arXiv:2603.01284v1 [cs.CV] 1 Mar 2026

FoSS：通过傅里叶—状态空间集成，在轨迹预测中建模长程依赖与多模态不确定性

Yizhou Huang[1]*、Gengze Jiang[1]、Yihua Cheng[2]、
Kezhi Wang[1]†伦敦布鲁内尔大学、 [2]伯明翰大学

摘要

准确的轨迹预测对于自动驾驶安全至关重要，但现有方法难以兼顾建模能力与计算效率。基于注意力的架构会随着智能体数量增加而产生二次复杂度，而循环模型则难以捕捉长程依赖和细粒度的局部动态。基于此，我们提出***FoSS***，一种双分支框架，将频域推理与线性时间序列建模相结合。频域分支通过离散傅里叶变换将轨迹分解为编码全局意图的幅度分量和捕捉局部变化的相位分量，随后采用渐进式螺旋重排序模块以保留谱序；两个选择性状态空间子模块 *Coarse2Fine-SSM* 和 *SpecEvolve-SSM* 在 $\mathcal{O}(N)$ 复杂度下细化频谱特征。与此同时，时域动态选择性状态空间模型以线性时间复杂度重构自注意力行为，从而保留长程时序上下文。交叉注意力层融合时序与频谱表示，可学习查询生成多个候选轨迹，而加权融合头则表征运动不确定性。在 *Argoverse 1* 和 *Argoverse 2* 基准上的实验表明，***FoSS***实现了最先进的准确度，同时将计算量降低了 *22.5%*，参数量减少了 *40%* 以上。全面的消融实验证实了各组件的必要性。

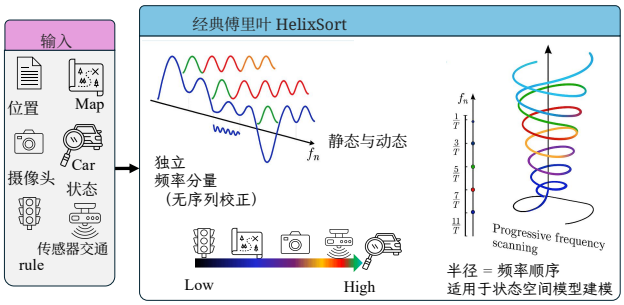

图1. 输入模态、经典傅里叶分解与我们的 HelixSort渐进式扫描之间的关系。交通规则和地图等低频输入编码了全局先验，而车辆状态和位置等高频输入则包含细粒度动态。HelixSort逐步扫描并对齐这些异构频率线索，形成有序频率序列，以便进行选择性状态空间建模。

1. 引言

准确的轨迹预测对于安全自动驾驶至关重要，尤其是在车辆与行人动态交互的密集多智能体环境中 [5, 12]。现代预测系统必须同时满足三项要求：对智能体之间的长程依赖进行建模、表示多模态未来以捕捉不确定性，以及在严格的实时约束下运行限制。尽管取得了显著进展，现有方法 [2, 4, 6, 27] 仍面临准确性、效率与可扩展性之间固有的权衡。仅依赖时间建模的方法往往会混淆全局运动模式与局部动态，而基于 Transformer 的架构 [14, 28, 40, 42, 44] 虽然能够实现高准确性，却要付出二次计算复杂度的代价，从而限制了其在资源受限系统中的部署。

针对这一问题，我们提出**FoSS**。据我们所知，这是大型自动驾驶轨迹预测中首批将频域分析与线性复杂度序列建模进行系统整合的方法之一。我们观察到，轨迹信号在频谱域和时域呈现互补结构：幅度谱编码全局运动趋势，而相位谱则捕捉细粒度的时间变化。然而，标准傅里叶表示缺乏有序的频率语义，导致序列模型难以有效处理频谱信息 [1, 22, 39]。此外，传统的时域与频谱特征融合策略 [29, 30] 常因尺度不匹配而影响优化稳定性。

我们的框架采用双分支架构来

*第一作者：byyh009@brunel.ac.uk
†通讯作者：kezhi.wang@brunel.ac.uk

保留并协同结合两种表示。该频域分支执行离散傅里叶分解，随后经由渐进式螺旋重排序模块（**HelixSort**）施加结构排序，从而使选择性状态空间模块，即*Coarse2Fine-SSM*和*SpecEvolve-SSM* 能够以线性计算复杂度细化频谱特征。该时域分支应用输入依赖的选择性状态空间模型 [9, 10, 31]以高效捕获长程依赖。一种交叉注意力机制融合两个分支，并通过归一化与残差连接解决特征尺度失配。最后，可学习的基于查询的解码器通过加权候选融合策略生成多模态未来轨迹。

在 Argoverse 1 [3] 与 Argoverse 2 [15, 37] 基准上的大量实验表明，FourierMamba-TP 在降低计算成本与参数量的同时达到了最先进的精度，验证了其在真实世界自主系统中的可扩展性。我们的贡献总结如下：1）我们提出一种双分支框架，将傅里叶域分解与选择性状态空间建模相结合，从而实现对全局与局部运动动态的解耦表示。2）我们设计了两个选择性状态空间模型子模块，即傅里叶由粗到细频率（Coarse2Fine-SSM）与频率通道演化（SpecEvolve-SSM），以线性计算复杂度细化频谱特征。3）我们改进了交叉注意力机制，以实现时序与频谱表示之间的稳定融合，从而支持高效的多模态轨迹预测。

2. 相关工作

2.1. 视觉任务中的傅里叶变换

傅里叶变换是一种经典工具，可将时域或空域信号分解为不同频率分量，将低频结构与高频细节和噪声分离。这些特性已在超分辨率 [21, 30]和图像去雨任务 [16, 29]中得到广泛探索。近期的多项研究 [1, 22, 39] 探讨了将基于傅里叶的表示融入轨迹预测。这些方法通常采用频谱分解，在频域中编码运动模式，从而有助于将全局趋势与局部变化分离。与以往基于傅里叶的方法不同，我们提出的螺旋扫描机制会将轨迹信号显式分解为低频全局趋势和高频局部变化。通过将其转换为状态空间模型（SSM）的结构化先验，我们提供了必要的频域引导，以补充状态空间模型对长程依赖关系的建模。这一设计不仅增强了对复杂运动模式的表示，还成功地将基于傅里叶的推理范式拓展至非线性轨迹预测领域。

2.2. 选择性状态空间模型（状态空间模型）

选择性状态空间模型通过融入依赖输入的参数化，扩展了传统线性状态空间模型，可根据不同输入动态调整状态转移 [9]。这一架构增强了模型捕捉序列长程依赖的能力 [26]。近期研究 [7, 17, 24, 34] 已成功将选择性状态空间模型应用于视频和语音等领域的长程建模任务，展现出其高效率和表征能力。在自动驾驶场景中，轨迹数据通常同时包含全局运动趋势和局部动态变化。选择性状态空间模型的输入自适应结构与这种双层信息高度契合，能够高效建模复杂的时空模式。仅依赖时间信息的模型难以同时捕捉全局运动趋势和局部动态。本文设计了一种双分支网络，在频域中嵌入选择性状态空间模型，对空间级和通道级频谱进行深度优化，从而获得更丰富的多模态表征，并实现比现有状态空间模型方法更有效的特征融合。

2.3. 多模态运动预测模型

基于深度学习的轨迹预测借助多种架构取得了显著进展，包括循环神经网络、长短期记忆网络和 Transformer。Social-LSTM 等早期研究及其扩展版本 [18, 25]采用循环结构来建模时间依赖关系，但在处理长序列时往往会遭遇梯度消失问题，从而限制了其在复杂驾驶场景中的效果。基于 Transformer 的方法 [14, 40, 44] 通过自注意力机制改进了交互建模，但其计算复杂度随序列长度呈二次增长，因此不太适用于涉及密集多智能体交互的实时应用。为应对未来运动的不确定性，一些研究 [12, 19, 32, 44] 提出了基于候选方案的预测方法，并在此基础上进行加权融合。尽管这些方法提升了多模态表达能力，但往往依赖启发式方法或静态融合权重，这可能会削弱模型在高度动态条件下的适应能力。

3. 方法

3.1. 问题表述

设历史轨迹序列表示为 $X \in \mathbb{R}^{T\times d}$，其中 T 是观测时间步数，d 是每个时间步的特征维度（例如，二维或三维位置、速度等）。轨迹预测的目标是

根据观测序列 X，估计未来 T' 个时间步的未来轨迹 $Y \in \mathbb{R}^{T' \times d}$。传统方法通常直接在时域中对 X 进行建模。然而，在处理复杂场景和多尺度信息时，它们往往难以同时捕捉全局运动趋势和局部动态细节。该方法以标准问题表述为基础，提出一种双分支方法，在频域中分解观测序列 X，并将其与时域信息融合以进行预测。这一设计旨在同时提高预测准确率和计算效率。

3.2. 使用状态空间模型进行傅里叶变换

在实际应用中，傅里叶变换涉及若干重要考量：**归一化因子**。因子 $\frac{1}{\sqrt{T}}$ 确保正向傅里叶变换与逆傅里叶变换之间的一致性。它保留了信号在时域和频域中的总能量。这一特性有助于维持数值稳定性，并促进对频域特征的可靠建模。**复数值表示**。由于 $F(X)(\omega) \in \mathbb{C}$ 是一个复数，其实部 $\Re\{F(X)(\omega)\}$ 和虚部 $\Im\{F(X)(\omega)\}$ 分别对应于原始信号在余弦基和正弦基上的投影。通过分析这两个分量，模型能够有效识别与不同频率相关的行为模式。这有助于将信号显式分解为全局结构和局部扰动。

FoSS 框架的核心思想是将时域信号表示为不同频率下正弦分量和余弦分量之和。这种分解能够根据频率特征分离全局趋势和局部变化。在轨迹建模任务中，这种表示有助于识别周期性行为模式和细粒度的动态变化。离散傅里叶变换（DFT）的标准形式[1] 如下所示

$$F(X)(\omega) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} X(t) e^{-j2\pi \frac{t\omega}{T}}, \omega \in \{0, 1, \ldots, T-1\}, \tag{1}$$

其中，$X(t) \in \mathbb{R}^{T \times d}$ 表示时间步 t 处的轨迹特征。ω 是频率索引，j 是虚数单位。

获得傅里叶变换结果 $F(X)(\omega) \in \mathbb{C}$ 后，我们进一步将其分解为幅度 $A(X)(\omega)$ 和相位 $P(X)(\omega)$ 两个分量，分别对轨迹中的全局趋势和局部动态行为进行建模。具体而言，对于每个频率索引 ω，

$$\begin{aligned} A(X)(\omega) &= \sqrt{\Re\{F(X)(\omega)\}^2 + \Im\{F(X)(\omega)\}^2}, \\ P(X)(\omega) &= \arctan\left(\frac{\Im\{F(X)(\omega)\}}{\Re\{F(X)(\omega)\}}\right), \end{aligned} \tag{2}$$

$A(X)$ 捕捉信号在不同频率上的能量分布，这通常与全局运动趋势和周期性行为模式相关。$P(X)$ 编码各频率分量的时间对齐情况，可揭示轨迹中细粒度的时间动态和局部扰动。

传统傅里叶变换的输出无法保持从低频到高频的连续排序，因为 DFT 存在对称性，即，ω 和 $T - \omega$ 对应相同的物理频率。为此，我们引入 HelixSort $(\cdot)$，这是一种受 JPEG 之字形编码（ [13]）启发的螺旋重排序模块。对于多维轨迹（$d > 1$），HelixSort 沿每个特征维度（$k \in \{1, \ldots, d\}$）独立运行。具体步骤如下：

1. 将一维离散傅里叶变换系数 $F^{(k)} \in \mathbb{C}^T$ 重塑为二维网格 $\mathcal{F}^{(k)} \in \mathbb{C}^{\sqrt{T} \times \sqrt{T}}$，并零填充至最接近的正方形维度，以保留频谱信息。2. 从频谱中心的直流分量（u_0, v_0）开始，按向外螺旋顺序遍历 $\mathcal{F}^{(k)}$。系数按频谱半径 $r = \sqrt{(u - u_0)^2 + (v - v_0)^2}$，升序排序，并列时按行主序打破平局。3. 记录索引映射 $\pi^{(k)}$，以生成重排后的一维序列 $\widehat{F}^{(k)} \in \mathbb{C}^T$。

由此得到一个非递减的半径序列，即，$\forall i < j$, $r_i \leq r_j$，它将低频分量（全局趋势）集中在序列起始处，并将高频分量（局部动态）置于序列末端。这种有序表示使序列状态空间模型能够以由粗到细的方式处理频谱信息。

与状态空间模型的兼容性。直接将状态空间模型应用于标准离散傅里叶变换的输出并非最优，因为交错排列的低频和高频系数迫使模型不断在全局推理与局部推理之间切换，扰乱了状态演化。HelixSort 通过提供有序的频谱序列解决了这一问题。状态空间模型（SSM）天然适合对一维序列进行长程建模。我们将 HelixSort 重新排序后的 $\widehat{\ }$ 序列 F 视为“人工时间轴”，据此定义递推关系：

$$h_{s+1} = A_s h_s + B_s \widehat{F}(s), \quad Y_{\text{freq}}(s) = C_s h_s + D_s \widehat{F}(s), \tag{3}$$

其中，索引 s 反映的是频率等级（通过谱半径 r_s），而非真实时间。由于低频分量会先被处理，状态空间模型会在其状态中累积全局

[1] 为简便起见，本文统一使用“FFT”一词，泛指离散傅里叶变换（DFT）及其快速计算实现，不区分二者在理论意义和算法意义上的差异。

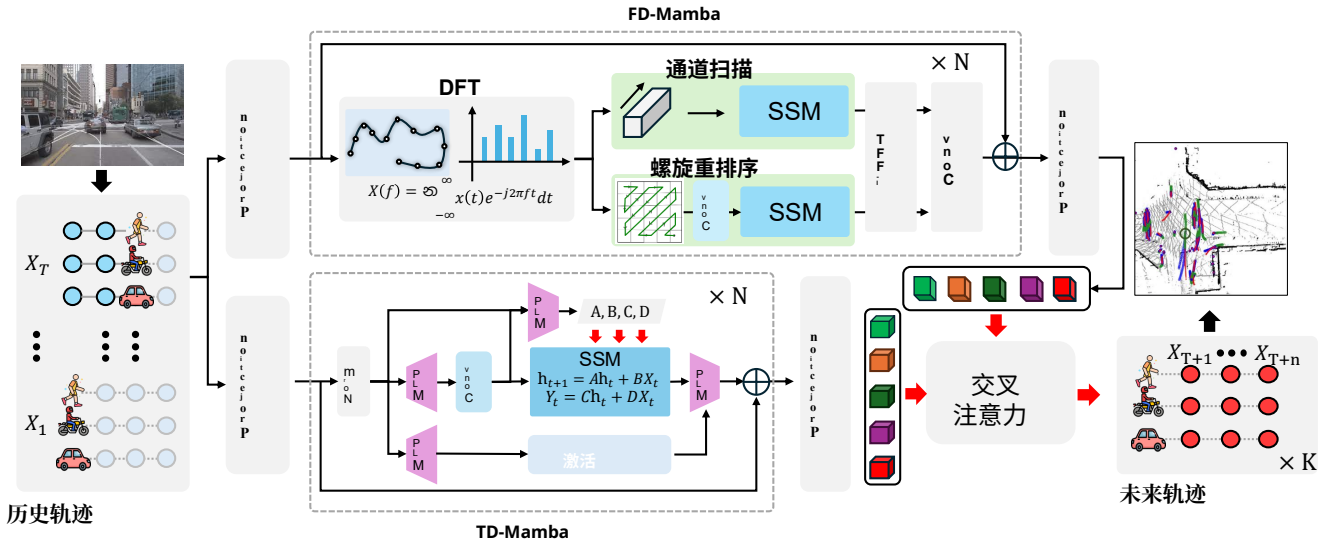

图2. 所提出框架的概览。历史轨迹经由两个并行分支处理：(i) 频率分支（FD-Mamba）执行离散傅里叶变换，并在两个选择性状态空间块中使用渐进式 HelixSort 对频谱进行重排。Coarse2Fine-SSM 用于空间交互，SpecEvolve-SSM 用于通道演化；(ii) 时间分支（TD-Mamba）将原始序列输入一个依赖输入的选择性状态空间模型，其中每个状态转移矩阵都根据当前观测值及其局部一维卷积特征动态生成。这种设计使时间动态能够随时间自适应演化，并以线性复杂度有效模拟自注意力行为。随后，通过交叉注意力层融合所得特征，再由可学习查询集解码出 K 个候选未来轨迹，其不确定性感知权重用于生成最终轨迹预测。

趋势。当高频分量随后到达时，模型已经构建了足够的上下文来调节短期细节——这一过程与常见的由粗到细推理策略相一致。此外，HelixSort 的单调性使参数 A_s、B_s 能够专门建模低频动态，同时对高频输入进行选择性门控，以抑制噪声。这种分离显著降低了状态转移矩阵 A_s的条件数，从而在训练过程中实现更稳定的梯度传播。

复杂度分析。 HelixSort 涉及一次快速傅里叶变换运算（$\mathcal{O}(T\log T)$）和索引重排（$\mathcal{O}(T)$）。它存储一个 T 维索引映射 π，不引入可学习参数，且开销可忽略：总浮点运算次数增加 $0.08 \pm 0.015\%$，内存占用为 <0.25MB（在 Argoverse 2 上测得，$T = 50$）。

3.3. 双分支网络架构

考虑到仅使用时域或频域方法对轨迹数据建模，可能不足以充分捕捉其多尺度特征，我们提出了一种双分支网络架构。如图1所示，该架构由时域分支和频域分支组成，分别从时间演化和频率结构的角度对轨迹进行建模与解读。

TD-Mamba：时域分支中的选择性状态空间建模。 在时域分支中，输入轨迹数据 $X \in \mathbb{R}^{T\times d}$ 直接作为输入。我们采用选择性状态空间模型，从中提取目标运动过程的动态演化特征并捕捉长程依赖。在传统线性状态空间模型中，状态更新通常采用以下公式：

$$h(t+1) = Ah(t) + BX(t),\ Y_{\text{time}}(t) = Ch(t) + DX(t), \tag{4}$$

其中 $h(t) \in \mathbb{R}^n$ 表示时间 t的隐状态，$A \in \mathbb{R}^{n\times n}$，$B \in \mathbb{R}^{n\times d}$，$C \in \mathbb{R}^{p\times n}$，$D \in \mathbb{R}^{p\times d}$ 分别表示状态转移矩阵、输入矩阵、输出矩阵和前馈矩阵。然而，对于复杂的轨迹数据，这种固定参数的方法通常难以充分抑制噪声并捕捉关键的动态信息。因此，我们在模型中引入了选择性状态空间模型，根据当前输入 $X(t)$及其局部卷积特征 $\tilde{X}(t)$ =Conv1D($X(t)$)动态调整每个参数矩阵。具体而言，矩阵参数定义为

$$\begin{aligned} A_t &= f_A(X(t), \tilde{X}(t)), \quad B_t = f_B(X(t), \tilde{X}(t)), \\ C_t &= f_C(X(t), \tilde{X}(t)), \quad D_t = f_D(X(t), \tilde{X}(t)). \end{aligned} \tag{5}$$

其中 f_A，f_B，f_C，f_D 均由轻量级多层感知机（MLP）实现，用于根据当前输入信息生成相应的参数。

因此，状态更新公式可改写为：

$$h(t+1) = A_t h(t) + B_t X(t), \quad Y_{\text{time}}(t) = C_t h(t) + D_t X(t). \tag{6}$$

这种输入依赖参数化使动态状态更新过程能够根据不同时间步的输入，自动且有选择地放大关键运动模式，同时抑制有害的噪声干扰。此外，在更新过程中，输入通过一维卷积（Conv1D）进行预处理，以提取局部短期˜特征。该卷积的输出 X（t）被整合到参数生成函数（f_A，f_B，f_C，f_D）中，从而增强模型对局部动态变化的敏感性。为确保数值稳定性，每个时间步的隐状态在经过SiLU激活后，还要进行层归一化，具体为：

$$\hat{h}(t+1) = \text{LayerNorm}(\text{SiLU}(h(t+1))), \tag{7}$$

在经过上述动态状态更新和归一化操作后，整个时域分支输出一系列特征 $Y_{\text{time}} \in \mathbb{R}^{T\times p}$。该特征不仅反映了运动序列的连续动态，还具备良好的抗干扰能力，为频域特征融合和候选轨迹生成提供有效的时间先验。

FD-Mamba：频域分支中的频谱重建与序列建模。为有效捕捉轨迹数据的全局运动趋势和局部动态信息，数据 X 首先通过公式（1）变换到频域。随后，分别通过公式（ A（X）计算幅度 P（X）和相位。2）

由于传统傅里叶变换输出的频率分量并未直接按照从低频到高频的顺序排列，渐进式螺旋重排序模块 HelixSort(·) 解决了这一局限。该模块首先根据各频率分量与频谱中心之间的距离，重新排列幅度 A（X）和相位 P（X），从而将低频信息集中在序列的开头部分，并随后依次排列高频信息。重新排列后的幅度和相位记为

$$\begin{aligned} A'(X)(\omega) &= \text{HelixSort}(A(X)(\omega)), \\ P'(X)(\omega) &= \text{HelixSort}(P(X)(\omega)). \end{aligned} \tag{8}$$

在此基础上，我们进一步引入两个子模块，从空间和通道两个角度深入建模频域特征：(a) 傅里叶空间交互状态空间模块（Coarse2Fine-SSM）和 (b) 傅里叶通道演化状态空间模块（SpecEvolve-SSM）。两个子模块并行作用于共享输入 F_{in}（经层归一化变换后得到 F_l），分别对空间和通道频谱依赖关系进行建模。

首先，在 Coarse2Fine-SSM 中，F_{in} 经过快速傅里叶变换，转换到频域，得到相应的幅度谱 A（F_l）和相位谱 P（F_l）。随后，渐进式螺旋重排序模块分别重新排列这两部分：

$$A'(F_l) = \text{HelixSort}(A(F_l)), \quad P'(F_l) = \text{HelixSort}(P(F_l)). \tag{9}$$

为进一步优化频域信息，该模块依次使用深度可分离卷积（DWConv）、SiLU 激活、选择性状态空间模型和层归一化处理重排后的序列。最后，通过逆快速傅里叶变换将处理后的特征还原至时域，再与经过 SiLU 激活的原始 F_l 逐元素相乘，形成具有空间交互性的表示：

$$F_f = \text{iFFT}\big(A'(F_l), P'(F_l)\big) \odot \text{SiLU}(F_l). \tag{10}$$

其次，SpecEvolve-SSM 模块沿通道维度处理频谱信息。具体而言，输入特征 $F_{in} \in \mathbb{R}^{H\times W\times C}$ 经过全局平均池化，得到 $F_g \in \mathbb{R}^{1\times 1\times C}$。对 F_g 应用傅里叶变换，得到逐通道频谱 $F(F_g)(z)$，其幅度为 $A(F_g)(z)$，相位为 $P(F_g)(z)$。为利用通道间相关性，一维渐进螺旋重排（通道扫描）按照频谱幅值升序重排系数 $|F(F_g)(z)|$。重排后的频谱 $A(F_g)'(z)$ 和 $P(F_g)'(z)$ 经逆快速傅里叶变换还原，得到处理后的通道特征。将其与经 SiLU 激活的 F_g 逐元素相乘，生成演化后的通道特征：

$$F_a = \text{iFFT}\big(A(F_g)'(z), P(F_g)'(z)\big) \odot \text{SiLU}(F_g). \tag{11}$$

演化后的特征 F_a 通过逐元素门控与 F_{in} 融合，从而获得增强表示：

$$F_{\text{enhance}} = F_a \odot F_{in}. \tag{12}$$

最后，将来自 Coarse2Fine-SSM 的输出 F_f 与来自 SpecEvolve-SSM 的输出 F_{enhance} 沿通道维度拼接，并通过一个线性投影层对齐二者的尺度和分布，得到最终的频域表示 F_{freq}。这保留了 Coarse2Fine-SSM 提取的空间交互线索，并融合了 SpecEvolve-SSM 生成的通道演化特征。因此，F_{freq}（t）同时捕获了全局运动趋势和局部动态细节，为后续特征融合和轨迹生成提供了强有力的频域先验。

3.4. 候选轨迹生成与损失函数

该模块介绍了基于交叉注意力的候选轨迹生成过程。它直接从融合时间域和频域线索的特征中提取显著信息，从而生成未来轨迹预测。具体而言，双分支融合特征 $Z \in \mathbb{R}^{T\times d}$ 与一组可学习的查询向量 $Q \in \mathbb{R}^{K\times d}$ 进行交互，其中 T 为序列长度， d 为特征维度，**K** 为候选轨迹的数量。查询向量经过初始化，而融合特征 Z 同时用作键和值，即 $K = Z$ 和 $V = Z$，其中 d_k 表示键维度。这种交叉注意力机制使每个查询都能聚合与目标运动相关的动态和结构信息，为候选轨迹生成上下文表示，记为： $Z_{atten} \in \mathbb{R}^{K\times d}$。获得注意力输出后，将特征张量 Z_{atten} 输入多层感知机，通过非线性变换将特征投影到候选轨迹的参数空间，记为：$\hat{} Y_k =$ 多层感知机$(Z_{\text{attn}}(k))$ $k = 1$，2，……，**K**。

为了在时域和频域共同约束预测质量，并优化候选轨迹生成与融合，本文提出了一种统一损失函数。首先，在时域中，使用 $\mathcal{L}1$ 损失直接衡量最终预测轨迹 $\hat{}Y$与真值 Y之间的差异。 $\mathcal{L}_{\text{time}} = \left\| \hat{}Y_{\text{final}} - Y \right\|_1$。其次，为确保频域的一致性，将 $\hat{}Y$和 Y均通过傅里叶变换进行转换，然后计算它们的频域表示之间的 $\mathcal{L}1$ 损失。 $\mathcal{L}_{\text{freq}} = \left\| F(\hat{}Y_{\text{final}}) - F(Y) \right\|_1$。最后，总损失定义为：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{time}} + \lambda \mathcal{L}_{\text{freq}} . \tag{13}$$

4. 实验

4.1. 实现细节

本工作中的所有实验均使用PyTorch Lightning框架实现。为进行精度比较，训练与评估在NVIDIA A100 GPU上开展。训练期间，批量大小设为128，初始学习率为0.001。采用Adam优化器进行权重更新，模型共训练50个训练轮次。根据验证性能应用动态学习率调度。若验证指标连续5个训练轮次未改善，则将学习率降至其原始值的10%。推理延迟与效率在NVIDIA RTX 3090 GPU上测量，该GPU可密切近似NVIDIA Jetson Orin等嵌入式设备的运行时特性。尽管RTX 3090提供更高的原始吞吐量，其与Jetson Orin共享的Ampere架构使这些结果能够作为真实边缘部署条件的可靠代理。

4.2. 数据集与评估指标

本研究在两个大规模数据集上进行了验证。Argoverse 1 数据集 [3], 的每个样本都提供 2 秒的观测轨迹，用于预测接下来的 3 秒。Argoverse 2 数据集 [15, 37] 提供 5 秒的历史数据，用于预测随后 6 秒的情况，其场景动态更加复杂。评估采用标准轨迹预测指标，包括所有时间步上的最小 ADE（minADE_K）以及最后一个时间步上的最小 FDE（minFDE_K）。此外，未命中率（MR_K）用于衡量最终点偏离超过 2 米的预测轨迹所占比例。还使用 Brier 分数变体（b-minFDE_K）评估预测置信度的校准情况。根据官方预测协议， K 的值设为 1 和 6。

4.3. 定量实验

比较涵盖了多个最先进的模型。所有基线模型均使用其公开发布的代码和官方超参数设置进行评估，或依据其原始论文复现。

Argoverse 2 上的结果。在 Argoverse 2 的 6 秒预测任务中（表1），**FoSS（本文方法）**取得了整体最佳性能。其 b-最小最终位移误差$_6$ （1.69），较 DenseTNT（1.98）提升了**14.6%**；最小平均位移误差$_6$（0.61），较 SceneTransformer（0.75）提升了**18.7%**。其最小最终位移误差$_6$ （1.07）较同一基线提升了**11.6%**，漏检率$_6$ （0.11）较 0.14 提升了**21.4%**。这些结果证实，所提出的频率分解和选择性状态空间模型显著提升了长时程轨迹预测的准确性。

Argoverse 1 上的结果。在 Argoverse 1 的 3 秒预测任务中（表2），**FoSS（本文方法）**的最小平均位移误差为$_1$ （1.67），相比 LaneGCN（1.92）降低了**13.0%**，最小最终位移误差为$_1$ （2.05），与 Wayformer（2.02， ∼1差异为 0.5%）相当。其漏检率为$_1$ （0.11），与现有最佳结果持平，表明其在短时域预测中具有很强的可靠性。总体而言，该方法在两个数据集上均实现了稳定提升，并在准确性与效率之间取得了良好平衡。

与其他时序骨干的兼容性。 我们进一步验证了频率分支[2]的即插即用 特性。**FD-Mamba** 通过替换时域分支，与两个时序编码器（Transformer 和 LSTM）结合，同时保持其他所有设置不变。这些变体对原始骨干进行微调，且无需任何

[2]报告混合变体仅用于展示 FD-Mamba 在不同架构中的适应性，并非声称其带来了额外的性能提升。

表1. Argoverse 2 数据集上的对比结果 [15, 37]。对于每项指标，最佳结果以**粗体**显示，下划线表示次优结果。值得注意的是，所报告的结果仅反映骨干网络的性能，不包括集成等工程技术带来的提升。

方法	b-最小最终位移误差 ↓	最小平均位移误差 ↓	最小最终位移误差 ↓	MR ↓	#参数 (M)↓
LaneGCN [19]	2.06	0.87	1.36	0.16	-
DenseTNT [11]	1.98	0.88	1.28	0.13	-
BANet [36]	1.92	0.71	1.36	0.19	9.49
SceneTransformer [23]	1.89	0.80	1.23	0.13	15.30
GOHOME [8]	1.86	0.89	1.29	0.08	-
HiVT [43]	1.84	0.77	1.17	0.13	-
MultiPath++ [33]	1.79	0.79	1.21	0.13	-
PAGA [38]	1.76	0.80	1.21	**0.11**	-
Wayformer [20]	1.74	0.77	1.16	0.12	-
QCNet [44]	1.91	0.65	1.29	0.16	7.71
DeMo [41]	1.84	0.63	1.17	0.13	5.92
FD-Mamba+Transformer	1.77	0.75	1.21	0.14	21.83
FD-Mamba+LSTM	1.91	0.85	1.28	0.13	7.89
FoSS（本文方法）	**1.69**	**0.61**	**1.07**	**0.11**	**4.18**

表 2. 在 Argoverse 1 数据集 [3] 上的对比结果。最佳结果以**粗体**显示，下划线表示次优结果。性能评估基于未经集成的骨干网络。

方法	最小平均位移误差$_1$ ↓	最小最终位移误差$_1$ ↓	漏检率$_1$ ↓
LaneGCN [19]	1.92	2.30	0.18
DenseTNT [11]	1.89	2.20	0.16
SceneTransformer [23]	1.89	2.13	0.13
HiVT [43]	1.84	2.05	0.13
MultiPath++ [33]	1.79	2.08	0.13
PAGA [38]	1.76	2.08	**0.11**
Wayformer [20]	1.74	**2.02**	0.12
QCNet [44]	1.76	2.08	**0.11**
DeMo [41]	**1.65**	3.89	0.14
FD-Mamba+Transformer	1.81	2.14	0.14
FD-Mamba+LSTM	1.88	2.27	0.15
FoSS（本文方法）	1.67	2.05	**0.11**

表 3. Argoverse 2 数据集上的消融实验 [15, 37]。频域分支指频域分支；P.H.R.M. 指渐进式螺旋重排序模块；傅里叶状态空间模型指 Coarse2Fine-SSM 和 SpecEvolve-SSM 子模块中的选择性状态空间模型；替换拼接+多层感知机是指在时域中用多层感知机替换交叉注意力以进行特征融合。表1进一步验证了频率分支与其他时序骨干（如 LSTM 和 Transformer）之间的普遍兼容性，展现了 FD-Mamba 在不同架构中的即插即用性能。

变体	最小平均位移误差 ↓	最小最终位移误差 ↓	MR ↓
w/o 频域分支	0.71	1.36	0.17
w/o 渐进式螺旋重排序模块	0.69	1.32	0.16
w/o 傅里叶状态空间模型	0.70	1.35	0.17
替换 Concat+多层感知机	0.69	1.33	0.16
完整模型	**0.65**	**1.29**	**0.15**

结构变化。在实证层面，**FD-Mamba**与 Transformer 架构天然契合，因为频谱先验能够补充逐token聚合；而我们的完整模型则专门针对状态空间模型进行了优化，采用动态参数更新和由粗到细频率路由。

效率与推理延迟。尽管其精度很高，**FoSS（本文）**仅包含**4.18M** 个参数，为所有对比模型中最少。如表4所示，其平均推理时间为**64ms**，快于 HiVT（82 ms）、SceneTransformer（76 ms）、QCNet（71 ms）和 MultiPath++（70 ms）。此外，它仅需**22.1G** 浮点运算次数，分别降低**51%** 和**19%**（相比 QCNet（45.3 G）和 Trajectory Mamba（27.3 G））。GANet 与 BANet 采用源自 LaneGCN 的卷积神经网络和基于图的编码器；尽管其浮点运算次数更低（15.79 G 和 11.93 G），精度却明显下降。总体而言，双分支设计实现了更优的权衡，兼顾精度、延迟与计算成本。

4.4. 定性分析

图3 展示了 Argoverse 2 数据集中的典型预测案例，涵盖了多种驾驶场景，例如掉头、变道、转弯和直线行驶。每个示例中，红线表示历史轨迹，橙线表示真实未来轨迹，绿线则表示 **FoSS** 的预测结果。如图所示，**FoSS**在大多数场景中都能生成平滑且准确的轨迹。即使在涉及急转弯或掉头的挑战性场景中，该模型也能输出多个高质量预测结果，紧密贴合车道几何，展现出强大的多模态推理能力，并遵循道路约束。在常见的变道场景中（图 3(f)），性能略有下降，预测轨迹出现轻微抖动。这可能是因为模型捕捉瞬时高频运动线索的能力有限。

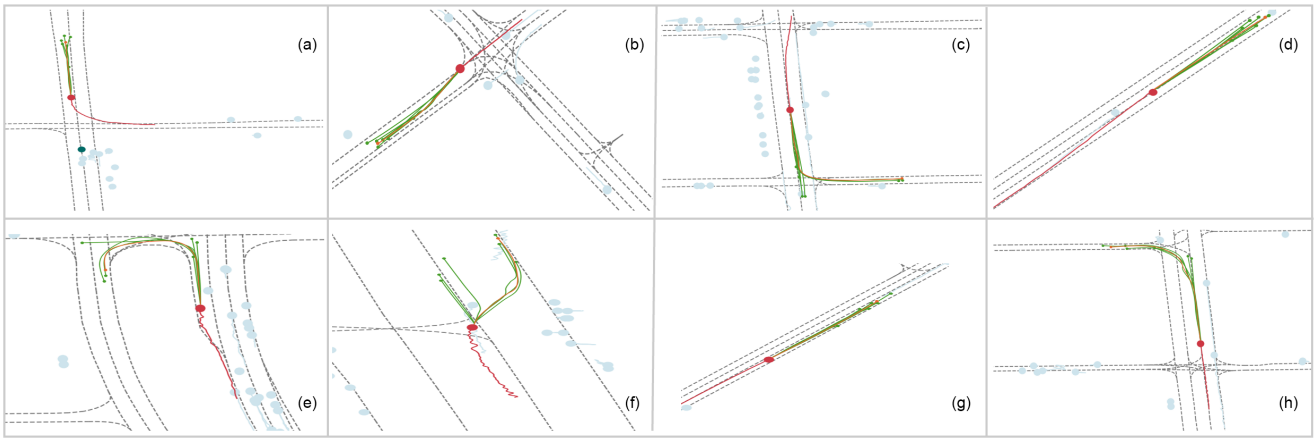

图 3. Argoverse 上的定性结果 2 [15, 37], 涵盖掉头（e）、变道（b、f）、转弯（a、c、h）和直行场景（d、g）。红色：历史轨迹；橙色：真值；绿色：预测轨迹。模型生成的轨迹平滑且多样，并与道路几何相一致；在频繁变道的情况下，轨迹会出现轻微抖动（f），这可能是由瞬态高频运动线索导致的。

方法	在线推理（毫秒）↓
HiVT [43]	82 ± 1
SceneTransformer [23]	76 ± 1
QCNet [44]	71 ± 1
MultiPath++ [33]	70 ± 1
FoSS（我们的方法）	$\mathbf{64 \pm 1}$

表 4. 在 NVIDIA RTX 3090 GPU 上评估推理延迟，其计算特性与 NVIDIA Jetson Orin 开发套件相近。性能指标越低越好。

方法	浮点运算次数（G）↓
GANet [35]	15.79
BANet [36]	**11.93**
QCNet [44]	45.3
Trajectory Mamba [12]	27.3
FoSS（我们的方法）	22.1

表5. Argoverse 上的计算效率 2 [15, 37]。所有实验均使用 NVIDIA 3090 GPU。

在频谱分解过程中，车道边界被表示为低频分量，这可能无法充分表征快速的横向机动。总体而言，这些定性结果证实了该模型的鲁棒性，以及其在复杂交通状况下生成多样且符合道路约束的未来轨迹的能力。

4.5. 消融实验

为量化各关键组件的贡献，我们在 Argoverse 2 数据集上进行了消融实验，结果汇总于表3。移除频域分支（频域分支）后，性能明显下降，最小平均位移误差$_6$ 从 0.65 增至 0.71，最小最终位移误差$_6$从 1.29 增至 1.36，这表明频率线索对于捕捉全局运动趋势以及减少长时域预测中的噪声至关重要。停用渐进式螺旋重排序模块（渐进式螺旋重排序模块）会导致性能中度下降（**+6.2% 的**最小平均位移误差，**+2.3% 的**最小最终位移误差），这表明有序的频谱遍历能够增强结构连贯性并促进细粒度运动提取。移除傅里叶状态空间模型后，最小平均位移误差/最小最终位移误差/漏检率的性能降至 0.70/1.35/0.17，凸显了选择性状态空间模型对于稳定实现频谱-时间融合的重要性。替换将交叉注意力融合替换为简单的拼接+多层感知机后，误差进一步增大（0.69/1.33/0.16），证实逐标记跨状态交互能够有效对齐时间和频率表示。总体而言，完整模型取得了最佳结果（最小平均位移误差$_6$=0.65、最小最终位移误差$_6$=1.29、漏检率$_6$=0.15），展现了频率-时间双分支建模的互补优势。

5. 结论与未来工作

本研究提出了 **FoSS**，这是一种双分支框架，将频域分解与输入依赖的状态空间建模相结合，以实现高效且富有表现力的轨迹预测器。通过跨域注意力和自适应轨迹融合，该方法在保持线性计算成本的同时，实现了稳健性与多模态性。在 Argoverse 1 和 2 上进行的大量实验表明，该方法取得了稳定提升，使最小平均位移误差和最小最终位移误差分别最高提升 **14.8%** 和**22.5%**，同时将推理延迟降低 **22%**，将参数量减少 **40%**。这些结果证实了该框架在复杂交通场景中的有效性。未来的工作将聚焦于涉及突发低频到高频运动变化的长尾边界情况，以确保其在高度动态环境中的稳健预测。

6. 致谢

本研究部分获得英国研究与创新署创新英国项目（10071278、10099265）的支持，同时也获得英国研究与创新署在地平线欧洲计划资助保障机制下的支持（EP/Y028031/1、EP/U536908/1）；相关工作属于欧洲委员会 MSCA COVER（101086228）和 ANT（101169439）项目的一部分。K. Wang 谨此感谢英国皇家学会产业研究员计划（IF\R2\23200104）提供的支持。Y. Huang 谨此感谢英国研究院提升研究文化基金提供的支持。

参考文献

[1] Defu Cao、Jiachen Li、Hengbo Ma 和 Masayoshi Tomizuka。用于轨迹预测的谱时序图神经网络。载于 2021 年 IEEE 国际机器人与自动化会议（ICRA），第 1839–1845 页。IEEE，2021。1、2[2] Sergio Casas、Cole Gulino、Simon Suo、Katie Luo、Renjie Liao 和 Raquel Urtasun。用于场景一致性运动预测的隐式潜变量模型。载于《计算机视觉——ECCV 2020：第 16 届欧洲会议》，英国格拉斯哥，2020 年 8 月 23—28 日，论文集，第 XXIII 卷第 16 部分，第 624–641 页。Springer，2020。1[3] Ming-Fang Chang、John W Lambert、Patsorn Sangkloy、Jagjeet Singh、Slawomir Bak、Andrew Hartnett、De Wang、Peter Carr、Simon Lucey、Deva Ramanan 和 James Hays。Argoverse：借助丰富地图进行三维跟踪与预测。载于计算机视觉与模式识别会议（CVPR），2019。2、6、7[4] Yuxiao Chen、Boris Ivanovic 和 Marco Pavone。Scept：用于规划的场景一致、基于策略的轨迹预测。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 17103–17112 页，2022。1[5] Yihua Cheng、Yaning Zhu、Zongji Wang、Hongquan Hao、Yongwei Liu、Shiqing Cheng、Xi Wang 和 Hyung Jin Chang。你在车里看到了什么？用于车内注视估计的综合视觉解决方案。载于 IEEE/CVF 计算机视觉与模式识别会议（CVPR）论文集，第 1556–1565 页，2024。1[6] Nachiket Deo、Eric Wolff 和 Oscar Beijbom。以车道图遍历为条件的多模态轨迹预测。载于机器人学习会议，第 203–212 页。PMLR，2022。1[7] Yassir Fathullah、Chunyang Wu、Yuan Shangguan、Junteng Jia、Wenhan Xiong、Jay Mahadeokar、Chunxi Liu、Yangyang Shi、Ozlem Kalinli、Mike Seltzer 等。用于语音识别的多头状态空间模型。arXiv 预印本 arXiv:2305.12498，2023。2[8] Thomas Gilles、Stefano Sabatini、Dzmitry Tsishkou、Bogdan Stanciulescu 和 Fabien Moutarde。GOHOME：用于未来运动估计的图导向热力图输出。载于 2022 年国际机器人与自动化会议（ICRA），第 9107–9114 页。IEEE，2022。7

[9] Albert Gu 和 Tri Dao。Mamba：采用选择性状态空间的线性时间序列建模。arXiv 预印本 arXiv:2312.00752，2023。2[10] Albert Gu、Karan Goel、Ankit Gupta 和 Christopher R´e。关于对角状态空间模型的参数化与初始化。《神经信息处理系统进展》，35:35971–35983，2022。2[11] Junru Gu、Chen Sun 和 Hang Zhao。DenseTNT：基于密集目标集的端到端轨迹预测。载于《IEEE/CVF国际计算机视觉会议论文集》，第15303–15312页，2021。7[12] Yizhou Huang、Yihua Cheng 和 Kezhi Wang。Trajectory Mamba：基于选择性状态空间模型的高效注意力-Mamba预测模型。载于《计算机视觉与模式识别会议论文集》，第12058–12067页，2025。1、2、8[13] ISO/IEC。ISO/IEC 10918-1:1994，1994。信息技术——连续色调静态图像的数字压缩与编码——要求与指南。3[14] Xiaosong Jia、Penghao Wu、Li Chen、Yu Liu、Hongyang Li 和 Junchi Yan。Hdgt：通过场景编码进行多智能体轨迹预测的异构驾驶图Transformer。《IEEE模式分析与机器智能汇刊》，45(11):13860–13875，2023。1、2[15] John Lambert 和 James Hays。信任，但要核实：用于高精地图变化检测的跨模态融合。载于《神经信息处理系统数据集与基准赛道论文集》（NeurIPS数据集与基准2021），2021。2、6、7、8[16] Dong Li、Yidi Liu、Xueyang Fu、Senyan Xu 和 Zheng-Jun Zha。Fouriermamba：将傅里叶学习与状态空间模型相结合，用于图像去雨。arXiv 预印本 arXiv:2405.19450，2024。2[17] Kunchang Li、Xinhao Li、Yi Wang、Yinan He、Yali Wang、Limin Wang 和 Yu Qiao。Videomamba：用于高效视频理解的状态空间模型。载于《欧洲计算机视觉会议》，第237–255页。Springer，2024。2[18] Lin Li、Wanzhong Zhao、Can Xu、Chunyan Wang、Qingyun Chen 和 Shijuan Dai。基于RNN的高速公路自动驾驶变道意图推断。《IEEE车辆技术汇刊》，70(6):5499–5510，2021。2[19] Ming Liang、Bin Yang、Rui Hu、Yun Chen、Renjie Liao、Song Feng 和 Raquel Urtasun。学习用于运动预测的车道图表示。载于《计算机视觉—ECCV 2020：第16届欧洲计算机视觉会议，英国格拉斯哥，2020年8月23日至28日，论文集》，第II卷16，第541–556页。Springer，2020。2、7[20] Nigamaa Nayakanti、Rami Al-Rfou、Aurick Zhou、Kratarth Goel、Khaled S Refaat 和 Benjamin Sapp。Wayformer：通过简单高效的注意力网络进行运动预测。载于《2023年IEEE机器人与自动化国际会议（ICRA）》，第2980–2987页。IEEE，2023。7[21] Hemkant Nehete、Amit Monga、Partha Kaushik 和 Brajesh Kumar Kaushik。基于傅里叶先验的图像复原两阶段架构。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第6014–6023页，2024。2

[22] Marion Neumeier、Andreas Tollk¨uhn、Michael Botsch 和 Wolfgang Utschick。用于车辆轨迹预测的多维图傅里叶变换神经网络。载于《2022年IEEE第25届智能交通系统国际会议（ITSC）》，第687–694页。IEEE，2022。1、2[23] Jiquan Ngiam、Benjamin Caine、Vijay Vasudevan、Zheng-dong Zhang、Hao-Tien Lewis Chiang、Jeffrey Ling、Rebecca Roelofs、Alex Bewley、Chenxi Liu、Ashish Venugopal 等。Scene Transformer：用于预测多个智能体轨迹的统一架构。arXiv 预印本 arXiv:2106 08417，2021。7、8。[24] Yuta Oshima、Shohei Taniguchi、Masahiro Suzuki 和 Yutaka Matsuo。状态空间模型遇上视频扩散模型：利用结构化状态空间实现高效的长期视频生成。arXiv 预印本 arXiv:2403.07711，2024。2[25] Edoardo Mello Rella、Jan-Nico Zaech、Alexander Liniger 和 Luc Van Gool。解码器融合RNN：用于轨迹预测的上下文感知与交互感知解码器。载于《2021年IEEE/RSJ智能机器人与系统国际会议（IROS）》，第5937–5943页。IEEE，2021。2[26] Farnoush Rezaei Jafari、Gr´egoire Montavon、Klaus-Robert M¨uller 和 Oliver Eberle。MambaLRP：解释选择性状态空间序列模型。《神经信息处理系统进展》，37:118540–118570，2024。2[27] Tim Salzmann、Boris Ivanovic、Punarjay Chakravarty 和 Marco Pavone。Trajectron++：利用异构数据进行动态可行的轨迹预测。载于《计算机视觉—ECCV 2020：第16届欧洲计算机视觉会议，英国格拉斯哥，2020年8月23日至28日，论文集》，第XVIII卷16，第683–700页。Springer，2020。1[28] Ari Seff、Brian Cera、Dian Chen、Mason Ng、Aurick Zhou、Nigamaa Nayakanti、Khaled S Refaat、Rami Al-Rfou 和 Benjamin Sapp。MotionLM：将多智能体运动预测视为语言建模。载于《IEEE/CVF国际计算机视觉会议论文集》，第8579–8590页，2023。1[29] Prasen Kumar Sharma、Sathisha Basavaraju 和 Arijit Sur。基于深度学习的图像去雨：利用离散傅里叶变换分析深度卷积神经网络对不相关变换域数据的行为。《The Visual Computer》，37(8):2083–2096，2021。1、2[30] Abhishek Kumar Sinha、S Manthira Moorthi 和 Debajyoti Dhar。NL-FFC：用于图像超分辨率的非局部快速傅里叶卷积。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第467–476页，2022。1、2[31] Jimmy Smith、Shalini De Mello、Jan Kautz、Scott Linderman 和 Wonmin Byeon。用于长程时空建模的卷积状态空间模型。《神经信息处理系统进展》，36，2024。2[32] Xiaolong Tang、Meina Kan、Shiguang Shan、Zhilong Ji、Jinfeng Bai 和 Xilin Chen。HPNet：利用历史预测注意力进行动态轨迹预测。载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集》，第15261–15270页，2024。2[33] Balakrishnan Varadarajan、Ahmed Hefny、Avikalp Srivastava、Khaled S Refaat、Nigamaa Nayakanti、Andre Cornman，

Kan Chen、Bertrand Douillard、Chi Pang Lam、Dragomir Anguelov等. Multipath++：面向行为预测的高效信息融合与轨迹聚合. 见2022年国际机器人与自动化会议（ICRA），第7814–7821页. IEEE，2022. 7, 8[34] Jue Wang、Wentao Zhu、Pichao Wang、Xiang Yu、Linda Liu、Mohamed Omar和Raffay Hamid. 面向长视频理解的选择性结构化状态空间. 见IEEE/CVF计算机视觉与模式识别会议论文集，第6387–6397页，2023. 2[35] Mingkun Wang、Xinge Zhu、Changqian Yu、Wei Li、Yuexin Ma、Ruochun Jin、Xiaoguang Ren、Dongchun Ren、Mingxu Wang和Wenjing Yang. GANet：面向运动预测的目标区域网络. 见2023年IEEE国际机器人与自动化会议（ICRA），第1609–1615页. IEEE，2023. 8[36] Zhepeng Wang、Feng Chen、Kanokphan Lertniphonphan、Siwei Chen、Jinyao Bao、Pengfei Zheng、Jinbao Zhang、Kaer Huang和Tao Zhang. Argoverse挑战赛关于统一的基于传感器的检测、跟踪与预测的技术报告，2023. 7, 8[37] Benjamin Wilson、William Qi、Tanmay Agarwal、John Lambert、Jagjeet Singh、Siddhesh Khandelwal、Bowen Pan、Ratnesh Kumar、Andrew Hartnett、Jhony Kaesemodel Pontes、Deva Ramanan、Peter Carr和James Hays. Argoverse 2：面向自动驾驶感知与预测的下一代数据集. 见神经信息处理系统数据集与基准分轨论文集（NeurIPS 2021数据集与基准），2021. 2, 6, 7, 8[38] F Alexander Wolf、Fiona K Hamey、Mireya Plass、Jordi Solana、Joakim S Dahlin、Berthold G¨ottgens、Nikolaus Rajewsky、Lukas Simon和Fabian J Theis. PAGA：图抽象通过单细胞的拓扑保持映射协调聚类与轨迹推断. 《基因组生物学》，20:1–9，2019. 7[39] Conghao Wong、Beihao Xia、Ziming Hong、Qinmu Peng、Wei Yuan、Qiong Cao、Yibo Yang和Xinge You. 纵向视角：一种基于傅里叶频谱的轨迹预测分层网络. 见欧洲计算机视觉会议，第682–700页. Springer，2022. 1, 2[40] Zhijie Yan、Pengfei Li、Zheng Fu、Shaocong Xu、Yongliang Shi、Xiaoxue Chen、Yuhang Zheng、Yang Li、Tianyu Liu、Chuxuan Li等. Int2：交叉路口处的交互式轨迹预测. 见IEEE/CVF国际计算机视觉会议论文集，第8536–8547页，2023. 1, 2[41] Bozhou Zhang、Nan Song和Li Zhang. Demo：将运动预测解耦为方向意图与动态状态. 见NeurIPS，2024. 7[42] Zhejun Zhang、Alexander Liniger、Christos Sakaridis、Fisher Yu和Luc V Gool. 基于相对位姿编码的异构折线Transformer实现实时运动预测. 《神经信息处理系统进展》，36:57481–57499，2023. 1[43] Zikang Zhou、Luyao Ye、Jianping Wang、Kui Wu和Kejie Lu. HiVT：面向多智能体运动预测的分层向量变换器. 见IEEE/CVF Con-

计算机视觉与模式识别会议，第 8823–8833 页，2022 年。7，8[44] Zikang Zhou、Jianping Wang、Yung-Hui Li 和 Yu-Kai Huang。以查询为中心的轨迹预测。载于 *IEEE/CVF* 计算机视觉与模式识别会议论文集，第 17863–17873 页，2023 年。1，2，7，8