多尺度、多物理场与耦合问题高并行系统研讨会 HPC Asia 2023，2023年2月27日至3月2日，新加坡

利用分布式训练加速领域感知深度学习模型

Aishwarya Sarkar、Chaoqun Lu 和 Ali Jannesari
爱荷华州立大学，美国{asarkar1、clu、jannesar}@iastate.edu

arXiv:2301.11787v1 [cs.LG] 25 Jan 2023

关键词： 领域知识、模型并行、分布式训练、知识引导神经网络

1 摘要与动机

数据生成技术的最新进展推动了地理时空数据的爆炸式增长。因此，在水文学、生态学和交通运输等领域，借助深度学习技术解读时空交互背后的复杂模式已成为当务之急。然而，不结合领域特定知识应用深度学习技术，往往难以取得理想的预测性能。其次，在大规模数据上训练此类模型需要大量计算资源。为应对这些挑战，我们提出了一种新颖的分布式领域感知时空网络，该网络利用领域特定知识来提升模型性能。该网络由像素贡献模块、分布式多头多通道卷积神经网络（CNN）空间模块以及循环时间模块组成。我们以水文学中的洪水预测为用例，检验所提出的方法。分析结果表明，该网络能够有效预测流域出口处流量测量值的高峰，最高加速比达 4.1 倍，预测性能最高提升至 93%。我们的方法整体加速比达到 12.6 倍，平均预测性能提升了 16%。我们在美国北部某州的 23 个流域数据集上进行了大量实验，并展示了研究结果。

领域感知深度学习（DL）模型并非只依赖手头的数据。对目标领域有所了解，正是它们脱颖而出的关键。缺少明确约束时，DL 模型容易做出错误预测。研究表明，在人体姿态估计 [1]、动物模型研究 [2]、乳腺癌诊断 [3]、地震反演 [4]等多种应用中，这些模型的表现优于数据驱动模型。为了让 DL 模型具备领域感知能力，研究人员尝试将其与领域专用模型紧密结合 [5]。尽管取得了这些进展，高效训练、测试和部署这些模型的瓶颈仍然是一项挑战。高性能计算系统近期取得的突破展现出巨大潜力，有望通过提升领域感知 DL 模型的可扩展性和成本效益，帮助其突破瓶颈 [6, 7]。

本十年全球洪水和极端降雨事件增加了 50% 以上，目前发生频率是 1980 年的四倍。越来越多的证据表明，预测并减轻洪水发生具有巨大潜力，有助于减少产量损失，

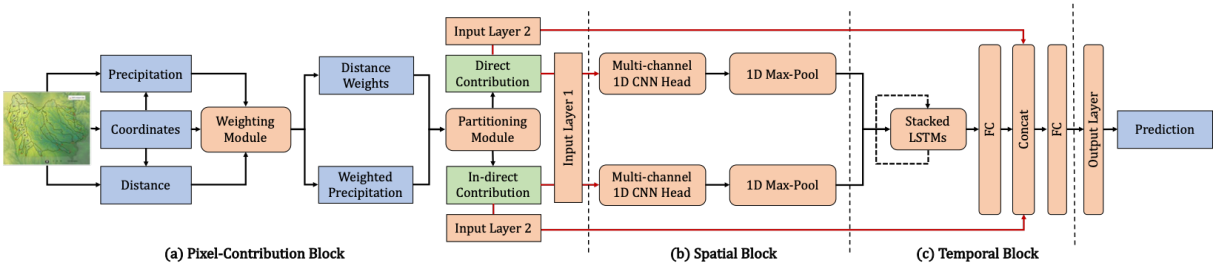


图 1：所提出的分布式时空网络 Dom-ST 的整体工作流程，包括 (a) 像素贡献（Pix-Con）模块、(b) 带并行卷积头的空间模块，以及循环式 (c) 时间模块。蓝色方框表示数据，橙色方框表示 Dom-ST 的各个模块。箭头表示推理过程中的数据流。

改善灾害风险管理并造福环境 [8, 9, 10]。本研究通过提出一种新方法，在一定程度上消除基于深度学习的像素化时空洪水预测模型的黑箱特性，从而应对这类模型面临的挑战。该方法利用像素化输入所蕴含的领域知识，学习各像素特有的时空贡献。我们还提出一种新颖的分布式训练方法，以加速并优化像素化领域感知时空网络的训练时间。总而言之，我们的贡献如下：

1. 一种领域感知分布式时空网络（Dom-ST），包含一个新颖的像素贡献模块（Pix-Con），用于计算像素特定权重，并根据时空输入对水流量的局部贡献对其进行转换（图 1）。2. Pix-Con 模块中的分区模块，可动态划分时空像素并将其分配到多个设备（图 1）。3. 一种新颖的领域引导的分布式训练方法，用于加速训练（图 2）。4. 一种利用领域重要的时间输入测量值（目标日）来增强 Dom-ST 领域感知能力的方法。5. 采用领域引导的分布式训练，在改善洪水预测方面取得最高达 93% 的竞争性结果，并实现 12.6 倍的整体加速比，

2 分布策略

由于一个地区的气候及其他环境因素（如土壤、土地覆盖、温度和人类影响）会因地点而异，因此，即使训练了一个全局水文模型，也会导致结果欠佳，或需要耗费大量计算资源。通过对目标区域进行建模所获得的知识，还可以进一步提升模型性能。此外，偏远地区的数据稀缺本身也是一项挑战，迁移学习方法 [11] 在一定程度上可以应对这一问题。由于多维时空数据规模庞大且复杂，在单台设备上基于深度学习方法训练水文模型，需要更多的时间和内存资源

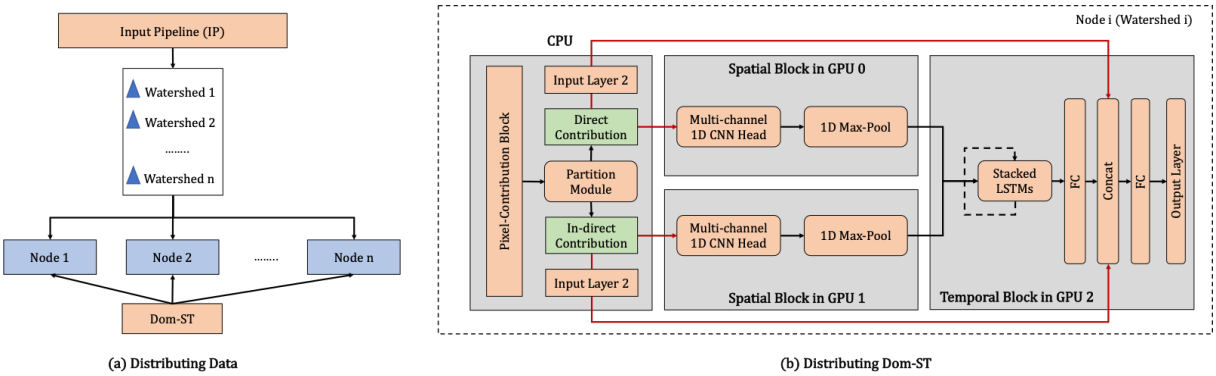


图 2: （a）输入流水线分发数据块（流域），并在多个节点上创建 Dom-ST 副本（左）。（b）每个节点使用主机（CPU）和多个设备（GPU）运行分布式 Dom-ST （右）。

。我们的分布式训练方法应对了这一挑战（图 2）。首先，在输入流水线（I.P.）中，将目标站点的训练数据划分为 23 个区域或流域；随后，输入流水线将这些流域分配给多个节点（图 2.a.），每个节点负责一个流域。输入流水线还会创建模型副本，并将其分配给相应节点上的各个流域。其次，在每个流域节点上，我们利用空间模块的并行架构，将模型划分并卸载到多个 GPU 上（图 2.b.）。卷积神经网络的每个头部都被拆分出来，并分别在单个 GPU 上进行训练，使各个头部能够并行训练，从而优化训练时间。空间模块完成前向传播后，其输出会传输到下一块 GPU；该 GPU 上运行着时间模块，其中包含堆叠的长短期记忆层以及模型后续的最终层。时间模块所在的 GPU 也会直接接收模型的第二个输入。

我们使用气候研究部提供的像素级每日降水量，以及每个像素到最近水源的距离作为输入，并以美国地质调查局的每日流量测量值 [12] 作为标签。我们选择纳什–萨克利夫效率（纳什效率系数） [13]作为模型评估指标。基线模型仅由空间模块和时间模块组成。空间模块由单头一维多通道一维卷积神经网络构成，不含任何 Pix-Con 模块，这意味着基线模型完全由数据驱动，不包含任何领域知识。我们将其称为单头。基线模型在单个 GPU 上运行。我们探究将目标日降水量作为额外输入后，模型性能如何超越基线。图3显示，在几乎所有流域（91%）中，将目标日降水量输入模型的最后几层（单头($+P$)）都能提升性能。作为山洪的主要成因，目标日降水量为 Dom-ST 提供了时间粒度更细的数据，作为领域特定线索来提升预测效果。接下来，我们评估所提出的 Dom-ST。该模型包含一个 Pix-Con 模块、一个分布式多头空间模块，以及一个接收目标日输入的时间模块。为便于比较，我们在图3中将其称为分布式多头（$+P$）。结果表明，分布式多头（$+P$）在几乎所有流域中的表现都优于单头和单头($+P$)。这表明 Dom-ST 比其对照模型更具领域感知能力。为评估输入流水线策略，我们比较单头$+P$与分布式多头（$+P$）的计算时间：两者先后分别按顺序（S）运行，再并行运行（IP-D）。表1显示了每种情况下 23 个流域的总训练时间。

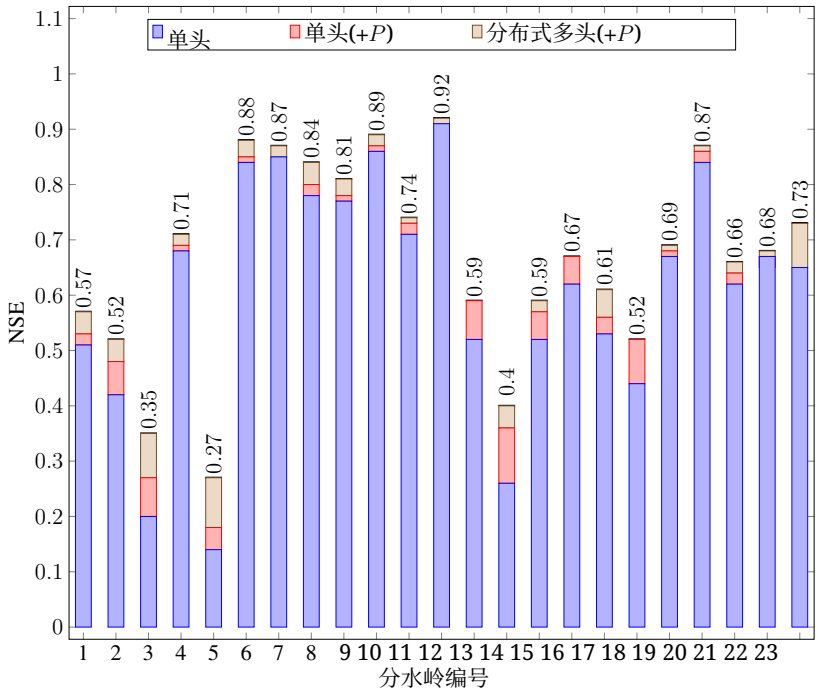


图3：从纳什效率系数 [14] 方面比较三种方法的预测性能——单头、带额外降水输入的单头（单头(+P)），以及带额外降水输入的分布式多头（分布式多头(+P)）。

表1：输入流水线策略（IP-D）的评估。

方法	时间（S）	时间（IP-D）	加速比
单头(+P)	9.96 小时	**1.18** 小时	8.5x
分布式多头(+P)	5.49 小时	**0.44** 小时	12.6 倍

3 结论

我们提出了一种新颖的分布式训练方法，用于加速具备领域感知能力的时空网络。通过严格的实验和消融研究，我们进一步探究领域知识如何影响模型预测，以及如何利用像素级贡献，将像素化时空数据融入其中，从而找到分布式训练方法。研究结果表明，在所有分析的流域中，使用 Dom-ST 平均可实现 2.02 倍加速比。Dom-ST 中的像素贡献模块利用领域知识，提升了模型的预测性能。我们的实验在单个流域上实现了最高 4.11 倍的加速比，总体加速比达到 12.6 倍，单个流域的纳什效率系数最高提升了 93%。

参考文献

[1] Guanghan Ning、Zhi Zhang 和 Zhiquan He。用于人体姿态估计的知识引导的深度分形神经网络。*IEEE*多媒体汇刊，20(5):1246–1259，2018。[2] Tao Zhong、Fenqiang Zhao、Yuchen Pei、Zhenyuan Ning、Lufan Liao、Zhengwang Wu、Yuyu Niu、Li Wang、Dinggang Shen、Yu Zhang 等。Dika-nets：领域不变

用于早期发育猕猴脑颅骨剥离的知识引导注意力网络。*NeuroImage*, 227:117649，2021。[3] Chen Chen、Yong Wang、Jianwei Niu、Xuefeng Liu、Qingfeng Li 和 Xuantong Gong。基于对比增强超声视频、由领域知识驱动的乳腺癌诊断深度学习。*IEEE*医学影像汇刊，40(9):2439–2451，2021。[4] Jian Sun、Kristopher A Innanen 和 Chao Huang。用于地震反演的物理引导深度学习：混合训练与不确定性分析。*Geophysics*，86(3):R303–R317，2021。[5] Aishwarya Sarkar、Jien Zhang、Chaoqun Lu 和 Ali Jannesari。Hydrodeep——一种用于地理时空数据分析的知识引导深度神经网络。*arXiv*预印本 *arXiv:2010.04328*，2020。[6] Sixing Yu、Murali Emani、Chunhua Liao、Pei-Hung Lin、Tristan Vanderbruggen、Xipeng Shen 和 Ali Jannesari。面向高性能计算中人工智能模型的无缝管理。*arXiv* 预印本 *arXiv:2212.06352*，2022。[7] Sixing Yu、Phuong Nguyen、Waqwoya Abebe、Wei Qian、Ali Anwar 和 Ali Jannesari。Spatl：用于异构联邦学习的显著参数聚合与迁移学习。载于*2022 SC22*：高性能计算、网络、存储与分析国际会议（*SC*），第495–508页。IEEE计算机学会，2022。[8] P. E. O'Connell、J. Ewen、G. O' Donnell 和 P. Quinn。农业土地利用管理与洪涝之间是否存在联系？水文学与地球系统科学，11(1):96–107，2007。[9] Roger Moussa、Marc Voltz 和 Patrick Andrieux。洪水事件期间农业管理的空间组织对农业流域水文行为的影响。*Hydrological Processes*，16(2):393–412，2002。[10] Howard Wheater 和 Edward Evans。土地利用、水资源管理与未来洪水风险。*Land Use Policy*，26:S251–S264，2009。Land Use Futures。[11] Aishwarya Sarkar、Jien Zhang、Chaoqun Lu 和 Ali Jannesari。基于网格的地理时空数据知识发现的迁移学习方法。载于*NeurIPS 2021*人工智能促进科学研讨会，2021。[12] Christopher S Jones、Jacob K Nielsen、Keith E Schilling 和 Larry J Weber。爱荷华州溪流硝酸盐与墨西哥湾。*PLOS ONE*，13(4):e0195930，2018。[13] Daniel N Moriasi、Jeffrey G Arnold、Michael W Van Liew、Ronald L Bingner、R Daren Harmel 和 Tamie L Veith。用于系统量化流域模拟精度的模型评估指南。*ASABE*会刊，50(3):885–900，2007。[14] J Eamonn Nash 和 Jonh V Sutcliffe。通过概念模型进行河流流量预报：第一部分——原理讨论。水文学杂志，10(3):282–290，1970。