算法? 《离群点集成：导论》，第207–274页。Springer（2017） 2. Ansari, A.F., Stella, L., Turkmen, A.C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S.S., Arango, S.P., Kapoor, S., Zschiegner, J., Maddix, D.C., Wang, H., Mahoney, M.W., Torkkola, K., Wilson, A.G., Bohlke-Schneider, M., Wang, B.：Chronos：学习时间序列的语言。《机器学习研究汇刊》（2024） 3. Blázquez-García, A., Conde, A., Mori, U., Lozano, J.A.：时间序列数据中的离群点/异常检测综述。《ACM计算综述》（CSUR）**54**（3），1–33（2021） 4. Boniol, P., Krishna, A.K., Bruel, M., Liu, Q., Huang, M., Palpanas, T., Tsay, R.S., Elmore, A., Franklin, M.J., Paparrizos, J.：VUS：用于时间序列异常检测的有效且高效的准确性度量。《VLDB期刊》**34**（3），32（2025）

5. Boniol, P., Paparrizos, J., Palpanas, T., Franklin, M.J.：SAND：流式子序列异常检测。《VLDB基金会会刊》**14**（10），1717–1729（2021） 6. Chandola, V., Banerjee, A., Kumar, V.：异常检测：综述。《ACM计算综述》（CSUR）**41**（3），1–58（2009） 7. Chen, Z., Chen, D., Zhang, X., Yuan, Z., Cheng, X.：使用Transformer学习图结构进行物联网中的多变量时间序列异常检测。《IEEE物联网期刊》**9**（12），9179–9189（2021） 8. Deng, A., Hooi, B.：多变量时间序列中的基于图神经网络的异常检测。《AAAI人工智能会议论文集》，第35卷，第4027–4035页（2021） 9. Ergen, T., Kozat, S.S.：使用LSTM神经网络进行无监督异常检测。《IEEE神经网络与学习系统汇刊》**31**（8），3127–3141（2019） 10. Garza, A., Challu, C., Mergenthaler-Canseco, M.：TimeGPT-1。arXiv预印本 arXiv:2310.03589（2023） 11. Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., Dubrawski, A.：MOMENT：一系列开放式时间序列基础模型。第四十一届国际机器学习会议（2024） 12. He, J., Zhou, C., Ma, X., Berg-Kirkpatrick, T., Neubig, G.：迈向参数高效迁移学习的统一视图。国际学习表征会议（2022） 13. Householder, A.S.：非对称矩阵的酉三角化。《ACM期刊》**5**（4），339–342（1958） 14. Hu, E.J., yelong shen, Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.：LoRA：大语言模型的低秩适配。国际学习表征会议（2022） 15. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T.：使用LSTM和非参数动态阈值检测航天器异常。第24届ACM SIGKDD知识发现与数据挖掘国际会议论文集，第387–395页（2018） 16. Jin, M., Zhang, Y., Chen, W., Zhang, K., Liang, Y., Yang, B., Wang, J., Pan, S., Wen, Q.：立场：大语言模型能告诉我们什么关于时间序列分析的信息。第四十一届国际机器学习会议（2024） 17. Li, D., Chen, D., Jin, B., Shi, L., Goh, J., Ng, S.K.：Mad-GAN：使用生成对抗网络进行时间序列数据的多变量异常检测。国际人工神经网络会议，第703–716页。Springer（2019） 18. Liu, F.T., Ting, K.M., Zhou, Z.H.：孤立森林。2008年第八届IEEE国际数据挖掘会议，第413–422页（2008） 19. Liu, H., Tam, D., Mohammed, M., Mohta, J., Huang, T., Bansal, M., Raffel, C.：少样本参数高效微调优于上下文学习，且成本更低。载于：Oh, A.H., Agarwal, A., Belgrave, D., Cho, K.（编），《神经信息处理系统进展》（2022） 20. Liu, Q., Paparrizos, J.：房间里的大象：迈向可靠的时间序列异常检测基准。第三十八届神经信息处理系统会议数据集与基准赛道（2024） 21. Liu, X., Liu, J., Woo, G., Aksu, T., Liang, Y., Zimmermann, R., Liu, C., Li, J., Savarese, S., Xiong, C., Sahoo, D.：Moirai-MoE：利用稀疏专家混合增强时间序列基础模型。第四十二届国际机器学习会议（2025）

22. Malhotra, P., Vig, L., Shroff, G., Agarwal, P., 等：用于时间序列异常检测的长短期记忆网络。论文集，第89卷，第94页（2015） 23. Paparrizos, J., Kang, Y., Boniol, P., Tsay, R.S., Palpanas, T., Franklin, M.J.：Tsb-uad：单变量时间序列异常检测端到端基准测试套件。VLDB基金会会刊**15**(8)，1697–1711（2022） 24. Park, D., Hoshi, Y., Kemp, C.C.：使用基于LSTM的变分自编码器进行机器人辅助喂食的多模态异常检测器。IEEE机器人与自动化快报**3**(3)，1544–1551（2018） 25. Qiu, Z., Liu, W., Feng, H., Xue, Y., Feng, Y., Liu, Z., Zhang, D., Weller, A., Schölkopf, B.：通过正交微调控制文本到图像扩散。第三十七届神经信息处理系统会议（2023） 26. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., 等：语言模型是无监督的多任务学习者。OpenAI博客**1**(8)，9（2019） 27. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.：探索统一文本到文本转换器的迁移学习极限。机器学习研究期刊**21**(140)，1–67（2020） 28. Sakurada, M., Yairi, T.：使用具有非线性降维功能的自编码器进行异常检测。MLSDA 2014感官数据分析机器学习第二届研讨会论文集，第4–11页（2014） 29. Shi, X., Wang, S., Nie, Y., Li, D., Ye, Z., Wen, Q., Jin, M.：Time-MoE：采用专家混合的十亿规模时间序列基础模型。第十三届国际学习表征会议（2025） 30. Shyalika, C., Bagga, H.K., Bhatt, A., Prasad, R., Ghazo, A.A., Sheth, A.：时间序列基础模型：它们在异常检测与预测中的作用。arXiv预印本 arXiv:2412.19286（2024） 31. Stickland, A.C., Murray, I.：BERT和PALs：用于高效多任务学习适配的投影注意力层。国际机器学习会议，第5986–5995页（2019） 32. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.：通过随机循环神经网络实现多变量时间序列的鲁棒异常检测。第25届ACM SIGKDD知识发现与数据挖掘国际会议论文集，第2828–2837页（2019） 33. Talukder, S.J., Yue, Y., Gkioxari, G.：TOTEM：用于通用时间序列分析的令牌化时间序列嵌入。机器学习研究汇刊（2024） 34. Tuli, S., Casale, G., Jennings, N.R.：Tranad：用于多变量时间序列数据异常检测的深度Transformer网络。VLDB基金会会刊**15**(6)，1201–1214（2022） 35. Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., Sahoo, D.：通用时间序列预测Transformer的统一训练。国际机器学习会议，第53140–53164页（2024） 36. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.：TimesNet：面向通用时间序列分析的时序二维变化建模。第十一届国际学习表征会议（2023） 37. Xu, J., Wu, H., Wang, J., Long, M.：Anomaly Transformer：通过关联差异进行时间序列异常检测。国际学习表征会议（2022） 38. Yuan, S., Liu, H., Xu, H.：通过Householder反射适配弥合低秩适配与正交适配之间的差距。第三十八届神经信息处理系统年度会议（2024）

39. Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., Xu, B.：Ts2vec：迈向时间序列的通用表征。AAAI人工智能会议论文集，第36卷，第8980–8987页（2022）
40. Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., Xu, B., Bai, J., Tong, J., Zhang, Q.：通过图注意力网络进行多变量时间序列异常检测。2020年IEEE国际数据挖掘会议，第841–850页（2020）