arXiv:2303.08046v2 [physics.ins-det] 1 Aug 2024

采用事件内感知生成对抗网络和自监督关系推理的超高颗粒度探测器模拟

BaranHashemi[1*]，Nikolai Hartmann[1]，SahandSharifzadeh[2]，James Kahn[3,4] 和 Thomas Kuhr[1]1*慕尼黑路德维希-马克西米利安大学物理学院，德国慕尼黑。[2]慕尼黑路德维希-马克西米利安大学计算机科学学院，德国。[3]亥姆霍兹人工智能，德国。[4]卡尔斯鲁厄理工学院（KIT）施泰因布赫计算中心（SCC），德国。

*通讯作者电子邮箱：
baran.hashemi@origins-cluster.de；

摘要

模拟高分辨率探测器响应是一项计算密集型任务，长期以来一直是粒子物理领域的一大挑战。尽管生成模型能够简化这一过程，但完整的超高颗粒度探测器模拟仍然十分困难，因为其中包含相关且细粒度的信息。为克服这些局限，我们提出了事件内感知生成对抗网络（IEA-GAN）。IEA-GAN 引入了一个关系推理模块，用于近似探测器模拟中的事件，并借助适当的关系归纳偏置，生成具有上下文信息的高分辨率完整探测器响应。IEA-GAN 还引入了自监督事件内感知损失和均匀性损失，显著提升了样本保真度和多样性。我们展示了 IEA-GAN 在 BelleII 实验中用于生成超高颗粒度像素顶点探测器（PXD）的传感器相关图像的应用，该探测器拥有超过 750 万个信息通道。这项工作的应用范围涵盖用于高颗粒度探测器模拟的基础模型，

例如高亮度大型强子对撞机（HL-LHC）上的探测器模拟，以及基于模拟的推断和细粒度密度估计。

关键词：高分辨率、高粒度探测器模拟，细粒度条件图像生成，关系归纳偏置，生成对抗网络，融合知识迁移的自监督学习

1 引言

高效快速模拟 [1–8] 活动促使粒子物理学界开始寻找更快速、存储效率更高的对撞物理实验模拟方法。模拟在多种下游任务中发挥着至关重要的作用，包括优化探测器几何结构、设计物理分析，以及寻找标准模型（SM）之外的新现象。生成对抗网络（GAN）的引入 [9] 彻底革新了图像数据的高效探测器模拟。

深度生成模型已广泛应用于粒子物理学，以实现大型强子对撞机的探测器模拟 [2–4、8、10–15]，主要用于量能器模拟或碰撞事件生成 [16–21]。此前生成高空间分辨率探测器响应的研究包括 Hashemi 等人提出的先验嵌入生成对抗网络（PE-GAN） [7],，该方法利用端到端嵌入来纳入有关探测器传感器的全局先验信息；以及 Srebre 等人的工作 [6] ，其中采用 Wasserstein 生成对抗网络 [22] 与梯度惩罚 [23] ，作为一种概念验证，在不进行条件约束的情况下生成高分辨率图像。对于中等粒度量能器模拟，近期方法 [8, 12],尝试采用多种类 GAN 和基于流的 [24]架构，模拟通道数不足 3 万；而用于高粒度量能器模拟的三维生成对抗网络仅处理 6.5 万个像素通道。尽管如此，这些研究对于未来探测器模拟所面临的巨大挑战仍只是浅尝辄止。以即将建成的高粒度量能器（HGCAL）为例，它是紧凑型缪子螺线管（CMS）实验中高亮度大型强子对撞机（HL-LHC） [25] 升级计划的一部分 [26]。HGCAL 预计将拥有 6.5 百万个探测器通道，分布在 50 层中，其复杂程度远超现有方法的处理能力，这凸显了开发更先进模拟方法的紧迫性。

学习生成超高分辨率探测器响应面临多项挑战。首先，我们通常处理的是空间不对称的高频击中图。对于当前用于高分辨率图像生成的最先进水平（SOTA）生成对抗网络配置，当判别器比生成器强得多时，假图像很容易与真实图像区分开来，进而导致判别器的梯度消失。由于生成不平衡的高频细节十分困难，这种情况在不对称的高分辨率图像中更为常见。另一方面，

判别器能力较弱则会导致模式崩溃，此时生成器贪婪地优化其损失函数，只生成少数几种模式来欺骗判别器。

此外，事件中的探测器响应（即粒子碰撞后单个读出窗口内的响应）在统计和语义上都具有相似性 [27] 。例如，类别内每幅图像的稀疏度（占有率）定义为非零像素所占的比例，它体现了不同探测器组件之间的统计相似性（见补充说明中的补充图）。此外，探测器响应图像在语义和视觉层面都极为相似 [27]，因此可以将其归类为细粒度图像。生成细粒度图像的目标是创建属于子类别的视觉对象。在计算机视觉中，类似的任务包括生成不同犬种或汽车型号的图像。细粒度图像分析中固有的类间差异小、类内差异大，使这一问题颇具挑战性 [28]。当前最先进的条件生成对抗网络模型关注类别层面和类内层面的图像相似性，其中考虑了图像内部的 [29],数据到类别 [30], 和数据到数据 [31] 关系。然而，在探测器模拟中，类别具有层级性且十分细粒度，因此，要区分语义和视觉上相似的生成类别更加困难。因此，前述模型在类间层面表现出严重的类别混淆 [32, 33] 。此外，由于事件中的信息来自探测器的单个读出窗口，该窗口中发生的过程会同时影响所有传感器，从而使它们之间产生相关性（见第2节）。本文将展示，这种细粒度的事件内相关性在下游物理分析中发挥着至关重要的作用。

为克服超高分辨率探测器模拟面临的所有这些挑战，我们提出了事件内感知生成对抗网络（IEA-GAN），这是一种新型深度生成模型，能够以最高保真度生成传感器/层依赖的探测器响应图像，同时满足所有相关指标。由于我们处理的是一组细粒度且（由各个事件）赋予上下文的图像，这些图像之间共享信息。首先，我们为判别器和生成器引入关系推理模块（RRM），以捕获类别间关系。然后，我们提出一种生成器损失函数，使其能够模仿判别器关于成对类别对应关系的知识 [34]。最后，我们为判别器引入辅助损失函数，通过施加信息均匀性条件 [35] 来利用其推理陪域，从而缓解模式崩溃问题并提升模型的泛化能力。IEA-GAN 在细粒度图像生成任务中，不仅捕获统计层面和语义层面的信息，还能捕获样本间的相关性。

我们展示了 IEA-GAN 在 Belle II [36] 的像素顶点探测器（PXD）超高维数据上的应用，该数据包含超过 750 万个像素通道，是目前空间分辨率最高的探测器模拟数据集

也是迄今为止使用深度生成模型分析过的数据集。随后，我们考察了多种评估指标，并表明在所有这些指标上，与其他用于高维图像生成的最先进水平深度生成模型相比，IEA-GAN 都与目标分布更加吻合。我们还进行了消融研究并探索了超参数，以深入了解该模型。

需要强调的是，我们的方法超越了现有量能器模拟模型的范畴。此类模型旨在捕捉逐层相关性 [38]。尽管这些现有模型确实考虑了同时依赖多个层的可观测量，但它们仅限于模拟簇射范围内受限且局部区域中源自单一粒子源的粒子簇射。相比之下，我们的方法涵盖整个事件的复杂性，能够处理多个粒子源，并涵盖全探测器模拟。通过这一视角，我们可以捕捉不同角度和层上的探测器传感器之间的相关性。由此，我们能够近似描述整个探测器中传感器间错综复杂且动态变化的相互作用，超越仅关注局部粒子簇射的模拟所存在的局限。

本文研究最具挑战性的探测器模拟问题，采用来自像素顶点探测器（PXD）的空间分辨率最高的数据集 [36], ，如图 1 所示。PXD 是 Belle II 最内侧的径迹子探测器 [37]。PXD 的配置如图 1 所示，由分布在两层探测器中的 40 个传感器组成。内层有 16 个传感器，外层有 24 个传感器。因此，每个事件包含 40 幅灰度图像，每幅图像的分辨率为 250×768 像素；如图1 所示，每个事件包含超过 750 万个像素通道。PXD 记录的背景特征构成了每个事件中大多数 PXD 击中，它们来自探测器中各种并非源自我们关注的物理过程（即信号过程）的过程。这些背景过程可分为束流诱发过程和依赖于光度的过程。束流诱发过程源于同步辐射，以及束流粒子与束流管中的残余气体、弯转磁铁或束团内粒子的碰撞。相比之下，依赖于光度的过程包括正负电子碰撞，并产生巴巴散射或双光子过程等物理过程。

此类高分辨率背景叠加的问题在于，其读出、存储和分发需要耗费大量资源 [39] 。例如，模拟单个事件所需的 PXD 背景叠加数据大小约为 200kB。这大约是所有其他探测器组件合计的每事件背景叠加数据大小的 $2N$ 倍 [40] ，其中 N 为 PXD 背景放大系数。因此，一种思路是在线模拟这些数据。然而，由于 Geant4 模拟单个事件需要约 1500 秒，所需模拟时间相当长，因此实时模拟本底事件并不可行。因此，尽管存储如此大量的

数据对于高分辨率探测器而言效率很低，我们提出使用 IEA-GAN 作为代理模型，实时生成本底特征。

本文展示了，通过将 IEA-GAN 应用于 Belle II 的 PXD，我们不仅能够将预先生成的背景数据的存储需求降低一半（见第 3 节），还能够实现线上模拟，如图所示。1 与旧有且不可行的 Geant4 方法相比，该方法大幅缩短了线上模拟所需的 CPU 时间。因此，终于可以将 IEA-GAN 用作在线代理模型，对超高颗粒度的 PXD 背景进行即时模拟；对于这种高分辨率探测器模拟而言，此前这项任务无法实现。因此，IEA-GAN 是一种可行的候选方案，能够应对即将到来的高亮度大型强子对撞机 [25] 时代的超高颗粒度需求。

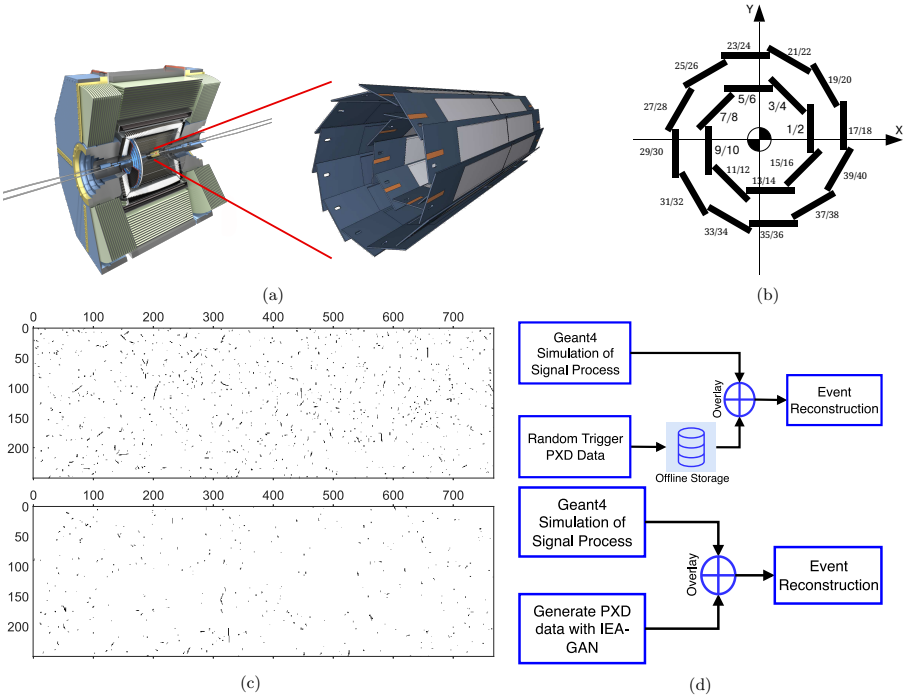


图 1：**像素探测器布局**（取自 [41]）像素探测器（PXD）是 Belle II 实验最内侧的子探测器（**a**），其传感器采用双层重叠结构（**b**）。传感器 7（上）和 25（下）的 PXD 图像示例（**c**）。（**d**）采用 Geant4 [42]（上）和 IEA-GAN（下）的事件生成流程。在分析过程中即时生成 PXD 数据，可避免将其离线存储。

2 结果

2.1 IEA-GAN 架构

生成对抗网络（GAN）是一种无监督的深度学习架构，由生成器和判别器两个网络组成，其目标是在双人极小极大问题中找到纳什均衡 [43]。如图所示，IEA-GAN2是一种基于自监督关系 grounding 的深度生成模型。

IEA-GAN 的判别器 D 将来自单个事件的一组探测器响应图像$\mathbf{x}_i \in \mathbf{R}^d$ 以自监督方式嵌入为全连接事件图中的输入节点。事件图是一种加权图，其中节点是事件中的嵌入探测器图像，边的权重由每个事件中探测器图像之间的相似度决定（参见第 4 节）。IEA-GAN 通过使用置换等变关系推理模块（RRM）进行上下文推理，来近似事件的概念（即事件图）。RRM 是一种兼容 GAN 的全连接多头图 Transformer 网络 [44–46]，它根据上下文相似度对事件中的图像令牌进行分组。RRM 中的注意力机制学习事件中样本之间的上下文相似度。为了进行多模态对比推理，判别器还将探测器的传感器嵌入作为类别令牌（参见第 4 节）。最后，正如第 4 节将详细讨论的那样，它通过将归一化图投影到超球面上，压缩图像和类别模态的信息。

为确保生成器 G 能够正确理解事件并捕捉事件内相关性，它首先在每个事件中从正态分布 $\mathcal{N}(0,\ 1)$中采样，将其作为随机自由度（Rdof），并用这个四维可学习 Rdof 修饰传感器嵌入（见第4节）。随后，为了对每个事件进行自监督上下文嵌入，RRM 会在此基础上发挥作用。值得注意的是，Rdof 与原始生成对抗网络的 [9] 高斯潜在向量不同。Rdof 可视为事件级可学习片段嵌入 [47] 或对词元嵌入的扰动 [48]，能够利用生成图像的多样性。将这些模块与 IEA 损失结合后，生成器便能洞察事件中的样本并建立其间的相关性，从而提升整体性能。

除了对抗损失外，IEA-GAN 还得益于一组自监督和基于对比的损失。模型通过基于代理的对比2C损失 [31] 理解探测器的几何结构，其中可学习代理是超球面上的传感器嵌入。此外，为了提升训练的多样性和稳定性，我们为判别器引入了均匀性损失。均匀性损失可以促使判别器对超球面的所有区域赋予相同权重 [35]，而不是只关注那些容易区分真实数据和伪造数据的区域。促使判别器施加均匀性不仅

能产生更多样、丰富的输出，还能缓解模式崩溃等问题。

IEA-GAN的另一个重要组成部分是IEA损失，它旨在解决条件生成模型中的类别混淆 [32, 33] 问题，适用于细粒度数据集。在IEA损失中，生成器通过二元信息传递，尝试模仿判别器对每个事件的理解，同时对判别器使用停止梯度（sg）。通过感知每个事件中条件的变化，这能够提升生成器在模拟过程中生成更多细粒度样本的能力。

2.2 IEA-GAN评估

我们的研究展示了IEA-GAN在生成超高粒度探测器响应方面的性能，并通过其在Belle II的像素顶点探测器（PXD）超高维数据上的成功应用加以证明；该数据包含超过750万个像素通道。此外，我们的研究结果表明，对于探测器模拟，Fr´echet初始距离（FID） [49] 和核 Inception 距离（KID） [50] 指标是非常通用的估计量，可与边缘分布结合使用，并且与其他图像级指标相关联。我们展示了，使用IEA-GAN能够捕获底层分布，从而生成并增强探测器响应信息，且与Geant4分布高度吻合。我们还发现，即使经过深入的超参数调优分析，最先进的高分辨率图像生成模型的表现仍不理想，相比之下逊色不少。

在评估中，我们采用两类指标：图像级和物理级。由于我们希望生成图像同时具备最佳的像素级属性、多样性和相关性，并且受计算资源限制，生成器的复杂度必须尽可能低，因此很难选择最佳迭代轮次来比较结果。为此，在所有比较中，我们均选取 FID 最佳时的模型权重。为了计算 FID 和 KID 分数，基于近期的 Clean-FID 项目 [51], ，我们在 PXD图像上对 Inception-V 3 [52] 模型进行了全面微调，因为 PXD图像与该模型初始训练所用的自然图像差异很大。微调的下游任务是多类别分类，涉及 40 个不同的传感器，模型由此获得了区分传感器的能力。换言之，Inception-V3 的分类任务是从 1 到 40 的范围内识别特定的传感器编号。这项任务要求模型通过辨别每个传感器固有的数据特征，识别每个数据样本所属的传感器。此过程也可用于任何其他探测器数据集。FID 用于衡量生成样本的表征与真实分布中样本表征之间的相似度。在大规模采样统计的基础上，对于 Inception 模型的每个隐藏激活，FID 会计算激活分布前两个矩之间的 Fr´echet 距离，也称为 Wasserstein-2距离。正如其在自然图像分析领域已被证明实用且有效，FID 的表现 [53]良好

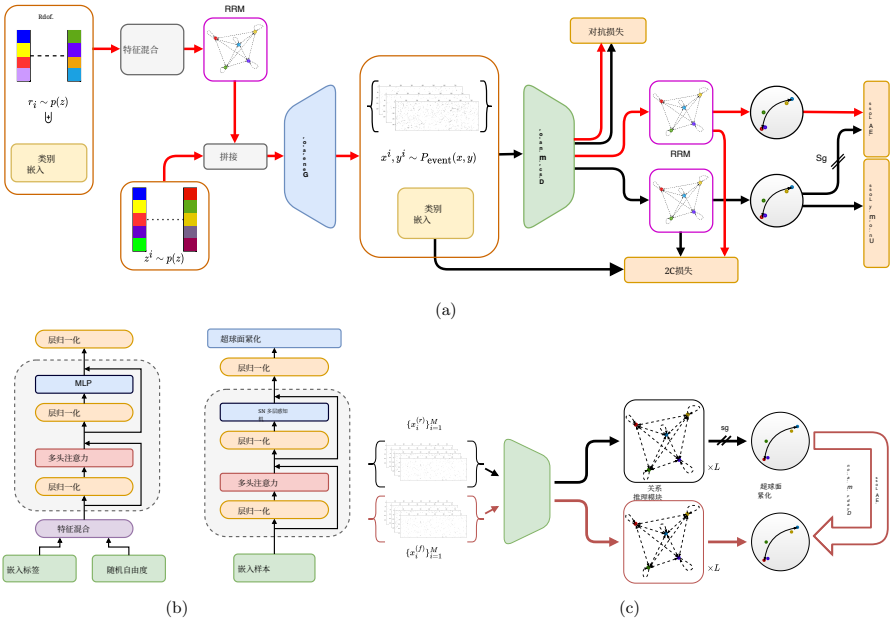


图 2：IEA-GAN 架构（a）及关系推理模块组件（b），以及 IEA自监督损失（c）。a Rdof 表示随机自由度，它将生成器的传感器/层嵌入与事件级可学习嵌入结合起来，后者负责生成器的事件内相关性。生成器和判别器中的关系推理模块（RRM）分别通过依据上下文相似度对类别/图像嵌入进行聚类，实现事件内推理。红线表示生成器的前向和反向传播。黑线表示判别器的前向和反向传播。判别器通过对抗损失（见公式（6））、2C损失（见公式（7））和均匀性损失（见公式（17））进行训练。另一方面，生成器使用对抗损失、2C损失和 IEA损失（见公式（16）），如（c）所示。Sg 表示对 IEA损失中判别器的梯度进行停止；该损失是一种面向生成器的自监督二元感知损失。b生成器（左）和判别器（右）的关系推理模块（RRM）在每次迭代时都会创建事件图。RRM 内部的注意力机制会学习事件中样本之间的上下文相似度程度。c IEA损失为生成器施加成对细粒度类别间模仿力。Sg 表示停止判别器的梯度，仅更新生成器的梯度。

在可区分性、多样性和鲁棒性方面表现出色，尽管它只对特征空间中分布的前两个矩进行建模。FID 分数越低，真实样本与生成样本的分布就越相似。核 Inception 距离（KID）是另一种与 FID 类似的指标，用于评估生成模型的质量。与 FID 不同，KID 使用核两样本检验，可对分布间距离进行无偏估计，并且在样本量较小时更具鲁棒性。

我们将 IEA-GAN 与另外三种模型（仅比较图像级指标）以及参考数据进行比较；该参考数据由 Geant4 模拟生成 [42] 。基线模型是条件图像生成领域的最先进水平模型：BigGAN-deep [54]和 ContraGAN [31]。我们还将 IEA-GAN 与此前用于 PXD 图像生成任务的研究成果进行比较：PE-GAN [7] 和 WGAN-gp [6] （仅比较 FID）。

表 1 显示，与其他模型相比，IEA-GAN 生成的图像具有最低的 FID 和 KID 分数，性能高出 42%。这表明，与其他模型生成的样本相比，我们的模型能够生成与目标数据更接近的合成样本。测试数据的低 FID 和 KID 值表明，模型已充分理解全部数据。在补充说明的表 2中，我们展示了 FID 对各种图像失真的敏感性，以及其与由相应传感器记录的底层物理之间可能存在的互操作性。为此，我们对图像施加受控的变化或抖动，并追踪这些变化对 FID 分数的影响。

表 1：各模型之间的 FID 和 KID 对比（基准测试中的所有模型均针对当前问题和数据集进行了高度调优），结果为六个随机种子下的平均值（使用不同随机种子训练的模型经过重新训练，并对六个随机种子下的结果取平均）。FID 和 KID 越低，图像质量和多样性越好。

	WGAN-gp	BigGAN-deep	ContraGAN	PE-GAN	IEA-GAN	测试数据
FID	12.09	4.40 ± 0.88	3.14 ± 0.74	2.61 ± 0.91	$\mathbf{1.50 \pm 0.16}$	2.4×10^{-5}
KID($\times 10^{-3}$)	9.6	3.1 ± 0.1	1.5 ± 0.2	2.1 ± 0.4	$\mathbf{1.0 \pm 0.2}$	7.6×10^{-1}

在像素级，有像素强度分布、占有率分布和平均占有率。像素强度分布定义了背景击中的能量分布。对于给定数量的事件，占有率分布和像素强度分布是在所有传感器上评估的，而平均占有率则对应于每个传感器在各事件中的稀疏度平均值。这些像素级信息至关重要，因为通过 basf2 软件进行物理分析时 [55], ，如果要使用图像并将提取的信息叠加到信号击中上，

表2：对 PXD 图像应用不同抖动方法后的 FID 分数，提供其数值的互操作性。

图像抖动	**FID**
None	2.4×10^{-5}
随机遮罩（死区）	14.58
随机噪声	87.23
随机旋转（30 度）	23.69
随机旋转（10 度）	2.81
随机平移 (0.1, 0.1)	1.99
随机剪切 (10, 10)	23.53
随机缩放	9.06
高强度电荷涂抹	3.16
低强度电荷涂抹	47.24

图像的稀疏度定义了每个传感器上背景击中的数量。像素强度分布、占有率分布以及每个传感器的平均占有率如图 3所示。IEA-GAN 模型的分布与参考结果最为吻合。

占有率的双峰分布源于探测器的几何结构：其传感器并非像量能器那样呈圆柱形，而是呈环形。这表明，要生成这种探测器特征，在几何结构和分辨率方面都颇具挑战。为了捕捉占有率分布正确的双峰性，RRM 和均匀性损失发挥着重要作用。通过在判别器中使用均匀性损失，可以促使生成器生成不会偏向某个特定模式或类别的样本，从而使生成样本的双峰分布更加宽广。

此外，通过利用考虑事件内样本间依赖关系和相关性的RRM模块，IEA-GAN在高能击中方面表现出更高的一致性，从而提高了低发生率区域中生成样本的多样性。

除了所有这些图像级指标外，我们还需要一种对事件内敏感的指标。上述所有指标在事件间样本置换下均具有等变性。换言之，如果在固定传感器编号的同时随机打乱事件间的样本，所有讨论过的指标都不会改变。因此，我们需要一种能够在事件空间中单独考察每个事件的上下文、并超越样本空间的指标。因此，我们计算生成事件总体中传感器占有率之间的Spearman相关性，

$$\mathbf{r}_s = \mathrm{Corr_p}(\mathrm{R}(\biguplus_{i=1}^{M=40} (\|\mathbf{x}_i\|_0)), \mathrm{R}(\biguplus_{i=1}^{M=40} (\|\mathbf{x}_i\|_0))), \tag{1}$$

其中，R(.) 是秩算子，其功能是为列表中的每个数字分配一个秩，如Spearman相关性的定义所示；$\mathrm{Corr_P}$(.) 是皮尔逊相关函数。$\biguplus$ 是表示拼接操作的不交并算子

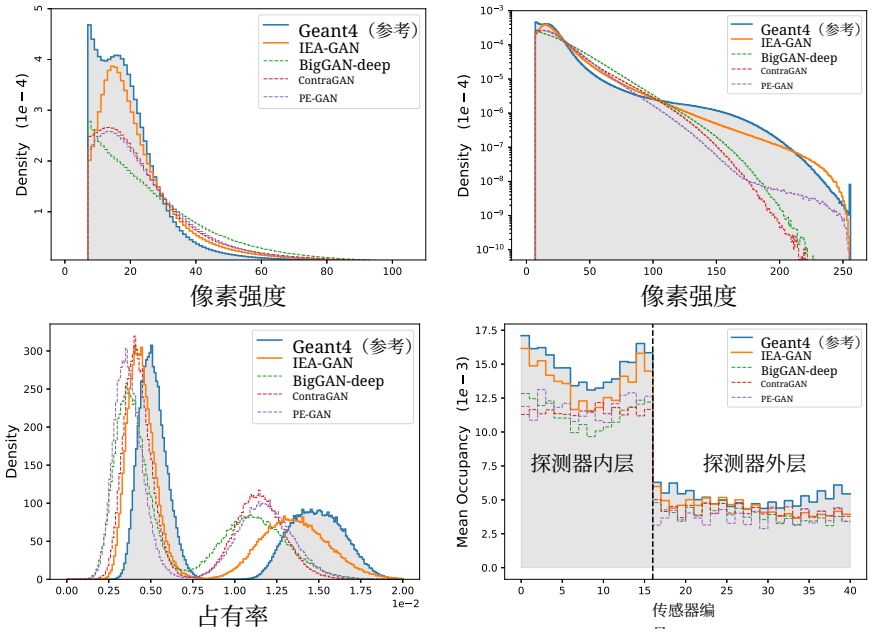


图 3：线性尺度（左上）和对数尺度（右上）下的像素强度分布、占有率分布（左下）以及每个传感器的平均占有率（右下），共 10000 个事件。

。下标为 0 的范数记作 $\|.\|_0$，即 L0 测度。它是一个用于统计向量中非零元素数量的函数。在本研究中，它用于计算传感器的占有率，即传感器图像中非零元素的数量 x_i。PE-GAN 生成的系数是范围 $[-0.2, 0.2]$内的随机值，而 IEA-GAN 的图像则显示出其生成图像之间存在有意义的相关性。尽管所需的相关性与参考值不同（如图4所示），IEA-GAN 仍能学习到层内传感器之间单调正相关、层间传感器之间主要为负相关的关系。

为了证明学到的相关性确实有意义，我们采用 Mantel 检验 [56, 57]，，这是一种针对两个距离/相关矩阵（不包括对角线部分）之间相关性的显著性检验。Mantel 检验通过比较两个矩阵中对应元素的每一对来进行。零假设是两组相关性之间不存在关系，而检验统计量是相关系数。我们使用置换检验评估观察到的相关性的显著性。具体做法是多次随机重排其中一个矩阵的元素，每次都重新计算检验统计量，然后判断观察到的检验统计量相对于这些检验统计量的零分布有多极端。

如果观察到的检验统计量非常极端，则 p值小于 0.05，零假设将被拒绝。对于 IEA-GAN，Mantel 检验结果显示，真实相关性为 $18 \pm 2\%$，经验p值为 0.0013。由于 p值小于 0.05，我们可以拒绝零假设，并观察到有显著证据表明两组矩阵之间存在相关性。这表明，在 Geant4样本中相关性较强的传感器类别，在 IEA-GAN 生成的样本中也往往存在相关性。相比之下，对于 PE-GAN，Mantel 检验结果显示，真实相关性为 0.2%，经验p值为 0.96，这支持零假设。

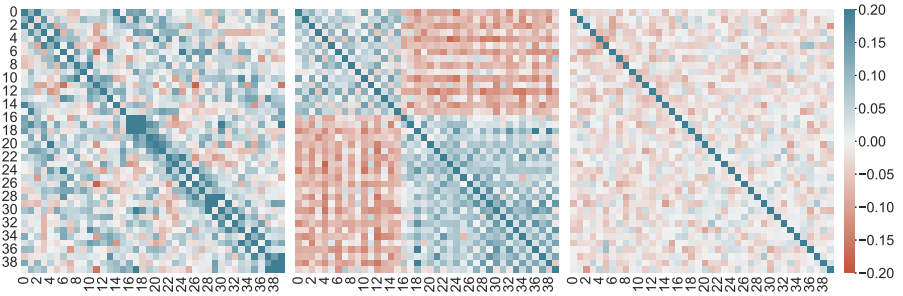


图 4：Geant4 传感器图像（左）、IEA-GAN 生成的传感器图像（中）与 PE-GAN 生成的传感器图像（右）的占有率之间的 Spearman相关性。

图像级指标反映了模拟的低层次质量，但我们还必须确认，从整个探测器的整体角度来看，所得模拟在物理合理性方面是可信的。为此，我们进行了径迹分析，以考察螺旋参数分辨率（HPR）。径迹重建质量和 HPR 会直接影响测量的精度与准确性。

在 Belle II 实验中，每次碰撞事件后，径迹在真空中的均匀磁场内传播，大致沿一条由五个螺旋线参数描述的螺旋轨迹运动 $\{d_0, z_0, \phi_0, \omega, \tan\lambda\}$ 相对于一个枢轴点 [58]。真实螺旋线参数与重建螺旋线参数之差定义了相应螺旋参数的分辨率。径迹参数分辨率受击中数量、击中强度以及潜在事件内相关性的影响。了解背景效应如何影响 HPR，有助于深入认识探测器的整体性能及其产生的数据质量。

在本研究中，我们在所有模拟中采用相同的事件生成和径迹重建方法（在所有模拟中使用同一组信号事件，以消除信号涨落的影响），确保各次模拟中的信号击中基本相同。因此，真实的径迹信息保持一致。于是，主要差异在于背景的来源。这种明确的区分使得所发现的任何差异

都可以主要归因于背景及其生成来源的不同，从而能够直接评估 IEA-GAN 模型相较于 Geant4 的质量和性能。需要注意的是，本研究的核心目标并非隔离或减轻背景噪声的影响，而是准确模拟并测量其对径迹重建效率的影响。背景会影响径迹追踪，使其错误地分配击中。背景过程可能在探测器模拟中产生并非实际粒子轨迹组成部分的额外击中。当这些伪击中被纳入径迹重建过程时，会导致径迹参数（螺旋线参数）分辨率不理想。因此，代理模型的有效性取决于它对 Geant4 模拟背景效应的复现能力。由此，我们旨在确保 IEA-GAN 能够生成影响径迹重建过程的 PXD背景，其作用方式与 Geant4 模拟的背景过程相似。

首先，作为物理动机，我们通过打乱 Geant4样本来强调事件内相关性的影响。换言之，我们表明，在物理分析中，事件内逐传感器相关性会影响径迹参数的性能。我们通过比较螺旋参数分辨率的标准差以及双样本柯尔莫哥洛夫–斯米尔诺夫检验（KS检验）的结果来进行考察 [59] ，比较对象为打乱后的 Geant4 PXD背景和未打乱的 Geant4 PXD背景。每个分辨率的标准差 Δ p（其中 p 为螺旋参数）按如下方式计算：

$$\sigma_{\mathrm{p}} = \sqrt{\frac{\sum_{i=1}^{n}(\Delta \mathrm{p}^{i} - \overline{\Delta \mathrm{p}})^{2}}{n}}, \tag{2}$$

其中， σ_p 是螺旋参数 p 的标准差， n 是重建径迹的数量， Δp 是分辨率均值。标准差的误差（误差范围，采用置信区间法） [60, 61] 为

$$\mathrm{Error} = \frac{\sqrt{\frac{n\sigma_{\mathrm{p}}^{2}}{\chi^{2}_{1-\alpha/2,n}}} - \sqrt{\frac{n\sigma_{\mathrm{p}}^{2}}{\chi^{2}_{\alpha/2,n}}}}{2}. \tag{3}$$

其中， σ_p 是根据公式（2）计算的标准差， n 是重建径迹的数量， $\chi^2_{\alpha/2,\ n}$ 和 $\chi^2_{1-\alpha/2,\ n}$ 是自由度为 n 时卡方分布的临界值， α 是95%置信区间的显著性水平。对于5000个事件，结果表明，对于动量大于0.4吉电子伏特的高动量径迹，有充分证据表明，丢失事件内逐传感器相关性会影响分辨率，进而影响 $\mathbf{d}_0$、 ϕ_0 和 ω 螺旋参数的精度。对于 $\mathbf{z}_0$ 和 $\tan\lambda$ 参数，分辨率的标准差没有显著差异。然而，这些参数的KS检验得到的p值较低，表明两种分布的形状存在较大差异。

对于每个螺旋参数而言，对于$\mathbf{d}_0$ 冲击参数，分辨率中显著的标准差表明，相关性的损失

会直接影响我们测量粒子在横向平面内最接近原点的位置的准确程度。失去有助于正确关联径迹击中的逐传感器相关性，会导致重建值的分布更加分散，如图 5 和表 3所示。这可能会影响后续分析，例如识别主顶点和次级顶点，尤其是在粒子偏转可以忽略不计的情况下（高动量区域）。

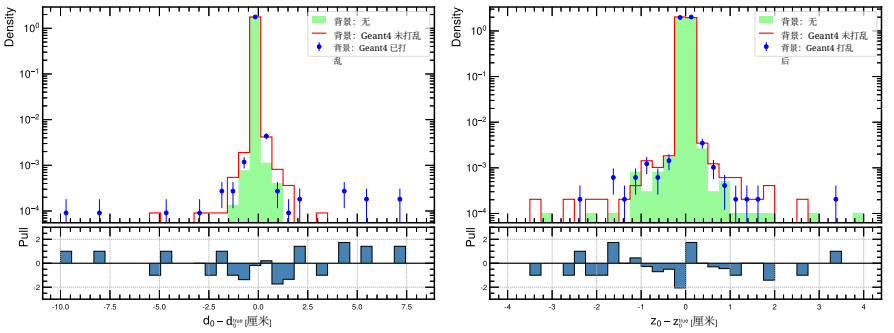


图 5：在高动量区域，存在相关（未打乱）背景和不相关（打乱）背景时， d_0 （左）和 z_0 （右）的分辨率拉偏图。拉偏值 $\pi(\Delta(p))$（其中 p 为螺旋参数）计算为 $\pi(\Delta(p)) = \frac{\Delta_p^{\text{Shuff.}} - \Delta_p^{\text{Unshuff.}}}{\sqrt{\sigma^2_{\Delta_p^{\text{Shuff.}}} + \sigma^2_{\Delta_p^{\text{Unshuff.}}}}}$。

对于 ϕ_0 参数，分辨率标准差的差异并不显著，且KS检验结果表明，缺少层间相关性不会显著影响高动量径迹方位角测量的分布和精度。打乱数据中 Δz_0 的标准差误差较高，这表明探测器层之间缺乏相关性会增加确定纵向相互作用点时的不确定性。高动量径迹在z方向上大幅偏离的可能性较低。结合不显著的KS检验结果，这表明缺乏层间相关性时，粒子轨迹在z方向上的重建方式存在根本差异。 ω 是粒子径迹曲率的度量，与粒子的动量成反比。对于高动量粒子，我们预期其曲率较小，因为高动量粒子的运动轨迹更趋于直线。 $\Delta\omega$ 的标准差在打乱数据和未打乱数据之间略有差异，但KS检验并未显示显著差异。这表明，尽管曲率分辨率的整体分布差异不显著，但曲率重建精度略有不同，这可能会影响后续依赖准确动量信息的物理分析

尽管分辨率差异并不显著，但 $\Delta\tan\ \lambda$的KS检验结果显著，这表明打乱数据和未打乱数据中高动量径迹的倾角分布存在差异。这可能意味着径迹的倾角（也与纵向动量有关）会受到层和传感器之间相关性损失的影响。

参数	标准差 ± 误差		KS统计量	p值
	打乱后的Geant4	未打乱的Geant4		
Δd_0 [厘米]	0.1343 ± 0.0007	0.0732 ± 0.0004	0.0067	0.7655
$\Delta\phi_0$ [弧度]	0.2158 ± 0.0011	0.1859 ± 0.0009	0.0066	0.7899
Δz_0 [厘米]	5.0076 ± 0.0253	4.9341 ± 0.0249	0.0152	0.0211
$\Delta\omega$ [厘米$^{-1}$]	0.0010 ± 0.0001	0.0008 ± 0.0001	0.0138	0.0485
ΔTan λ	0.0388 ± 0.0002	0.0382 ± 0.0002	0.0167	0.0086

表3：标准差（使用公式（2）和公式（3））以及打乱和未打乱的Geant4数据在5个螺旋线参数上的KS检验结果。

现在，我们针对 5000 个高动量径迹（$P_T > 0.4$吉电子伏特），比较 IEA-GAN 与 PE-GAN（图像级指标综合表现第二好的模型）在全部五个螺旋参数分辨率方面的表现，如图6 和表4所示。在低动量区域，各模型的分辨率表现相当。 $\Delta\phi$ 的尾部行为来自卷曲径迹，这类径迹的方向发生了互换。我们经过细致比较发现，IEA-GAN 模型生成的这些参数的标准差比 PE-GAN 模型更接近 Geant4 参考值，且在所有情况下均表现更优。此外，柯尔莫哥洛夫–斯米尔诺夫检验结果进一步印证了我们的发现：IEA-GAN 模型的 p值更高，因此与 Geant4 参考值更吻合。另一个有趣的观察是，与打乱后的Geant4相比，IEA-GAN 在$\mathbf{z}_0$、 ω和 Tan λ 分辨率方面的 KS 检验 p值更高，且$\mathbf{d}_0$ 重建精度更高。考察 IEA-GAN 对$\mathbf{d}_0$ 和$\mathbf{z}_0$ 的重建精度，还可以得出结论：尽管 IEA-GAN 只捕捉到了较弱的相关性，但下游物理分析中的径迹重建仍能从这种相关性中获益。

分析结果显示，IEA-GAN 与 Geant4 在尾部区段（标准差）以及分辨率精度方面具有良好的一致性，而 Geant4 与无背景之间的最大差异也出现在这些方面。因此，IEA-GAN 不仅在图像级上与 Geant4 高度一致，在径迹重建过程中也能保持恰当的重建物理行为。

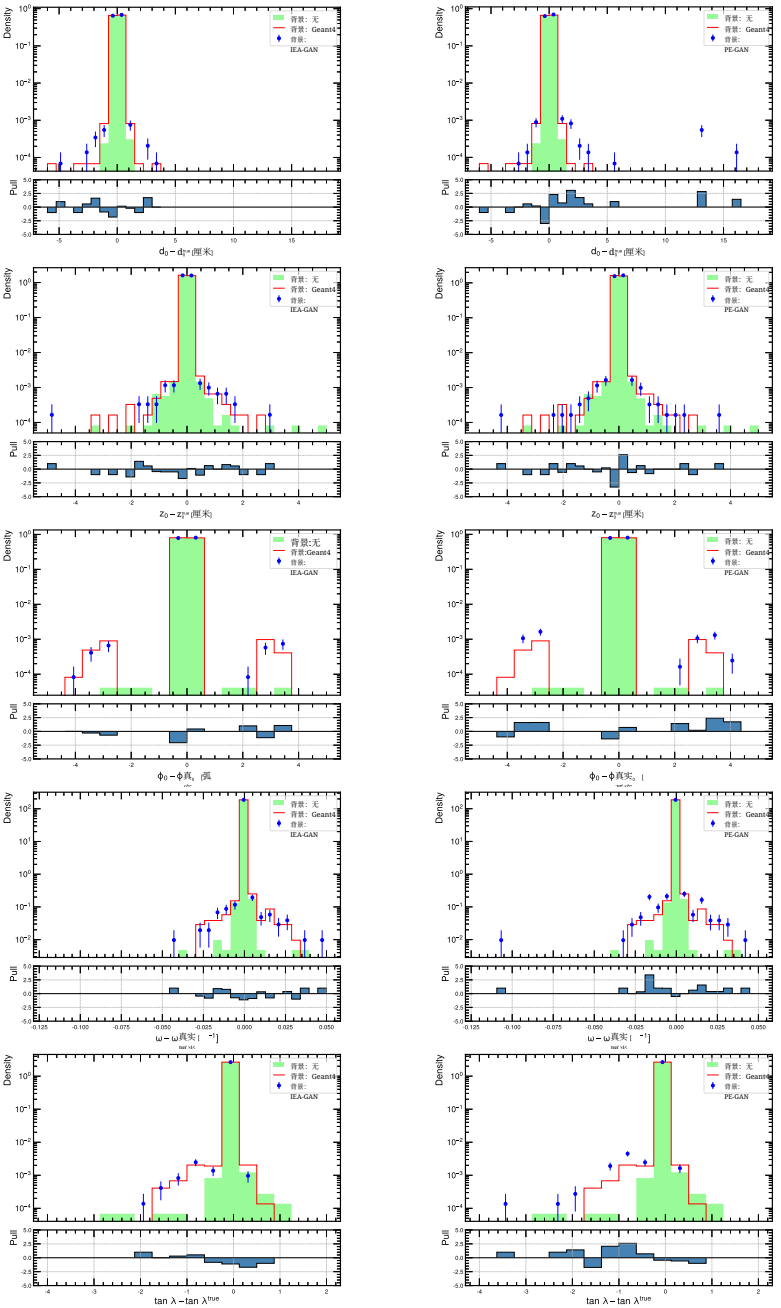


图6：比较 d_0、z_0、 ϕ_0、 ω 和 $\tan\lambda$ 的螺旋参数分辨率的拉偏图。对于每个参数，左图对应 IEA-GAN 模拟的背景，右图对应 PE-GAN 模拟的背景。

模型	参数	标准差 ± 误差		KS统计量	p值
		模型	Geant4		
PEGAN	Δd_0 [厘米]	0.1709 ± 0.0009	0.0732 ± 0.0004	0.0156	0.0164
	$\Delta \phi_0$ [弧度]	0.2207 ± 0.0011	0.1859 ± 0.0009	0.0120	0.1193
	Δz_0 [厘米]	6.9073 ± 0.0349	4.9341 ± 0.0249	0.0183	0.0029
	$\Delta \omega$ [厘米$^{-1}$]	0.0014 ± 0.0001	0.0008 ± 0.0001	0.0116	0.1425
	ΔTan λ	0.0579 ± 0.0003	0.0382 ± 0.0002	0.0179	0.0037
IEA-GAN	Δd_0 [厘米]	0.0762 ± 0.0004	0.0732 ± 0.0004	0.0104	0.2373
	$\Delta \phi_0$ [弧度]	0.1905 ± 0.0010	0.1859 ± 0.0009	0.0109	0.1939
	Δz_0 [厘米]	5.1467 ± 0.0261	4.9341 ± 0.0249	0.0073	0.6814
	$\Delta \omega$ [厘米$^{-1}$]	0.0010 ± 0.0001	0.0008 ± 0.0001	0.0103	0.2537
	Δ棕褐色 λ	0.0412 ± 0.0002	0.0382 ± 0.0002	0.0068	0.7538

表4：PE-GAN 和 IEA-GAN 模型在 5 个螺旋线参数上的标准差和 KS 检验结果与 Geant4 参考值的比较，针对高动量径迹。

3 讨论

在本研究中，我们提出了一系列稳健方法，利用事件内感知生成对抗网络（IEA-GAN）完成超高分辨率、细粒度、相关的探测器响应生成与条件采样任务。IEA-GAN 不仅能捕捉成对类别间关系，还能展现生成的探测器图像之间可解释的事件内相关性，而其他模型无法捕捉任何相关性。为此，我们提出了关系推理模块（RRM）和 IEA损失，并采用深度度量学习中的均匀性损失。RRM 为事件中的样本引入了自监督的关系上下文嵌入，这与 GAN 训练策略兼容，而这项任务极具挑战性。它根据对撞机事件中图像固有的相关性对其进行动态聚类，最终近似一个碰撞事件。我们的 IEA损失是一种由判别器监督的损失，有助于生成器就判别器所确定的每个事件中样本之间的成对关系达成共识。最后，我们证明，均匀性损失对于判别器最大化嵌入空间中信息熵的同质性至关重要，从而帮助模型克服模式崩溃，并更好地捕捉生成占有率的双峰性。

因此，与此前最先进水平相比，所有指标均有所提升：FID 分数达到 1。50，提升超过 42%；KID分数达到 0。0010，如表 1所示。使用 IEA-GAN 还可将存储需求减少两个数量级以上。此外，由于 CPU 加速显著（ $\times 147$ 如表 5 所示），现在可以将 IEA-GAN 用作在线代理模型，即时完成超高颗粒度的 PXD探测器背景模拟；此前，这项任务

对于这类高分辨率探测器模拟而言，此前无法实现。因此，IEA-GAN 作为一种能够生成超过 7。5M 个信息通道的代理模型，将成为适用于高亮度大型强子对撞机时代超高颗粒度需求的候选方案。此外，我们已证明，在探测器模拟中应用 FID 和 KID 指标，是评估深度生成模型性能的有力工具。我们还阐明了事件内逐传感器相关性在下游物理分析中的关键作用。因此，我们发现，尽管 IEA-GAN 仅捕捉到较弱的相关性，但在某些指标上仍优于 PE-GAN，甚至胜过事件间打乱（无相关性）的 Geant4。我们还深入研究了 RRM、IEA 损失和均匀性损失的最优设计及超参数。值得注意的是，许多现有的量能器模拟模型确实会考虑同时依赖多个层的可观测量。这些模型通常侧重于模拟来自单个粒子原点的粒子簇射，并仅关注簇射所在的小范围区域；这确实捕捉了局部范围内的层间相关性。相比之下，我们的方法将这一概念扩展至整个事件，考虑多个粒子原点，并覆盖整个 PXD 探测器；在其读出窗口内，不同方向和深度（各种角度和层）之间的相关性变得至关重要。鉴于 PXD 独特的拓扑结构和几何结构，这一区别至关重要，使得全探测器模拟更加全面，能够近似整个探测器范围内事件内部的复杂相互作用，而不只是局部粒子簇射。此外，尽管以往研究可能在其框架中隐式考虑了逐层相关性，但我们的研究明确评估并比较了这些相关性及其对下游物理分析的影响。

对于粒子物理实验而言，结果的物理解释在很大程度上依赖于数据潜在相关结构的捕捉能力，这一点至关重要。传感器占有率之间的真实相关性由模拟中潜在的物理过程决定。虽然实际相关性与 IEA-GAN 捕捉到的相关性有所不同，但该模型正在学习与探测器几何结构相关的模式，这些模式基于相关性在 PXD 探测器结构中的表现。具体而言，模型捕捉到了一种明显的正向逐层相关性，尤其是第一层中传感器 $0-15$ 之间以及第二层中传感器 $16-39$ 之间的相关性。这种鲜明的模式反映了模型对 PXD 环面几何结构的逐层理解，尽管它与 Geant4数据中观察到的实际相关性部分不同。这表明，内层和外层之间的占有率差异可能是模型学到的一项主要特征，并可能妨碍模型学习更细微的相关性。因此，尽管 IEA-GAN 能够为理解数据中的相关性和模式提供有价值的见解，但仍须结合领域知识来解读其结果。为了缓解

表5：IEA-GAN 和 PE-GAN 生成器在英特尔至强银牌 4108 1.80GHz 单核（中央处理器）和配备 32 GB 内存的英伟达 V100（图形处理器）上的计算性能，与 Geant4 进行比较。对于生成模型，均值和标准差是基于每组 10 000 个事件计算得出的，这意味着模型逐个生成这些事件，而不是一次批量生成 10 000 个。Geant4 的时间指的是逐个事件即时运行背景过程模拟所需的理论时间。Geant4 的存储消耗对应于存储 10000 个模拟事件（每个事件包含 1 倍 PXD 背景信息）的空间；对于代理模型（即 IEA-GAN 和 PE-GAN），存储专指模型的权重。

硬件	模拟器	时间/事件 [秒]	存储 [Mb]	加速比
CPU	Geant4	≈ 1500	≈ 2000	1
	PE-GAN	11.781 ± 0.357	≈ 47	$\approx \times 127$
	IEA-GAN	10.159 ± 0.208	≈ 47	$\approx \times \mathbf{147}$
GPU	PE-GAN	0.090 ± 0.010	≈ 47	$\approx \times 16667$
	IEA-GAN	0.070 ± 0.006	≈ 47	$\approx \times 21429$

这种差异，我们预计，直接在判别器的 RRM模块中引入扰动，将改善其上下文理解能力，从而提升事件内相关性。例如，对样本进行随机遮罩 [62]或事件间置换 [63]，并要求 RRM模块预测扰动后样本的表征，将提升模型的鲁棒性。

这项工作对高粒度、快速且高效的探测器响应和对撞机事件模拟具有重要影响。由于这些任务需要生成细粒度、事件内相关的数据，我们相信，事件内感知生成对抗网络（IEA-GAN）能够为所有粒子物理实验和模拟提供稳健且可控的采样，例如 Belle II [37]和大型强子对撞机 [67]上的探测器模拟 [11, 64] 和事件生成 [19, 20, 65, 66]。尤其是，预计高亮度大型强子对撞机（HL-LHC） [25] 的积分亮度将达到大型强子对撞机设计值的 10 倍。例如，即将投入使用、拥有约 6.5M 个通道的高粒度量能器（HGCAL），以及高亮度大型强子对撞机 [68] 上拥有约 1M 个信息通道的 ITk 三维像素探测器，都将大幅增加几何结构和精度的复杂性，导致探测器模拟 [69] 所需的时间和存储空间急剧增加。因此，我们需要更有效、更高效的高分辨率探测器模拟。IEA-GAN 有望成为模拟相应的全分辨率、高粒度探测器特征的候选方案，并具备生成超过 7.5M 个像素通道的能力。尽管如此，我们的方法为应对任何超高粒度实验的复杂性奠定了基础，但这仅仅是迈出的第一步。我们的工作旨在为此类进展铺平道路，提供一个能够初步应对超高粒度挑战的框架。

最后，IEA-GAN 也有望应用于蛋白质设计。蛋白质设计是生成全新氨基酸序列、以获得具有特定功能的蛋白质的过程，例如提高稳定性和可折叠性、获得新的结合特异性或酶活性 [70]。蛋白质可根据其二级结构特征（如 α螺旋和 β折叠）的排列方式和连接关系分为不同类别。我们开发的事件内感知方法将事件视为更高层次的特征类别，也可用于细粒度密度估计 [71] ，以生成新的可折叠蛋白质 [72–74] 结构，其中类别层面的推理至关重要。

4 方法

4.1 生成对抗网络理论

生成对抗网络是生成模型，旨在尽可能忠实地学习并生成输入分布。对于条件生成对抗网络 [75], ，目标是根据标签生成特征。双玩家生成对抗网络模型在两个合成智能体之间引入了零和博弈：生成器网络 G 和判别器网络 D。

定义 4.1 **(**原始生成对抗网络**)**。 给定生成器 G，一个函数 $\mathrm{G}:\mathbb{R}^d \to \mathbb{R}^n$，它将从某个分布中采样的潜变量 $\mathbf{z} \in \mathbb{R}^d$ 映射为一个 n 维可观测量；以及判别器 D，一个函数 $\mathrm{D}: \mathbb{R}^n \to [0,1]$，它接收生成的图像 $\mathbf{x} \in \mathbb{R}^n$ 并为其分配一个概率。二者构成以下双玩家极小极大博弈的参与者，其价值函数为$\mathrm{V}(\mathrm{D}, \mathrm{G})$ *[9],*

$$\min_{\mathrm{G}} \max_{\mathrm{D}} \; \mathrm{V}(\mathrm{D},\mathrm{G}) = \mathbb{E}_{\mathbf{x}\sim\mathbb{R}^n}[\log \mathrm{D}(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim\mathbb{R}^d}[\log(1-\mathrm{D}(\mathrm{G}(\mathbf{z}))]. \tag{4}$$

在介绍原始生成对抗网络之后，大量研究致力于提升其收敛性和稳定性，因为通常生成对抗网络训练是一项极易出现不稳定的任务。对于特定领域的任务，这需要大量调整超参数。人们引入了许多技巧、模型附加组件和结构性改动。近期一项全面的研究催生了一个极具实力的最先进水平模型 BigGAN-deep [54], ，该模型采用了对抗损失的合页损失变体 [76],

$$\mathcal{L}_{\mathrm{D}}^{\mathrm{hinge}} = -\mathbb{E}_{\mathbf{x}\sim\mathbb{R}^n}[\min(0,-1+\mathrm{D}(\mathbf{x}))] - \mathbb{E}_{\mathbf{z}\sim\mathbb{R}^d}[\min(0,-1-\mathrm{D}(\mathrm{G}(\mathbf{z}))] \;, \tag{5}$$

$$\mathcal{L}_{\mathrm{G}}^{\mathrm{hinge}} = -\mathbb{E}_{\mathbf{z}\sim\mathbb{R}^d}[\mathrm{D}(\mathrm{G}(\mathbf{z}))]. \tag{6}$$

自条件生成对抗网络基于图像标签的方案问世以来，人们提出了许多用于捕捉类别条件的方法 [75]。

其主要思路是将条件先验信息的嵌入注入生成器，然后最小化判别器类别识别输出与实际标签之间的特定度量。例如，ACGAN [77]通过引入辅助分类器来捕捉数据到类别关系。ProjGAN [30] 则通过内积将类别嵌入投影到判别器的输出上，从而对对抗损失产生影响，也试图捕捉这些数据到类别关系。近期的 ContraGAN [31]引入了源自 NT-Xent损失的 2C损失，并融入度量学习或自监督学习（SSL）的概念，以捕捉数据到数据关系或类内关系 [78], 以及基于代理的自监督学习,

$$\ell_{2C}(\mathbf{x}_i, \mathbf{y}_i) = -\ \log\left(\frac{\exp\left(\mathrm{S}_c(\mathbf{h}(\mathbf{x}_i)^\top \mathbf{e}(\mathbf{y}_i))\right) + \sum_{k=1}^{m} \mathbf{1}_{k=i}.\exp\left(\mathrm{S}_c(\mathbf{h}(\mathbf{x}_i)^\top \mathbf{h}(\mathbf{x}_k))\right)}{\exp\left(\mathrm{S}_c(\mathbf{h}(\mathbf{x}_i)^\top \mathbf{e}(\mathbf{y}_i))\right) + \sum_{k=1}^{m} \mathbf{1}_{k\neq i}.\exp\left(\mathrm{S}_c(\mathbf{h}(\mathbf{x}_i)^\top \mathbf{h}(\mathbf{x}_k))\right)}\right). \quad (7)$$

这里，$\mathbf{x}_i \in \mathbf{x}$ 表示图像，$\mathbf{y}_i \in \mathbf{y}$ 表示对应的标签，$\mathrm{S}_c(.,.)$ 是相似度度量，$\mathbf{h}(.)$ 是图像嵌入的输出，而$\mathbf{e}(.)$ 是类别嵌入的输出。尽管 ContraGAN 能够通过此损失捕捉属于同一类别的图像的类内特征，但它容易产生类别混淆 [32, 33] ，因为不同类别之间也可能表现出相似性，这是由于它们在嵌入空间中的向量表示可能并非彼此正交，而这正是我们在细粒度数据集中要处理的问题。

4.2 关系推理

Transformer [79] 广泛应用于各种场景。然而，它们在生成对抗网络中的应用，要么是在图像流形上学习像素之间的长程交互 [29, 80]，要么是通过纯视觉Transformer架构的生成对抗网络 [81] ，在这类网络中，生成器和判别器完全基于视觉Transformer构建。鉴于 Transformer 的训练众所周知地十分困难 [83] ，且在确定最佳学习率调度、预热策略、衰减设置和梯度裁剪时与具体任务无关，因此在 GAN 的学习机制中融合并调整 Transformer编码器是一项极具挑战性的任务。本文成功地将一个适配 GAN 训练方案的基于 Transformer 的模块融入判别器的图像模态和生成器的类别模态，且没有遇到上述任何问题。

定义 4.2 (注意力)。 *Transformer* 使用自注意力机制，其数据为($\mathbf{K}$、$\mathbf{Q}$、$\mathbf{V}$、A)。向量空间$\mathbf{K} \in \mathbb{R}^{N\times d_k}$、$\mathbf{Q} \in \mathbb{R}^{N\times d_k}$和$\mathbf{V} \in \mathbb{R}^{N\times d_v}$ 分别是键、查询和值的集合。双线性映射a:$\mathbf{K} \times \mathbf{Q} \to \mathbb{R}^{N\times N}$ 是键与查询之间的相似度函数。注意力A定义为

$$\mathrm{A}(\mathbf{K}, \mathbf{Q}, \mathbf{V}) := \mathrm{Softmax}(\mathrm{a}(\mathbf{K}, \mathbf{Q}))\mathbf{V}\ , \quad (8)$$

其中， d_k 和 d_v 是相应向量空间的维度。

标准 Transformer 中使用的注意力映射 [79]采用缩放点积作为键与查询之间的双线性映射，形式如下

$$\mathrm{A}(\mathbf{K},\mathbf{Q},\mathbf{V}) := \mathrm{Softmax}(\frac{\mathbf{K}\mathbf{Q}^T}{\sqrt{d_k}}))\mathbf{V} \ . \tag{9}$$

归一化因子 $\frac{1}{\sqrt{d_k}}$ 可缓解大输入下的梯度消失。多头机制并非只计算一次注意力，而是通过可学习映射$\mathbf{W}_i^k$、$\mathbf{W}_i^q$ 和$\mathbf{W}_i^v$，并行多次计算经过线性变换的键、查询和值的缩放点积注意力。随后，将h 个头各自独立产生的注意力输出进行聚合，并通过$\mathbf{W}^p$投影回所需的维度数，

$$\mathrm{MultiHead}(\mathbf{K},\mathbf{Q},\mathbf{V}) := [\bigoplus_{i=1}^{h} \mathbf{H}_i]\mathbf{W}^p \ , \tag{10}$$

其中$\mathbf{H}_i$ 由A($\mathbf{K}\mathbf{W}_i^k$， $\mathbf{Q}\mathbf{W}_i^q$， $\mathbf{V}\mathbf{W}_i^v$)给出。将自注意力机制用于处理词元序列时，Transformer能够判断序列中的其他所有词元相对于目标词元的重要性，然后利用这些权重构建每个词元的特征。

事件近似。 一个事件，即粒子碰撞后单个读出窗口内的数据，包含40幅图像，每幅图像都是尺寸为 256×768的传感器命中图（图像）。因此，每个事件代表一轮探测器特征采集。为了近似表示事件这一概念，在每次迭代中，IEA-GAN都应接收一个包含40幅传感器图像的事件。因此，我们通过传感器类型 $[[1,40]]$, 对模型进行条件化，该类型可视为角度和半径条件的组合。这些条件必须作为可学习的词元 输入模型，因为它们并非绝对，而是依赖上下文。我们无法预先定义一个事件中样本节点之间有意义的稀疏连接。例如，不同传感器图像之间的关系会因事件而异，尽管从累积角度看，它们遵循特定的分发。因此，模型必须从上下文数据中学习任何动态继承的条件（通过关系推理模块）。

为了建模每个事件中不同探测器传感器之间基于上下文的相似性，而非它们的绝对属性，我们必须使用置换等变的 [84, 85] 关系模块，以编码输入集合中各元素之间的成对对应关系。例如，最大池化（如 DeepSets [86]）和自注意力 [79] 是基于集合的问题中常见的置换等变模块。对集合中的所有词元对执行注意力计算，以识别哪些词元对最值得关注，这使得

像 BERT 这样的 Transformer [47] 能够学习特定于上下文的句法，因为多头注意力中的不同注意力头可能会关注不同的句法属性 [87, 88]。

因此，我们采用带有加权求和池化的自注意力机制作为信息路由方式，以处理输入集合中各元素之间有意义的连接，并构建事件图。事件中的每个样本都被视为全连接事件图中的一个节点，其中边表示可学习的相似度。每个事件中的样本经过我们的关系推理模块（RRM）的消息传播步骤，该模块是一个兼容生成对抗网络的全连接多头图变换器网络 [44–46]。

关系推理模块。 关系推理模块（RRM）专为兼容 GAN 训练策略而设计，能够捕获上下文化嵌入，并根据图像或事件中类别标记的内在相似度对其进行聚类。

令 $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_m\}$ 表示每个事件中采样图像的集合，其中 $\mathbf{x}_i \in \mathbb{R}^d$，且 $\mathbf{y} = \{\mathbf{y}_1, ..., \mathbf{y}_m\}$ 表示标签集合，其中 $y_i \in [\![1, 40]\!]$ 对应 40 个探测器（PXD）传感器。我们还定义了两个线性超球投影微分同胚，$\mathbf{h}_x : \mathbb{R}^k \to \mathbb{S}^n$ 和 $\mathbf{h}_y : \mathbb{Z} \to \mathbb{S}^n$，它们分别将图像嵌入流形和标签集合映射到单位n维球面。单位n维球面是点的集合 $\mathbb{S}^n = \{s \in \mathbb{R}^{n+1} \mid \|s\|_2 = 1\}$，它始终是凸且连通的。关系推理模块采用预归一化Transformer [79] 的一种变体，并配有点积多头注意力模块，使得

$$\mathbf{p}_i'^{(l)} = \mathbf{p}_i^{(l)} + \sum_{k=1}^{h} \sum_{j=1}^{m} a_{ij}^{(l,k)} \mathbf{W}_{\mathrm{SN}}^{(l)} \mathbf{LN}(\mathbf{p}_j^{(l)}) \,, \tag{11}$$

$$\mathbf{p}_i^{(L)} = \mathbf{h}_x^{\mathrm{LN}} \left(\mathbf{LN}(\bigcirc_{l=0}^{L} (\mathbf{p}_i'^{(l)} + \mathcal{F}_{\mathrm{SN}}[\mathbf{LN}(\mathbf{p}_i'^{(l)})])) \right) \,, \tag{12}$$

其中，$\mathbf{p}_i'^{(l)} \in \mathbb{R}^k$ 是判别器对 RRM 第l 层中每张图像生成的嵌入。LN 是层归一化函数 [89]，h 是公式（10）中定义的头数。$\mathcal{F}[\,.\,]$ 是一个双层多层感知机函数，定义为 $\mathcal{F}_{\mathrm{SN}}[\mathbf{p}_i^{(l)}]$ =ReLU($\mathbf{p}_i^{(l)}$ $\mathrm{W}^{(l,1)}$ SN)$\mathrm{W}^{(l,2)}$ SN，其中采用谱归一化 [90]。logits $a_{ij}^{(l,k)}$ 是双线性函数的归一化注意力权重，用于监测公式（9）中定义的第 l 层、第k 个头中图像嵌入之间的二元交互。公式（11）中的 $\mathrm{W}^{(l)}$ SN 是公式（10）中定义的第 l 层可学习的多头投影器，并采用谱归一化。通过组合所有层的 L 个函数$\bigcirc_{l=0}^{L} \Phi^l := \phi_{w_L} \circ ... \circ \phi_{w_0}[\mathbf{p}_i^{(l=0)}] \in \mathbb{R}^{m \times k}$得到的输出，会送入层归一化层，其中 $\Phi^l = \mathbf{p}_i'^{(l)} + \mathcal{F}[\mathrm{LN}(\mathbf{p}_i'^{(l)})]$。公式（12）中的 $\mathbf{h}_x^{\mathrm{LN}}(\,.\,)$ 是超球面紧化过程；在此过程中，向量经层归一化后，被标准化到单位 n 维球面 $\mathbb{S}^n$ 上。

对于判别器，该模块将图像嵌入集合用作全连接事件图中的输入节点，对它们应用点积自注意力，然后通过注意力消息传递更新每个样本或节点的嵌入，如图右侧所示 2。最后，它通过 L2 归一化 [91]将归一化图投影到超球面上，从而对信息进行紧化。将事件中的样本嵌入单位超球面，可带来多项益处。在人脸验证和人脸识别等现代机器学习任务中 [91], ，点积无处不在，而固定范数向量已知能够提升训练稳定性。在我们的场景中，这可以避免判别器发生梯度爆炸。此外，由于 S^n 与 $\mathbb{R}^n$ 的一点紧化同胚，当类别在 n 维球面这一紧凸流形上密集分布时，它们是线性可分的，而在欧几里得空间中则并非如此 [92]。

对于生成器的 RRM，考虑到稳定性问题，我们采用上述点积多头注意力模块的一个更简单版本，省略最后的超球面紧化，如图2左侧所示。它会为每个事件找到一个可学习的上下文嵌入，并通过特征混合层将其融合到每个类别标记中；该层是一个矩阵分解线性层$\mathbf{W}_{\mathrm{SN}}$(。)。形式化地，

$$\mathbf{q}_i^{(0)} = \mathbf{W}_{\mathrm{SN}}(\mathbf{r}_i \uplus \mathbf{e}_i) \ , \tag{13}$$

$$\mathbf{q}_i'^{(l)} = \mathbf{q}_i^{(l)} + \sum_{k=1}^{M}\sum_{j=1}^{m} a_{ij}^{(l,k)} \mathbf{W}^{(l)} \mathbf{LN}(\mathbf{q}_j^{(l)}) \ , \tag{14}$$

$$\mathbf{q}_i^{L} = \mathbf{LN}\left(\bigcirc_{l=0}^{L}(\mathbf{q}_i'^{(l)} + \mathcal{F}[\mathbf{LN}(\mathbf{q}_i'^{(l)})])\right) \ , \tag{15}$$

其中，$\mathbf{e}_i : \mathbb{Z} \to \mathbb{R}^t$ 是生成器的嵌入层为每个类别标记生成的嵌入。logits $a_{ij}^{(l,k)}$ 是双线性函数的归一化注意力权重，用于监测事件嵌入中类别之间的二元交互；这些嵌入定义于第 l 层和第 k 个头中，如公式 (9) 所定义。公式 (14) 中的 $\mathbf{W}^{(l)}$ 是第 l 层上可学习的多头投影器，如公式 (10) 所定义。所有层通过 L 个泛函 $\bigcirc_{l=0}^{L}\Phi^l := \phi_{w_L} \circ ... \circ \phi_{w_0}[\mathbf{q}_i^{(l=0)}] \in \mathbb{R}^{m \times t}$ 组合后的输出，会进入一个层归一化层，其中 $\Phi^l = \mathbf{q}_i'^{(l)} + \mathcal{F}[\mathbf{LN}(\mathbf{q}_i'^{(l)})]$ ，如公式 (15) 所示。

生成器的一个输入是嵌入标签，可将其视为刚性的令牌嵌入，并将其学习为每个传感器的全局表征偏置。由于每个事件的传感器条件会作为一个整体发生变化，因此仅使用类别嵌入（如条件生成对抗网络中所用的那样）[75], 是不够的，因为无法学习基于上下文的信息。因此，在每个事件中，生成器都会从每事件共享分布中采样，将其作为随机自由度（Rdof）。Rdof 是从每个类别的共享正态分布中抽取的随机样本，$\mathbf{r}_i \sim \mathcal{N}(0,\ 1)$，为生成器引入四维可学习自由度，见公式（13）。通过这种方式，我们确保

生成器能够感知事件内局部变化，最终使生成图像之间产生事件内相关性。Rdof 既可解释为对令牌嵌入的扰动 [48] ，也可解释为事件级片段嵌入 [47], ，从而增强生成图像的多样性。

4.3 事件内感知损失

受自监督学习 [93],的启发，为了以显式方式将判别器的事件内上下文知识迁移到生成器，我们为生成器引入事件内感知（IEA）损失，以捕获类别间关系，

$$\ell_{\mathrm{IEA}}(\mathbf{x}_r, \mathbf{x}_f) = \sum_{i,j} D_{\mathrm{KL}} \left(\sigma \left(\mathbf{h}(\mathbf{x}_i^{(r)})^\top \mathbf{h}(\mathbf{x}_j^{(r)}) \right) \middle\| \sigma \left(\mathbf{h}(\mathbf{x}_i^{(f)})^\top \mathbf{h}(\mathbf{x}_j^{(f)}) \right) \right), \quad (16)$$

其中，$\mathbf{x}_r = \{\mathbf{x}_i^{(r)}\}_{i=1}^m$ 表示真实图像集合，$\mathbf{x}_f = \{\mathbf{G}(\mathbf{z}^i, \mathbf{y}^i, \mathbf{r}^i) = \mathbf{x}_i^{(f)}\}_{i=1}^m$ 表示生成图像集合。softmax函数 $\sigma : \mathbb{R}^m \to [0,1]^m$对图像嵌入之间的点积自注意力进行归一化。映射$\mathbf{h} : \mathbb{R}^k \to \mathbb{S}^n$ 是判别器的单位超球面投影。因此，点积等价于余弦距离。 $D_{\mathrm{KL}}(.||.)$是Kullback-Leibler（KL）散度 [94] ，其作用于两个 $m \times m$ 取值均位于闭单位区间内的矩阵（这是由softmax函数所致）。因此，在此处使用KL散度是自然的，因为我们希望比较同一事件中的两个概率密度。我们还测试了补充说明中报告的其他距离函数。通过考虑事件中每个样本之间的线性交互 [34]并为它们的相似度赋予权重，生成器能够模仿每个事件内细粒度的类别间关系，并将这些信息融入其RRM模块，如图2所示。

在最小化该损失以训练生成器时（对判别器执行停止梯度），我们通过鼓励生成器寻找图像之间更细致的成对连接，并对细微差异保持敏感，从而对生成器的事件内感知施加判别器监督惩罚。最终，我们希望最大化真实图像和生成图像嵌入在两个单位超球面上的数据点共识。

4.4 均匀性损失

另一个关键的损失函数源自对比表征学习。在学习图像之间细粒度类别间关系的任务中，我们还希望确保特征向量具有尽可能高的超球面多样性。因此，通过对单位超球面上的特征向量施加均匀性条件，它们能够保留尽可能多的信息，因为均匀分布具有较高的熵。这一思想源自汤姆孙问题 [95], ：该问题旨在将 N 个电子以尽可能均匀的方式分布在单位球面上，使其达到势能最小的静态平衡

。为此，我们引入均匀性度量 [35]，该度量基于高斯势核，

$$\mathcal{L}_{\text{uniform}}(\mathbf{x};\ s) = \log \mathbb{E}_{\mathbf{x}_i,\mathbf{x}_j \sim p_{\text{event}}}[\exp(s\|\mathbf{h}(\mathbf{x}_i) - \mathbf{h}(\mathbf{x}_j)\|_2^2)]. \tag{17}$$

最小化判别器的这一损失时，它会尝试使那些聚类效果不佳、因而彼此不相似的样本之间保持均匀距离。换言之，我们最终希望利用黎曼 s 核 [96] ，并以 $s = -2$ 作为测量测地线相似度的指标，达到最大的测地线分离，从而在超球面上保留最多的信息。因此，从渐近角度看，这对应于超球面 [97]上的均匀分布。该损失有助于捕捉平均占用分布的精确分布，并平衡关系推理模块的类间拉力。因此，它不仅有助于生成更多样、更丰富的输出，还能防止模式崩溃或过拟合等问题。

4.5 模型细节与超参数

在本研究中，我们使用了一个包含 40 000 个蒙特卡罗模拟事件的数据集 [40]，其中 35 000 个用于训练，5000 个用于模型选择（验证），另有一组独立的 10 000 个事件作为测试集，用于评估最终模型的性能。值得注意的是，对于 $\mathcal{O}(10^7)$ 个数据通道而言，这个数据集相当小，不足以用于训练深度生成模型。每个事件的数据由 40 张灰度 256×768 零填充图像组成。为满足训练需求，这些图像在原始尺寸 250×768 的两侧进行零填充，使其尺寸可被 16 整除。

为了使用 RRM 捕捉图像之间的事件内互信息，并近似表示事件这一概念，模型在每次迭代中对整个事件进行采样（并生成）。这种方法确保我们的分析中每个事件都由一组相关的 40 张不同图像组成。所有超参数均根据模型在验证集上的稳定性和性能进行选择。生成器和判别器的学习率为5×10^{-5}，并采用每类单样本采样器。生成器的关系推理模块包含两个头，以及一层非谱归一化的消息传播层，其嵌入维度为 128，并采用 ReLU 非线性。生成器 RRM 的输入是通过谱归一化线性层，将类别令牌与 4 个随机自由度混合后得到的嵌入。

对于判别器，RRM 有四个头，每个头包含一层经过谱归一化的消息传播层，嵌入维度为 1024（即超球面维度），并采用 ReLU 非线性激活函数。所有生成器和判别器模块均使用正交初始化 [98]。对于公式（16）中的 IEA损失，系数 $\lambda_{\text{IEA}} = 1.0$（在补充说明中的补充算法 1 里突出显示）效果最佳。公式（17）定义的均匀性损失，在$\lambda_{\text{uniform}} = 0.1$时贡献最稳定。对于判别器和生成器的骨干网络，

我们使用 BigGAN-deep [54]，仅在判别器的第 32 个通道处加入非局部块。由于在生成对抗网络训练中没有有意义的方法来定义最小损失，因此我们以 FID 发散作为停止点。

5 消融研究及我们尝试过但未奏效的方法

对于 IEA损失，我们测试了多种损失函数，以实现最佳稳定性，如图 7所示。其中一些方法各有优缺点。我们研究了 Kullback-Leibler（KL）散度 [94]，L1 损失、Huber 损失 [99]，和 L2 损失。KL 散度在捕捉真实与生成自相似性之间的差异时更加稳定，并且对离群值更具鲁棒性。

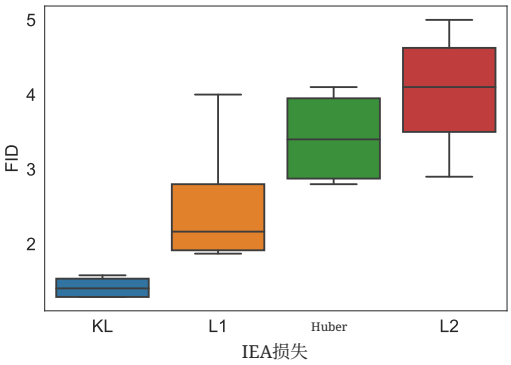


图 7：不同 IEA 损失之间 FID 的比较

我们测试了 IEA损失和均匀性损失的一系列系数。对于作为 IEA损失的 KL 散度，我们尝试了这些值 $\{0.1、1、5、10\}$ 并选定了 1。对于 L1 损失以及 IEA损失，最佳 λ_{IEA} 值为 10。对于均匀性损失，我们测试了这些值 $\{0.01、0.1、0.5、0.75、1、5、10\}$ 并选定了 0。1。此外，不使用 IEA损失和均匀性损失的 IEA-GAN，缺少对生成器的一致性最大化惩罚以及判别器的信息最大化。我们的研究表明，单独使用其中任一种损失都会导致训练不稳定、发散和保真度下降，如表 6所示。

对于超球面维度，我们探测了这些值 $\{512, 768, 1024, 2048\}$，并选定 1024。维度小于 512 时，判别器无法收敛。我们还调整了超球面投影层的位置，将其分别置于多头注意力 [79]之前和之后。超球面投影的最佳位置是在多头注意力和两层多层感知机之后。此外，对于超球面投影，我们还尝试了逆

表 6：比较 IEA-GAN、仅使用 RRM 的 IEA-GAN、仅使用均匀性损失的 IEA-GAN，以及同时使用 IEA损失的 IEA-GAN 的 FID，结果为六个随机种子的平均值。

	IEA-GAN	仅使用 RRM	RRM 与均匀性	RRM 与 IEA损失
FID	$\mathbf{1.50 \pm 0.16}$	2.74 ± 0.62	2.29 ± 0.14	3.42 ± 0.52

立体投影 $h: \mathbb{R}^N \to \mathbb{S}^N/\{p\}$ 以 n 维球面上的北极作为 p [100]，而不是采用 L2 紧化。该映射是保角的，因此局部保留了数据点之间的角度。结果更加稳定，但如表 7所示，采用 L2 紧化时的平均 FID 更好。

表 7：比较 IEA-GAN 判别器的两种不同超球面投影的 FID，结果为六个随机种子的平均值。

	L2紧化	逆立体投影
FID	$\mathbf{1.50 \pm 0.16}$	2.01 ± 0.07

在 RRM 中，我们尝试使用 GeLU [101] 非线性激活函数替代 ReLU，结果表明后者效果更好。我们还在多头注意力前后都加入了层归一化。预归一化版本似乎稳定性更好，也更能适应生成对抗网络 [75] 训练中的各种复杂情况。关于 RRM 的另一项观察涉及线性层的权重归一化。我们发现，对于判别器，谱归一化多层感知机效果最佳。对于生成器，对线性层应用谱归一化 [90] 会使训练不稳定。关于 RRM 对生成器标签嵌入的影响，我们观察到，若不采用 RRM，判别器中的 RRM 也会变得不稳定，训练会很早就发散。

对于随机自由度（Rdof），我们首先使用输入生成器的随机向量，并在其上应用 RRM。然而，FID [49] 未能降至 20 以下，且不存在相关性。因此，我们引入了用于事件生成的独立随机采样，并对维度 $\{2, 4, 8, 16, 32\}$进行了探测。4 个自由度是最优选择。我们观察到，随着 Rdof 维度增加，生成图像之间的事件内相关性会逐渐消失。我们还测试了用于事件 Rdof 采样的均匀分布，但未能得到稳定的结果。我们测试了多种将 Rdof 融合到类别嵌入中的方法，例如可学习神经网络层（矩阵分解）、拼接、求和，以及使用带有非线性部分的多层感知机，但最终选择了可学习神经网络层（矩阵分解）作为特征混合层。

我们还考察了 G 和 D 使用不同学习率组合的情况。采用 TTUR 会导致严重的模式崩溃。因此，我们使用

G 和 D 使用相同的学习率。我们对 $\{1 \times 10^{-5}, 2.5 \times 10^{-5}$、$5 \times 10^{-5}$、$7.5 \times 10^{-5}$、$1 \times 10^{-4}\}$ 进行了测试，并选择了 5×10^{-5}。对于骨干模型，BigGAN-deep 的浅层版本 [54],BigGAN [54], 导致了模式崩溃；因此，我们选择了 BigGAN-deep。

6 补充图和补充表

以下图 8在引言中被引用，算法1 在主稿的方法部分中突出显示。

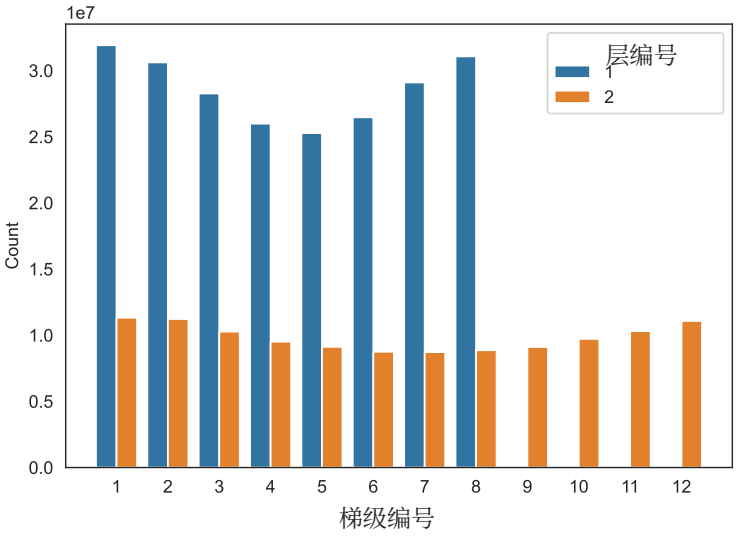


图 8：每层和传感器（梯级）的 PXD 击中数。内层和外层传感器的占有率各不相同。它们之间的全局不对称性源于 PXD 图像对 ϕ 的依赖。

算法 1 事件内感知生成对抗网络

要求： 生成器和判别器参数 θ_G, θ_D，事件内感知系数 λ_{IEA}，均匀性系数 λ_{uniform} 和超参数 s，Adam 超参数 α, β_1, β_2，事件规模 M，每次生成器迭代的判别器迭代步数 N_D

对于 训练迭代次数 **执行**
  对于 $t = 1, \ldots, N_D$ **执行**
    采样 $\{\mathbf{z}^i\}_{i=1}^{M} \sim \mathrm{p}(\mathbf{z})$,
    $\{\mathbf{x}^i, \mathbf{y}^i\}_{i=1}^{M} \sim \mathrm{p}_{\text{event}}(\mathbf{x}, \mathbf{y})$, $\{\mathbf{r}^i\}_{i=1}^{M} \sim \mathrm{p}_{\text{Rdof}}(\mathbf{z})$ ▷ 事件采样。
    对于 $i = 1, \ldots, M$ **执行**
      $\ell_{\text{Dhinge}}^{(i)} \leftarrow \ell_{\text{D}_{\text{铰链}}}(\mathbf{x}^{(i)}; \mathrm{G}(\mathbf{z}^i, \mathbf{y}^i, \mathbf{r}^i))$
    结束循环
    $\mathcal{L}_{\text{D}_{\text{铰链}}} \leftarrow \frac{1}{M}\sum_{i=1}^{M} \ell_{\text{Dhinge}}^{(i)}$
    $\mathcal{L}_{\text{uniform}} \leftarrow \mathcal{L}_{\text{uniform}}(\mathbf{x}; s)$ ▷ 均匀性损失。
    $\mathcal{L}_{2C}^{\text{real}} \leftarrow \frac{1}{M}\sum_{i=1}^{M} \ell_{2C}(\mathbf{x}^i, \mathbf{y}^i)$
    $\theta_D \leftarrow Adam(\mathcal{L}_{\text{D}_{\text{hinge}}} + \lambda_{2C}\mathcal{L}_{2C}^{\text{real}} + \lambda_{\text{uniform}}\mathcal{L}_{\text{uniform}}, \alpha, \beta_1, \beta_2)$
  结束循环
  采样 $\{\mathbf{z}^i\}_{i=1}^{M} \sim \mathrm{p}(\mathbf{z})$,
  采样 $\{\mathbf{r}^i\}_{i=1}^{M} \sim \mathrm{p}_{\text{Rdof}}(\mathbf{z})$ ▷ 事件采样。
  对于 $i = 1, \ldots, M$ **执行**
    $\ell_{\text{Ghinge}}^{(i)} \leftarrow \ell_{G_{\text{铰链}}}(\mathrm{G}(\mathbf{z}^i, \mathbf{y}^i, \mathbf{r}^i))$
  结束循环
  $\mathcal{L}_{\text{G}_{\text{铰链}}} \leftarrow \frac{1}{M}\sum_{i=1}^{M} \ell_{\text{Ghinge}}^{(i)}$
  $\mathcal{L}_{\text{IEA}} \leftarrow \frac{1}{M}\sum_{i=1}^{M} \ell_{\text{IEA}}(\mathrm{G}(\mathbf{z}^i, \mathbf{y}^i, \mathbf{r}^i), \mathbf{x}^i)$ ▷ 事件内感知损失。
  $\mathcal{L}_{2C}^{\text{fake}} \leftarrow \frac{1}{M}\sum_{i=1}^{M} \ell_{2C}(\mathrm{G}(\mathbf{z}^i, \mathbf{y}^i, \mathbf{r}^i), \mathbf{y}^i)$
  $\theta_G \leftarrow \mathrm{Adam}(\mathcal{L}_{\text{G}_{\text{hinge}}} + \lambda_{2C}\mathcal{L}_{2C}^{\text{fake}} + \lambda_{\text{IEA}}\mathcal{L}_{\text{IEA}}, \alpha, \beta_1, \beta_2)$
结束循环

声明

数据可用性

本研究所用数据可在 Zenodo [102]。

代码可用性

本研究的代码可在 GitHub [103]获取。

参考文献

[1] Paganini, M., de Oliveira, L., Nachman, B.：利用生成对抗网络加速科学研究：在多层量能器中的三维粒子簇射上的应用。《物理评论快报》**120**卷（4），042003（2018 年）。https://doi.org/10.1103/PhysRevLett.120.042003 2

[2] Vallecorsa, S.：用于快速模拟的生成模型。《物理学杂志：会议系列》**1085**(2), 022005（2018 年）。https://doi.org/10.1088/1742-6596/1085/2/022005 2

[3] Paganini, M., de Oliveira, L., Nachman, B.：CaloGAN：利用生成对抗网络，在多层电磁量能器中模拟三维高能粒子簇射。《物理评论D辑：粒子与场》**97**(1),014021（2018 年）。https://doi.org/10.1103/PhysRevD.97.014021

[4]de Oliveira, L., Paganini, M., Nachman, B.：在生成对抗网络加速的电磁量能器模拟中控制物理属性。《物理学杂志：会议系列》**1085**(4), 042017（2018 年）。https://doi.org/10.1088/1742-6596/1085/4/042017 2

[5] Erdmann, M., Geiger, L., Glombitza, J., Schmidt, D.：在对抗网络中使用 Wasserstein距离生成和改进粒子探测器模拟。《大科学计算与软件》**2**(1), 4 (2018 年)。https://doi.org/10.1007/s41781-018-0008-x

[6] Srebre, M., Schmolz, P., Hashemi, B., Ritter, M., Kuhr, T.：使用生成对抗网络生成 Belle II 像素探测器背景数据。EPJ网络会议**245**, 02010 (2020年)。https://doi.org/10.1051/epjconf/202024502010 2,9

[7] Hashemi, B., Hartmann, N., Kuhr, T., Ritter, M., Srebre, M.：在 Belle II 使用生成对抗网络生成像素探测器背景。EPJ网络会议**251**, 03031 (2021年)。https://doi.org/10.1051/epjconf/202125103031 2,9

[8]Buhmann, E., Diefenbacher, S., Eren, E., Gaede, F., Kasieczka, G.,Korol, A., Kr¨uger, K：GettingHigh：采用高速高保真模拟高颗粒度量能器。《大科学计算与软件》**5**(1), 13 (2021)。https://doi.org/10.1007/s41781-021-00056-0 2[9]Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde- Farley, D., Ozair, S., Courville, A., Bengio, Y.：生成对抗网络。载于：《神经信息处理系统大会论文集》，第27卷。Curran Associates, Inc.，线上（2014）。

https://proceedings.neurips.cc/paper
files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html
访问日期：2024-04-012，6，20

[10] Belayneh, D., Carminati, F., Farbin, A., Hooberman, B., Khattak, G., Liu, M., Liu, J., Olivito, D., Pacela, V.B., Pierini, M., Schwing, A., Spiropulu, M., Vallecorsa, S., Vlimant, J.-R., Wei, W., Zhang, M.: 深度学习量热测量：用于对撞机物理的粒子模拟与重建。《European Physical Journal C》**80**(7), 688（2020年）。https://doi.org/10.1140/epjc/s10052-020-8251-9 2

[11] Khattak, G.R., Vallecorsa, S., Carminati, F., Khan, G.M.：利用生成对抗网络快速模拟高粒度量能器。《欧洲物理学杂志C》**82**(4), 386 (2022). https://doi.org/10.1140/epjc/s10052-022-10258-4 2，19[12]Krause, C., Pang, I., Shih, D.：面向CaloChallenge数据集1的CaloFlow。arXiv。说明：32页，18幅图，v2：更新了π介子评估（2023）。https://doi.org/10.48550/arXiv.2210.14245 2[13] Buhmann, E.,Diefenbacher, S., Eren, E., Gaede, F., Hundhausen, D.,Kasieczka, G.,Korcari,W., Korol,A.,Kr¨uger, K.,McKeown,P.,Rustige, L.：生成模型生成的快速、精确电磁簇射和强子簇射。EPJ网络会议**251**, 03049 (2021). https://doi.org/10.1051/epjconf/202125103049[14] Mikuni, V., Nachman, B.：用于量能器簇射模拟的基于分数的生成模型。《物理评论D辑》**106**(9), 092009 (2022). https://doi.org/10.1103/PhysRevD.106.092009[15] Krause, C., Shih, D.：CaloFlow II：利用归一化流实现更快速且同样精确的量能器簇射生成。arXiv。说明：24页，15幅图，4张表；v2：与录用版本一致（2023）。https://doi.org/10.48550/arXiv.2110.11377 2[16] Hashemi, B., Amin, N., Datta, K., Olivito, D., Pierini, M.：利用生成对抗网络构建大型强子对撞机分析专用数据集。arXiv。说明：14页，11幅图（2019）。https://doi.org/10.48550/arXiv.1901.05282 2[17] Di Sipio, R., Giannelli, M.F., Haghighat, S.K., Palazzo, S.：DijetGAN：一种用于模拟大型强子对撞机上量子色动力学双喷注事例的生成对抗网络方法。《高能物理学杂志》**2019**(8), 110(2019). https://doi.org/10.1007/JHEP08(2019)110[18] Mart´ınez, J.A., Nguyen, T.Q., Pierini, M., Spiropulu, M., Vlimant, J.- R.：用于大型强子对撞机全事例模拟的粒子生成对抗网络

及其在堆积描述中的应用。《物理学杂志：会议系列》**1525**(1)，012081(2020)。https://doi.org/10.1088/1742-6596/1525/1/012081

[19] Alanazi, Y., Sato, N., Ambrozewicz, P., Hiller-Blin, A., Melnitchouk, W., Battaglieri, M., Liu, T., Li, Y.：机器学习物理事件生成综述。载于：第二十九届国际人工智能联合会议，第 5 卷，第 4286–4293 页（2021年）。https://doi.org/10.24963/ijcai.2021/588 19

[20] Butter, A., Plehn, T., Winterhalder, R.：如何生成大型强子对撞机事件。Sci- Post Physics **7**(6)，075（2019 年）。https://doi.org/10.21468/SciPostPhys.7.6.075 19

[21] Otten, S., Caron, S., de Swart, W., van Beekveld, M., Hendriks, L., van Leeuwen, C., Podareanu, D., Ruizde Austri, R., Verheyen, R.：基于深度生成模型和密度信息缓冲区的物理事件生成与统计抽样。《自然·通讯》**12**(1)，2985（2021年）。https://doi.org/10.1038/s41467-021-22616-z 2

[22] Arjovsky, M., Chintala, S., Bottou, L.：Wasserstein 生成对抗网络。载于：第34届国际机器学习会议论文集，第214–223页。PMLR，线上（2017年）。https://proceedings.mlr.press/v70/arjovsky17a.html 访问日期：2024-04-01 2

[23] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.：改进Wasserstein生成对抗网络的训练。载于：第31届国际神经信息处理系统会议论文集。NIPS'17，第5769–5779页。Curran Associates Inc.，美国纽约州雷德胡克（2017年）2

[24]Rezende, D., Mohamed, S.：使用归一化流进行变分推断。载于：第32届国际机器学习会议论文集，第1530–1538页。PMLR，线上（2015年）。https://proceedings.mlr.press/v37/rezende15.html 访问日期：2024-04-01 2

[25] Aberle, O., B´ejar Alonso, I., Br¨uning, O., Fessia, P., Rossi, L., Tavian, L., Zerlauth, M., Adorisio, C., Adraktas, A., Ady, M., Albertone, J., Alberty, L., Alcaide Leon, M., Alekou, A., Alesini, D., Ferreira, B.A., Lopez, P.A., Ambrosio, G., Andreu Munoz, P., Anerella, M., Angal- Kalinin, D., Antoniou, F., Apollinari, G., Apollonio, A., Appleby, R., Arduini, G., Alonso, B.A., Artoos, K., Atieh, S., Auchmann, B., Badin, V., Baer, T., Baffari, D., Baglin, V., Bajko, M., Ball, A., Ballarino, A., Bally, S., Bampton, T., Banfi, D., Barlow, R., Barnes, M., Barranco, J., Barthelemy, L., Bartmann, W., Bartosik, H., Barzi, E., Battistin, M., Baudrenghien, P., Alonso, I.B., Belomestnykh, S., Benoit, A., Ben- Zvi, I., Bertarelli,A., Bertolasi,S.,Bertone, C.,Bertran, B.,Bestmann,

P., Biancacci, N., Bignami, A., Bliss, N., Boccard, C., Body, Y., Bor- burgh, J., Bordini, B., Borralho, F., Bossert, R., Bottura, L., Boucherie, A., Bozzi, R., Bracco, C., Bravin, E., Bregliozzi, G., Brett, D., Broche, A., Brodzinski, K., Broggi, F., Bruce, R., Brugger, M., Br¨uning, O., Buffat, X., Burkhardt, H., Burnet, J., Burov, A., Burt, G., Cabezas, R., Cai, Y., Calaga, R., Calatroni, S., Capatina, O., Capelli, T., Car- don, P., Carlier, E., Carra, F., Carvalho, A., Carver, L.R., Caspers, F., Cattenoz, G., Cerutti, F., Chanc´e, A., Rodrigues, M.C., Chemli, S., Cheng, D., Chiggiato, P., Chlachidze, G., Claudet, S., Coello De Por- tugal, JM., Collazos, C., Corso, J., Costa Machado, S., Costa Pinto, P., Coulinge, E., Crouch, M., Cruikshank, P., Cruz Alaniz, E., Czech, M., Dahlerup-Petersen, K., Dalena, B., Daniluk, G., Danzeca, S., Day, H., De Carvalho Saraiva, J., De Luca, D., De Maria, R., De Rijk, G., De Silva, S., Dehning, B., Delayen, J., Deliege, Q., Delille, B., Delsaux, F., Denz, R., Devred, A., Dexter, A., Di Girolamo, B., Dietderich, D., Dilly, J.W., Doherty, A., Dos Santos, N., Drago, A., D.Drskovic, Ramos, D.D., Ducimeti`ere, L., Efthymiopoulos, I., Einsweiler, K., Esposito, L., Esteban Muller, J., Evrard, S., Fabbricatore, P., Farinon, S., Fartoukh, S., Faus-Golfe, A., Favre, G., Felice, H., Feral, B., Ferlin, G., Ferracin, P., Ferrari, A., Ferreira, L., Fessia, P., Ficcadenti, L., Fiotakis, S., Fis- carelli, L., Fitterer, M., Fleiter, J., Foffano, G., Fol, E., Folch, R., Foraz, K., Foussat, A., Frankl, M., Frasciello, O., Fraser, M., Menendez, P.F., Fuchs, J.-F., Furuseth, S., Gaddi, A., Gallilee, M., Gallo, A., Alia, R.G., Gavela, H.G., Matos, J.G., Garcia Morales, H., Valdivieso, A.G.-T., Garino, C., Garion, C., Gascon, J., Gasnier, Ch., Gentini, L., Gentsos, C., Ghosh, A., Giacomel, L., Hernandez, K.G., Gibson, S., Ginburg, C., Giordano, F., Giovannozzi, M., Goddard, B., Gomes, P., Gonzalez De La Aleja Cabana, M., Goudket, P., Gousiou, E., Gradassi, P., Costa, A.G., Grand-Cl´ement, L., Grillot, S., Guillaume, JC., Guinchard, M., Hagen, P., Hakulinen, T., Hall, B., Hansen, J., Heredia Garcia, N., Herr, W., Herty, A., Hill, C., Hofer, M., H¨ofle, W., Holzer, B., Hopkins, S., Hrivnak, J., Iadarola, G., Infantino, A., Bermudez, S.I., Jakobsen, S., Jebramcik, M.A., Jenninger, B., Jensen, E., Jones, M., Jones, R., Jones, T., Jowett, J., Juchno, M., Julie, C., Junginger, T., Kain, V., Kaltchev, D., Karastathis, N., Kardasopoulos, P., Karppinen, M., Keintzel, J.,Ker- sevan, R., Killing, F., Kirby, G., Korostelev, M., Kos, N., Kostoglou, S., Kozsar, I., Krasnov, A., Krave, S., Krzempek, L., Kuder, N., Kur- tulus, A., Kwee-Hinzmann, R., Lackner, F., Lamont, M., Lamure, A.L., m, L.L., Lazzaroni, M., Le Garrec, M., Lechner, A., Lefevre, T., Leuxe, R., Li, K., Li, Z., Lindner, R., Lindstrom, B., Lingwood, C., L¨offler, C., Lopez, C., Lopez-Hernandez, LA., Losito, R., Maciariello, F., Mac- intosh, P., Maclean, E.H., Macpherson, A., Maesen, P., Magnier, C., Durand, H.M., Malina, L., Manfredi, M., Marcellini, F., Marchevsky, M., Maridor, S., Marinaro, G., Marinov, K., Markiewicz, T., Marsili, A., Martinez Urioz, P., Martino, M., Masi, A., Mastoridis, T., Mattelaer, P.,

May, A., Mazet, J., Mcilwraith, S., McIntosh, E., Medina Medrano, L., Mejica Rodriguez, A., Mendes, M., Menendez, P., Mensi, M., Mereghetti, A., Mergelkuhl, D., Mertens, T., Mether, L., M´etral, E., Migliorati, M., Milanese, A., Minginette, P., Missiaen, D., Mitsuhashi, T., Modena, M., Mokhov, N., Molson, J., Monneret, E., Montesinos, E., Moron-Ballester, R., Morrone, M., Mostacci, A., Mounet, N., Moyret, P., Muffat, P., Muratori, B., Muttoni, Y., Nakamoto, T., Navarro-Tapia, M., Neupert, H., Nevay, L., Nicol, T., Nilsson, E., Ninin, P., Nobrega, A., Noels, C., Nolan, E., Nosochkov, Y., Nuiry, FX., Oberli, L., Ogitsu, T., Ohmi, K., R, O., Oliveira, J., Orlandi, Ph., Ortega, P., Osborne, J., Otto, T., Palumbo, L., Papadopoulou, S., Papaphilippou, Y., Paraschou, K., Par- ente, C., Paret, S., Park, H., Parma, V., Pasquino, Ch., Patapenka, A., Patnaik, L., Pattalwar, S., Payet, J., Pechaud, G., Pellegrini, D., Pepin- ster, P., Perez, J., Espinos, J.P., Marcone, A.P., Perin, A., Perini, P., Persson, T.H.B., Peterson, T., Pieloni, T., Pigny, G., Pinheiro de Sousa, J.P., Pirotte, O., Plassard, F., Pojer, M., Pontercorvo, L., Poyet, A., Prelipcean, D., Prin, H., Principe, R., Pugnat, T., Qiang, J., Quaranta, E., Rafique, H., Rakhno, I., Duarte, D.R., Ratti, A., Ravaioli, E., Ray- mond, M., Redaelli, S., Renaglia, T., Ricci, D., Riddone, G., Rifflet, J., Rigutto, E., Rijoff, T., Rinaldesi, R., Riu Martinez, O., Rivkin, L., Rodriguez Mateos, F., Roesler, S., Romera Ramirez, I., Rossi, A., Rossi, L., Rude, V., Rumolo, G., Rutkovksi, J., Sabate Gilarte, M., Sabbi, G., Sahner, T., Salemme, R., Salvant, B., Galan, F.S., Santamaria Garcia, A., Santillana, I., Santini, C., Santos, O., Diaz, P.S., Sasaki, K., Savary, F., Sbrizzi, A., Schaumann, M., Scheuerlein, C., Schmalzle, J., Schmick- ler, H., Schmidt, R., Schoerling, D., Segreti, M., Serluca, M., Serrano, J., Sestak, J., Shaposhnikova, E., Shatilov, D., Siemko, A., Sisti, M., Sitko, M., Skarita, J., Skordis, E., Skoufaris, K., Skripka, G., Smekens, D., Sobiech, Z., Sosin, M., Sorbio, M., Soubelet, F., Spataro, B., Spiezia, G., Stancari, G., Staterao, M., Steckert, J., Steele, G., Sterbini, G., Struik, M., Sugano, M., Szeberenyi, A., Taborelli, M., Tambasco, C., Rego, R.T., Tavian, L., Teissandier, B., Templeton, N., Therasse, M., Thiesen, H., Thomas, E., Toader, A., Todesco, E., Tom´as, R., Toral, F., Torres-Sanchez, R., Trad, G., Triantafyllou, N., Tropin, I., Tsinga- nis, A., Tuckamantel, J., Uythoven, J., Valishev, A., Van Der Veken, F., Van Weelderen, R., Vande Craen, A., Vazquez De Prada, B., Velotti, F., Verdu Andres, S., Verweij, A., Shetty, N.V., Vlachoudis, V., Volpini, G., Wagner, U., Wanderer, P., Wang, M., Wang, X., Wanzenberg, R., Wegscheider,A., Weisz, S.,Welsch, C., Wendt, M., Wenninger,J., Weter- ings, W., White, S., Widuch, K., Will, A., Willering, G., Wollmann, D., Wolski, A., Wozniak, J., Wu, Q., Xiao, B., Xiao, L., Xu, Q., Yakovlev, Y., Yammine, S., Yang, Y., Yu, M., Zacharov, I., Zagorodnova, O., Zan- nini, C., Zanoni, C., Zerlauth, M., Zimmermann, F., Zlobin, A., Zobov, M., Zurbano Fernandez, I.: 高 l uminosity 大型强子对撞机 (HL- LHC)：技术设计报告。技术报告，欧洲核子研究中心，日内瓦 (2020)。

https://doi.org/10.23731/CYRM-2020-0010 2, 5, 19

[26] CMS 端盖量能器第二阶段升级。欧洲核子研究中心，日内瓦 (2017年)。https://doi.org/10.17181/CERN.IV8M.1JY2 2[27] Deselaers, T., Ferrari, V.：ImageNet 中的视觉与语义相似性。载于：CVPR 2011，第 1777–1784 页 (2011)。https://doi.org/10.1109/CVPR. 2011.5995474 3[28] Wei, X.-S., Song, Y.-Z., Aodha, O.M., Wu, J., Peng, Y., Tang, J., Yang, J., Belongie, S.：深度学习的细粒度图像分析：综述。《IEEE 模式分析与机器智能汇刊》44(12)，8927–8948 (2022 年)。https://doi.org/10.1109/TPAMI.2021. 3126648 3[29] Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.：自注意力生成对抗网络。载于：第 36 届国际机器学习会议论文集，第 7354–7363 页。PMLR，线上 (2019)。https://proceedings.mlr.press/v97/zhang19d.html 访问日期：2024-04-01 3, 21[30] Miyato, T., Koyama, M.：带投影判别器的条件生成对抗网络。arXiv。说明：作为会议论文发表于 ICLR 2018 (2018)。https: // doi.org/10.48550/arXiv.1802.05637 3, 21[31] Kang, M., Park, J.：ContraGAN：用于条件图像生成的对比学习。载于：Larochelle, H., Ranzato, M., Hadsell, R., Bal- can, M.F., Lin, H. (编)《神经信息处理系统进展》，第 33 卷，第 21357–21369 页。Curran Associates, Inc.，线上 (2020)。https://proceedings.neurips.cc/paper files/paper/2020/file/f490c742cd8318b8ee6dca10af2a163f-Paper.pdf 3, 6, 9, 21[32] Kang, M., Shim, W., Cho, M., Park, J.：重启 ACGAN：采用稳定训练的辅助分类器生成对抗网络。载于：《神经信息处理系统进展》，第 34 卷，第 23505–23518 页。Curran Associates, Inc.，线上 (2021年)。https://proceedings.neurips.cc/paper files/paper/2021/hash/c5ab6cebaca97f7171139e4d414ff5a6-Abstract.html 访问日期：2024-04-01 3, 7, 21[33] Rangwani, H., Mopuri, K.R., Babu, R.V.：分类器参与训练的类别平衡生成对抗网络。载于：第三十七届人工智能不确定性会议论文集，第 1618–1627 页。PMLR，线上 (2021年)。https://proceedings.mlr.press/v161/rangwani21a.html 访问日期：2024-04-01 3, 7, 21[34] Cao, L.：复杂交互的耦合学习。《信息处理与管理》51(2)，167–186 (2015)。https://doi.org/10.1016/

j.ipm.2014.08.007 3, 25

[35] Wang, T., Isola, P.：通过超球面上的对齐与均匀性理解对比表征学习。载于：第37届国际机器学习会议论文集，第9929–9939页。PMLR，线上（2020年）。https://proceedings.mlr.press/v119/wang20k.html 访问日期：2024-04-01 3，6，26

[36] Mueller, F.：Belle II 像素顶点探测器（PXD）的一些方面。Journal of Instrumentation9(10)，10007（2014）。
https://doi.org/10.1088/1748-0221/9/10/C10007 3，4

[37] Abe, T., Adachi, I., Adamczyk, K., Ahn, S., Aihara, H., Akai, K., Aloi, M., Andricek, L., Aoki, K., Arai, Y., Arefiev, A., Arinstein, K., Arita, Y., Asner, D.M., Aulchenko, V., Aushev, T., Aziz, T., Bakich, A.M., Bal- agura, V., Ban, Y., Barberio, E., Barvich, T., Belous, K., Bergauer, T., Bhardwaj, V., Bhuyan, B., Blyth, S., Bondar, A., Bonvicini, G., Bozek, A., Bracko, M., Brodzicka, J., Brovchenko, O., Browder, T.E., Cao, G., Chang, M.-C., Chang, P., Chao, Y., Chekelian, V., Chen, A., Chen, K.- F., Chen, P., Cheon, B.G., Chiang, C.-C., Chistov, R., Cho, K., Choi, S.-K., Chung, K., Comerma, A., Cooney, M., Cowley, D.E., Critchlow, T., Dalseno, J., Danilov, M., Dieguez, A., Dierlamm, A., Dillon, M., Dingfelder, J., Dolenec, R., Dolezal, Z., Drasal, Z., Drutskoy, A., Dun- gel, W., Dutta, D., Eidelman, S., Enomoto, A., Epifanov, D., Esen, S., Fast, J.E., Feindt, M., Garcia, M.F., Fifield, T., Fischer, P., Flanagan, J., Fourletov, S., Fourletova, J., Freixas, L., Frey, A., Friedl, M., Fruehwirth, R., Fujii, H., Fujikawa, M., Fukuma, Y., Funakoshi, Y., Furukawa, K., Fuster, J., Gabyshev, N., Cueto, A.G.d.V., Garmash, A., Garrido, L., Geisler, C., Gfall, I., Goh, Y.M., Golob, B., Gorton, I., Grzymkowski, R., Guo, H., Ha, H., Haba, J., Hara, K., Hara, T., Haruyama, T., Hayasaka, K., Hayashi, K., Hayashii, H., Heck, M., Heindl, S., Heller, C., Hemperek, T., Higuchi, T., Horii, Y., Hou, W.-S., Hsiung, Y.B., Huang, C.-H., Hwang, S., Hyun, H.J., Igarashi, Y., Iglesias, C., Iida, Y., Iijima, T., Imamura, M., Inami, K., Irmler, C., Ishizuka, M., Itagaki, K., Itoh, R., Iwabuchi, M., Iwai, G., Iwai, M., Iwasaki, M., Iwasaki, M., Iwasaki, Y., Iwashita, T., Iwata, S., Jang, H., Ji, X., Jinno, T., Jones, M., Julius, T., Kageyama, T., Kah, D.H., Kakuno, H., Kamitani, T., Kanazawa, K., Kapusta, P., Kataoka, S.U., Katayama, N., Kawai, M., Kawai, Y., Kawasaki,T., Kennedy,J.,Kichimi, H.,Kikuchi, M.,Kiesling, C., Kim, B.K., Kim, G.N., Kim, H.J., Kim, H.O., Kim, J.-B., Kim, J.H., Kim, M.J., Kim, S.K., Kim, K.T., Kim, T.Y., Kinoshita, K., Kishi, K., Kisielewski, B., van Dam, K.K., Knopf, J., Ko, B.R., Koch, M., Kodys, P., Koffmane, C., Koga, Y., Kohriki, T., Koike, S., Koiso, H., Kondo, Y., Korpar, S., Kouzes, R.T., Kreidl, C., Kreps, M., Krizan, P., Krokovny, P., Krueger, H., Kruth, A., Kuhn, W., Kuhr, T., Kumar, R., Kumita, T., Kupper, S., Kuzmin, A., Kvasnicka, P., Kwon, Y.-J., Lacasta, C.,

Lange, J.S., Lee, I.-S., Lee, M.J., Lee, M.W., Lee, S.-H., Lemarenko, M., Li, J., Li, W.D., Li, Y., Libby, J., Limosani, A., Liu, C., Liu, H., Liu, Y., Liu, Z., Liventsev, D., Virto, A.L., Makida, Y., Mao, Z.P., Mari- nas, C., Masuzawa, M., Matvienko, D., Mitaroff, W., Miyabayashi, K., Miyata, H., Miyazaki, Y., Miyoshi, T., Mizuk, R., Mohanty, G.B., Moha- patra, D., Moll, A., Mori, T., Morita, A., Morita, Y., Moser, H.-G., Martin, D.M., Mueller, T., Muenchow, D., Murakami, J., Myung, S.S., Nagamine, T., Nakamura, I., Nakamura, T.T., Nakano, E., Nakano, H., Nakao, M., Nakazawa, H., Nam, S.-H., Natkaniec, Z., Nedelkovska, E., Negishi, K., Neubauer, S., Ng, C., Ninkovic, J., Nishida, S., Nishimura, K., Novikov, E.,Nozaki, T.,Ogawa, S., Ohmi, K., Ohnishi, Y.,Ohshima, T., Ohuchi, N., Oide, K., Olsen, S.L., Ono, M., Ono, Y., Onuki, Y., Ostrowicz, W., Ozaki, H., Pakhlov, P., Pakhlova, G., Palka, H., Park, H., Park, H.K., Peak, L.S., Peng, T., Peric, I., Pernicka, M., Pestot- nik, R., Petric, M., Piilonen, L.E., Poluektov, A., Prim, M., Prothmann, K., Regimbal, K., Reisert, B., Richter, R.H., Riera-Babures, J., Ritter, A., Ritter, A., Ritter, M., Roehrken, M., Rorie, J., Rosen, M., Rozan- ska, M., Ruckman, L., Rummel, S., Rusinov, V., Russell, R.M., Ryu, S., Sahoo, H., Sakai, K., Sakai, Y., Santelj, L., Sasaki, T., Sato, N., Sato, Y., Scheirich, J., Schieck, J., Schwanda, C., Schwartz, A.J., Schwenker, B., Seljak, A., Senyo, K., Seon, O.-S., Sevior, M.E., Shapkin, M., She- balin, V., Shen, C.P., Shibuya, H., Shiizuka, S., Shiu, J.-G., Shwartz, B., Simon, F., Simonis, H.J., Singh, J.B., Sinha, R., Sitarz, M., Smerkol, P., Sokolov, A., Solovieva, E., Stanic, S., Staric, M., Stypula, J., Suetsugu, Y., Sugihara, S., Sugimura, T., Sumisawa, K., Sumiyoshi, T., Suzuki, K., Suzuki, S.Y., Takagaki, H., Takasaki, F., Takeichi, H., Takubo, Y., Tanaka, M., Tanaka, S., Taniguchi, N., Tarkovsky, E., Tatishvili, G., Tawada, M., Taylor, G.N., Teramoto, Y., Tikhomirov, I., Trabelsi, K., Tsuboyama, T., Tsunada, K., Tu, Y.-C., Uchida, T., Uehara, S., Ueno, K., Uglov, T., Unno, Y., Uno, S., Urquijo, P., Ushiroda, Y., Usov, Y., Vahsen, S., Valentan, M., Vanhoefer, P., Varner, G., Varvell, K.E., Vazquez, P., Vila, I., Vilella, E., Vinokurova, A., Visniakov, J., Vos, M., Wang, C.H., Wang, J., Wang, M.-Z., Wang, P., Wassatch, A., Watan- abe, M., Watase, Y., Weiler, T., Wermes, N., Wescott, R.E., White, E., Wicht, J., Widhalm, L., Williams, K.M., Won, E., Xu, H., Yabsley, B.D., Yamamoto, H., Yamaoka, H., Yamaoka, Y., Yamauchi, M., Yin, Y., Yoon, H., Yu, J., Yuan, C.Z., Yusa, Y., Zander, D., Zdybal, M., Zhang, Z.P., Zhao, J., Zhao, L., Zhao, Z., Zhilich, V., Zhou, P., Zhu- lanov, V., Zivko, T., Zupanc, A., Zyukova, O.: Belle II 技术设计报告. arXiv. 注：由 Z. Dole 编辑\v{z}al 和 S. Uno (2010)。
https://doi.org/10.48550/arXiv.1011.0352 3, 4, 19

[38]Diefenbacher, S., Eren, E., Gaede, F., Kasieczka, G., Krause, C., Shekhzadeh, I., Shih, D.：L2LFlows：生成高保真度的三维量能器图像。Journal of Instrumentation18(10), 10017 (2023)。
https://doi.org/10.1088/1748-0221/18/10/P10017 4

[39] Kim, D.Y., Ritter, M., Bilka, T., Bobrov, A., Casarosa, G., Chilikin, K., Ferber, T., Godang, R., Jaegle, I., Kandra, J., Kodys, P., Kuhr, T., Kvasnicka, P., Nakayama, H., Piilonen, L., Pulvermacher, C.,San- telj, L., Schwenker, B., Sibidanov, A., Soloviev, Y., Stariˇc, M., Uglov, T.: Belle II 软件系统的模拟库。《物理学杂志：会议系列》 **898**(4), 042043 (2017年)。https://doi.org/10.1088/1742-6596/898/4/042043 4

[40]Kuhr, T.: Belle II 的计算。《物理学杂志：会议系列》**331**(7), 072021 (2011)。https://doi.org/10.1088/1742-6596/331/7/072021 4, 26

[41] Belle II / 高能加速器研究机构：`https://www.belle2.org/archives/`版权所有：(C) Belle II / 高能加速器研究机构。(2024)。5

[42] Agostinelli, S., Allison,J., Amako, K., Apostolakis, J., Araujo, H., Arce,P., Asai, M., Axen, D., Banerjee, S., Barrand, G., Behner, F.,Bel-lagamba, L.,Boudreau,J., Broglia, L., Brunengo, A., Burkhardt, H.,Chauvie, S.,Chuma, J., Chytracek, R., Cooperman, G., Cosmo, G.,Degtyarenko, P., Dell'Acqua,A., Depaola, G.,Dietrich, D., Enami, R.,Feliciello,A., Ferguson, C., Fesefeldt, H., Folger, G.,Foppiano, F.,Forti,A., Garelli, S.,Giani, S., Giannitrapani, R.,Gibin, D., G´omez Cade-nas, J.J., Gonz´alez,I.,Gracia Abril, G., Greeniaus, G.,Greiner,W.,Grichine, V., Grossheim, A., Guatelli, S., Gumplinger, P.,Hamatsu,R., Hashimoto,K.,Hasui,H.,Heikkinen, A., Howard, A., Ivanchenko,V., Johnson,A., Jones, F.W., Kallenbach, J., Kanaya, N., Kawabata,M., Kawabata, Y., Kawaguti, M., Kelner, S., Kent, P., Kimura,A.,Kodama,T.,Kokoulin,R.,Kossov,M.,Kurashige, H., Lamanna, E.,Lamp´en, T., Lara, V., Lefebure, V., Lei, F., Liendl, M., Lockman, W., Longo, F., Magni, S., Maire, M., Medernach, E., Minamimoto, K., Mora de Freitas, P., Morita, Y., Murakami, K., Nagamatu, M., Nartallo, R., Nieminen, P., Nishimura, T., Ohtsubo, K., Okamura, M., O'Neale, S., Oohata, Y., Paech, K., Perl, J., Pfeiffer, A., Pia, M.G., Ranjard, F., Rybin, A., Sadilov, S., Di Salvo, E., Santin, G., Sasaki, T., Savvas, N., Sawada, Y., Scherer, S., Sei, S., Sirotenko, V., Smith, D., Starkov, N., Stoecker, H., Sulkimo, J., Takahata, M., Tanaka, S., Tcherniaev, E., Safai Tehrani, E., Tropeano, M., Truscott, P., Uno, H., Urban, L., Urban, P., Verderi, M., Walkden, A., Wander, W., Weber, H., Wellisch, J.P., Wenaus, T., Williams, D.C., Wright, D., Yamada, T., Yoshida, H., Zschiesche, D.: Geant4——一个模拟工具包。《核仪器与方法物理研究 A 部分：加速器、能谱仪、探测器及相关设备》**506**(3), 250–303 (2003).https://doi.org/10.1016/S0168-9002(03)01368-8 5, 9

[43] 纳什, J.: 非合作博弈。《数学年刊》**54**(2),286–295(1951)https://arxiv.org/abs/1969529。https://doi.org/10.2307/

1969529 6

[44] Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., Pascanu, R.: 关系归纳偏置、深度学习与图网络。arXiv（2018 年）。https://doi.org/10.48550/arXiv.1806.01261 6，23

[45] Sharifzadeh, S.,Baharlou, S.M.,Tresp,V.：基于注意力的分类：利用先验知识进行场景图分类。AAAI人工智能大会论文集35(6), 5025–5033（2021年）。https://doi.org/10.1609/aaai.v35i6.16636

[46]Locatello, F., Weissenborn, D., Unterthiner, T., Mahendran, A., Heigold, G., Uszkoreit, J., Dosovitskiy, A., Kipf, T.：基于槽注意力的以物体为中心的学习。载于：神经信息处理系统进展，第 33 卷，第 11525–11538 页。Curran Associates, Inc.，线上（2020年）。
https://proceedings.neurips.cc/paper/2020/hash/8511df98c02ab60aea1b2356c013bc0f-Abstract.html 访问日期：2024-04-016，23[47] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.：BERT：用于语言理解的深度双向变换器预训练。载于：Burstein, J., Doran, C., Solorio, T.（编）《2019年北美计算语言学协会分会会议论文集：人类语言技术》，第 1 卷（长短篇论文），第 4171–4186 页。计算语言学协会，明尼阿波利斯，明尼苏达州（2019 年）。https://doi.org/10.18653/v1/N19-1423 6，23，25[48] Zhang, D., Yang, Z.：用于句子分类的词嵌入扰动。arXiv（2018 年）。https://doi.org/10.48550/arXiv.1804.08166 6，25[49] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.：采用双时间尺度更新规则训练的生成对抗网络收敛至局部纳什均衡。载于：第 31 届国际神经信息处理系统大会论文集。NIPS'17，第 6629–6640 页。Curran Associates Inc.，美国纽约州雷德胡克（2017年）7，28[50] Bi´nkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.：揭开最大均值差异生成对抗网络的奥秘。arXiv。备注：发表于 ICLR 2018： https://openreview.net/forum?id=r1lUOzWCW（2021年）。https://doi.org/10.48550/arXiv.1801.01401 7

[51] Parmar, G., Zhang, R., Zhu, J.-Y.：关于混叠缩放及生成对抗网络评估中令人意外的细微之处。载于：2022年 IEEE/CVF 计算机视觉与模式识别会议（CVPR），第 11400–11410 页。IEEE 计算机学会，线上（2022 年）。https://doi.org/10.1109/CVPR52688.2022.01112 7

[52] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.：重新思考用于计算机视觉的 Inception 架构。载于：2016 年 IEEE 计算机视觉与模式识别会议（CVPR），第 2818–2826 页（2016 年）。https://doi.org/10.1109/CVPR.2016.308 7

[53] Xu, Q., Huang, G., Yuan, Y., Guo, C., Sun, Y., Wu, F., Weinberger, K.：生成对抗网络评估指标的实证研究。arXiv。注：arXiv 管理员备注：本文与其他作者的 arXiv:1802.03446 存在文本重叠（2018 年）。https://doi.org/10.48550/arXiv.1806.07755 7

[54] Brock, A., Donahue, J., Simonyan, K.：用于高保真自然图像合成的大规模生成对抗网络训练。arXiv（2019 年）。https://doi.org/10.48550/arXiv.1809.11096 9, 20, 27, 29

[55] Kuhr, T., Pulvermacher, C., Ritter, M., Hauth, T., Braun, N.：《Belle II 核心软件》。面向大科学的计算与软件**3**(1), 1（2018 年）。https://doi.org/10.1007/s41781-018-0017-9 9

[56]Mantel, N.：疾病聚集的检测与广义回归方法。《癌症研究》**27**(2), 209–220（1967 年）11

[57] Sokal, R.R., Rohlf, F.J.：《生物统计学》。W. H. Freeman，ISBN 978-0-7167- 2411-7（1995 年）11

[58] Kou, E., Urquijo, P., Altmannshofer, W., Beaujean, F., Bell, G., Beneke, M., Bigi, I.I., Bishara, F., Blanke, M., Bobeth, C., Bona, M., Brambilla, N., Braun, V.M., Brod, J., Buras, A.J., Cheng, H.Y., Chiang, C.W., Ciu- chini, M., Colangelo, G., Crivellin, A., Czyz, H., Datta, A., De Fazio, F., Deppisch, T., Dolan, M.J., Evans, J., Fajfer, S., Feldmann, T., God- frey, S., Gronau, M., Grossman, Y., Guo, F.K., Haisch, U., Hanhart, C., Hashimoto, S., Hirose, S., Hisano, J., Hofer, L., Hoferichter, M., Hou, W.S., Huber, T., Hurth, T., Jaeger, S., Jahn, S., Jamin, M., Jones, J., Jung, M., Kagan, A.L.,Kahlhoefer, F., Kamenik, J.F., Kaneko, T., Kiyo, Y., Kokulu, A., Kosnik, N., Kronfeld, A.S., Ligeti, Z., Logan, H., Lu, C.D., Lubicz, V., Mahmoudi, F., Maltman, K., Mishima, S., Misiak, M., Moats, K., Moussallam, B., Nefediev, A., Nierste, U., Nomura, D., Offen, N., Olsen, S.L., Passemar, E., Paul, A., Paz, G., Petrov, A.A., Pich, A., Polosa, A.D., Pradler, J., Prelovsek, S., Procura, M., Ricciardi, G., Robinson, D.J., Roig, P., Rosiek, J., Schacht, S., Schmidt-Hoberg, K., Schwichtenberg, J., Sharpe, S.R., Shigemitsu, J., Shih, D., Shimizu, N.,

Shimizu, Y., Silvestrini, L., Simula, S., Smith, C., Stoffer, P., Straub, D., Tackmann, F.J., Tanaka, M., Tayduganov, A., Tetlalmatzi-Xolocotzi, G., Teubner, T., Vairo, A., van Dyk, D., Virto, J., Was, Z., Watanabe, R., Watson, I., Westhoff, S., Zupan, J., Zwicky, R., Abudin´en, F., Adachi, I., Adamczyk, K., Ahlburg, P., Aihara, H., Aloisio, A., Andricek, L., Anh Ky, N., Arndt, M., Asner, D.M., Atmacan, H., Aushev, T., Aushev, V., Ayad, R., Aziz, T., Baehr, S., Bahinipati, S., Bambade, P., Ban, Y., Barrett, M., Baudot, J., Behera, P., Belous, K., Bender, M., Bennett, J., Berger, M., Bernieri, E., Bernlochner, F.U., Bessner, M., Besson, D., Bettarini, S., Bhardwaj, V., Bhuyan, B., Bilka, T., Bilmis, S., Bilokin, S., Bonvicini, G., Bozek, A., Braˇcko, M., Branchini, P., Braun, N., Briere, R.A., Browder, T.E., Burmistrov, L., Bussino, S., Cao, L., Caria, ˇ G., Casarosa, G., Cecchi, C., Cervenkov, D., Chang, M.-C., Chang, P., Cheaib, R., Chekelian, V., Chen, Y., Cheon, B.G., Chilikin, K., Cho, K., Choi, J., Choi, S.-K., Choudhury, S., Cinabro, D., Cremaldi, L.M., Cuesta, D., Cunliffe, S., Dash, N., de la Cruz Burelo, E., de Lucia, E., De Nardo, G., De Nuccio, M., De Pietro, G., De Yta Hernandez, A., Deschamps, B., Destefanis, M., Dey, S., Di Capua, F., Di Carlo, S., Dingfelder, J., Doleˇzal, Z., Dom´ınguez Jim´enez, I., Dong, T.V., Dos- sett, D., Duell, S., Eidelman, S., Epifanov, D., Fast, J.E., Ferber, T., Fiore, S., Fodor, A., Forti, F., Frey, A., Frost, O., Fulsom, B.G., Gabriel, M., Gabyshev, N., Ganiev, E., Gao, X., Gao, B., Garg, R., Garmash, A., Gaur, V., Gaz, A., Geßler, T., Gebauer, U., Gelb, M., Gellrich, A., Getzkow, D., Giordano, R., Giri, A., Glazov, A., Gobbo, B., Godang, R., Gogota, O., Goldenzweig, P., Golob, B., Gradl, W., Graziani, E., Greco, M., Greenwald, D., Gribanov, S., Guan, Y., Guido, E., Guo, A., Halder, S., Hara, K., Hartbrich, O., Hauth, T., Hayasaka, K., Hayashii, H., Hearty, C., Heredia De La Cruz, I., Hernandez Villanueva, M., Her- shenhorn, A., Higuchi, T., Hoek, M., Hollitt, S., Hong Van, N.T., Hsu, C.-L., Hu, Y., Huang, K., Iijima, T., Inami, K., Inguglia, G., Ishikawa, A., Itoh, R., Iwasaki, Y., Iwasaki, M., Jackson, P., Jacobs, W.W., Jaegle, I., Jeon, H.B., Ji, X., Jia, S., Jin, Y., Joo, C., K¨unzel, M., Kadenko, I., Kahn, J., Kakuno, H., Kaliyar, A.B., Kandra, J., Kang, K.H., Kato, Y., Ka wasaki, T., Ketter, C., Khasmidatul, M., Kichimi, H., Kim, J.B., Kim, K.T., Kim, H.J., Kim, D.Y., Kim, K., Kim, Y., Kimmel, T.D., Kindo, H., Kinoshita, K., Konno, T., Korobov, A., Korpar, S., Kotchetkov, D., Kowalewski, R., Kriˇzan, P., Kroeger, R., Krohn, J.-F., Krokovny, P., Kuehn, W., Kuhr, T., Kulasiri, R., Kumar, M., Kumar, R., Kumita, T., Kuzmin, A., Kwon, Y.-J., Lacaprara, S., Lai, Y.-T., Lalwani, K., Lange, J.S., Lee, S.C., Lee,J.Y., Leitl, P., Levit, D., Levonian, S., Li, S., Li, L.K., Li, Y., Li, Y.B., Li, Q., Li Gioi, L., Libby, J., Liptak, Z., Liventsev, D., Longo, S., Loos, A., Lopez Castro, G., Lubej, M., Lueck, T., Luetticke, F., Luo, T., M¨uller, F., M¨uller, T., MacQueen, C., Maeda, Y., Maggiora, M., Maity, S., Manoni, E., Marcello, S., Marinas, C., Martinez Hernan- dez, M., Martini, A., Matvienko, D., McKenna, J.A., Meier, F., Merola,

M., Metzner, F., Miller, C.,Miyabayashi, K., Miyake, H., Miyata, H.,Mizuk, R., Mohanty,G.B., Moon, H.K., Moon, T., Morda, A., Morii, T.,Mrvar, M., Muroyama, G.,Mussa, R., Nakamura, I., Nakano, T., Nakao,M., Nakayama, H.,Nakazawa, H., Nanut, T., Naruki, M.,Nath, K.J.,Nayak, M., Nellikunnummel,N.,Neverov, D., Niebuhr, C., Ninkovic,J., Nishida, S.,Nishimura,K., Nouxman, M., Nowak, G., Ogawa, K.,Onishchuk,Y., Ono, H., Onuki, Y., Pakhlov, P., Pakhlova, G., Pal, B.,Paoloni, E.,Park, H., Park, C.-S., Paschen, B., Passeri, A., Paul, S.,Pedlar, T.K.,Perell´o, M., Peruzzi, I.M., Pestotnik, R., Piilonen, L.E., Podesta Lerma, L., Popov, V., Prasanth, K., Prencipe, E., Prim, M., Purohit, M.V., Rabusov, A., Rasheed, R., Reiter, S., Remnev, M., Resmi, P.K., Ripp-Baudot, I., Ritter, M., Ritzert, M., Rizzo, G., Rizzuto, L., Robertson, S.H., Rodriguez Perez, D., Roney, J.M., Rosenfeld, C., Rosto- myan, A., Rout, N., Rummel, S., Russo, G., Sahoo, D., Sakai, Y., Salehi, M., Sanders, D.A., Sandilya, S., Sangal, A., Santelj, L., Sasaki, J., Sato, Y., Savinov, V., Scavino, B., Schram, M., Schreeck, H., Schueler, J., Schwanda, C., Schwartz, A.J., Seddon, R.M., Seino, Y., Senyo, K., Seon, O., Seong, I.S., Sevior, M.E., Sfienti, C., Shapkin, M., Shen, C.P., Shimo- mura, M., Shiu, J.-G., Shwartz, B., Sibidanov, A., Simon, F., Singh, J.B., Sinha, R., Skambraks, S., Smith, K., Sobie, R.J., Soffer, A., Sokolov, A., Solovieva, E., Spruck, B., Stanič, S., Starič, M., Starinsky, N., Stolzen- berg, U., Stottler, Z., Stroili, R., Strube, J.F., Stypula, J., Sumihama, M., Sumisawa, K., Sumiyoshi, T., Summers, D., Sutcliffe, W., Suzuki, S.Y., Tabata, M., Takahashi, M., Takizawa, M., Ta mponi, U., Tan, J., Tanaka, S., Tanida, K., Taniguchi, N., Tao, Y., Taras, P., Tejeda Munoz, G., Tenchini, F., Tippawan,U., Torassa, E., Trabelsi, K., Tsuboyama, T., Uchida, M., Uehara, S., Uglov, T., Unno, Y., Uno, S., Ushiroda, Y., Usov, Y., Vahsen, S.E., van Tonder, R., Varner, G., Varvell, K.E., Vinokurova, A., Vitale, L., Vos, M., Vossen, A., Waheed, E., Wakeling, H., Wan, K., Wang, M.-Z., Wang, X.L., Wang, B., Warburton, A., Webb, J., Wehle, S., Wessel, C., Wiechczynski, J., Wieduwilt, P., Won, E., Xu, Q., Xu, X., Yabsley, B.D., Yamada, S., Yamamoto, H., Yan, W., Yan, W., Yang, S.B., Ye, H., Yeo, I., Yin, J.H., Yonenaga, M., Yoshinobu, T., Yuan, W., Yuan, C.Z., Yusa, Y., Zakharov, S., Zani, L., Zeyrek, M., Zhang, J., Zhang, Y., Zhang, Y., Zhou, X., Zhukova, V., Zhulanov, V., Zupanc, A.: Belle II物理学手册。《理论与实验物理进展》 **2020年**(2), 029201(2020). https://doi.org/10.1093/ptep/ptaa00812

[59]Massey Jr, F.J.：拟合优度的柯尔莫哥洛夫–斯米尔诺夫检验。美国统计学会期刊 **46**(253), 68–78 (1951)13

[60] Devore, J.L., Berk, K.N.：《现代数学统计及其应用》。施普林格，ISBN 978-1-4614-0390-6（2011）13

[61] Ramachandran, K.M., Tsokos, C.P.：《数学统计及其在 R 中的应用》。Academic Press，ISBN 978-0-12-817816-4（2020）13[62] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.：RoBERTa：一种鲁棒优化的 BERT 预训练方法。arXiv（2019）。https://doi.org/10.48550/arXiv.1907.11692 19[63] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q.V.：XLNet：面向语言理解的广义自回归预训练。在：第 33 届神经信息处理系统国际会议论文集，第 5753–5763 页。Curran Associates Inc.，美国纽约州雷德胡克（2019）19[64] Hariri, A., Dyachkova, D., Gleyzer, S.：用于高能物理快速探测器模拟的图生成模型。arXiv。说明：修订参考文献并更正错别字（2021）。https://doi.org/10.48550/arXiv.2104.01725 19[65] Verheyen, R.：使用满射归一化流进行事件生成与密度估计。SciPost Physics**13**（3），047（2022）。https://doi.org/10.21468/SciPostPhys.13.3.047 19[66] Butter, A., Plehn, T., Schumann, S.,Badger, S., Caron, S., Cranmer, K.,Di Bello, F.A.,Dreyer,E.,Forte, S., Ganguly, S., Gon˙calves, D., Gross,E., Heimel, T.,Heinrich,G., Heinrich, L., Held, A., H¨oche, S., Howard,J.N., Ilten, P.,Isaacson,J.,Janßen, T., Jones, S., Kado, M., Kagan, M.,Kasieczka, G.,Kling, F., Kraml, S., Krause, C., Krauss, F., Kr¨oninger, K., Barman, R.K., Luchmann, M., Magerya, V., Maˆıtre, D., Malaescu, B., Maltoni, F., Martini, T., Mattelaer, O., Nachman, B., Pitz, S., Rojo, J., Schwartz, M., Shih, D., Siegert, F., Stegeman, R., Stienen, B., Thaler, J., Verheyen, R., Whiteson, D., Winterhalder, R., Zupan, J.：机器学习与大型强子对撞机事件生成。SciPost Physics**14**（4），079（2023）。https://doi.org/10.21468/SciPostPhys.14.4.079 19[67] Evans, L., Bryant, P.：大型强子对撞机。Journal of Instrumentation**3**（08），08001（2008）。https://doi.org/10.1088/1748-0221/3/08/S08001 19[68] Terzo, S., Boscardin, M., Carlotto, J., Dalla Betta, G.-F., Darbo, G., Dorholt, O., Ficorella, F., Gariano, G., Gemme, C., Giannini, G., Grin- stein, S., Heggelund, A., Huiberts, S., Kok, A., Koybasi, O., Lapertosa, A., Lauritzen, M.E., Manna, M., Mendicino, R., Oide, H., Pellegrini, G., Povoli, M., Quirion, D., Rohne, O.M., Ronchin, S., Sandaker, H., Abdulla Samy, M.A., Stugu, B., Vannoli, L.：用于升级 ATLAS 内层径迹探测器的新型三维像素传感器。Frontiers in Physics**9**（2021）。https://doi.org/10.3389/fphy.2021.624668 19

[69] Pedro, K.：《CMS模拟的当前与未来性能》。EPJ网络会议 214, 02036（2019）。https://doi.org/10.1051/ epjconf/201921402036 19[70] Huang, P.-S., Boyken, S.E., Baker, D.：从头蛋白质设计走向成熟。《自然》537(7620)，320–327（2016）。https://doi.org/10. 1038/nature19946 20[71] Liu, Q., Xu, J., Jiang, R., Wong, W.H.：使用深度生成神经网络进行密度估计。美国国家科学院院刊 118(15)，2101344118（2021）。https://doi.org/10.1073/pnas.2101344118 20[72]Repecka, D., Jauniskis, V., Karpus, L., Rembeza, E., Rokaitis, I., Zrimec, J., Poviloniene, S., Laurynenas, A., Viknander, S., Abuajwa, W., Savolainen, O., Meskys, R., Engqvist, M.K.M., Zelezniak, A.：利用生成对抗网络拓展功能蛋白质序列空间。《自然·机器智能》3(4)，324–333（2021）。https://doi.org/10. 1038/s42256-021-00310-5 20[73]Anand, N., Huang, P.：蛋白质结构的生成建模。载于：Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R.（编），《神经信息处理系统进展》，第31卷。Curran Associates, Inc.，线上（2018）。https://proceedings.neurips.cc/paper files/paper/2018/file/afa299a4d1d8c52e75dd8a24c3ce534f-Paper.pdf[74] Strokach, A., Kim, P.M.：用于蛋白质设计的深度生成建模。《结构生物学当前观点》72, 226–236（2022）。https://doi. org/10.1016/j.sbi.2021.11.008 20[75] Mirza, M., Osindero, S.：条件生成对抗网络。arXiv -（2014）。https://doi.org/10.48550/arXiv.1411.1784 20, 24, 28[76] Lim, J.H., Ye, J.C.：几何生成对抗网络。arXiv（2017）。https://doi.org/10. 48550/arXiv.1705.02894 20[77]Odena, A., Olah, C., Shlens, J.：基于辅助分类器生成对抗网络的条件图像合成。载于：第34届国际机器学习会议论文集，第2642–2651页。PMLR，线上（2017）。https://proceedings.mlr.press/v70/odena17a.html 访问日期：2024-04-01 21[78] Chen, T., Kornblith, S., Norouzi, M., Hinton, G.：视觉表征对比学习的简单框架。载于：第37届国际机器学习会议论文集。IC ML'20，第119卷，第1597–1607页。JMLR.org，线上（2020）21

[79] Vaswani, A., Shazeer,N., Parmar, N., Uszkoreit, J., Jones, L., Gomez,A.N., Kaiser, Ł., Polosukhin, I.：注意力就是你所需要的一切。载于：第31届国际神经信息处理系统会议论文集。NIPS'17，第6000–6010页。Curran Associates Inc.，美国纽约州雷德胡克（2017）21，22，23，27

[80] Hudson, D.A.,Zitnick, L.：生成对抗变换器。载于：第38届国际机器学习会议论文集，第4487–4499页。PMLR，线上（2021年）。https://proceedings.mlr.press/v139/hudson21a.html 访问日期：2024-04-01 21

[81] Jiang, Y., Chang,S., Wang, Z.：TransGAN：两个纯 Transformer 即可构建强大的生成对抗网络，并实现扩展。载于：神经信息处理系统进展，第34卷，第14745–14758页。Curran Associates, Inc.，线上（2021年）。https://proceedings.neurips.cc/paperfiles/paper/2021/hash/7c220a2091c26a7f5e9f1cfb099511e3-Abstract.html访问日期：2024-04-01 21

[82] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.：《图像识别领域，一张图像胜过 16×16 个词：大规模视觉 Transformer》。arXiv。注：微调代码和预训练模型可在 https://github.com/google-research/vision transformer 获取。国际学习表征会议最终定稿版本包含两处小修改：1）在附录中新增了有关 CLS 与 GAP 分类器的讨论；2）修正了图5和表 6 中 exaFLOPs计算的一处误差（模型的相对性能基本不受影响）（2021年）。https://doi.org/10.48550/arXiv.2010.11929 21

[83] Liu, L., Liu, X., Gao, J., Chen, W., Han, J.：理解训练 Transformer 的难度。载于：Webber, B., Cohn, T., He, Y., Liu, Y.（编）《2020年自然语言处理实证方法会议论文集》（EMNLP），第 5747–5763 页。计算语言学协会，线上（2020年）。https://doi.org/10.18653/v1/2020.emnlp-main.463 21

[84]Guttenberg, N., Virgo, N., Witkowski, O., Aoki, H., Kanai, R.：应用于动力学预测的置换等变神经网络。arXiv。备注：7 页，4 幅图（2016 年）。https://doi.org/10.48550/arXiv.1612.04530 22

[85]Ravanbakhsh, S., Schneider, J., P´oczos, B.：通过参数共享实现等变性。载于：第34届国际机器学习会议论文集，第2892–2901页。PMLR，线上（2017年）。https://proceedings.mlr.press/v70/ravanbakhsh17a.html访问日期：2024-04-0122

[86] Zaheer, M., Kottur, S., Ravanbhakhsh, S., P´oczos, B., Salakhutdinov, R., Smola, A.J.：深度集合。载于：第31届国际神经信息处理系统会议论文集。NIPS'17，第3394–3404页。Curran Associates Inc.，美国纽约州雷德胡克（2017年）22[87] Clark, K., Khandelwal, U., Levy, O., Manning, C.D.：BERT关注什么？对BERT注意力的分析。载于：Linzen, T., Chrupa-la, G., Belinkov, Y., Hupkes, D.（编）《2019年ACL BlackboxNLP研讨会论文集：分析和解读自然语言处理神经网络》，第276–286页。计算语言学协会，意大利佛罗伦萨（2019年）。https://doi.org/10.18653/v1/W19-4828 23[88]Yun, C., Bhojanapalli, S., Rawat, A.S., Reddi, S.J., Kumar, S.：Transformer是序列到序列函数的通用逼近器吗？arXiv。说明：23页，ICLR 2020最终会议版本（2020年）。https://doi.org/10.48550/arXiv.1912.10077 23[89] Ba, J.L., Kiros, J.R., Hinton, G.E.：层归一化。arXiv（2016 年）。https://doi.org/10.48550/arXiv.1607.06450 23[90] Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.：生成对抗网络的谱归一化。arXiv。说明：作为会议论文发表于 ICLR 2018（2018 年）。https://doi.org/10.48550/arXiv.1802.05957 23, 28[91]Wang, F., Xiang, X., Cheng, J., Yuille, A.L.：NormFace：用于人脸验证的 L2 超球面嵌入。载于：第25届ACM多媒体国际会议论文集。MM '17，第1041–1049页。计算机协会，纽约州纽约（2017年）。https://doi.org/10.1145/3123266.3123359 24[92] Gorban, A.N., Tyukin, I.Y.：随机分离定理。《神经网络》94，255–259（2017年）。https://doi.org/10.1016/j.neunet.2017.07.014 24[93]Rani, V., Nabi, S.T., Kumar, M., Mittal, A., Kumar, K.：自监督学习：简明综述。《计算工程方法档案》30(4)，2761–2775（2023）。https://doi.org/10.1007/s11831-023-09884-2 25[94] Kullback, S., Leibler, R.A.：论信息与充分性。《数理统计年刊》22(1)，79–86（1951）。https://doi.org/10.1214/ aoms/1177729694 25, 27[95] Thomson, J.J.：XXIV. 论原子结构：对若干以相等间隔排列在圆周上的微粒的稳定性和振荡周期的研究；并附应用

关于原子结构理论的研究结果。《伦敦、爱丁堡和都柏林哲学杂志与科学杂志》**7**(39), 237– 265 (1904)。https://doi.org/10.1080/14786440409463107 25

[96] Liu, W., Lin, R., Liu, Z., Liu, L., Yu, Z., Dai, B., Song, L.：学习趋向最小超球面能量。载于：第 32 届神经信息处理系统国际会议论文集。NIPS'18，第 6225–6236 页。Curran Associates Inc.，美国纽约州雷德胡克（2018 年）26[97] Kuijlaars, A., Saff, E.：球面上最小离散能量的渐近性质。《美国数学会会刊》**350**(2), 523– 538 (1998)。https://doi.org/10.1090/S0002-9947-98-02119-9 26[98] Saxe, A.M., McClelland, J.L., Ganguli, S.：深度线性神经网络学习非线性动力学的精确解。arXiv。说明：提交至 ICLR2014。根据审稿意见修订（2014）。https://doi.org/10.48550/arXiv.1312.6120 26[99] Huber, P.J.：位置参数的稳健估计。《数理统计年鉴》**35**(1), 73–101 (1964)。https://doi.org/10.1214/aoms/1177703732 27[100] Eybpoosh, K., Rezghi, M., Heydari, A.：将逆立体投影应用于流形学习和聚类。《应用智能》**52**(4), 4443–4457 (2022)。https://doi.org/10.1007/s10489-021-02513-028 [101]Hendrycks, D., Gimpel, K.：高斯误差线性单元（高斯误差线性单元）。arXiv。说明：2016 年草稿的精简版本（2023）。https://doi.org/10.48550/arXiv.1606.08415 28[102]Hashemi, B.：超高颗粒度像素顶点探测器（PXD）特征图像。载于：机器学习与物理科学，神经信息处理系统大会 *2022*），Zenodo，版本 v1。https://doi.org/10.5281/zenodo.8331919（2023）30[103]Hashemi, B.：Hosein47/IEA-GAN：IEA-GAN v1。载于：《自然通讯》，Zenodo。https://doi.org/10.5281/zenodo.11070305 （2024）30

致谢

本研究得到合作项目 IDT-UM（用于探索宇宙与物质的创新数字技术）和 KISS 联盟的支持；这两个项目由德国联邦教育与研究部（BMBF）以及德国的

卓越战略资助——EXC 2094 ORIGINS——390783311。James Kahn 的工作得到亥姆霍兹协会倡议与网络基金旗下亥姆霍兹人工智能平台项目的支持。BH 和 SS 谨向 Volker Tresp 致谢，感谢他提出的宝贵意见，这些意见丰富了本研究。我们也感谢 David Katheder 协助准备 GitHub 仓库。我们感谢慕尼黑路德维希-马克西米利安大学和粒子与天体物理计算中心（C2PAP）的同事，他们提供的专业知识和计算资源为本研究提供了极大帮助。

作者贡献

B.H. 设计并实现了研究的核心构想，开展了所有实验运行，进行了下游物理分析，并撰写了稿件。N.H.、S.S. 和 T.K. 在不同阶段为完善研究构想作出了贡献。N.H. 参与了部分代码、图表、数据分析流程以及结果验证工作。T.K. 指导了物理分析和结果验证。J.K. 在各个阶段指导并参与了稿件和图表的编写与制作。所有作者均审阅了稿件并提出了意见。

竞争性利益

作者声明不存在竞争性利益。