arXiv:2304.05293v1 [physics.ins-det] 11 Apr 2023

用于带电粒子追踪的等变图神经网络

丹尼尔·默南[1],萨凡纳·泰斯[2],阿梅亚·特特[3]
[1] 劳伦斯伯克利国家实验室科学数据部，美国加利福尼亚州伯克利94720[2] 哥伦比亚大学数据科学研究所，美国纽约州纽约市10027[3] BITS皮拉尼物理系，KK比拉尔果阿校区，祖阿里纳加尔，果阿403726，印度

电子邮件：dtmurnane@lbl.gov

摘要。 图神经网络（GNN）凭借其提升准确性和可扩展性的潜力，已在高能物理（HEP）领域受到广泛关注。然而，其资源密集型特性和复杂操作推动了对对称等变架构的开发。在本工作中，我们介绍了 EuclidNet，一种用于带电粒子径迹重建的新型对称等变GNN。EuclidNet 利用碰撞事件的图表示，并针对探测器束流线轴强制施加旋转对称性，从而得到更高效的模型。我们在 TrackML 数据集上将 EuclidNet与最先进的交互网络进行了基准比较；该数据集模拟了高亮度大型强子对撞机（HL-LHC）预期出现的高堆积条件。结果表明，在较小的模型规模下（< 1000 参数），EuclidNet 达到了接近最先进水平的性能，优于非等变基准模型。这项研究为未来探索用于 HEP 实验中粒子径迹重建的资源效率更高的 GNN 模型铺平了道路。

1 引言

近年来，图神经网络（GNN）在高能物理（HEP）分析中的应用显著增加 [1]。这些研究表明，GNN 有望大幅提升准确率，并且能够轻松扩展以处理海量数据。然而，许多此类架构要么包含大量参数，要么涉及复杂的图操作和卷积，导致 GNN 部署时资源消耗大、耗时长。许多真实世界的数据集（包括 HEP 实验的数据集）都呈现出已知的对称性，可据此构建替代方案，避免使用计算成本高昂且不受约束的架构。通过利用特定问题中固有的对称性，我们可以对神经网络施加相对于某个对称群中变换的等变性约束，将其函数空间限制在相关候选函数之内。这种方法有双重优势：引入等变性可为神经网络加入归纳偏置，而等变模型也可能比非等变模型更节省资源 [2–4]。本文提出一种用于带电粒子追踪的新型对称等变图神经网络架构。我们使用碰撞事件的图表示，提出 EuclidNet，通过将输入的追踪特征标量化来确保旋转对称性。具体而言，鉴于探测器围绕束流线（z）轴具有对称性，我们构建了一个对 SO(2) 旋转群等变的 EuclidNet 表述。与当前最先进的（SoTA）交互网络相比，EuclidNet 在较小模型规模下也能达到接近 SoTA 的性能。

我们研究 EuclidNet 和 InteractionNet 的分布外推理性能，以解释模型性能与模型规模之间的关系。因此，我们不仅揭示了完全等变模型性能上限的线索，也揭示了非等变模型所学习到的对称性。代码和模型已公开发布于
https://github.com/ameya1101/equivariant-tracking。

2 理论与背景

在本节中，我们将介绍对称群等变性的概念，并简要介绍图神经网络理论。

2.1 等变性

形式上，如果 T_g： $X \to X$ 是抽象对称群 $g \in G$ 在向量空间 X 上的一组变换，则函数 ϕ：$X \to Y$ 被定义为关于 g 等变的，当且仅当在输出空间 S_g 上存在一组等价的变换 $Y \to Y$，使得:

$$\phi(T_g(\mathbf{x})) = S_g\phi(\mathbf{x}) \tag{2.1}$$

如果一个模型由一组对 G具有等变性的函数 ϕ 组成，则称该模型对群 G 具有等变性。在本文中，我们将讨论限制在与 SO(2) 群对应的旋转等变性上。例如，设 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}2, \cdots, \mathbf{x}_M)$ 是嵌入 $n-$ 维空间中的一组 M 点，并且 $\phi(\mathbf{x}) = \mathbf{y} \in \mathbb{R}^{M\times n}$ 是经过变换后的点集。对于一个正交旋转矩阵$Q \in \mathbb{R}^{n\times n}$，如果 $Q\mathbf{y} = \phi(Q\mathbf{x})$，则 ϕ 对旋转具有等变性。

2.2 图神经网络

考虑一个图 $\mathcal{G} = (V, E)$，其节点为 $v_i \in V$，边为 $e_{ij} \in E$。图神经网络是一种置换不变的深度学习架构，用于处理图结构数据 [5]。GNN 通常由多个层组成，每层执行图卷积操作，其定义如下 [6]:

$$\mathbf{m}_{ij} = \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, a_{ij}) \tag{2.2}$$

$$\mathbf{m}_i = \sum_{j\in\mathcal{N}(i)} \mathbf{m}_{ij} \tag{2.3}$$

$$\mathbf{h_i^{l+1}} = \phi_h(\mathbf{h}_i^l, \mathbf{m}_i) \tag{2.4}$$

其中，$\mathbf{h}_i^l \in \mathbb{R}^k$ 是节点 v_i 在第 l层的 k维嵌入。a_{ij} 是边属性。$\mathcal{N}(i)$ 表示节点 v_i的邻居集合。最后， ϕ_e 和 ϕ_h 是边操作和节点操作，使用多层感知机（MLP）进行近似。$\mathbf{m}ij$ 称为消息，在节点 v_i 和v_j之间传递。

3 网络架构

本节介绍 EuclidNet 的架构。如图 1所示，EuclidNet 由欧几里得等变块（EEB）以及一些编码层和解码层堆叠而成。EuclidNet 的架构与[7]中介绍的 LorentzNet 非常相似。本节中， $\phi(a, b, \ldots\ldots, f)$ 表示将量 a 至 f 拼接后再传入 ϕ。

输入层。 网络的输入是三维命中位置。对于 SO(2) 群，标量是命中位置的 $z-$坐标，而二维坐标构成向量。标量通过嵌入层投影到嵌入空间，然后传入第一个

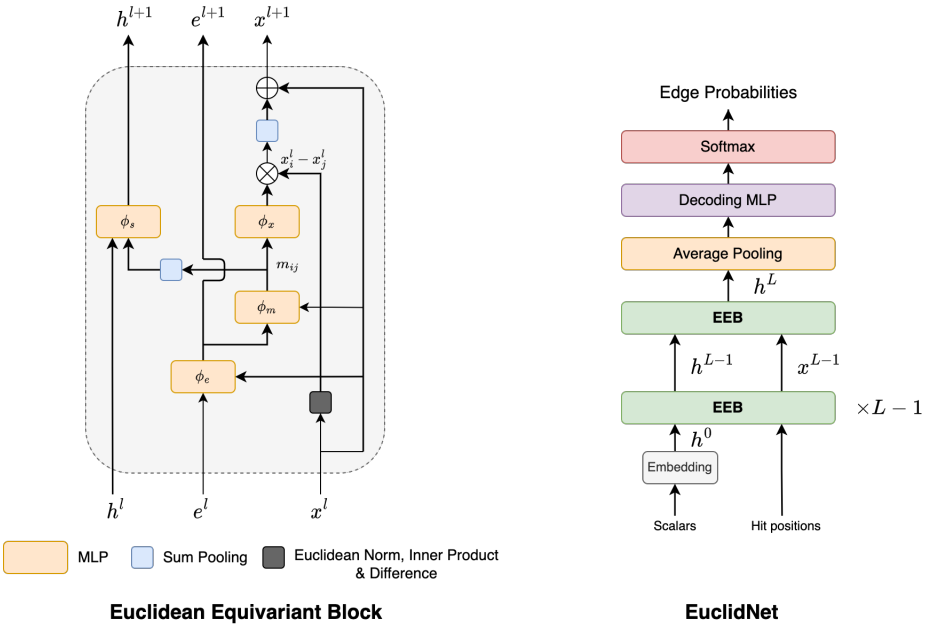

图1：（**左**）欧几里得等变块（EEB）的结构。（**右**）EuclidNet 的网络架构。

等变模块。

欧几里得等变块。 按照惯例，我们用 $h^l = (h_1^l, h_2^l, \ldots, h_N^l)$ 表示第 $l-$ 个 EEB层中的节点嵌入标量，用 $x^l = (x_1^l, x_2^l, \ldots, x_N^l)$ 表示坐标嵌入向量。 x^0 对应命中位置， h^0 对应标量变量的嵌入输入。待传递的消息 m_{ij}^l 构造如下：

$$m_{ij}^l = \phi_m\left(h_i^l, h_j^l, e_{ij}^l, \psi(\|x_i^l - x_j^l\|^2), \psi(\langle x_i^l, x_j^l \rangle)\right) \tag{3.1}$$

$$e_{ij}^l = \phi_e\left(\psi(\|x_i^l - x_j^l\|^2), \psi(\langle x_i^l, x_j^l \rangle), e_{ij}^{l-1}\right) \tag{3.2}$$

其中， $\phi_m(\cdot)$是一个神经网络，而 $\psi(\cdot) = \text{sgn}(\cdot)\log(|\cdot| + 1)$ 将广泛分布中的大数值进行归一化，以便于训练。 ϕ_m 的输入还包含欧几里得点积$\langle x_i^l, x_j^l \rangle$。命中之间的欧几里得距离 $\|x_i^l - x_j^l\|^2$ 是一个重要特征，我们将其纳入以便于训练。 e_{ij}^l 是由多层感知机学习得到的边显著性权重。

我们采用 [7] 中的点积注意力的欧几里得形式作为聚合函数，其定义如下：

$$x_i^{l+1} = x_i^l + c \sum_{j \in \mathcal{N}(i)} \phi_x(m_{ij}^l) \cdot (x_i^l - x_j^l) \tag{3.3}$$

其中 $\phi_x(\cdot) \in \mathbb{R}$ 是由多层感知机建模的标量函数。引入超参数 c 以控制更新的尺度。标量特征 h_i^l的更新如下：

$$h_i^{l+1} = h_i^l + \phi_h\left(h_i^l, \sum_{j \in \mathcal{N}(i)} m_{ij}^l\right) \tag{3.4}$$

解码层。经过 L 个 EEB 堆叠后，我们对节点嵌入 $h^L = (h_1^L, h_2^L, ..., h_N^L)$ 进行解码。消息传递确保向量嵌入中包含的信息传播到标量，因此无需同时对二者进行解码。一个包含三个全连接层、后接 softmax 函数的解码模块用于为每个轨迹片段生成真值分数。

4 结果

4.1 数据集

在本研究中，我们在 TrackML数据集上测试所开发的跟踪模型。该数据集是为 TrackML 粒子跟踪挑战赛开发的质子-质子碰撞事件模拟集合 [8]。事件平均生成 200 次堆积相互作用，用于模拟 HL-LHC 预期的高堆积条件。每个事件都包含三维命中位置，以及生成这些命中的粒子的真实信息。在本研究中，我们仅讨论像素层，这些像素层由最内侧区域中高度细粒度的四个桶部层和十四个端盖层组成。每个事件的跟踪器命中通过边构建算法转换为命中图。除横向动量、噪声和同层筛选器用于调节命中数量外，图的边还必须满足关于某些几何量的约束。在本研究中，我们使用 [9] 中的几何图构建策略来生成图，每个数据集事件对应 $p_T^{\min} = 1.5$ GeV。

4.2 实验

我们针对三种不同的隐藏通道数（8、16 和 32）训练 EuclidNet 和交互网络。在 TrackML 数据集上获得的结果汇总于表 1 中。我们使用 ROC 曲线下面积（AUC）评估模型，这是一种分类问题中常用的指标。我们还报告了模型参数数量，以及生成的事件图的纯度（图中真实边占边总数的比例）和效率（所有真实轨迹段中被正确分类的比例）。

表 1：EuclidNet 与交互网络（IN）在 TrackML 数据集上的性能比较。EuclidNet 和 IN 的结果均为 5 次运行的平均值

N_{hidden}	**模型**	**参数量**	**AUC**	**效率**	**纯度**
8	EuclidNet	967	$\mathbf{0.9913 \pm 0.004}$	$\mathbf{0.9459 \pm 0.022}$	$\mathbf{0.7955 \pm 0.040}$
	InteractionNet	1432	0.9849 ± 0.006	0.9314 ± 0.021	0.7319 ± 0.052
16	EuclidNet	2580	0.9932 ± 0.003	0.9530 ± 0.014	$\mathbf{0.8194 \pm 0.033}$
	InteractionNet	4392	0.9932 ± 0.004	$\mathbf{0.9575 \pm 0.019}$	0.8168 ± 0.073
32	EuclidNet	4448	0.9941 ± 0.003	0.9547 ± 0.019	0.9264 ± 0.023
	InteractionNet	6448	$\mathbf{0.9978 \pm 0.003}$	$\mathbf{0.9785 \pm 0.022}$	$\mathbf{0.9945 \pm 0.043}$

对于较小的模型规模，EuclidNet 的表现优于无约束的交互网络。通用 TrackML 探测器近似满足旋转对称性，而利用这种对称性，初步看来能够构建出准确的边分类图神经网络。然而，当潜在空间更大时，强制施加旋转对称性所带来的性能提升会趋于平稳。此时，无约束网络的表现更好。我们推测，与 ATLAS 和 CMS 等真实探测器类似，TrackML 数据集的旋转对称性可能仅为近似对称，这是由探测器分段的颗粒度以及探测器周围磁场的不均匀性造成的。为研究这一可能性，我们使用数据集经 $\theta \in [0, \pi/4, \pi/2, 3\pi/4, \pi]$ 旋转后得到的一组数据来训练这两个模型，并在这组旋转数据上分别对每个实例进行推理

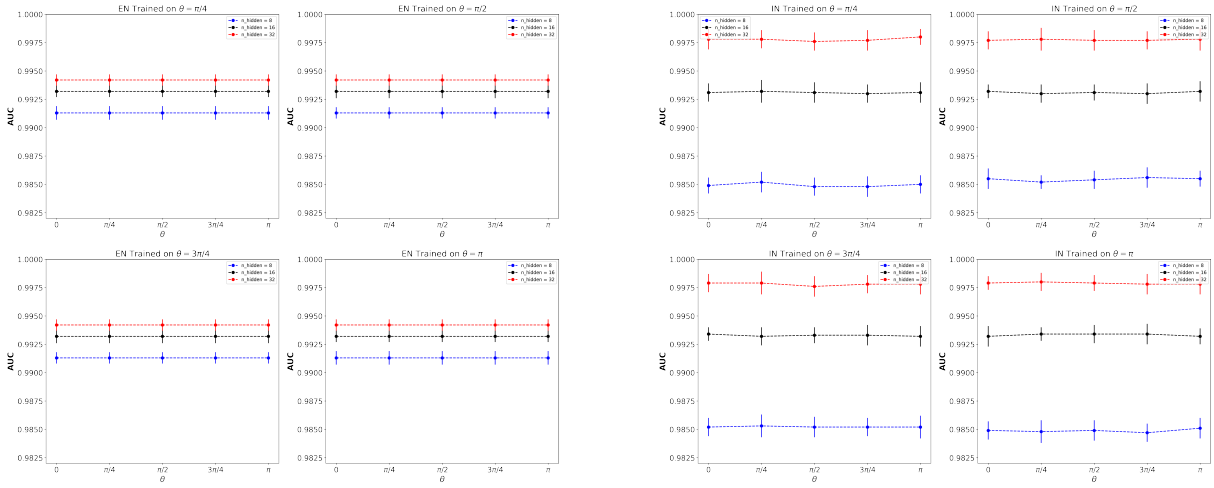

图 2：对于在数据集旋转角度为 $\theta \in [0, \pi/4, \pi/2, 3\pi/4, \pi]$ 的一组旋转变换上训练的模型，推理时的曲线下面积（AUC）随输入空间旋转角度 θ 的变化曲线：（**左**）EuclidNet；（**右**）交互网络。结果基于 5 次独立推理运行计算得出。

。这样，我们就能判断无约束网络是否学到了特定于探测器材料和磁场朝向的函数。这项研究的结果总结见图 2。

在图 2 中，我们观察到，无论模型基于数据集的哪组旋转变换进行训练，SO(2) 等变 EuclidNet 和交互网络都能稳健地应对输入空间中的旋转。表 1 中观察到的总体趋势在此处也得到了重复：潜在空间更大的模型会产生更高的曲线下面积（AUC）分数。尽管还需要进一步研究来解释无约束网络的优越性能，但我们推测，这可能是数据集固有的近似旋转对称性所造成的结果。在这种情况下，表达能力较弱的 EuclidNet 将无法完全匹配事件的近似对称性，因此只有在潜在空间大小不足以让交互网络捕捉完整对称集时，其性能才会超过交互网络。然而，当潜在空间足够大时，像交互网络这样的无约束网络似乎能够轻松学习数据集中的近似对称性和非对称修正。未来需要开展更深入的研究，利用可解释人工智能技术以及“LieGG”等方法 [10] 来确定 EuclidNet 和交互网络实际学习到的对称性，从而查明产生这一行为的原因。

5 结论与未来工作

在本研究中，我们提出了 EuclidNet——首个用于粒子跟踪问题的欧几里得旋转等变图神经网络（GNN）。在模型规模较小（< 1000 参数）时，该 SO(2) 等变模型的表现略优于基准模型（曲线下面积 $= 0.9849$），其曲线下面积为 $= 0.9913$。然而，对于粒子跟踪问题，我们发现，在模型规模较大时，未受约束的模型仍然优于等变架构。要具体确定这一结果背后的原因，还需要开展更多研究。未来可能的研究方向包括探究该问题的等变性并识别其中可能存在的非等变方面，以及调查所学习到的对称性的质量。不过，如果数据集本身包含非等变特征，那么任何对称模型都很可能始终表现不佳。在这种情况下，[4] 中这类放宽了对称性架构所施加严格约束的模型，或许能够学习

跟踪数据集的非等变方面。

致谢
本研究通过 IRIS-HEP 获得美国国家科学基金会合作协议 OAC-1836650 的支持。本研究使用了国家能源研究科学计算中心（NERSC）的资源。该中心是美国能源部科学用户设施，位于劳伦斯伯克利国家实验室，并依据合同编号 DE-AC02-05CH11231 运营。

附录 A 训练细节

EuclidNet 和交互网络均使用二元交叉熵损失函数。所有模型均使用 Adam 优化器进行优化，[11] 该优化器在 PyTorch 中实现，学习率为 $\eta = 1\times10^{-3}$，动量系数为（β_1，β_2）$=(0.9,0.999)$。学习率每 100 个训练轮次衰减为原来的 0.3。两个模型分别在单张 NVIDIA A100 GPU 上训练 200 个训练轮次，并采用耐心值为 50 个训练轮次的早停策略。两个模型的总训练时间通常为 2 小时。

附录 B 等变性测试

我们还测试了 EuclidNet 的等变性，发现除浮点数值误差外，它确实对 SO(2) 变换保持等变。对于给定的变换 $R \in SO(2)$，我们比较 $R_\theta\phi(\mathbf{x})$ 和 $\phi(R_\theta\mathbf{x})$，其中$\mathbf{x}$ 为击中坐标，而 $\phi(\cdot)$ 是一个 EuclidNet 实例。在图 B1中，我们绘制了相对偏差，其定义如下：

$$\varepsilon(\theta) = \frac{\langle\phi(R_\theta\mathbf{x})\rangle - \langle R_\theta\phi(\mathbf{x})\rangle}{\langle R_\theta\phi(\mathbf{x})\rangle} \tag{B.1}$$

其中，$\langle\cdot\rangle$ 是对测试数据集中的 500 个跟踪事件求得的均值。我们发现，旋转导致的相对偏差为 $< 10^{-8}$。

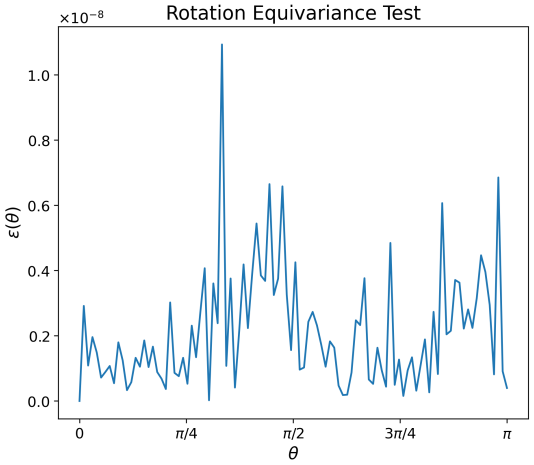

图 B1：输出节点向量表示相对于绕束流线（z 轴）旋转的相对偏差。

参考文献

[1] Thais S、Calafiura P、Chachamis G、DeZoort G、Duarte J、Ganguly S、Kagan M、Murnane D、Neubauer M S 和 TeraoK 2022年 *2022年 Snowmass 夏季研究* （预印本 2203.12852）

[2] Satorras V G、Hoogeboom E 和 Welling M 2021年 国际机器学习大会（PMLR），第 9323–9332 页[3] Bogatskiy A 等 2022年 2022年 Snowmass 夏季研究（预印本 2203.06153）[4] Murnane D、Thais S 和 Wong J 2023年 《物理学杂志：会议系列》2438 012121，网址：https://doi.org/10.1088%2F1742-6596%2F2438%2F1%2F012121[5] Kipf T N 和 Welling M 2016年（预印本 1609.02907）[6] Gilmer J、Schoenholz S S、Riley P F、Vinyals O 和 Dahl G E 2017年 国际机器学习大会（PMLR），第 1263–1272 页[7] Gong S、Meng Q、Zhang J、Qu H、Li C、Qian S、Du W、Ma Z M 和 Liu T Y 2022年 高能物理学杂志 07 030（预印本 2201.08187）[8]Amrouche S 等 2019年 跟踪机器学习挑战赛：精度阶段（预印本 1904.06778）[9] DeZoort G、Thais S、Duarte J、Razavimaleki V、Atkinson M、Ojalvo I、Neubauer M 和 Elmer P 2021年 计算、软件与大科学 5 26（预印本 2103.16701）[10] Moskalev A、Sepliarskaia A、Sosnovik I 和 Smeulders A 2023年 Liegg：研究学习得到的李群生成元（预印本 2210.04345）[11] Kingma D P 和 Ba J 2014年（预印本 1412.6980）