arXiv:2302.00410v1 [cond-mat.soft] 1 Feb 2023

# 利用深度学习推断单条轨迹的逐点扩散性质

Borja Requena,[1] Sergi Masó,[2] Joan Bertran,[2] Maciej Lewenstein,[1,3] Carlo Manzo,[2] 和 Gorka Muñoz-Gil[4],*[1]
*ICFO – 光子科学研究所，巴塞罗那科学技术研究所，Av. Carl Friedrich Gauss 3, 08860* 卡斯特尔费尔德斯（巴塞罗那），西班牙[2]科学、技术与工程学院，维克大学—加泰罗尼亚中央大学（*UVic-UCC*），*C. de la Laura,13, 08500* 维克，西班牙[3]*ICREA，Pg. Lluís Companys 23, 08010* 巴塞罗那，西班牙[4]理论物理研究所，因斯布鲁克大学，*Technikerstr. 21a, A-6020* 因斯布鲁克，奥地利

为了表征生物环境中控制粒子扩散的机制，准确确定其扩散性质至关重要。为此，我们提出一种机器学习方法，可在实验时间分辨率下表征性质随时间变化的扩散过程。我们的方法在单轨迹层面运行，在轨迹的每个时间步预测所关注的性质，例如扩散系数或反常扩散指数。这样，轨迹中扩散性质的变化会自然地体现在预测结果中，从而无需对系统进行任何先验了解或假设即可完成表征。我们首先在多种条件下模拟的合成轨迹上对该方法进行基准测试。结果表明，我们的方法能够成功表征扩散系数或反常扩散指数的突变和连续变化。最后，我们利用该方法分析活细胞中两种膜蛋白的单分子扩散实验：病原体识别受体 DC-SIGN 和整合素 $\alpha 5\beta 1$。该分析使我们能够以前所未有的精度表征物理参数和扩散状态，为揭示其潜在机制带来新的认识。

## I. 引言

光学成像技术的进步，使得人们能够观测活体生物系统中的单分子 [1]。结合粒子追踪算法后，这些技术可以以纳米级精度追踪单个分子、病毒和细胞器的运动，从而能够研究复杂生物环境中的运输机制。通过对轨迹进行生物物理表征，我们可以提取有意义的参数来描述物理和生物过程。然而，由于这些过程具有随机性，且实验存在成像噪声和发射体光物理性质等缺陷，要准确量化轨迹仍是一项挑战 [2]。

细胞系统中一种普遍存在的扩散特征是运动的时间依赖性变化 [3]。通常，这些变化与和其他组分的瞬时相互作用有关 [4–6] 这会导致某个参数（例如扩散系数）突然变化，该参数可能在离散 [7], 或连续的一系列水平之间切换 [8–10]。此外，这些变化还可能引起平滑变化，例如与环境时空异质性相关的变化 [11]。图 1A 示意性地展示了经历此类扩散的轨迹示例。

具有时变扩散性质的轨迹给表征单个粒子的运动带来了额外挑战，人们已采用不同方法应对这一问题。对于呈现突变的轨迹，分割方法 [3, 12–14] 是一种有效策略但无法处理长程相关性，而且由于时间平均，往往只能提供有限的时间分辨率。另一方面，依赖模型的方法（如隐马尔可夫模型（HMM））在描述异质扩散方面相当成功 [15–17], 但这类方法需要预先了解涉及的扩散状态及其动力学方案。近年来，数据驱动方法展现出从单个随机轨迹中提取信息的卓越能力，即使扩散性质发生变化也不例外 [18–20]。

在这项工作中，我们提出了 STEP，这是一种基于最先进的深度学习架构的方法，能够在无需任何先验信息的情况下，从单条轨迹中提取逐点扩散特征（见图 1B）。STEP 融合了序列到序列学习的最新进展， [21], 这类方法在自然语言处理任务及其他领域均展现出令人瞩目的成果 [22–24]。它结合了卷积层 [25] 和注意力层 [26], 以应对短程和长程相关性的存在，在任意长度的轨迹以及有噪声的情况下都能提供出色的性能。

文章结构如下：首先，我们展示 STEP 在模拟数据上预测扩散性质的能力，包括扩散系数和反常扩散指数；这些模拟数据再现了与实验相关的情景。随后，我们分析扩散系数平滑变化的模拟轨迹，表明 STEP 能够正确识别预期的标度关系。最后，我们使用 STEP 研究两个实验数据集，这些数据集通过单分子活细胞成像实验的追踪获得，记录了两种膜受体的运动：*i*) 病原体识别受体 DC-SIGN，它与扩散系数的随机变化有关 [10]；以及 *ii*) $\alpha 5\beta 1$ 整合素，

* munoz.gil.gorka@gmail.com

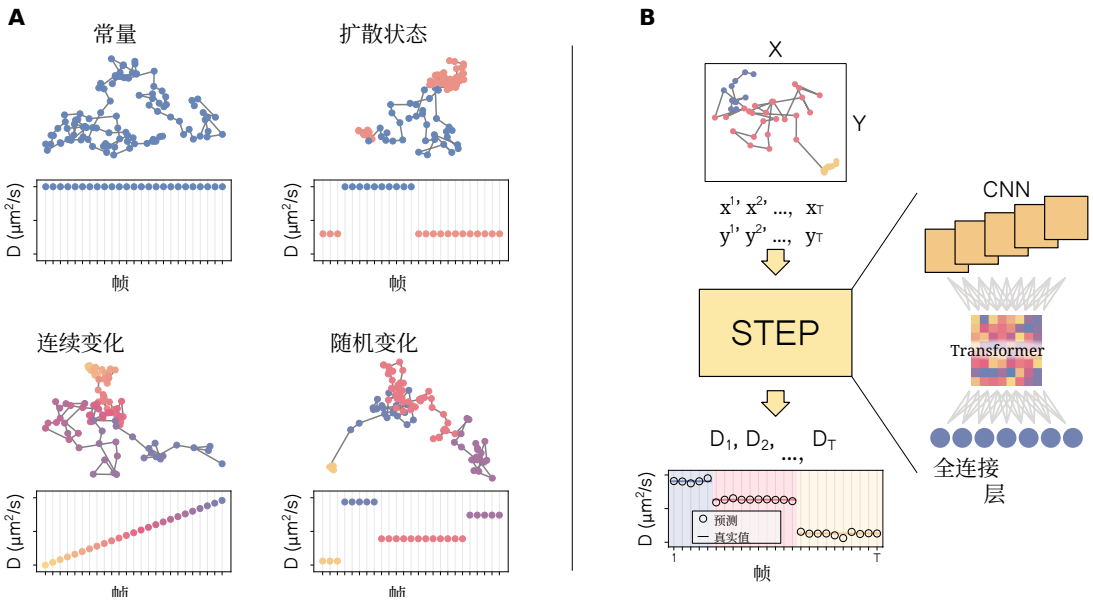


图1。**异质轨迹与 STEP 流程**。（**A**）轨迹示例及其对应的扩散系数 $D$ 随时间的变化情况，包括：常量 $D$；在固定 $D$ 的离散状态集合内变化； $D$ 连续且单调地变化；随机 $D$ 之间的切换。（**B**）STEP 流程示意图：输入轨迹被送入该架构，该架构由一组卷积神经网络、一个 Transformer 编码器和一个逐点前馈层组成。模型的输出是对感兴趣扩散参数的逐点预测（本例中为 $D$）。

预计其会因与细胞骨架和细胞外基质结合而暂时停滞 [27]。

## II. 结果

### STEP 架构

近年来，我们见证了人们投入巨大努力，开发用于研究扩散过程的深度学习方法 [18]。以往的研究通常聚焦于表征单条轨迹的扩散性质，即为每条输入轨迹预测一个整体或平均扩散参数 [28–32]。近期研究已证明，这种方法适用于在不同实验情境中研究复杂现象 [33, 34]。借助 STEP，我们提出了一种序列到序列方法，将位置坐标转换为扩散性质，并在输入轨迹的每个时间步提供逐点预测 [20]。STEP 可将任意长度的输入轨迹 $T$ 转换为一个由 $T$ 元素构成的序列，其中包含所关注的扩散性质，如图1B 所示。虽然仅凭单个位移实际上不可能表征扩散，但 STEP 会利用整条轨迹作为上下文，对每个点进行预测。

这种方法使我们能够研究扩散特性随时间以不同模式变化的轨迹：从扩散特性为常量的轨迹，到在离散扩散状态之间切换的轨迹，或扩散参数连续变化的轨迹。随时间变化（示例见图 1A）。与以往需要专家输入才能选择合适方法的研究不同，STEP 可无缝应用于任何扩散数据。重要的是，它不依赖先验假设，例如变点的数量 [35, 36]或预期扩散状态的性质 [37]。

用于扩散表征的先进架构采用了截然不同的方法 [38–40]。一些模型通过结合循环神经网络（RNNs）来捕捉长程相关性 [30, 36]，并利用卷积神经网络（CNNs）提取局部特征，从而取得了出色的性能 [34, 41, 42]。在 STEP 中，我们采用类似的方案，但以 Transformer编码器取代 RNN [26], ——这是序列建模的前沿架构；此外，我们堆叠 XResNet模块 [43]来构建卷积部分（见方法）。通过卷积神经网络提取局部特征，并将其输入 Transformer，从而同时处理局部和长程特征（见图 1B）。最后，我们使用一个由非线性神经元构成的逐点全连接层，以获得合适的输出维度。方法的详细说明见附录B1。

### 扩散性质的逐点预测

我们首先在一项任务中验证 STEP：从模拟轨迹中逐点推断扩散系数。这些轨迹再现了扩散系数发生突变的瞬态布朗运动。扩散系数可在 $D \in [10^{-3}, 10^{3}]$ 范围内随机变化，且

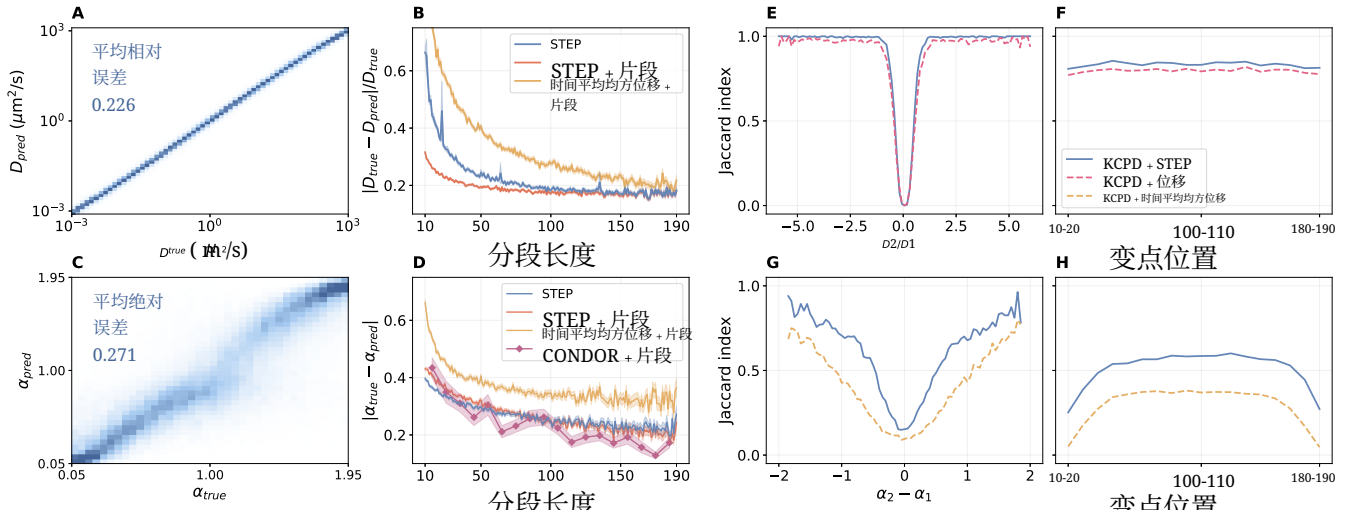


图 2. **时变扩散性质预测。(A)** 预测的扩散系数 $D$ 与真实值的二维直方图。整个测试集上的平均相对误差（MRE）为 0.226。**(B)** $D$ 的预测相对误差随分段长度的变化。**(C)** 分数布朗运动片段预测的反常扩散指数 $\alpha$ 与真实值的二维直方图。使用所有扩散模型时，测试集上的平均绝对误差（MAE）为 0.271。**(D)** $\alpha$ 预测的平均绝对误差（MAE）随分段长度的变化。**(E** 至**H)** 变点检测问题的杰卡德指数随以下因素的变化：**(E)** 连续片段 $D$s 之间的比率，**(F)** $D$的变点位置，**(G)** 连续片段 $\alpha$s 之间的差异，以及 **(H)** $\alpha$的变点位置。

每种扩散系数下的驻留时间均取自指数分布。模拟和数据集的更多细节见附录 B3。真实值与预测扩散系数的二维直方图显示，STEP 能够在模拟中涵盖的整个$D$ 数值范围内精确确定扩散系数（图2A），总体相对误差为 $|D_{\text{true}} - D_{\text{pred}}|/D_{\text{true}} = 0.226$。

为进一步探究 STEP 的性能，我们计算了相对误差随分段长度（即每个扩散系数对应的驻留时间）的变化情况。正如预期，分段长度越长，误差越小（图 2B 中的蓝线）。为了将 STEP 估算扩散系数的精度作为基线进行比较，我们采用时间平均均方位移（TA-MSD）的线性拟合（见附录 A）。然而，STEP 能够从异质轨迹中获得 $D$ 的逐点预测，而 TA-MSD 需要常量 $D$ 的轨迹，因此我们向其提供了预分段轨迹。尽管存在这一劣势，STEP 仍能给出更好的结果（图 2B 中的蓝线与黄线）。值得注意的是，向 STEP 提供预分段轨迹后，结果还能进一步改善：在分段长度较短时，误差几乎降低了一半（图 2B 中的红线），展现出卓越的预测能力。

接下来，我们考察 STEP 预测反常扩散指数的能力 $\alpha$。我们考虑根据不同的底层模型进行模拟的异质轨迹（更多细节见附录A），这些轨迹由若干片段组成，且 $\alpha$ 在 $\alpha \in [0.05, 2]$范围内突然变化。真实值与预测值的二维直方图 $\alpha$ 显示，对于经历分数布朗运动（FBM）的轨迹（图2C），STEP 能够成功预测反常扩散指数，平均绝对误差为 $|\alpha_{\text{true}} - \alpha_{\text{pred}}| = 0.271$，其结果与该任务中得分最高的方法相当 [36, 41]，这些方法参加了反常扩散挑战赛 [18]。

在图2D 中，我们进一步报告了推断 $\alpha$ 时的平均绝对误差，并将其作为分段长度的函数，与通过在对数空间中对时间平均均方位移进行线性拟合以及 CONDOR [35]，（AnDi 挑战赛相应任务中的最佳方法） [18]的结果进行比较。如前所述，这两种方法使用的是预分段轨迹，而 STEP 则直接处理异质轨迹。STEP 的表现远优于时间平均均方位移方法，与 CONDOR 相当。对于这一任务，向 STEP 提供预分段数据略微提升了其在长片段上的表现，但反而降低了其在短片段上的表现。这一结果表明，分段长度比精确获知片段边界更为重要，而且 STEP 能有效地结合局部信息与全局信息。

在附录 C中，我们扩展了对 STEP 性能的评估，考察其与轨迹的定位精度以及驻留时间/段数的关系，结果表明，该方法可有效应用于广泛的实验条件。此- 外，在附录 D中，我们扩展了关于反常扩散的分析，展示了 STEP 性能与底层扩散模型及定位精度之间的关系。

### 检测异质轨迹中的扩散变点

对于扩散性质发生突变的轨迹，准确知晓这些性质发生变化的点，对于推断系统的时间性质和动力学速率，以及完整表征其 underlying 物理过程至关重要。尽管 STEP 不会显式检测变点，但其输出能够精确估计预计会发生变化的扩散性质，从而相较于使用原始数据，简化变点检测及定位任务。为突出这一能力，我们比较了采用最先进的核变点检测（KCPD）方法 [44, 45]分别应用于 STEP 预测结果和轨迹位移时间序列时所获得的结果。为评估该方法的性能，我们计算了杰卡德指数（JI），并将位于阈值距离 $\mathcal{E}$ 以内的任何变点预测视为真正例。

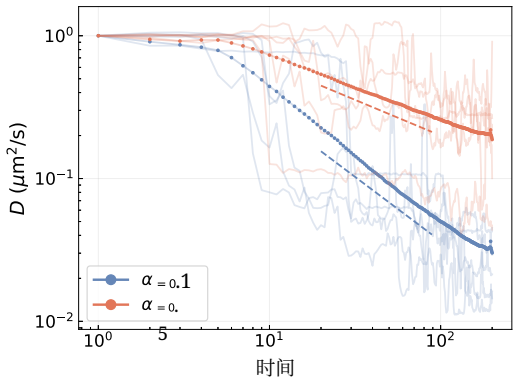


图 3. **扩散性质的连续变化**。具有 $\alpha = 0.1$ （蓝色）和 0.5（红色）的两组缩放布朗运动轨迹的扩散系数预测。细实线表示示例单条轨迹的预测，粗实线表示对 3000 条轨迹取平均后的结果，虚线表示每组理论预期的标度关系。

我们首先量化该方法检测执行布朗运动的异质轨迹中扩散系数变化的性能。我们使用一个基准数据集，其中的轨迹均包含一个变点，且 $\mathcal{E} = 5$。通过对 STEP 的预测应用 KCPD算法，即使相邻片段的扩散系数仅相差一个数量级，我们也能以很高的准确率成功检测出变点，如图 2E（蓝线）所示。此外，该方法对于轨迹中变点位置的变化也具有鲁棒性（图 2F（蓝线））。相比之下，直接对轨迹位移应用 KCPD（紫色虚线）时，我们观察到在整个扩散系数比范围内，性能均有所下降。平均而言，STEP 将误差降低了 20%，平均 JI 达到 0.833，而原始位移的平均 JI 为 0.796。

我们采用类似的分析方法，检测分数布朗运动轨迹中反常扩散指数的变化。由于反常扩散指数是一种渐近性质，无法从原始数据中轻易计算，因此，为了建立基线，我们使用滑动窗口对时间平均均方位移在双对数坐标下进行线性拟合，窗口长度为 30 个时间步长，计算得到 $\alpha$ ，然后将其输入 KCPD 算法（黄色虚线）。不出所料，分段参数之间的差异越大，我们就越能能够检测变点，如图2G 所示。与基线方法相比，使用 STEP 可将误差降低 30%，在 $\mathcal{E} = 20$ 条件下，两者的平均 JI 分别达到 0.515 和 0.297。不过，性能图显示，发现 $\alpha$ 中的变化比发现 $D$ 中的变化更困难。此外，我们还观察到，当变点接近轨迹边缘时，性能会下降（图 2H）。在这些情况下，我们处理的是较短的片段，其反常扩散指数可能难以确定，因为这取决于长程相关性的产生。

### 揭示扩散性质的连续变化

在生物学背景下考察异质轨迹时，人们通常预期，粒子扩散过程中各项性质呈分段常量，并可能突然发生变化，例如与其他生物组分发生特定相互作用所致。然而，分子拥挤和浓度梯度会导致扩散性质随时间连续变化。由于实验的时空分辨率有限，或缺乏专门的轨迹分析方法，这些变化可能难以检测。STEP 以无模型的方式预测逐点扩散性质，因此天然具备进行此类分析的能力，即使未经专门训练也能做到。

为了评估 STEP 在扩散性质平滑变化的轨迹上的表现，我们依靠缩放布朗运动（SBM）的模拟 [11]。SBM 轨迹的特征是扩散系数随时间变化，并遵循幂律关系 $D(t) \sim t^{\alpha-1}$，其中 $\alpha$ 是反常扩散指数。关于模拟的更多细节见附录B3。

在图3中，我们展示了对轨迹每个时间步的扩散系数所作的预测，其$\alpha = 0.1$ 和0.5。阴影线表示对各条轨迹单独作出的预测；尽管存在波动，但已显现出下降趋势。对具有相同 $\alpha$ 的 3000 条轨迹取平均值（粗线），则显示出正确的标度关系，与预期的幂律（虚线）相符。此外，对 $\alpha$ 的推断正确地给出了近似恒定的值

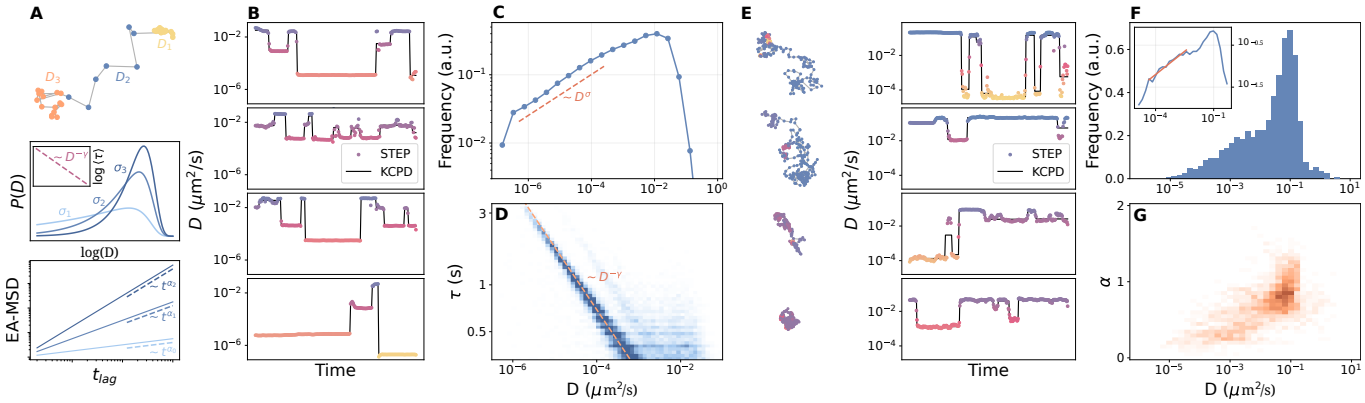


图4. **病原体识别受体 DC-SIGN 在随机扩散状态间的切换。** **(A)** ATTM模型的特征：一条扩散系数发生变化的示例轨迹、扩散系数的分布，以及扩散系数与驻留时间之间的关系。**(B)** 将 STEP 应用于模拟的 ATTM 轨迹（点）所得的扩散系数预测，以及变点分析结果（黑线）。**(C)** 通过**(B)**中所述分析得到的 $D$ 分布，显示了小 $D$时预期的幂律行为。**(D)** 通过**(B)**中所述分析得到的 $D$ 与驻留时间 $\tau$ 之间的关系，显示了预期的幂律行为。**(E)** DC-SIGN 实验轨迹示例，以及对应的$D$ 预测（点）和变点分析结果（黑线）。**(F)** 实验轨迹得到的 $D$ 分布直方图。插图：小 $D$时的幂律拟合。**(G)** 实验轨迹得到的 $D$ 和 $\alpha$ 的二维直方图。

贯穿整个轨迹，符合预期（更多细节请参见附录D），并表明对于反常扩散的变化，假阳性率较低。

### 根据正常扩散性质的变化表征反常扩散

为检验 STEP 分析实验轨迹的潜力，我们用它研究中国仓鼠卵巢细胞中表达的病原体识别受体 DC-SIGN 的运动 [46]。此前对这些实验的分析发现，扩散系数变化会导致反常扩散和弱遍历性破缺 [10]。这种行为可用退火跃迁时间模型（ATTM）来描述 [47], ，其主要特征示意总结于图4A。

简而言之，ATTM 描述布朗扩散的扩散系数随机切换 $D$，其取值从具有幂律行为的分布中采样 $D^{\sigma-1}$，适用于较小的 $D$，而在较大的 $D \to \infty$处则快速衰减。该模型还假设 $D$ 与驻留时间 $\tau$ 之间存在如下形式的相关性 $\tau(D) \sim D^{-\gamma}$，据此预测反常扩散指数 $\alpha = \sigma/\gamma$。因此，正确表征 $D$ 和 $\tau$ 的分布，对于证实其与底层模型的相容性至关重要。在原始研究中，通过变点分析检测扩散率变化 [48] ，但该方法的灵敏度和时间分辨率不足以对这一行为进行全面考察。

为了证明 STEP 能够更好地表征这些数据，我们首先使用模拟的 ATTM 轨迹。我们对轨迹应用前文介绍的 KCPD 算法，对 STEP 的预测结果进行分段，如图 4B 所示。因此，我们为每个片段分配一个 $D$，取该片段预测值的平均值，并分配一个 $\tau$。我们成功恢复了 $D$ 的幂律行为（图 4C），以及 $\tau$ 与 $D$ 之间的幂律关系（图 4D）。

然后，我们将这一方法应用于文献 [10]中的 DC-SIGN 轨迹。结果证实，不同片段之间存在扩散率变化，这些片段具有近似恒定的扩散系数和可变的持续时间，如图4E 所示。有趣的是，与先前的分析相比，我们的方法揭示出的变点数量多了一倍。

对轨迹分段获得的 $D$ 分布跨越了多个数量级，如图4F 的直方图所示。对于较小的 $D$，它在近三个数量级范围内呈现出与指数为 $\sigma \approx 0.37$的幂律相符的行为（图4F 的插图），与 ATTM 相符。值得注意的是，在原始文章中无法直接验证这一行为。原则上，我们的方法可以验证$D$ 与驻留时间之间的相关性，正如我们在模拟中所展示的那样。然而，这项工作受到可变轨迹长度 [49]以及统计数据不足的限制，尤其是对于具有较小 $D$ 的片段。

作为进一步检验，我们使用 STEP 预测反常扩散指数。我们通过对每个片段的预测取平均值，来为单个 $\alpha$ 赋予一个值。图 4G 中报告的结果显示， $D$与 $\alpha$ 之间存在有趣的相关性，这表明扩散模式更加复杂：反常扩散也会出现在单个片段的层面。

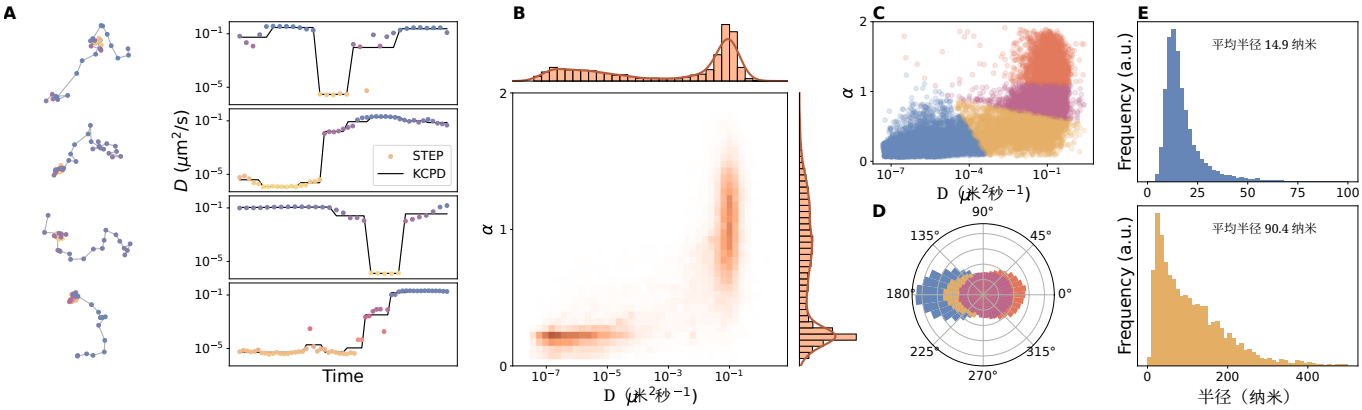


图5. **整合素的多状态扩散 $\alpha5\beta1$。(A)** 整合素 $\alpha5\beta1$ 的实验轨迹示例，以及针对 $D$ （点）得到的相应预测和变点分析结果（黑线）。**(B)** $D$ 和 $\alpha$ 的二维直方图及其相应的边缘分布。**(C)** 在分段层面对 $D$ 和 $\alpha$ 的预测结果进行散点图展示，并根据使用$k$均值算法进行的聚类分析着色。**(D)** 如**(C)**中所示获得的四个片段簇的转向角分布。**(E)** 表现出受限扩散的各簇的受限半径分布。

### 表征多状态扩散过程

我们使用 STEP 分析整合素 $\alpha5\beta1$ 在 HeLa 细胞细胞膜中扩散的实验轨迹。整合素是黏着斑中的细胞外基质（ECM）跨膜受体，它们在力学上连接 ECM 与细胞质中的肌动蛋白丝，并激活参与细胞迁移、增殖或细胞凋亡的信号通路 [27]。整合素 $\alpha5\beta1$ 的动力学受到其与纤连蛋白及肌动蛋白结合蛋白相互作用的影响 [49, 50]。据报道，它的运动会在快速自由扩散、缓慢自由扩散和固定化之间切换，同时还表现出由肌动蛋白驱动的向后运动。

我们使用 STEP 预测整合素$\alpha5\beta1$ 轨迹的扩散系数和反常扩散指数。接着，我们同时对这两项预测应用 KCPD 方法，对轨迹进行分割。通过这种方式，我们为每个片段赋予一个唯一的$D$ 和 $\alpha$ ，具体取片段内的平均预测值。图 5A 展示了结果示例，图 $D$ 和 $\alpha$ 的联合分布见图 5B。目视检查图 5B 可见两个主要簇，分别以 $(D = 10^{-6}\mu\text{m}^2/\text{s},\ \alpha = 0.25)$ 和 $(D = 0.1\,\mu\text{m}^2/\text{s},\ \alpha = 1)$ 为中心。对同一组参数在逐点层面（分割前）计算得到的二维直方图，与图 5B 相比没有明显差异。

然而，使用 $k$-均值算法对数据进行无监督聚类，并结合肘部法则进行模型选择 [51] ，揭示了四个具有不同运动特征的片段簇（图5C）。前两个簇表现出较为受限的运动，整合素有 40% 的时间处于 $D = 1.2^{-5}\mu\text{m}^2/\text{s}$ 和$\alpha = 0.23$状态，另有 14% 的时间处于 $D = 0.06\,\mu\text{m}^2/\text{s}$ 和$\alpha = 0.46$状态。对于这两个簇，角度分布在连续步之间的分布显示，峰值集中在 180°，这表明由于在潜在边界处发生反射而产生了向后运动，如图5D 所示。第一簇的受限半径中位数为14。9nm（标准差为 ±13.3nm），与这些实验的定位精度相当。因此，我们可以将其归因于蛋白质固定。第二簇在大小分布较广的区域内表现出受限运动（见图5E），其半径中位数为 90。4 nm（标准差为 ±98.2nm）。第三簇占总记录时间的 29%，其运动与布朗运动仅有轻微偏差，表现为 $\alpha = 0.88$ ，且角度分布近乎均匀；其平均值为 $D = 0.10\,\mu\text{m}^2/\text{s}$，接近文献中通常报道的该蛋白质数值。有趣的是，分析还发现了第四群体，占总记录时间的 20%，表现出超扩散，其 $\alpha = 1.3$ 和$D = 0.14\,\mu\text{m}^2/\text{s}$，并且连续步之间的运动方向具有持续性（图5D）。

## III. 讨论

在这项工作中，我们介绍了 STEP，这是一种机器学习方法，可在每个时间步从单条轨迹预测扩散性质。该方法结合了多种先进的机器学习架构，并考虑不同时间尺度上的相关性。所提出的方法尤其适用于分析经历异质运动的粒子轨迹，因为这类运动的扩散性质会随时间发生变化。此外，该方法不需要预先了解底层物理过程，也不需要知道扩散变化发生时所对应的时间分辨率。

为说明 STEP 的预测能力，我们在各种条件下的模拟轨迹上对其进行了基准测试。我们展示了其预测分段常量扩

散性质的能力，例如在含噪且较短的轨迹中预测扩散系数或反常扩散指数。此外，我们证明 STEP 能够提高变点检测算法的准确性，从而检测扩散变化发生的时间。重要的是，我们还证明了该方法适用于研究扩散的连续变化。为进一步展示该方法的潜在应用，我们研究了通过活细胞单分子成像实验追踪质膜蛋白获得的轨迹。首先，我们表征了病原体识别受体 DC-SIGN 的运动；已有研究表明，该受体的扩散系数会发生随机变化。我们的分析证实了这一假设，并提高了检测这些变化的准确性。此外，我们的结果提示可能存在更为复杂的现象，有待进一步研究。其次，我们研究了整合素 $\alpha_5\beta_1$的扩散。与先前研究一致，我们的分析证实了不同扩散模式的存在，并能够根据扩散系数、反常扩散指数以及空间约束程度对这些模式进行精确分类。

我们认为，STEP 是一类新型机器学习算法的第一步，可通过序列到序列方法研究动态系统。聚焦于对感兴趣属性的瞬时预测，使我们能够在实验时间分辨率下表征轨迹，而无需进行平均和滤波，并最大限度地减少分析所需的先验知识。因此，利用 STEP 获取的信息能够以前所未有的分辨率揭示扩散性质，从而深入了解各种系统中潜在的物理过程。

## 材料与方法

### 机器学习模型

在扩散现象中，我们面对的是复杂的统计信号，这些信号可能呈现多种类型的时间相关性。因此，我们需要一种能够捕捉不同时间尺度上相关性的机器学习（ML）模型。此外，即使在同一实验中，我们也常会遇到长度差异很大的轨迹。因此，模型必须不受长度限制，以确保其具有尽可能广泛的适用性。在这项工作中，我们提出采用一种基于序列到序列学习进展的机器学习技术 [21], ，这种技术在自然语言处理任务及其他领域都取得了令人瞩目的成果 [52–54]。与翻译任务类似，我们实现了一种模型，以粒子轨迹为输入，并在每个时间步输出感兴趣的扩散性质。这样，输入和输出的长度相同。

在我们的框架中，输入轨迹$\mathbf{x}$是一个 $d$-维度任意的向量 $T$，其元素$x_t$对应于每个时间步的粒子位置，$t \in [1, T]$。在本研究中，我们考虑二维轨迹（$d = 2$），但该方法很容易推广到任意维度。另一方面，输出是一个长度为 $T$ 的一维向量，其元素包含所关注扩散属性的值。例如，如果我们研究扩散系数沿轨迹的变化，则输出元素就是每个时间步的扩散系数， $D_t$。有关该流程的示意图，请参见图1 。如果我们关注的属性不止一个，可以扩展输出维度，例如预测$[D_t, \alpha_t]$，也可以为每个属性分别建立独立模型。本研究采用了第二种方案。

我们提出了一种结合卷积与自注意力机制的架构。有趣的是，近期的研究已展示了类似策略，既包括监督方法 [20]，也包括无监督方法 [55]。首先，输入轨迹经过一系列卷积层处理，以捕捉短程相关性。我们按照 XResNet 架构 [43]实现这些卷积层。随后，处理结果输入 Transformer编码器 [26], ，以捕捉全局相关性。最后，我们使用逐点全连接层，将输出维度调整至所需大小。

详细的架构和训练过程说明请参见附录B 。我们还提供了一个完整的程序库，其中包含代码以及如何复现本研究结果的详细说明，见参考文献 [56]。

### 细胞培养与接种

在涉及整合素 $\alpha5\beta1$ 的活细胞单分子成像实验中，HeLa 细胞培养于 DMEM 培养基（Gibco，11960-044）中，并添加 10%（v/v）胎牛血清（FBS，Sigma）。使用聚合酶链式反应（PCR）（Biotools 试剂盒，4542）检测细胞是否受到支原体污染。进行荧光成像时，将纤连蛋白（FN，Sigma，F2008）涂布于玻底培养皿（IBIDI，81158）上：将 $10\,\mu$g/mL 的 FN 加至玻璃表面，在37 ℃ 下放置 1 h，随后加入 2 mg/mL 牛血清白蛋白，在37 ℃ 下封闭 1 h。细胞以每皿 $5\times10^4$ 个细胞的密度接种，并培养 24 h 后使用。

### 半抗体片段的制备

半抗体片段采用与 [6] 中类似的方案制备。简而言之，将小鼠抗人整合素 $\alpha5$ 抗体（50 $\mu$L；BD Biosciences，610633）在室温下置于 PBS 中透析过夜（ThermoFisher，Slide-A-Lyze MINI 透析装置，2K），以置换商业缓冲液。随后，在室温下用 1 mM 二硫苏糖醇处理抗体 30 分钟，

然后在室温下使用 Slide-A-Lyze MINI 透析装置 2K 对其进行再次透析 4 小时，以去除二硫苏糖醇。为避免还原后的抗体重新缔合，在 4 ℃ 下搅拌孵育 1 小时，以 20 mM 碘乙酰胺封闭巯基。随后通过在 4 ℃ 下透析过夜，从反应液中去除碘乙酰胺。最后，在室温下搅拌 30 分钟，用 10 倍摩尔过量的 EZ-Link Sulfo-NHS_LC_Biotin（Thermo Scientific）对还原后的抗体进行生物素化，并在 4 ℃ 下保存至使用。

**单分子标记**

生物素化半抗体片段与链霉亲和素包被量子点（QD655 链霉亲和素偶联物，Invitrogen，Q10123mp）偶联。细胞用洗涤缓冲液（含 6% 牛血清白蛋白的磷酸盐缓冲盐水）清洗 3 次，并在洗涤缓冲液中使用半抗体-量子点（约 1 nM）进行标记（每个培养皿 200 $\mu$L），于 37 ℃ 下孵育 15 分钟，随后清洗两次。

**活细胞单分子成像**

使用 Leica DMi8 荧光显微镜进行成像。样品采用全内反射荧光（TIRF）几何构型进行照明。使用连续波激光器（Obis，Coherent，$\lambda$=488nm，$<$1kW/cm$^2$）进行激发。荧光通过油浸物镜（Leica，100X，NA=1.47）和科学级互补金属氧化物半导体相机（Photometrics 95B）进行记录，并配备适当的滤光片（Chroma）。影片以 33 Hz 的帧率记录。记录期间，使用显微镜环境舱（Okolab）使细胞保持在 5% $CO_2$ 气氛中。

**单粒子跟踪**

使用 u-track 进行粒子检测和跟踪 [57]。检测（高斯混合模型拟合）和跟踪参数根据对所得检测与跟踪结果的目视检查及性能诊断进行优化。所有图像和数据分析任务均在 MATLAB 2020a 及更高版本（The MathWorks，美国马萨诸塞州纳蒂克）中完成。视频使用 Bio-Formats [58]加载到 MATLAB 中。


[1] Leonhard Möckl、Don C Lamb 和 Christoph Bräuchle。超分辨荧光显微镜：2014年诺贝尔化学奖得主 Eric Betzig、Stefan Hell 和 William E. Moerner。《德国应用化学》国际版，53(51):13972–13977，2014年。[2] Carlo Manzo 和 Maria F Garcia-Parajo。单粒子追踪研究进展综述：从方法到生物物理学见解。《物理学进展报告》，78(12):124601，2015年。[3] Shuhui Yin、Nancy Song 和 Haw Yang。单粒子轨迹中速度和扩散系数变化点的检测。《生物物理学杂志》，115：229，2018年 2 月 217( ): – 。[4] Suvrajit Saha、Il-Hyung Lee、Anirban Polley、Jay T Groves、Madan Rao 和 Satyajit Mayor。动态皮层肌动蛋白的活性影响GPI锚定蛋白的扩散。《细胞分子生物学》，26(22):4033–4045，2015年。[5] Nirmalya Bag、Shuangru Huang 和 Thorsten Wohland。静息状态和配体结合状态下表皮生长因子受体的质膜组织。《生物物理学杂志》，109(9):1925–1936，2015年。[6]Shalini T Low-Nam、Keith A Lidke、Patrick J Cutler、Rob C Roovers、Paul MP van Bergen en Henegouwen、Bridget S Wilson 和 Diane S Lidke。Erbb1二聚化受到结构域共限制的促进，并通过配体结合得到稳定。《自然：结构与分子生物学》，18(11):1244–1249，2011年。[7] Adal Sabri、Xinran Xu、Diego Krapf 和 Matthias Weiss。阐明哺乳动物细胞细胞质中异质反常扩散的起源。《物理评论快报》，125(5):058101，2020年。

[8] Jae-Hyung Jeon、Matti Javanainen、Hector Martinez- Seara、Ralf Metzler 和 Ilpo Vattulainen。脂质双层中的蛋白质拥挤导致磷脂和蛋白质出现非高斯反常侧向扩散。*Physical Review X*，6(2):021006，2016年。[9] Thomas J Lampo、Stella Stylianidou、Mikael P Back- lund、Paul A Wiggins 和 Andrew J Spakowitz。细胞质中的RNA-蛋白质粒子表现出非高斯次扩散行为。《生物物理学杂志》，112 542，2017年 3 532( ): – 。[10] Carlo Manzo、Juan A Torreno-Pina、Pietro Massignan、Gerald J Lapeyre Jr、Maciej Lewenstein 和 Maria F Garcia Parajo。随机扩散率导致活细胞中受体运动的弱遍历性破缺。*Physical Review X*，5(1):011021，2015年。[11] Jae-Hyung Jeon、Aleksei V Chechkin 和 Ralf Metzler。缩放布朗运动：一种具有时间依赖扩散率的悖论过程，用于描述反常扩散。*Physical Chemistry Chemical Physics*，16(30):15811–15817，2014年。[12] Anthony R Vega、Spencer A Freeman、Sergio Grinstein 和 Khuloud Jaqaman。用于瞬态迁移率分析的多步轨迹分割和运动分类。《生物物理学杂志》，114(5):1018–1025，2018年。[13] Marloes Arts、Ihor Smal、Maarten W Paul、Claire Wyman 和 Erik Meijering。利用深度学习和矩标度谱进行粒子迁移率分析。*Scientific Reports*，9(1):1–10，2019年。[14] Yann Lanoiselée 和 Denis S Grebenkov。通过局部凸包方法揭示单粒子轨迹中的间歇性特征。《物理评论*E*》，96 022144，2017年 2( ): 。

[15]Jonathan E Bronson、Jingyi Fei、Jake M Hofman、Ruben L Gonzalez Jr 和 Chris H Wiggins。从生物物理时间序列中学习速率和状态：一种用于模型选择和单分子FRET数据的贝叶斯方法。《生物物理学杂志》，97(12):3196–3205，2009年。[16] Fredrik Persson、Martin Lindén、Cecilia Unoson 和 Jo- han Elf。从单分子追踪数据中提取细胞内扩散状态和转变速率。《自然·方法》，10(3):265–269，2013年。[17]Nilah Monnier、Zachary Barry、Hye Yoon Park、Kuan- Chung Su、Zachary Katz、Brian P English、Arkajit Dey、Keyao Pan、Iain M Cheeseman、Robert H Singer 等人。推断活细胞中的瞬态粒子运输动力学。《自然·方法》*style>，12(9):838–840，2015*年。*Gorka Muñoz-Gil*、*Giovanni Volpe*、*Miguel Angel Garcia-March*、*Erez Aghion*、*Aykut Argun*、*Chang Beom Hong*、*Tom Bland*、*Stefano Bo*、*J Alberto Conejero*、*Nicolás Fir-bas* 等人。对反常扩散解码方法进行客观比较。《自然·通讯》，12(1):1– 16，2021年。[19]Hippolyte Verdier、François Laurent、Alhassan Cassé、Christian L Vestergaard 和 Jean-Baptiste Masson。具有线性计算复杂度的分数布朗运动变分推断。《物理评论*E*》，106(5):055311，2022年。[20]Jesús Pineda、Benjamin Midtvedt、Harshith Bachi- manchi、Sergio Noé、Daniel Midtvedt、Giovanni Volpe 和 Carlo Manzo。几何深度学习揭示微观运动的时空特征。《自然·机器智能》，第1–12页，2023年。[21] Ilya Sutskever、Oriol Vinyals 和 Quoc V Le。利用神经网络进行序列到序列学习。载于 Z. Ghahramani、M. Welling、C. Cortes、N. Lawrence 和 K.Q. Weinberger 编，《神经信息处理系统进展》，第27卷。Curran Associates, Inc.，2014年。[22] Tom Brown、Benjamin Mann、Nick Ryder、Melanie Sub- biah、Jared D Kaplan、Prafulla Dhariwal、Arvind Nee- lakantan、Pranav Shyam、Girish Sastry、Amanda Askell 等人。语言模型是少样本学习者。神经信息处理系统进展，33 1901，2020年 1877: – 。[23]Scott Reed、Konrad Zolna、Emilio Parisotto、Ser- gio Gomez Colmenarejo、Alexander Novikov、Gabriel Barth-Maron、Mai Gimenez、Yury Sulsky、Jackie Kay、Jost Tobias Springenberg 等人。一种通才智能体。*arXiv*预印本 *arXiv:2205.06175*，2022年。[24] Ross Taylor、Marcin Kardas、Guillem Cucurull、Thomas Scialom、Anthony Hartshorn、Elvis Saravia、Andrew Poulton、Viktor Kerkez 和 Robert Stojnic。Galactica：一种面向科学的大语言模型。*arXiv*预印本 *arXiv:2211.09085*，2022年。[25]Yann LeCun、Bernhard Boser、John S Denker、Don- nie Henderson、Richard E Howard、Wayne Hubbard 和 Lawrence D Jackel。将反向传播应用于手写邮政编码识别。《神经计算》，1(4):541–551，1989年。[26]Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。注意力机制即全部所需。神经信息处理系统进展，30，2017年。[27] Pakorn Kanchanawong 和 David A Calderwood。整合素介导的细胞-细胞外基质黏附的组织结构、动力学和力学调控。《自然综述：分子》细胞生物学，第 1–20 页，2022年。[28] Gorka Muñoz-Gil、Miguel Angel Garcia-March、Carlo Manzo、José D Martín-Guerrero 和 Maciej Lewenstein。通过机器学习表征单条轨迹。新物理学杂志，22(1):013010，2020年。[29] Naor Granik、Lucien E Weiss、Elias Nehme、Maayan Levin、Michael Chein、Eran Perlson、Yael Roichman 和 Yoav Shechtman。利用深度学习表征单粒子扩散。生物物理学杂志，117(2):185– 192，2019年。[30] Stefano Bo、Falko Schmidt、Ralf Eichhorn 和 Giovanni Volpe。使用循环神经网络测量反常扩散。物理评论 *E*，100:010102，2019年7月。[31] Patrycja Kowalek、Hanna Loch-Olszewska 和 Janusz Szwabiński。单粒子追踪数据中的扩散模式分类：基于特征的方法与深度学习方法。物理评论*E*，100 032410，2019 3( ): 。[32] Henrik Seckler 和 Ralf Metzler。贝叶斯深度学习在反常扩散分析误差估计中的应用。自然·通讯，13(1):6717，2022年。[33]Vida Jamali、Cory Hargus、Assaf Ben-Moshe、Amirali Aghazadeh、Hyun Dong Ha、Kranthi K Mandadapu 和 A Paul Alivisatos。深度学习辅助分析揭示液相透射电子显微镜中的反常纳米粒子表面扩散。美国国家科学院院刊，118(10):e2017616118，2021年。[34] Gorka Muñoz-Gil、Catalina Romero-Aristizabal、Nicolas Mateos、Felix Campelo、Lara I de Llobet Cucalon、Miguel Beato、Maciej Lewenstein、Maria F Garcia-Parajo 和 Juan A Torreno-Pina。随机粒子解离调节活细胞中转录因子凝聚体的生长动力学和大小。美国国家科学院院刊，119(31):e2200667119，2022年。[35] Alessia Gentili 和 Giorgio Volpe。利用深度学习表征反常扩散的经典统计学（CONDOR）。物理学杂志 *A* 数学通论，54(31):314003，2021年。[36] Aykut Argun、Giovanni Volpe 和 Stefano Bo。使用循环神经网络对反常扩散进行分类、推断和分割。物理学杂志 *A*：数学与理论，54(29):294003，2021年。[37] Héctor Buena Maizón 和 Francisco J Barrantes。一种基于深度学习的膜蛋白反常扩散建模方法：以烟碱型乙酰胆碱受体为例。生物信息学简报，23 10 2021 1 435 ( )，。bbab 。[38]Hippolyte Verdier、Maxime Duval、François Laurent、Alhassan Cassé、Christian L. Vestergaard 和 Jean- Baptiste Masson。使用图神经网络学习反常随机游走的物理性质。物理学杂志 *A*：数学与理论，54(23):234001，2021年5月。[39] Carlo Manzo。用于从单条轨迹表征反常扩散的极限学习机（ANDI-ELM）。物理学杂志 *A*：数学与理论，54(33):334002，2021年。[40] Dezhong Li、Qiujin Yao 和 Zihan Huang。基于 WaveNet 的深度神经网络用于反常扩散表征（WADNet）。物理学杂志 *A*：数学与理论，54(40):404003，2021年9月。[41] Òscar Garibo-i Orts、Alba Baeza-Bosca、Miguel A Garcia-March 和 J Alberto Conejero。用于反常扩散的高效循环神经网络方法

单粒子的短时且含噪轨迹。物理学杂志 *A*：数学与理论，54 504002，2021 50( ): 。[42] Nicolás Firbas、Òscar Garibo i Orts、Miguel Angel Garcia-March 和 J Alberto Conejero。通过卷积变换器表征反常扩散。物理学杂志 *A*：数学与理论，56:014001，2023年。[43] Tong He、Zhi Zhang、Hang Zhang、Zhongyue Zhang、Jun- yuan Xie 和 Mu Li。用于卷积神经网络图像分类的技巧集。载于*IEEE/CVF*计算机视觉与模式识别会议，第 558–567 页，2019年。[44] A. Celisse、G. Marot、M. Pierre-Jean 和 G.J. Rigaill。基于再生核的多变点检测新型高效算法。计算统计与数据分析，128:200–220，2018年。[45] Sylvain Arlot、Alain Celisse 和 Zaid Harchaoui。通过模型选择实现的核多变点算法。机器学习研究杂志，20(162):1– 56，2019年。[46] Carlo Manzo、Juan A Torreno-Pina、Ben Joosten、Inge Reinieren-Beeren、Emilio J Gualda、Pablo Loza-Alvarez、Carl G Figdor、Maria F Garcia-Parajo 和 Alessandra Cambi。C型凝集素 DC-SIGN 的颈部区域调节其在抗原呈递细胞表面的时空组织及病毒结合能力。生物化学杂志，287(46):38946–38955，2012年。[47] P Massignan、C Manzo、JA Torreno-Pina、MF García-Parajo、M Lewenstein 和 GJ Lapeyre Jr.。非均匀介质中的布朗运动导致非遍历亚扩散。物理评论快报，112(15):150603，2014年。[48] D Montiel、H Cang 和 H Yang。定量表征单粒子追踪研究中动力学行为的变化。物理化学杂志 *B*，110(40):19763–19770，2006年。[49]Taka A Tsunoyama、Yusuke Watanabe、Junri Goto、Kazuma Naito、Rinshi S Kasai、Kenichi GN Suzuki、Takahiro K Fujiwara 和 Akihiro Kusumi。超长时间单分子追踪揭示动态锚定诱导的整合素功能。自然·化学生物学，14(5):497–506，2018年。[50]Olivier Rossier、Vivien Octeau、Jean-Baptiste Sibarita、Cecile Leduc、Béatrice Tessier、Deepak Nair、Volker Gatterdam、Olivier Destaing、Corinne Albiges-Rizo、Robert Tampé 等。整合素 $\beta 1$ 和 $\beta 3$ 在黏着斑内部呈现不同的动态纳米尺度组织。自然细胞生物学，14(10):1057–1067，2012年。[51] Ville Satopaa、Jeannie Albrecht、David Irwin 和 Barath Raghavan。在大海捞针：检测系统行为中的膝点。载于*2011*年第*31*届分布式计算系统国际会议研讨会，第 166–171 页。IEEE，2011年。[52]Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell、Sandhini Agarwal、Ariel Herbert-Voss、Gretchen Krueger、Tom Henighan、Rewon Child、Aditya Ramesh、Daniel Ziegler、Jeffrey Wu、Clemens Winter、Chris Hesse、Mark Chen、Eric Sigler、Mateusz Litwin、Scott Gray、Benjamin Chess、Jack Clark、Christopher Berner、Sam McCandlish、Alec Radford、Ilya Sutskever 和 Dario Amodei。语言模型是少样本学习者。载于 H. Larochelle、M. Ranzato、R. Hadsell、M.F. Balcan 和 H. Lin，编者，神经信息处理系统进展，第33卷，第1877–1901页。Curran Associates, Inc.，2020年。[53] Jean-Baptiste Alayrac、Jeff Donahue、Pauline Luc、Antoine Miech、Iain Barr、Yana Hasson、Karel Lenc、Arthur Mensch、Katie Millican、Malcolm Reynolds、Roman Ring、Eliza Rutherford、Serkan Cabi、Tengda Han、Zhi-tao Gong、Sina Samangooei、Marianne Monteiro、Jacob Menick、Sebastian Borgeaud、Andrew Brock、Aida Nematzadeh、Sahand Sharifzadeh、Mikolaj Binkowski、Ricardo Barreira、Oriol Vinyals、Andrew Zisserman 和 Karen Simonyan。Flamingo：用于小样本学习的视觉语言模型，2022年。[54]Scott Reed、Konrad Zolna、Emilio Parisotto、Sergio Gomez Colmenarejo、Alexander Novikov、Gabriel Barth-Maron、Mai Gimenez、Yury Sulsky、Jackie Kay、Jost Tobias Springenberg、Tom Eccles、Jake Bruce、Ali Razavi、Ashley Edwards、Nicolas Heess、Yutian Chen、Raia Hadsell、Oriol Vinyals、Mahyar Bordbar 和 Nando de Freitas。通才智能体，2022年。[55] Hélène Kabbech 和 Ihor Smal。使用无监督深度学习方法识别跟踪应用中的扩散状态。载于*2022*年 *IEEE* 第*19*届生物医学成像国际研讨会（*ISBI*），第1–4页。IEEE，2022年。[56] Borja Requena 和 Gorkal Muñoz-Gil。Borjarequena/step，2022年12月。[57]Khuloud Jaqaman、Dinah Loerke、Marcel Mettlen、Hirotaka Kuwata、Sergio Grinstein、Sandra L Schmid 和 Gaudenz Danuser。活细胞延时序列中的稳健单粒子跟踪。《自然·方法》，5 702，2008 8 695( ): – 。[58]Melissa Linkert、Curtis T Rueden、Chris Allan、Jean-Marie Burel、Will Moore、Andrew Patterson、Brian Loranger、Josh Moore、Carlos Neves、Donald MacDonald 等。元数据至关重要：现实世界中的图像数据访问。《细胞生物学杂志》，189(5):777–782，2010年。[59] Ralf Metzler、Jae-Hyung Jeon、Andrey G Cherstvy 和 Eli Barkai。反常扩散模型及其性质：单粒子追踪百年之际的非平稳性、非遍历性和老化。*Physical Chemistry Chemical Physics*，16(44):24128–24164，2014年。[60] Joseph Klafter 和 Igor M Sokolov。随机游走入门：从工具到应用。牛津大学出版社，2011年。[61] Harvey Scher 和 Elliott W. Montroll。非晶固体中的反常渡越时间弥散。*Phys. Rev. B*，12:2455–2477，1975年9月。[62] Benoit B. Mandelbrot 和 John W. Van Ness。分数布朗运动、分数噪声及其应用。《*SIAM*评论》，10(4):422–437，1968年。[63] J. Klafter 和 G. Zumofen。哈密顿系统中的莱维统计。*Phys. Rev. E*，49:4873–4877，1994年6月。[64] S. C. Lim 和 S. V. Muniandy。用于建模反常扩散的自相似高斯过程。*Phys. Rev. E*，66:021114，2002年8月。[65] Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。PyTorch：一种命令式风格的高性能深度学习库。载于 H. Wallach、H. Larochelle、A. Beygelzimer、F. d'Alché-Buc、E. Fox 和 R. Gar-

nett，编者，神经信息处理系统进展，第32卷。Curran Associates, Inc.，2019年。[66] Jeremy Howard 和 Sylvain Gugger。fastai：一种分层深度学习 API。*Information*，11(2)，2020年。[67] Charles Truong、Laurent Oudre 和 Nicolas Vayatis。离线变点检测方法综述。信号处理，167:107299，2020年。[68] Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。载于*ICLR*（海报），2015年。[69] Leslie N. Smith 和 Nicholay Topin。超收敛：利用大学习率快速训练神经网络。载于 Tien Pham 编，面向多域行动的人工智能与机器学习应用，第11006卷，第369–386页。国际光学与光子学学会，SPIE，2019年。[70] Geoffrey E. Hinton、Nitish Srivastava、Alex Krizhevsky、Ilya Sutskever 和 Ruslan R. Salakhutdinov。通过防止特征检测器的协同适应来改进神经网络，2012年。[71] Stephen Hanson 和 Lorien Pratt。比较通过反向传播进行最小网络构建的偏差。载于 D. Touretzky 编，神经信息处理系统进展，第1卷。Morgan-Kaufmann，1988年。[72] Anders Krogh 和 John Hertz。简单的权重衰减可以提升泛化能力。载于 J. Moody、S. Hanson 和 R.P. Lippmann 编，神经信息处理系统进展，第4卷。Morgan-Kaufmann，1991年。[73] Gorka Muñoz-Gil、Borja Requena、Giovanni Volpe、Miguel Angel Garcia-March 和 Carlo Manzo。AnDiChallenge/ANDI_数据集：Challenge 2020 发布版，2021年5月。

## 致谢

作者感谢 Montserrat Masoliver-Prieto 和 Marta Cullell-Dalmau 在实验流程方面给予的宝贵帮助。GMG 感谢欧盟（ERC、QuantAI，项目编号 10105529）以及奥地利科学基金（FWF）通过 SFB BeyondC F7102 提供的支持。CM 感谢由 MCIN/AEI/10.13039/501100011033 和“ESF 投资你的未来”资助的 RYC-2015-17896 项目、由 MCIN/AEI/10.13039/501100011033/ 和 FEDER“ERDF 建设欧洲之路”资助的 BFU2017- 85693-R 和 PID2021-125386NB-I00 项目，以及由加泰罗尼亚自治区政府资助的 AGAUR 2017SGR940 项目提供的支持。BR 和 ML 感谢以下机构和项目提供的支持：ERC AdG NOQIA；西班牙科学与创新部国家研究局（PGC2018-097027-B-I00/10.13039/501100011033、CEX2019-000910-S/10.13039/501100011033、国家计划 FIDEUA PID2019-106901GB-I00、FPI、QUANTERA MAQS PCI2019-111828-2、QUANTERA DYNAMITE PCI2022-132919、I+D+I 项目“合作挑战”QUSPIN RTC2019-007196-7）；MICIIN 获得欧盟下一代欧盟（PRTR- C17.I1）和加泰罗尼亚自治区政府的资金支持；Cellex基金会；Mir-Puig基金会；加泰罗尼亚自治区政府（欧洲社会基金、FEDER 和 CERCA计划，AGAUR 2021 SGR 01452 号资助、QuantumCAT \ U16-011424（由加泰罗尼亚 ERDF 运营计划 2014-2020 共同资助）；巴塞罗那超级计算中心 MareNostrum（FI-2022-1-0042）；欧盟地平线 2020 FET-OPEN OPTOlogic（资助编号 899794）；欧盟地平线欧洲计划（资助协议 101080086 — NeQST）、波兰国家科学中心（Symfonia Grant 资助编号 2016/20/W/ST4/00314）；ICFO 内部“QuantumGaudi”项目；欧盟玛丽·斯克沃多夫斯卡-居里研究与创新计划地平线 2020 项目资助协议编号 101029393（STREDCH）和编号 847648（“La Caixa”青年领袖奖学金 ID100010434：LCF/BQ/PI19/11690013、LCF/BQ/PI20/11760031、LCF/BQ/PR20/11770012、LCF/BQ/PR21/11840013）。然而，本研究中表达的观点和意见仅代表作者的观点，并不一定反映欧盟、欧洲气候、基础设施和环境执行机构（CINEA）或任何其他资助机构的立场。欧盟和任何资助机构均不对此承担责任

## 附录 A：扩散性质

本节简要介绍正常扩散和反常扩散的一些主要特征。关于该领域全面而精彩的入门介绍，请参阅文献 [59, 60] 。

扩散轨迹通常通过其均方位移（MSD）来描述。对于布朗运动，均方位移随时间呈线性标度关系，即 MSD $\propto Dt$，其中 $D$ 为扩散系数。然而，均方位移也可能偏离这种线性标度关系，从而与时间呈幂律关系，即 MSD $\propto K_\alpha t^\alpha$，其中 $\alpha$ 定义为反常扩散指数， $K_\alpha$ 为有效扩散系数。前者可用于区分正常（或布朗）扩散（$\alpha = 1$）和反常扩散（$\alpha \neq 1$）。

反常扩散的出现可能与截然不同的现象有关，从扩散粒子运动中相关性的产生，到时空异质性的存在。为解释其中大多数现象，我们遵循文献 [18] ，考虑五种反常扩散模型及其反常扩散指数的特定取值范围：退火瞬态时间模型（ATTM） [47] ，其取值为 $\alpha \in [0.05, 1]$；连续时间随机游走（CTRW） [61] ，其取值为 $\alpha \in [0.05, 1]$；分数布朗运动（FBM） [62] ，其取值为 $\alpha \in [0.05, 1.95]$；莱维行走（LW） [63] ，其取值为 $\alpha \in [1.05, 2]$；以及缩放布朗运动（SBM） [64] ，其取值为 $\alpha \in [0.05, 2]$。

## 附录 B：机器学习流程

这里，我们详细说明了为获得本文所述结果而采用的机器学习方法

在本研究中始终如此。

正如我们在正文中简要提到的，我们训练了两个不同的模型：一个用于扩散系数任务，另一个用于反常扩散指数任务。关于扩散系数和反常扩散指数预测的结果见第 II节。我们按照相同的原则实现这两个模型，仅有细微差异。本节将介绍我们采用的模型架构和训练过程，并在必要时指出模型之间的差异。

我们提供了源代码，并附有关于如何复现结果的详细说明，见 [56]。我们大量使用 $PyTorch$ [65] 和 $fastai$ [66] 程序库来实现模型架构和训练过程。核变点检测方法（KCPD）使用 $ruptures$ Python 程序库 [67]实现。

### 1. 架构细节

我们提出采用一个以轨迹 $\mathbf{x}$ 为输入，并在每个时间步输出目标扩散性质的模型。输入轨迹是一个 $d$维向量，长度任意 $T$，其元素 $x_t$对应每个时间步的粒子位置 $t$。输出则是一个长度为 $T$的一维向量，其元素对应每个时间步关注的扩散属性，例如， $D_t$ 在扩散系数的情况下。关于维度的更多细节请参见图 6 。在本文中，我们主要考虑维度为 $d = 2$的轨迹。

我们提出的模型由三个主要模块组成：用于处理输入轨迹的初始卷积部分；基于自注意力的部分，以前一部分提取的特征为输入；以及一个浅层的逐点全连接前馈模块，用于提供所需的输出维度。整个架构与长度无关，因此能够处理任意长度的轨迹。

**卷积模块 –** 第一个主要卷积模块通过多个卷积滤波器扩展轨迹维度，从而为后续层提供基于短程相关性的、更丰富的嵌入表示。

我们遵循 XResNet [43] 架构构建该模块。如图 6 所示，该模块由一个初始卷积层（通常称为 $stem$）和一系列 残差块 组成，每个残差块都包含一个带跳跃连接的卷积层。我们使用核大小为 3、步幅为 1 的一维卷积，以保持轨迹大小不变。不过，跳跃连接中的核大小为 1，可起到恒等映射或缩放因子的作用；其主要目的是匹配残差块两条路径上的张量形状，具体说明见下文。该模块可接收一批大小为 [批次_大小 $\times T \times d$] 的输入轨迹，并输出一批大小为 [批次_大小 $\times T \times$ 嵌入_大小]的特征。

在整个架构中，我们在每个卷积层后紧接着添加一个批归一化层；默认情况下，我们使用整流线性单元（ReLU）作为激活函数，最后的输出层除外。

为了得到第 II 节中的结果，我们在主干网络中使用一个卷积层和一个 ReLU 激活函数。我们使用 64 个滤波器来预测扩散系数，使用 32 个滤波器来预测反常扩散指数。随后，我们添加了三个残差块，分别包含 128、256 和 512 个滤波器。因此，嵌入_大小 $= 512$。在这些模块中，卷积路径包含两个卷积层：第一个层增大嵌入维度，第二个层则保持维度不变。在跳跃连接中，我们仅使用一个卷积层来增大嵌入维度，使其与卷积路径的维度相匹配。我们在模块末尾实现 ReLU 激活函数，即在将两条路径的输出相加之后。

**自注意力模块 –** 我们使用自注意力机制处理卷积模块提取的特征，使模型能够捕获长程相关性。

更确切地说，我们实现了一个*Transformer*编码器，其引入方式见参考文献 [26]。如图6所示，编码器块包含两个主要部分，二者都采用跳跃连接，并在两条路径求和后进行层归一化。第一部分是一个以输入为输入的多头注意力层；第二部分则包含一对逐点前馈层，它们分别作用于输入张量中的每个元素。此外，我们可以在第一个编码器块之前加入位置编码，为轨迹中每个元素的相对位置提供 Information。该模块可以处理一批嵌入，同时保持其维度不变。因此，输入和输出的大小均为[批次_大小 $\times T \times$ 嵌入_大小]。

为得到第 II 节中的结果，我们使用了四个带有八个头的变换器编码器块，应用于多头注意力层中。逐点前馈部分添加了两个全连接层，每层包含 嵌入_大小为 的神经元，即本例中为 512 个。有趣的是，我们发现，在卷积之后，位置编码对结果的影响非常小。因此，为了简洁起见，我们没有使用位置编码来获得本文报告的结果。

**前馈模块 –** 最后一个主要部分是一个浅层前馈全连接网络，它逐元素作用于前一个模块提取的特征。我们会根据具体任务对这一部分进行定制，以通过适当的维度获得所需的输出。

例如，在回归任务中，输出维度为 1，我们会在末尾使用缩放Sigmoid激活函数，并留出一定余量来定义输出范围，例如，$\log D \in (-3.1, 3.1)$，$\alpha \in (0, 2.05)$。这段余量使 Sigmoid 在达到饱和之前能够取到所需的值。假设有一个分类任务（例如，如参考文献 [18] 中所做的那样，对扩散模型进行分类），

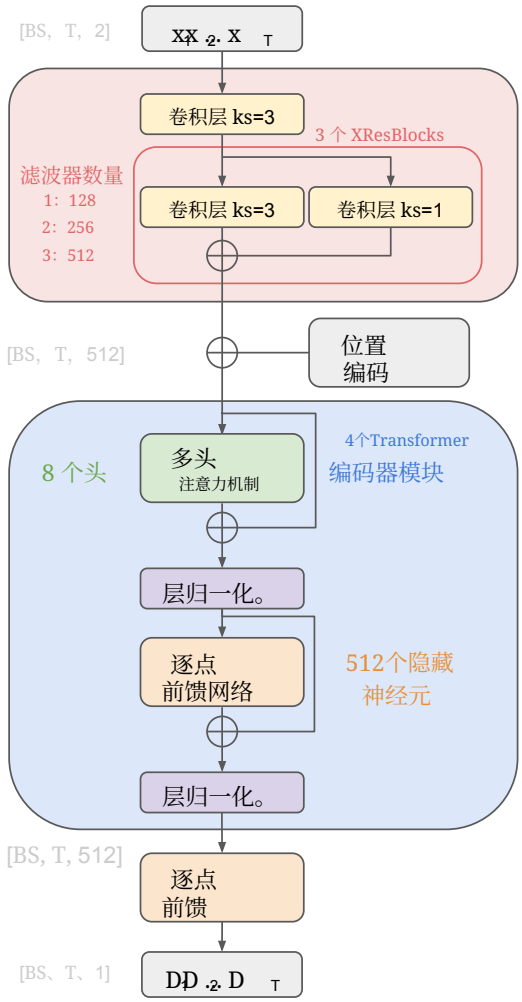


图6. 机器学习架构表示

最终维度即类别数量，我们使用 Softmax激活函数。然后，通过选择激活值最大的类别来获得预测结果。因此，我们可以处理大小为 [batch_size $\times$ $T$ $\times$ embedding_size] 的特征批次，并输出大小为 [batch_size $\times$ $T$ $\times$ num_class] 的预测结果。如果 num_class $>$ 1，如分类任务中的情况，我们会执行一个额外的后处理步骤，得到大小为 [batch_size $\times$ $T$ $\times$ 1] 的输出，其中包含各时间步对应的预测结果。

### 2. 训练过程

我们对两个模型都采用标准的基于梯度的训练过程。两者之间的唯一区别在于训练数据及其处理方式。主要训练循环包括：

1. 预测一批训练数据的数值。
2. 计算损失函数与真实值之间的差异。3. 根据损失的梯度更新模型参数。

我们使用包含128条轨迹的训练批次和L1损失函数，该函数对应于平均绝对误差。形式化地，

$$\mathcal{L}_{\mathrm{MAE}}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} |y_i - f(\mathbf{x}_i)| \,, \tag{B1}$$

其中，$f(\mathbf{x}_i)$ 表示一个包含 $n$个样本的训练批次中第 $i$条轨迹的预测。

为执行参数更新，我们使用 Adam [68]优化器。我们使用 fastai [66] 程序库，通过学习率查找工具选择学习率，其典型数量级为 $10^{-4}$。随后，我们按照参考文献 [69] 中提出的单周期策略，对训练批次中的学习率及其动量都采用调度方案。我们持续训练模型，直到验证集上的性能趋于稳定，通常训练十到二十轮。

为进一步防止过拟合并增强模型泛化能力，我们使用随机失活 [70]和权重衰减 [71, 72]。此外，我们以不同强度向轨迹中添加高斯局部化噪声，作为一种数据增强形式。

### 3. 训练、验证和测试我们的模型

为了妥善评估我们的模型，我们生成了多个相互独立的数据集。我们使用其中一个来训练和验证模型，并使用其他数据集在未见场景中测试模型。本研究中报告的所有结果均使用测试集获得；这些测试集旨在评估模型的不同方面。

表I中，我们详细介绍了用于训练、验证和测试模型的数据集。这些数据集包含模拟轨迹，以及每个时间步对应的标签。根据研究的是正常扩散还是反常扩散，我们采用两种主要方法来模拟轨迹。下面将说明这两种情况下数据的生成方式。

虽然正常扩散和反常扩散轨迹的模拟与标注方式存在一些差异，但它们也有若干共同点。例如，所有片段的扩散特性均为常量，且长度至少为十个时间步。

**布朗运动** – 我们将不相关高斯噪声作为轨迹位移，以此模拟布朗运动轨迹。通过高斯噪声的标准差来控制每个时间步的扩散系数，这对应于 $\sqrt{2D}$。这样，我们便能轻松生成任意长度的片段

| Task | 模型 | $D$ | $\alpha$ | $\sigma_{\text{noise}}$ | 轨迹长度 | 片段 | 片段长度 | | Size |
|---|---|---|---|---|---|---|---|---|---|
| **训练 BM** | 布朗运动 | $[-3,3]$ | 1 | $[-6,2]$ | 200 | $[2,5]$ | $[10,190]$ | | 100,000 |
| **训练 AnDi** | 全部反常 | 1 | $[0.05,2]$ | 0.1 | 200 | $[2,5]$ | $[10,190]$ | | 100,064 |
| 图 2A 和 B | 布朗运动 | $[-3,3]$ | 1 | 0 | 200 | $[2,5]$ | $[10,190]$ | | 48,000 |
| 图 2C，图 9<br>图 8C 和 D | 全部反常 | 1 | $[0.05,2]$ | $[-5,2]$ | 200 | $[2,5]$ | $[10,190]$ | | 49,994 |
| 图 2D，图 8A 和 B | 全部反常 | 1 | $[0.05,2]$ | $\{0,0.1\}$ | 200 | $[2,5]$ | $[10,190]$ | | 50,000 |
| 图 2E 和 F | 布朗运动 | $[-3,3]$ | 1 | 0 | 200 | 2 | $[10,190]$ | | 50,000 |
| 图 2G 和 H | FBM | 1 | $[0.05,1.95]$ | 0 | 200 | 2 | $[10,190]$ | | 40,000 |
| 图 3 | SBM | 1 | $\{0.1,0.5\}$ | 0 | 200 | 1 | 200 | | 6,000 |
| 图 4B、C 和 D | ATTM | (-6.7, 0) | 0.75 | 0 | 200 | $[1,51]$ | $[1,200]$ | | 10,000 |
| 图 4E、F 和 G | (实验) | | | | $[200,2000]$ | | | | 755 |
| 图 5 | (实验) | | | | $[20,500]$ | | | | 4734 |
| 图 7A | 布朗运动 | $[-3,3]$ | 1 | 0 | $[20,660]$ | $[1,11]$ | $\{20,40,60\}$ | | 22,000 |
| 图 7B | 布朗运动 | $[-3,3]$ | 1 | 0 | $[40,660]$ | $[2,11]$ | $\{20,40,60\}$ | | 20,000 |
| 图7C | 布朗运动 | $[-3,3]$ | 1 | $[-6,0]$ | 200 | $[2,5]$ | $[10,190]$ | | 48,000 |

表 I. 本文报告的所有结果所用数据集的详细信息。 $D$ 和 $\sigma_{\text{noise}}$ 的取值范围采用对数$_{10}$ 尺度，并在所示范围内以 0.05 为间隔取 $\alpha$ 个值。我们将 20% 的训练数据用于验证。从包含 199,976 条轨迹的测试集中抽取两个相互独立的子样本：一个用于图 2C、图 8C 和 D、图 9，另一个用于图 2D、图 8A 和 B。在图 2D 中，我们考虑无噪声轨迹，但会对平坦的连续时间随机游走段添加噪声 $\sigma_{\text{noise}} = 0.1$，以避免时间平均均方位移方法出现数值不稳定性。为生成图 2C、图 8C 和 D 以及图 9 中的含噪轨迹，我们为数据集中的每条轨迹添加 128 个随机定位噪声水平，最终得到约 $6.4 \times 10^6$ 条轨迹。图 2C 包含在图 9 中，因此使用相同的数据。在图 3 中，我们将 $\alpha$ 的两个取值均匀分配给所有轨迹。在图 4B、C 和 D 中，我们使用具有 $\sigma = 0.3$ 和 $\gamma = 0.4$ 的 ATTM 轨迹（见 [10, 47],，请勿与 $\sigma_{\text{noise}}$混淆）。我们通过随机抽样 $D$ 并相应地抽取片段长度来模拟这些轨迹；此处所列数值来自对所得轨迹的分析，这些轨迹平均包含 18 个不同片段。对于图 4E、F 和 G，我们考虑了 755 条病原体识别受体 DC-SIGN 的实验轨迹，每条轨迹包含 200 至 2000 帧，采样频率为 60 Hz。图 5 中的所有结果均基于整合素 $\alpha5\beta1$ 的 4734 条轨迹得出，每条轨迹包含 20 至 500 帧，采样频率为 33 Hz。图 7B 所用数据集是图 7A 数据集的一个子集。在图 7C 中，我们使用与图 2A 和 B 相同的轨迹，并为每条轨迹添加 128 个随机定位噪声水平，最终得到 $6.144 \times 10^6$ 条含噪轨迹。

沿轨迹具有常量扩散系数 $D$。最后，我们对位移进行累积求和，得到轨迹坐标，并减去初始位置，使其从原点开始。

我们考虑跨越六个数量级的扩散系数 $D \in [10^{-3}, 10^3]$。不过，我们将其对数作为回归任务的标签，从而使得 $y_i \in [-3,3]$ 在每个时间步都成立。这大大简化了问题，并使我们能够在所有数量级范围内保持一致的性能。

此外，我们可以通过添加标准差为 $\sigma_{\text{noise}}$ 的高斯噪声来模拟实验中的定位噪声。在训练期间，我们将此作为一种数据增强形式，并借此研究模型对噪声的鲁棒性。更多详情请参见表 I 。

**反常扩散 –** 为模拟反常扩散轨迹，我们考虑附录中介绍的五种扩散模型A 及其各自的反常扩散指数范围。我们按照参考文献补充材料中详述的相同流程，为每种模型生成完整轨迹 [18] ，并使用作者提供的程序库 [73]。随后，为获得异质轨迹，我们将其拆分为片段并组合起来。我们施加以下条件：任意两个连续片段至少在扩散模型或反常扩散指数上有所不同。最后，我们添加标准差为 $\sigma_{\text{noise}}$ 的高斯定位噪声。然后，用所得位移的标准差对其进行归一化，并减去初始位置，以确保轨迹从原点开始。

因此，在每个时间步，我们都有两个标签：反常扩散指数，以及生成相应片段所用的扩散模型。这样，我们就可以使用相同的数据，同时对反常扩散指数进行回归，并对扩散模型进行分类。不过，在这项工作中，我们主要关注前者。此外，我们对所有数据集进行平衡，使得在所有时间步上，反常扩散指数和扩散模型的分布均匀。

### 附录 C：扩散系数预测

在第II节中，我们研究 STEP 是否能够准确推断每个时间步的扩散系数，并检测扩散行为的变化。此处，我们通过考察影响性能的其他因素，对正文中的分析进行补充，包括轨迹中的片段数以及实验装置中常见的定位噪声。我们将

考虑这些因素后的结果见图7。

我们研究片段数量对轨迹表征的影响。为此，我们固定分段长度，并生成包含一个至十一个片段（零至十个变点）的轨迹，因此轨迹长度差异很大（详情见附录 B3）。在图 7A 中可以看到，相对误差随片段数量略有增加，但其影响远小于分段长度的影响。例如，相较于八个连续、每段包含 40 个时间步长的片段，包含 20 个点的单个片段更难表征。重要的是，即使存在 10 个变点，STEP 的表现仍远优于将 TA-MSD方法应用于大小相同、无变化的片段；例如，图 7A 中分段长度为 20 时的整条曲线，都明显低于图 2C 中分段长度为 20 时的 TA-MSD 数据点。

在图 2E 和 F 中，我们展示了如何将 STEP 与 KCPD 方法结合，以检测扩散变化。在图 7B 中，我们展示了性能随片段数量的变化情况。可以看到，最短的片段最难表征。不过，对于足够长的片段，分段长度的重要性会降低，因为长度为 40 和 60 时的曲线表现相当相似。轨迹很长（11 个片段）时，STEP 在较短片段上的得分更高。这表明，轨迹中每增加一个变点，都会带来数量相近的误差来源；而随着轨迹变长，沿轨迹累积的误差最终会占据主导。尽管如此，即便在最具挑战性的情况下（11 个片段），STEP 仍能正确检测出绝大多数点。

最后，我们研究了该方法对噪声的鲁棒性，这是实验轨迹的一项重要特征。在实际场景中，轨迹会受到定位噪声的影响，通常将其建模为方差为 $\sigma^2_{\text{noise}}$ 的高斯噪声，并叠加到轨迹上。由于我们考察轨迹上尺度差异很大的扩散系数，在图7C 中，我们绘制了误差随噪声标准差与扩散系数之比的变化情况。可以看到，即使噪声水平远超相关实验场景中通常存在的水平（通常为 $\sigma_{\text{noise}}/D < 10^{-1}$），STEP 的表现仍远优于已知片段的 TA-MSD方法。令人惊讶的是，即使误差极大（$\sigma_{\text{noise}}/D > 10$），STEP 仍能正确提取恒定片段的扩散系数（红线），展现出出色的抗噪能力。

### 附录 D：各种扩散模型的反常扩散指数预测

在第 II节中，我们简要展示了如何使用 STEP 研究在不同反常扩散状态间随机切换的粒子。在这里，我们全面评估了该方法用于此任务的适用性。我们使用 STEP 来预测轨迹中每个时间步的反常扩散指数 $\alpha$；这些轨迹由反常扩散指数和扩散模型均为常量的片段组成，具体内容详见附录 B3。

在正文中，我们已经研究了 $\alpha$ 预测的平均绝对误差（MAE）如何随分段长度变化，并将 STEP 与两种基准方法进行了比较。作为另一种基准，我们计算了所有轨迹和时间步长上的平均绝对误差：STEP 为 0.271，STEP（已知片段）为 0.275，TA-MSD（已知片段）为 0.368；对于长度相同的轨迹，CONDOR 在 AnDi 挑战赛中取得了 0.237 的平均绝对误差 [18]。

此外，我们还可以按扩散模型分别考察性能。图 8A 报告了每种反常扩散模型所含所有片段的平均绝对误差。我们观察到，某些模型之间存在明显差异，其中连续时间随机游走、分数布朗运动和莱维行走片段的误差最低。此外，我们还提供了所有扩散模型的预测值和真实值 $\alpha$ 直方图，见图 9。

特别是，缩放布朗运动（SBM）片段的平均绝对误差显著高于其他模型。先前的研究中也观察到了这一现象（例如参见文献 [18]的图 2d），不过这里的差异更大。仔细检查发现，最大的误差来自较短的片段，这与图 2D 的结果一致。这是合理的，因为 SBM 中的老化是反[59]常扩散的来源（参见第 II节），因此需要更长的片段才能准确表征。也可以合理地预期，每当 $\alpha \in (0, 2)$ 接近其范围边界，而预测值位于相反一侧时，误差就会最大。图 8B 显示，对于较小和较大的 $\alpha$，它们显然是图 8A 所示整体平均绝对误差的主要贡献项。

STEP 显示出明显的预测倾向 $\alpha \sim 0.8$ 对于 SBM 片段，正如图9最右列所示，这与图8B 中的结果一致。噪声的存在会增强这一行为，这表明模型难以在较短片段中识别出任何明确的行为，而这类片段恰恰也是最常见的。

有趣的是，我们在 CTRW 段中也发现了类似趋势：模型倾向于预测 $\alpha \simeq 0.25$，这对应于几乎不动的粒子。CTRW 轨迹的特征是在随机时间发生跳跃，从而产生粒子不发生移动的片段，这类片段通常被称为等待时间。因此，我们的异质轨迹中的许多 CTRW 段由于长度较短，并未显示出任何运动，而是对应于一段等待时间。因此，模型不可能正确预测 $\alpha$，因为它没有任何可供利用的信息。

在较小程度上，我们还发现模型会预测 ATTM 段中较低的反常扩散指数 $\alpha \sim 1$。在 $\alpha$ 较小的 ATTM 轨迹中，我们会遇到扩散系数较低的超长片段。与连续时间随机游走的情况类似，我们在

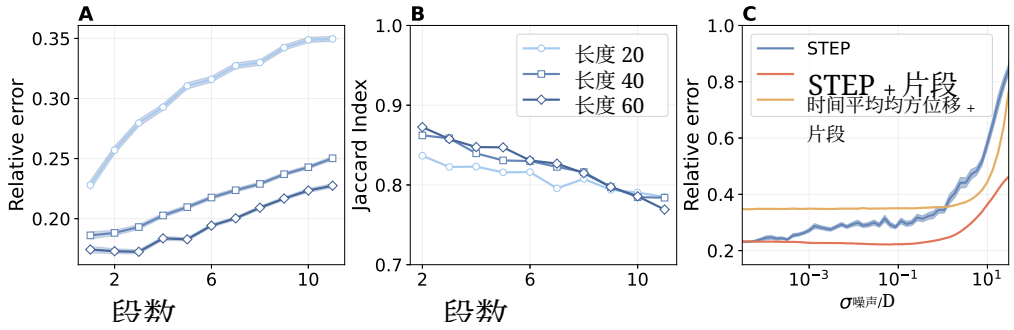


图7. **STEP 在每个时间步预测扩散系数的表现。(A)** STEP 的相对误差随片段数变化的情况，展示了三种不同分段长度。**(B)** STEP 的预测准确率随片段数变化的情况，展示了三种不同分段长度。**(C)** STEP（蓝色）、已知片段的 STEP（红色）以及已知片段的时间平均均方位移（黄色）的相对误差，随定位噪声标准差 $\sigma$ 与扩散系数 $D$之比变化的情况。

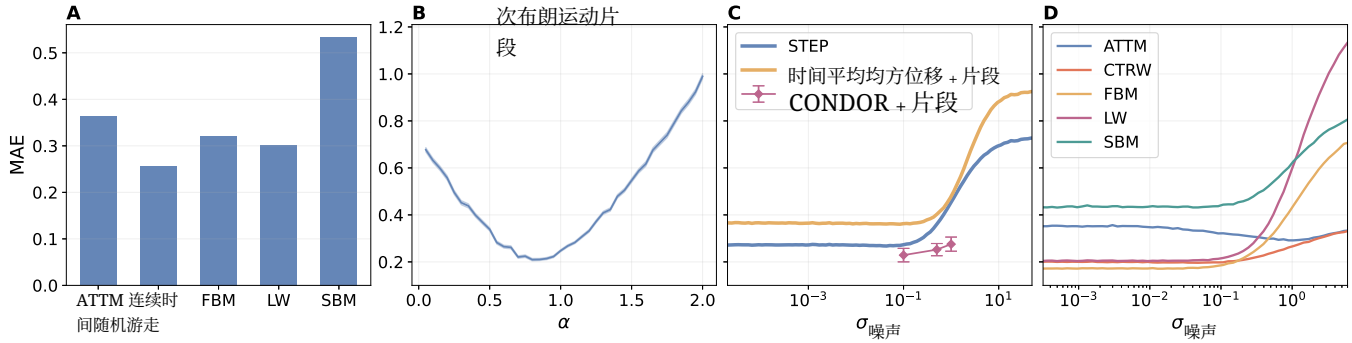


图8. **STEP 的预测表现 $\alpha$ 以平均绝对误差（MAE）衡量。**我们考察以下反常扩散模型：ATTM、连续时间随机游走、分数布朗运动、莱维行走和 SBM（见附录 B3）。**(A)** 各扩散模型的预测 MAE。**(B)** SBM 片段的 MAE 随 $\alpha$ 变化的情况。**(C)** STEP、时间平均均方位移和 CONDOR 的 MAE 随定位噪声变化的情况；后两者使用已知片段。**(D)** 按扩散模型分类后，STEP 的 MAE 随定位噪声变化的情况。因此，**(C)** 中的蓝线对应于本面板所示各条曲线的平均值。

这些异质轨迹的长片段中会遇到具有唯一扩散系数的部分，因此在观测时间窗内表现得如同布朗运动。因此，在这些情况下，预测 $\alpha \sim 1$ 是正确的。

最后，我们研究了这些方法抵抗定位噪声的能力，附录 C中也进行了同样的研究。我们绘制了平均绝对误差随噪声标准差变化的曲线$\sigma_{\text{noise}}$ ，见图 8C。我们发现，在噪声达到相当高的水平之前，所有方法的性能都保持稳定。与前文一样，尽管 CONDOR 的优势在于预先知道片段，STEP 的表现仍与之相当。

正如本节中所见，与某些扩散模型相比，另一些模型更难以表征；定位噪声对它们的影响也各不相同，如图 8D 和图 9所示。尽管噪声水平升高总体上会产生负面影响，但我们发现，莱维行走片段受到的影响最大，而连续时间随机游走片段的表现几乎不受影响。总体而言，误差在超过$\sigma_{\text{noise}} \sim 2\times10^{-1}$后开始显著增大，这相当于严苛的实验条件。有趣的是，在有限范围内，随着噪声增加，ATTM 片段的平均绝对误差反而有所下降。

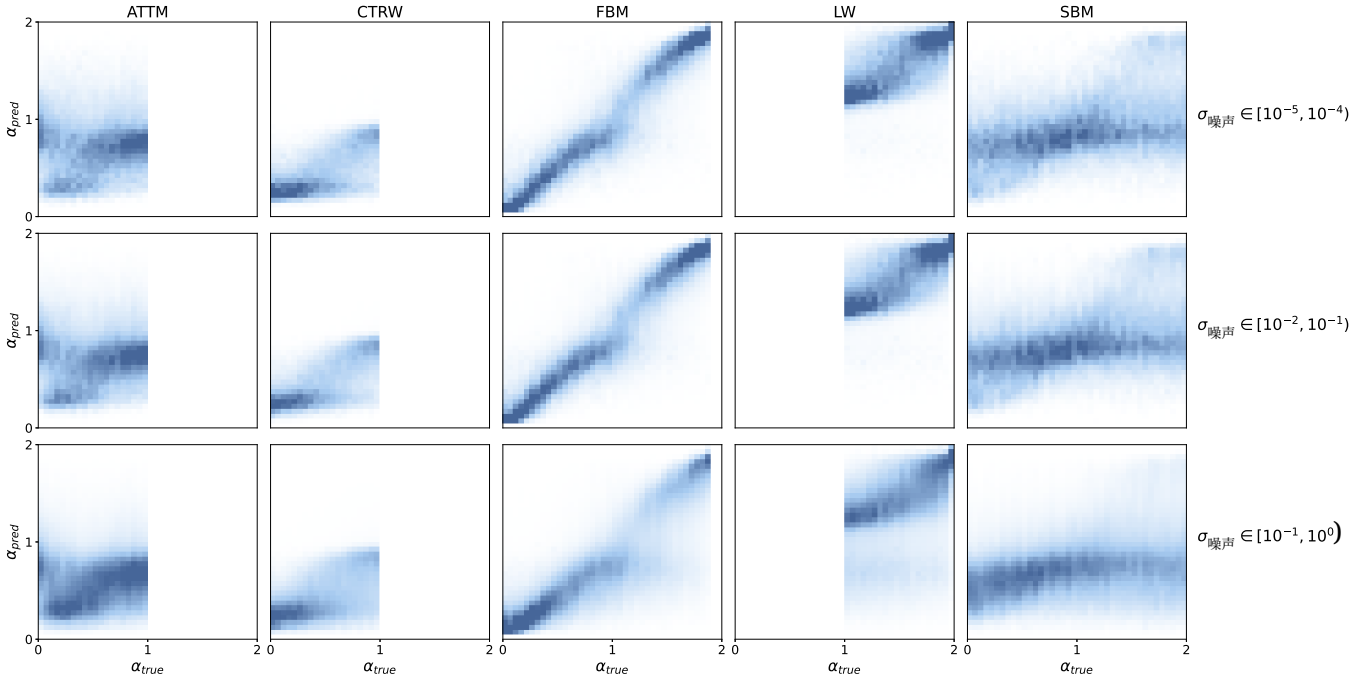


图9。 **预测值与真实反常扩散指数**。 二维直方图展示了不同扩散模型的真实反常扩散指数和预测反常扩散指数。每一列对应一种扩散模型在三种不同定位噪声水平下的直方图： $\sigma_{\text{noise}} \in [-5, -4)$ （上）， $\sigma_{\text{noise}} \in [-2, -1)$ （中），以及 $\sigma_{\text{noise}} \in [-1, 0)$ （下）。低噪声分数布朗运动直方图（第一行，中间列）与图2C。