

Integrating spoken instructions into flight trajectory prediction to optimize automation in air traffic control

将语音指令整合到飞行轨迹预测中以优化空管自动化

I Introduction

I Introduction

研究背景

行业背景：

- 航空运输业蓬勃发展，交通流量激增，管制员(ATCo)工作负荷加重。

安全隐患：

- 据研究，70% 的航空事故与人为因素(Human Factors)有关。
- 主要人为因素相关的风险来源：ATC通信过程中的错误（错误决策、听错、误解）。

关键痛点：

- 现有空中交通管制系统(ATC)是“人在环路(Human-in-the-Loop)”的。系统本身未能考虑到人类意图(通过语音指令的控制意图来体现)对空中交通预测的影响。
- 空中交通管制系统(ATC)仍无法处理语音指令，缺乏自动化手段来检测因误解指令导致的轨迹偏差。
- 预测空中交通管制指令对实时交通的影响，成为检测和防范管制过程中人为错误风险的有效途径

I Introduction

核心任务: 指令驱动的飞行轨迹预测。提出一种自动化范式，将管制意图整合到信息处理回路中。

- 输入: 历史轨迹观测序列 + 对应的文本化语音指令
- 输出: 未来一段时间内的飞行轨迹 (经度、纬度、高度等)

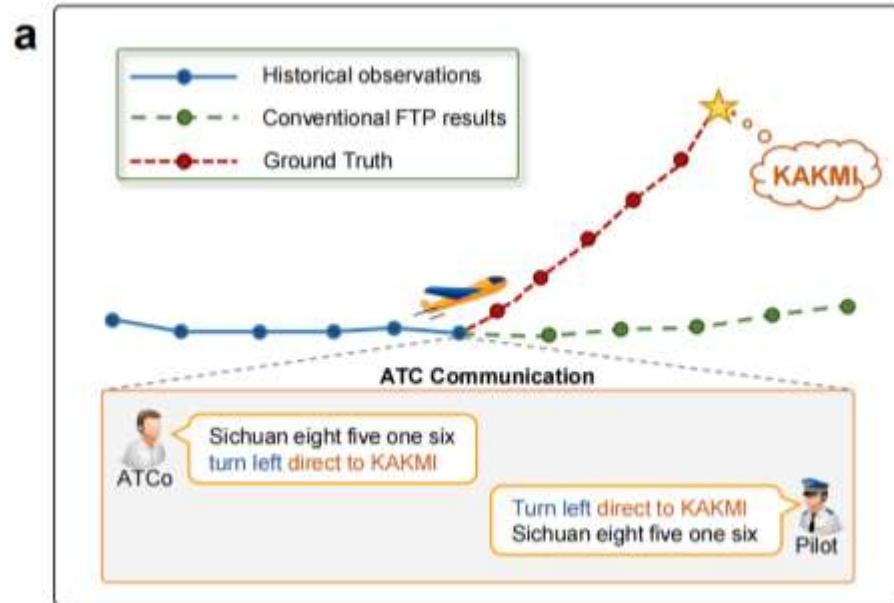
$$p_t = [\text{Lon}_t, \text{Lat}_t, \text{Alt}_t, \text{Vx}_t, \text{Vy}_t, \text{Vz}_t]$$

传统方法的局限: 仅依赖历史轨迹数据，无法感知未来的控制意图，导致在机动飞行阶段预测延迟大、不可靠。

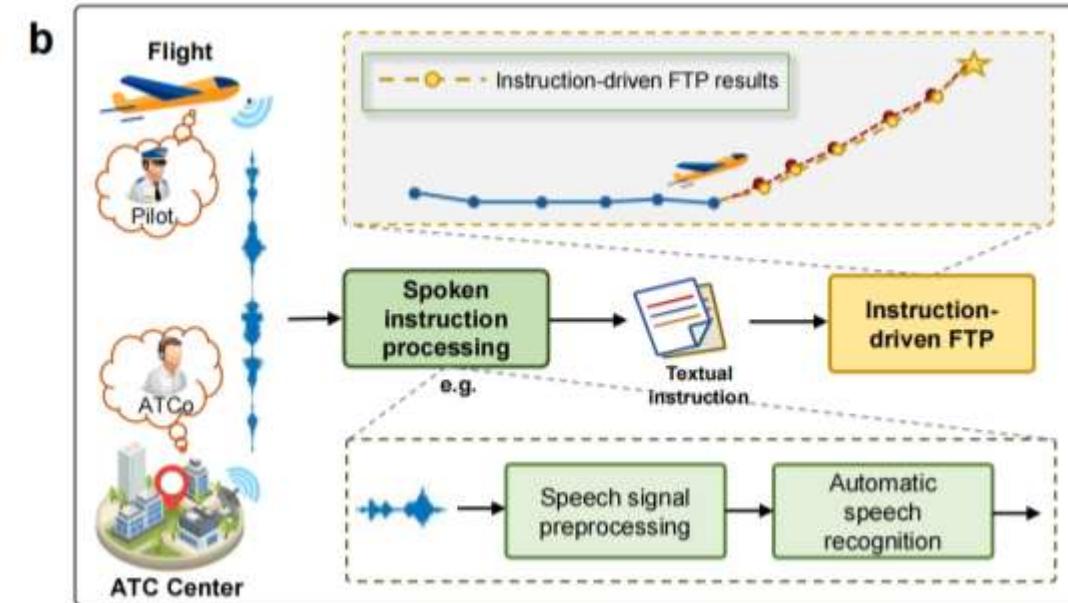
主要挑战:

- 模态差异: 轨迹是时空序列数据，语音指令是非结构化文本，异构。
- 数据稀缺: 虽然轨迹数据和语音数据单独都很丰富，但“轨迹-指令”严格对齐的成对数据 (Paired Data) 非常稀缺且标注昂贵。

$$P_{t+1:t+n} = \{p_{t+1}, p_{t+2}, \dots, p_{t+n}\} = \mathcal{F}(O_{t-k+1:t})$$



$$P_{t+1:t+n} = \{p_{t+1}, p_{t+2}, \dots, p_{t+n}\} = \mathcal{F}(O_{t-k+1:t}, \text{SI})$$



II Methods

II Methods

提出的框架：SIA-FTP (Spoken Instruction-Aware Flight Trajectory Prediction)

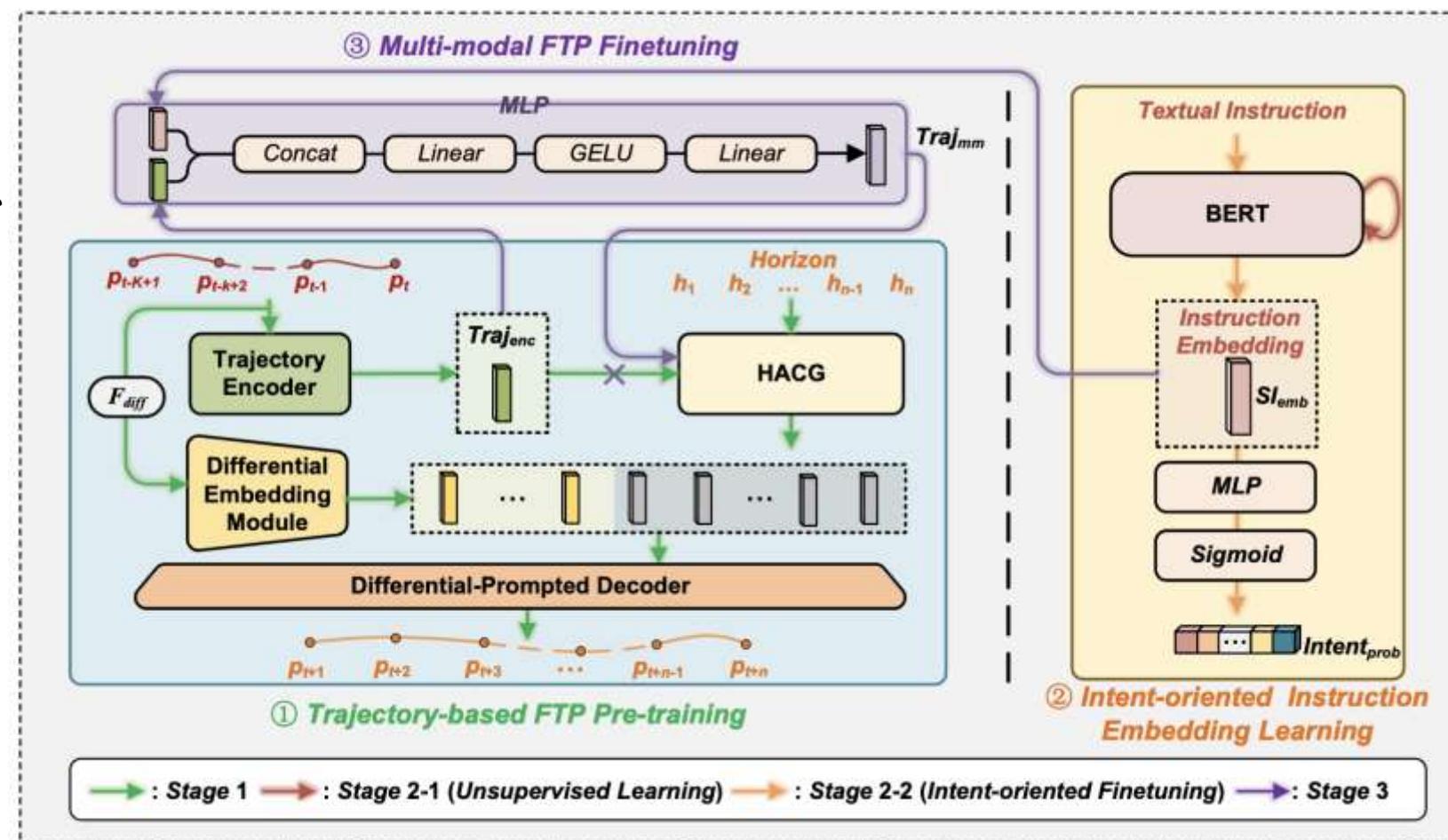
解决方案：提出3阶段渐进式多模态

学习范式

- 利用大规模单模态数据（轨迹、文本）进行预训练。
- 利用少量“轨迹-指令”成对数据进行多模态微调。

三个阶段：

- 基于轨迹的FTP预训练。
- 意图导向的指令嵌入学习，多标签意图识别(IID)。
- 多模态FTP微调。



II Methods

Stage1 : 基于轨迹的FTP预训练 (Trajectory-based FTP Pre-training)

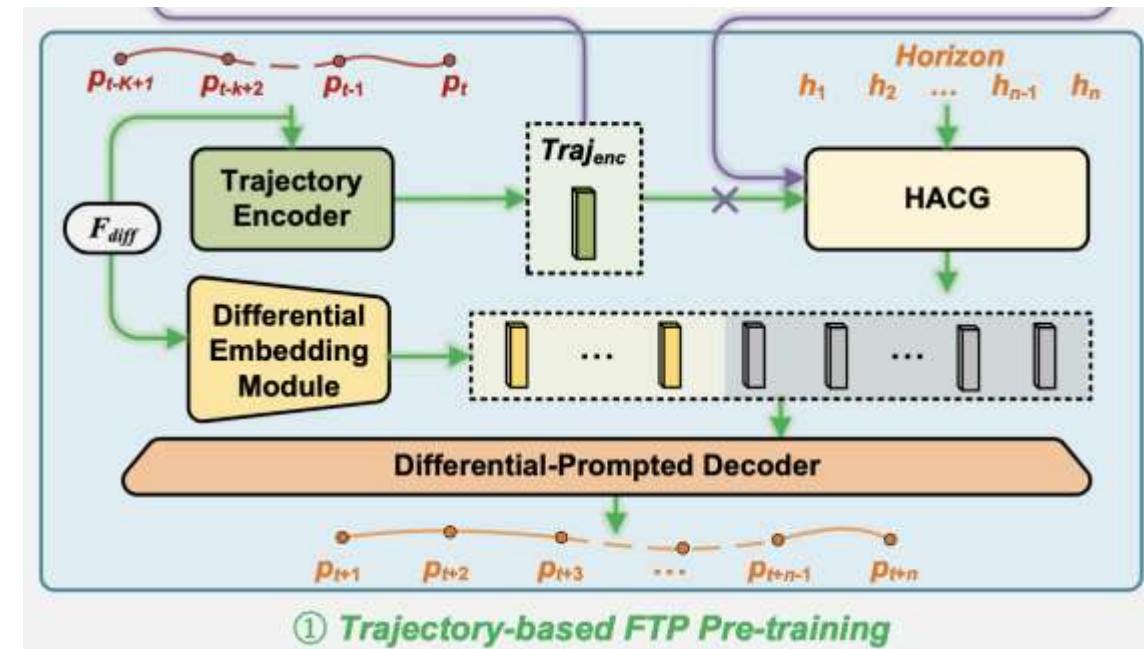
目的: 利用海量无标签的历史轨迹数据, 学习飞机的基本时空运动模式 (物理规律)

输入数据: 仅使用单模态的历史轨迹数据 ($0_{t-k+1:t}$)

基座模型: FlightBERT++

- 特点: 非自回归的多步预测模型(非一步步递归预测, 而是一次性直接输出未来多个时间步的轨迹)
- 由三个模块组成: 轨迹编码器 (TE)、视距感知上下文生成器 (HACG)、差分提示解码器 (DPD)
- BCE 损失函数: 衡量“预测概率”与“真实类别值”之间的差距

作用: 建立一个对常规飞行模式有认知的“基础模型”, 此时暂不考虑语音指令。



II Methods

Stage2：意图导向的指令嵌入学习（Intent-oriented instruction embedding learning）

目的：解决自然语言指令的理解问题，提取具有判别性的意图特征。

子步骤 2-1：无监督文本表示学习

- 使用 BERT 模型，在大量未标注的空管文本上进行 Masked Language Model 预训练。

子步骤 2-2：多标签意图识别微调（Intent-oriented Finetuning）

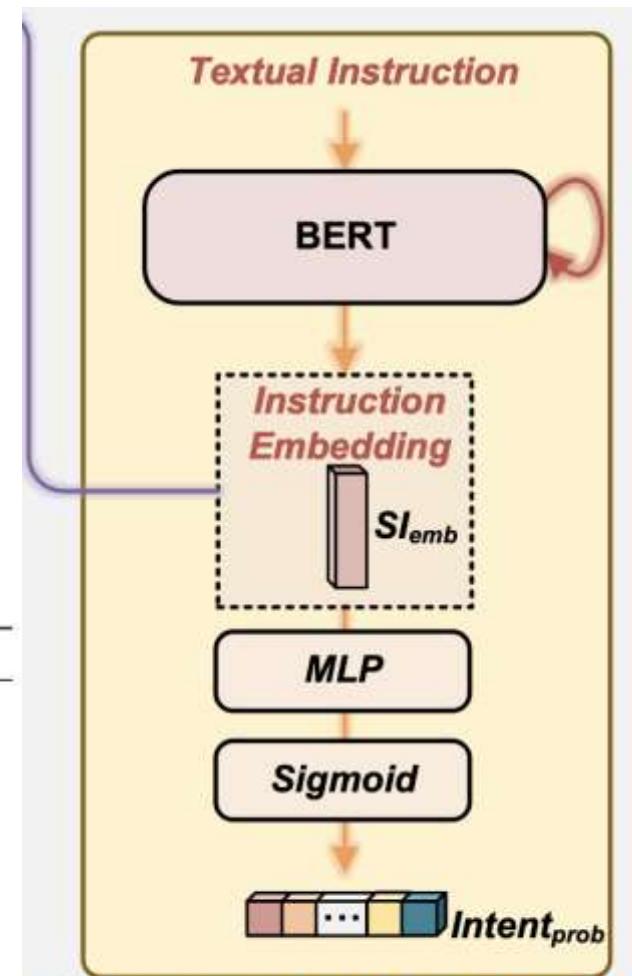
- 设计一个多标签意图识别（IID）任务。
- 6类指令意图：高度调整（ALT_ADJ）、航迹偏置（OFFSET）、取消偏置（CANOFF）、直飞（FLYTO）、SPD_ADJ（速度变化）、HEAD_ADJ & FLYTO（航向变化）

输入：文本化指
令

输出：获得高质
量的句子级指令

嵌入向量（SI_{emb}）

No.	Intents	Descriptions	Example
1	ALT_ADJ	The instruction indicates the altitude changes.	Air China nine three one climb maintain one two thousand two hundred meters
2	OFFSET	The instruction indicates the offset of the track.	Qatari eight four seven four offset three miles right of the track due to weather
3	CANOFF	The instruction indicates cancel offset.	Loulan two six one six can cancel offset
4	FLYTO	The instruction indicates direct to the waypoint.	Dynasty five nine nine seven direct to marso
5	SPD_ADJ	The instruction indicates speed changes.	West China six two five tree increase speed mach point seven six
6	HEAD_ADJ & FLYTO	The instruction indicates heading changes.	Sichuan eight seven four five turn left direct to ubdid



② **Intent-oriented Instruction Embedding Learning**

II Methods

Stage3：多模态FTP的微调 (Multi-modal FTP finetuning)

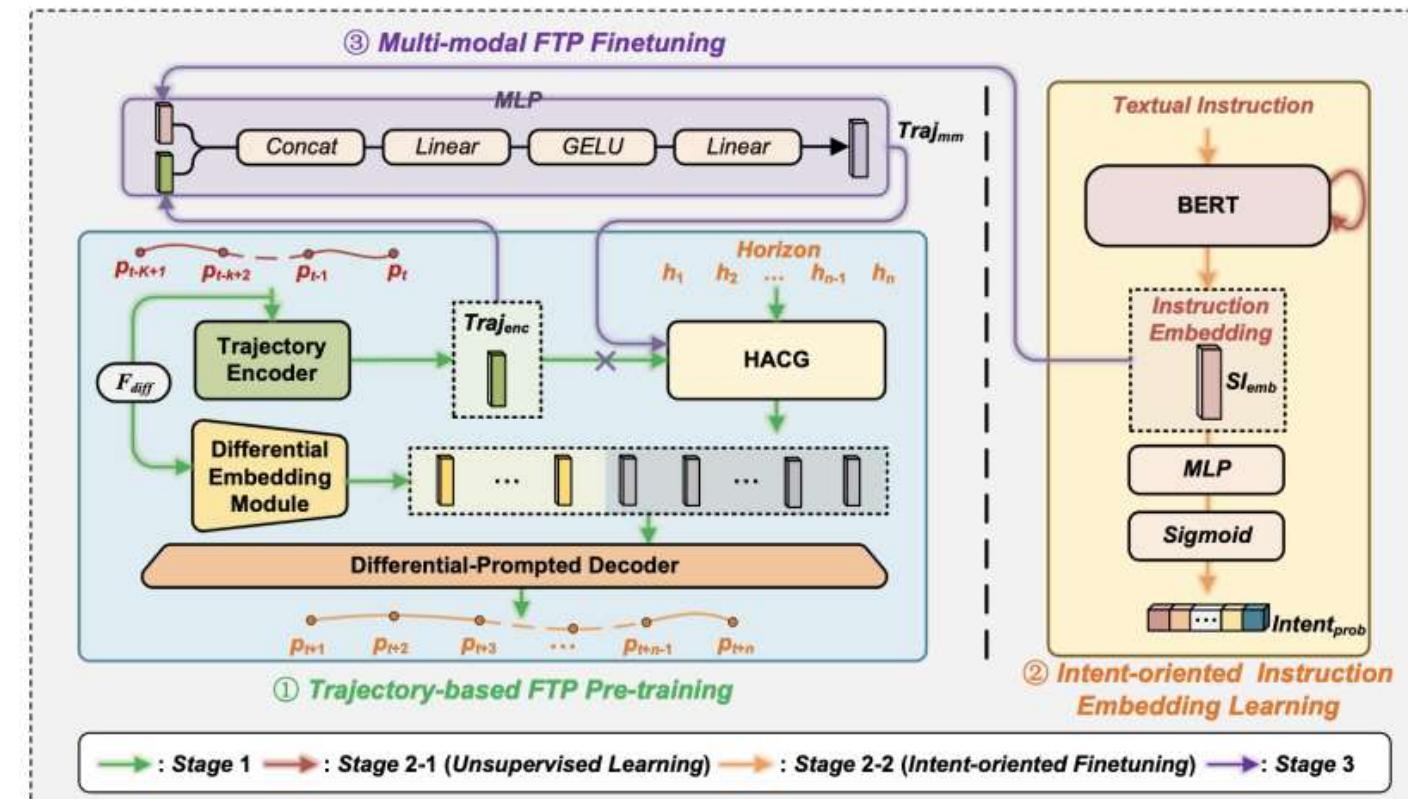
目的：利用少量的“轨迹-指令”成对数据，教会模型如何根据指令调整轨迹预测。

融合机制：

- 简单的拼接： $\mathbf{J}_{mm} = \text{Concat}[\mathbf{Traj}_{enc}, \mathbf{SI}_{emb}]$
+ MLP 映射 $\mathbf{Traj}_{mm} = \text{MLP}(\mathbf{J}_{mm})$
- 将融合后的特征输入到预训练好的解码器中。

训练策略：

- 使用少量的“轨迹-指令”配对数据。
- 联合优化，让模型学习“听到什么指令，应该产生什么样的轨迹变化”



III Results

III Results

实验设置与数据集

数据集: M2ATS（来自2021年2月19日至2月27日中国真实空管系统的数据），涵盖飞行轨迹数据、飞行计划、空域信息以及带有标准标注的空管通信语音

- 划分：轨迹子集(10,362)、文本指令子集(113,862)和轨迹-指令子集(7520)。前7天的数据作为训练集，剩余两天的数据分别用于验证和测试

对比基线 (Baselines):

- 传统迭代预测模型：LSTM、Transformer、Kalman-Filter、FlightBERT、WTFTP
- 直接预测式：LSTM+Attention, FlightBERT++

评价指标：

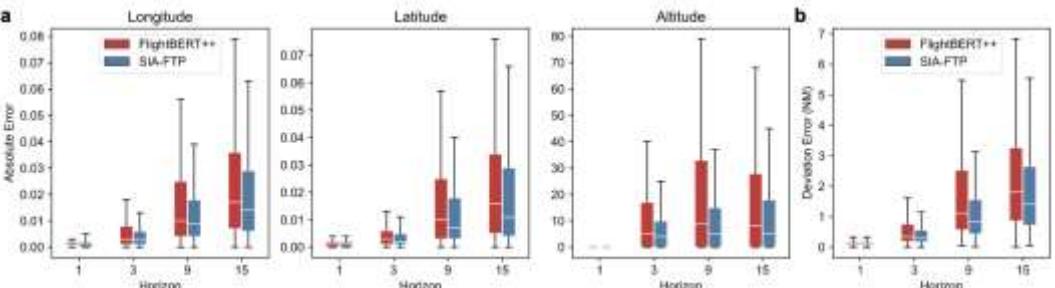
- MAE、MAPE、RMSE（针对经纬度高度）
- MDE (Mean Deviation Error): 预测轨迹点与实际轨迹点在三维空域中的欧几里得距离，这是最直观的指标。

III Results

定量实验结果

结论：SIA-FTP 在所有预测步长上均优于基线模型。

- 在长时预测（9步~15步，3~5分钟）上优势最明显。
- 相比最佳基线（FlightBERT++），在15步预测上的 MDE 降低了 20% 以上。
- 证明了引入语音指令有效纠正了长时预测中的趋势偏差。
- 误差分布的箱线图：SIA-FTP 误差更低且更稳定



Style	Methods	Horizon	MAE ↓			MAPE (%) ↓			RMSE ↓			MDE ↓
			Lon	Lat	Alt	Lon	Lat	Alt	Lon	Lat	Alt	
Iterative	LSTM	1	0.0050	0.0055	2.15	0.0046	0.0205	0.24	0.0071	0.0078	3.24	0.48
		3	0.0063	0.0066	4.37	0.0058	0.0246	0.48	0.0093	0.0097	7.36	0.58
		9	0.0138	0.0135	13.41	0.0128	0.0502	1.47	0.0251	0.0246	21.38	1.23
		15	0.0253	0.0260	23.63	0.0234	0.0969	2.57	0.0479	0.0587	37.80	2.31
	Transformer	1	0.0032	0.0041	1.69	0.0030	0.0150	0.19	0.0044	0.0051	2.82	0.33
		3	0.0065	0.0082	4.10	0.0060	0.0305	0.45	0.0095	0.0111	7.35	0.66
		9	0.0172	0.0208	12.12	0.0160	0.0771	1.32	0.0282	0.0309	21.71	1.70
		15	0.0292	0.0321	24.59	0.0271	0.1192	2.70	0.0505	0.0611	41.51	2.72
	Kalman-Filter	1	0.0044	0.0047	3.40	0.0041	0.0177	0.38	0.0165	0.0087	5.53	0.41
		3	0.0084	0.0089	7.15	0.0077	0.0331	0.79	0.0296	0.0164	11.79	0.77
		9	0.0266	0.0250	19.87	0.0247	0.0928	2.21	0.0871	0.0462	32.12	2.29
		15	0.0639	0.0592	37.53	0.0594	0.2199	4.15	0.2512	0.1208	74.96	5.50
	FlightBERT	1	0.0029	0.0018	1.63	0.0027	0.0068	0.18	0.0235	0.0311	12.33	0.25
		3	0.0047	0.0032	3.54	0.0043	0.0121	0.39	0.0309	0.0181	16.33	0.39
		9	0.0136	0.0100	12.51	0.0126	0.0373	1.39	0.0381	0.0286	28.83	1.13
		15	0.0287	0.0192	34.18	0.0267	0.0715	3.78	0.0733	0.0430	55.60	2.24
	WTFTP	1	0.0019	0.0016	1.10	0.0017	0.0060	0.12	0.0031	0.0025	2.08	0.16
		3	0.0032	0.0030	5.02	0.0029	0.0112	0.56	0.0052	0.0054	8.77	0.28
		9	0.0136	0.0129	19.07	0.0127	0.0481	2.10	0.0247	0.0253	30.32	1.21
		15	0.0260	0.0240	27.64	0.0241	0.0897	3.07	0.0492	0.0455	41.67	2.25
Direct	LSTM+Attention	1	0.0059	0.0060	2.19	0.0054	0.0225	0.25	0.0084	0.0091	2.98	0.53
		3	0.0058	0.0057	4.21	0.0054	0.0216	0.47	0.0082	0.0086	7.19	0.52
		9	0.0124	0.0115	14.41	0.0115	0.0428	1.58	0.0243	0.0235	25.42	1.09
		15	0.0201	0.0176	20.07	0.0186	0.0658	2.19	0.0425	0.0406	35.67	1.71
	FlightBERT++	1	0.0018	0.0014	1.02	0.0017	0.0055	0.12	0.0033	0.0027	2.94	0.16
		3	0.0037	0.0032	4.95	0.0035	0.0122	0.56	0.0064	0.0061	9.22	0.33
		9	0.0120	0.0105	13.80	0.0112	0.0392	1.53	0.0236	0.0203	25.14	1.03
		15	0.0180	0.0163	17.37	0.0167	0.0609	1.92	0.0348	0.0309	33.53	1.55
	SIA-FTP	1	0.0018	0.0014	0.85	0.0016	0.0053	0.10	0.0030	0.0027	2.57	0.15
		3	0.0034	0.0028	3.42	0.0032	0.0105	0.38	0.0061	0.0053	6.91	0.29
		9	0.0089	0.0078	7.51	0.0083	0.0290	0.83	0.0168	0.0160	14.53	0.77
		15	0.0141	0.0124	9.75	0.0131	0.0464	1.07	0.0285	0.0253	19.30	1.22

III Results

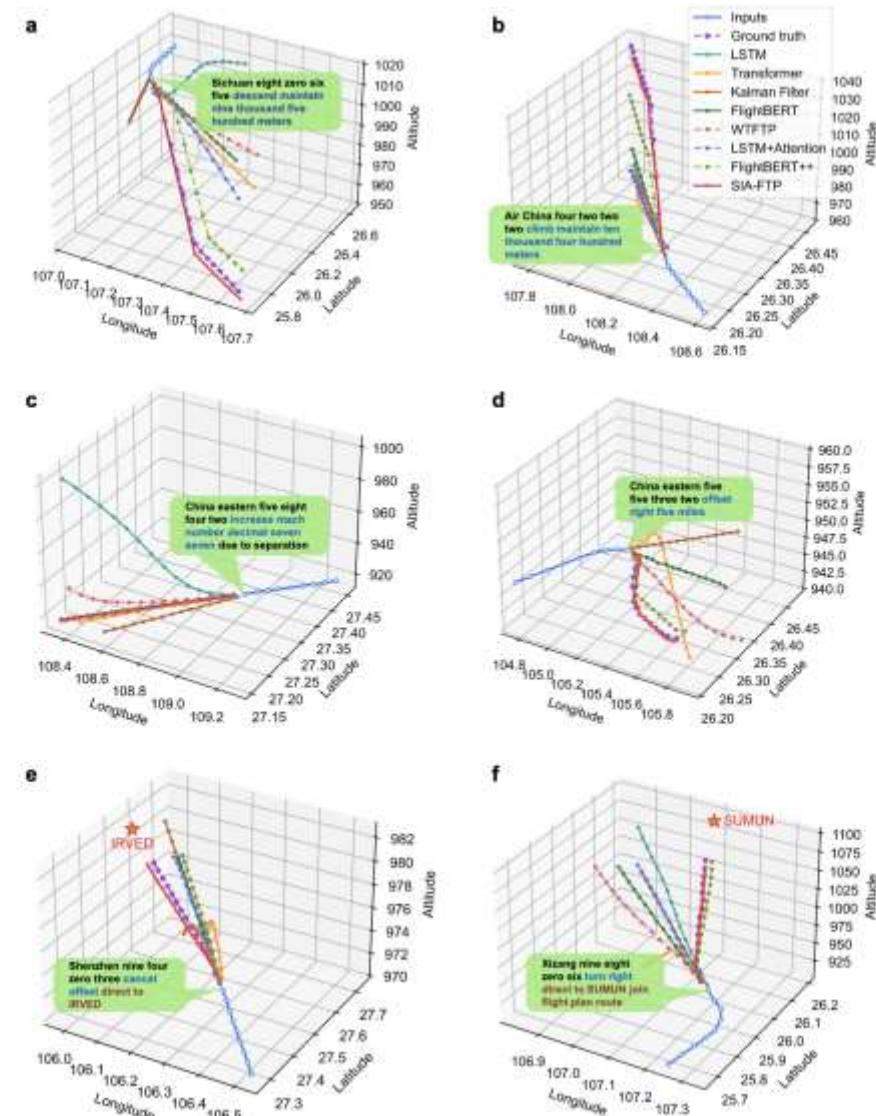
可视化与定性分析

直观展示模型在不同指令下的预测效果。

案例：

- 高度调整 (ALT ADJ)：基线模型往往预测“保持高度”，而 SIA-FTP 能根据指令预测出“下降/爬升”的趋势 (Figure a, b)
- 偏置 (OFFSET)：SIA-FTP 能预测出航路偏置的操作 (Figure d)
- 复合指令：如“右转 + 直飞”，SIA-FTP 能同时捕捉多个意图 (Figure e)

结论：SIA-FTP 具备机动意图感知能力。



III Results

消融实验

验证三阶段训练的必要性：

- A1 (w/o Stage 2-2)：去掉意图识别微调 \rightarrow 性能下降。
- A2 (w/o Stage 2)：完全去除第二阶段(意图导向的指令嵌入学习) \rightarrow 性能下降。

结论：

- 引入空管语音指令对于提高预测高机动飞行轨迹是有显著帮助的
- 完整的第 2 阶段（包括无监督预训练和有监督意图微调）有助于提高预测精度

Table 2: The experimental results of the ablation study. Notation w/o indicates the specific module is not included in the SIA-FTP.

Exp.	Methods	Horizon	MAE ↓			MAPE (%) ↓			RMSE ↓			MDE ↓
			Lon	Lat	Alt	Lon	Lat	Alt	Lon	Lat	Alt	
A1	SIA-FTP (w/o Stage 2-2)	1	0.0020	0.0016	0.79	0.0019	0.0060	0.10	0.0037	0.0030	2.35	0.17
		3	0.0037	0.0031	4.14	0.0034	0.0114	0.46	0.0066	0.0057	7.46	0.32
		9	0.0094	0.0085	8.34	0.0088	0.0317	0.92	0.0183	0.0167	15.50	0.83
		15	0.0149	0.0139	10.55	0.0139	0.0519	1.15	0.0298	0.0271	20.34	1.32
A2	SIA-FTP (w/o Stage 2)	1	0.0018	0.0015	0.76	0.0017	0.0057	0.09	0.0032	0.0030	2.37	0.16
		3	0.0034	0.0029	3.92	0.0031	0.0110	0.44	0.0058	0.0055	7.25	0.30
		9	0.0087	0.0088	7.99	0.0081	0.0329	0.88	0.0161	0.0180	15.48	0.80
		15	0.0137	0.0142	9.90	0.0127	0.0528	1.08	0.0262	0.0278	21.36	1.27

↓ represents minimization indicators.

III Results

可解释性研究

嵌入可视化 (t-SNE):

- 对比 BERT 预训练、IID 微调、SIA-FTP 微调后的指令嵌入分布。
- 结论：经过 Stage 2 和 Stage 3 训练后，不同意图的指令在特征空间中区分度越来越高，聚类更加紧凑。甚至能区分同一意图下的不同参数（如：下降到8400m vs 下降到8900m）。

注意力权重 (Token Weights):

- 可视化模型关注指令中的哪些词。
- 热力图显示，模型不仅关注意图词（如 "turn left"），还

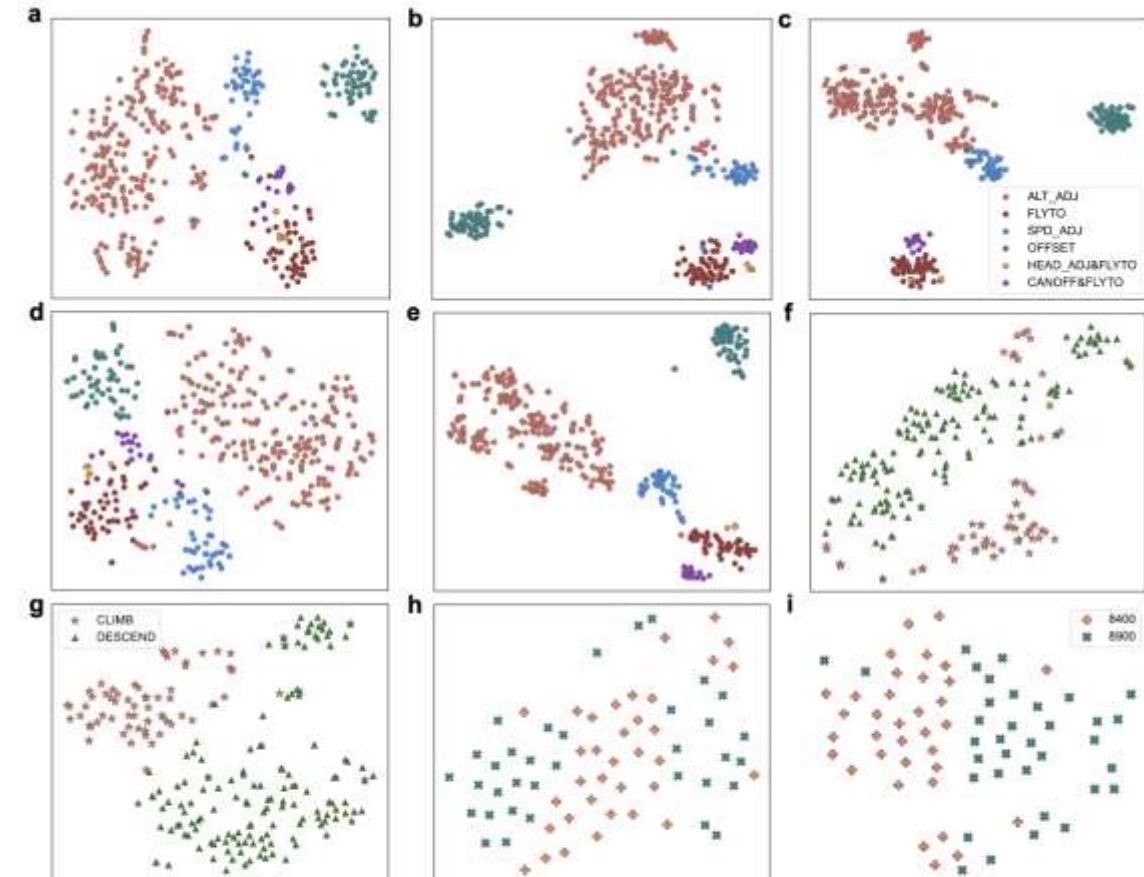
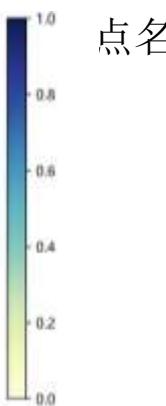
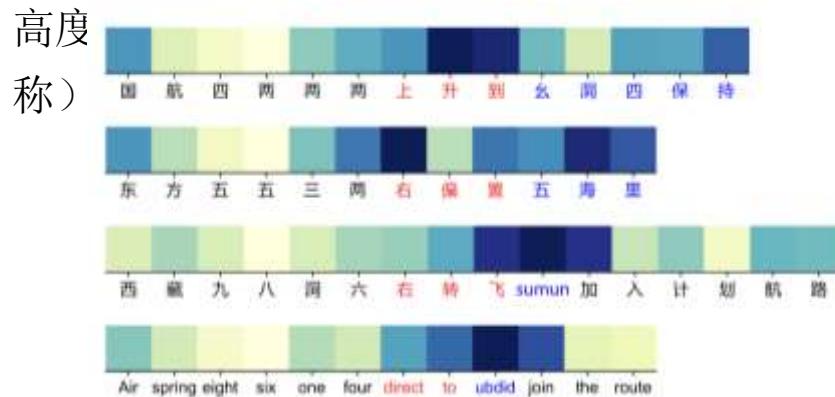


图4：文本指令嵌入可视化展示。**a-e**不同模型通过 TSNE 工具呈现的文本指令嵌入可视化结果（含不同操作意图）。**a:** BERT 模型（阶段2-1）。**b:** IID 模型（阶段2-2）。**c:** SIA-FTP 模型（阶段1至3）。**d:** A1模型（未使用阶段2-2的SIA-FTP）。**e:** A2模型（未使用阶段2的SIA-FTP）。**f-g**不同模型通过 TSNE 工具呈现的文本指令嵌入可视化结果（含爬升与下降意图）。**f:** IID 模型（阶段2-2）。**g:** SIA-FTP 模型（阶段1至3）。**h-i**不同模型通过 TSNE 工具呈现的文本指令嵌入可视化结果（含不同下降意图）。**h:** IID 模型（阶段2-2）。**i:** SIA-FTP 模型（阶段1至3）。源数据文件已作为源数据文件提供。

III Results

泛化性与真实场景应用

骨干网络泛化：将 SIA-FTP 的核心思想应用到 LSTM+Attention 架构上，依然取得了显著提升（Figure 6a）。

语言模型泛化：将 BERT 换成 LSTM 或 Transformer 处理文本，依然有效（Figure 6b）。

真实场景测试 (ASR Cascade)：

- 前端接入 ASR（自动语音识别）系统，使用识别出的文本作为输入。
- 结果：性能仅微弱下降，证明模型对 ASR 噪声具有鲁棒性，具备实战价值。

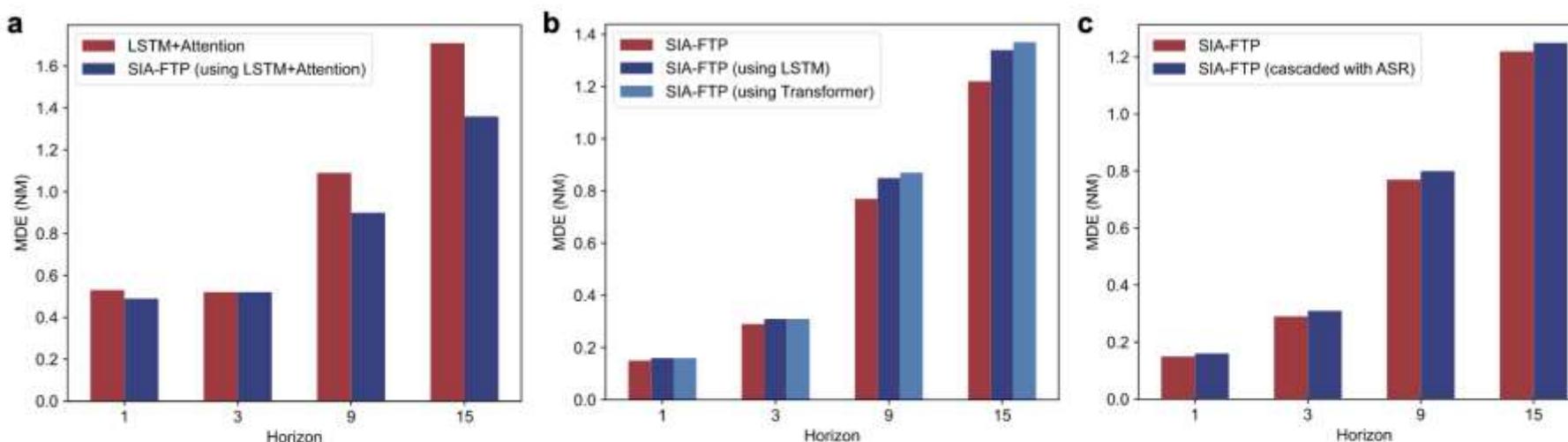


图6：在1、3、9和15个预测时间范围内的 MDE 比较，用于泛化研究。**a** LSTM +注意力机制和 SIAFTP 在第一阶段使用 LSTM +注意力机制的 MDE 结果。**b** SIA-FTP、第二阶段使用 LSTM 的SIA-FTP以及第二阶段使用Transformer的SIA-FTP的 MDE 结果。**c** SIA-FTP 和与ASR系统级联的SIA-FTP的 MDE 结果。源数据文件已提供。

IV Discussion

IV Discussion

总结与贡献

主要贡献:

- 提出了 Instruction-driven FTP 范式，填补了空管自动化中“人为意图”缺失的空白。
- 设计了 SIA-FTP 框架，通过 3阶段渐进式学习 解决了模态鸿沟和数据稀缺问题。
- 提出了意图导向的指令嵌入方法，有效提取机动意图和参数。
- 实现了对机动飞行的精准预测（MDE 降低 20%+）

实际意义:

- 提高飞行轨迹预测精度。
- 支持下游应用：冲突检测、指令执行监控、人为错误预警（如果预测轨迹与实际严重不符，可能意味着飞行员误操作或指令有误）

局限性:

- 对极其复杂的复合指令（多个动作同时发生）区分度仍有提升空间。

谢 谢 !