2025 中国智能车未来挑战赛离线测试比赛赛题

一、比赛整体的简要介绍

无人驾驶是人工智能重要应用场景之一，是近年来的国际研究热点，发展无人驾驶技术也已经上升到国家战略高度。得益于我国良好的基础设施建设，通信传输能力与协同感知能力深度融合的网联智能方案成为我国无人驾驶的重要发展方向。

在网联智能无人驾驶的技术路线下，如何实现智能车之间高效的信息共享、如何实现精确鲁棒的智能车协同定位与感知成为关键问题。复杂交通场景通信与感知算法挑战赛旨在挖掘多模态感知信息对通信传输的增强作用和多车通信对协同感知的提升作用。

复杂交通场景通信与感知算法挑战赛分为五个赛题。

赛题一为预训练 LLM4PG 赋能的智能车射频地图构建。射频地图是表征电磁信号传播衰减的重要指标，其精确生成对空中出租车网络设计及智能车轨迹优化具有关键意义。在复杂城镇与郊区环境中，建筑遮挡、交通动态及天气变化等因素使传统射频地图构建方法难以兼顾精度与泛化能力。本赛题要求选手基于预训练 LLM4PG（Adapting Large Language Model for Pathloss Map Generation），充分利用地面车辆与空中出租车等多车多路端的多模态感知数据（包括 RGB 图像、深度图像、LiDAR 点云及毫米波雷达点云），探索预训练 LLM4PG 在跨模态特征融合与时空场景理解方面的潜力，提升在复杂环境下智能车射频地图构建能力。

赛题二为无线基座模型 WiFo 赋能的多模态感知辅助智能车射频链路预测。在复杂车联网/低空经济环境中，低开销和高精度的 CSI 获取对于维持通信链路

的稳定性并降低通信系统能耗至关重要。智能车射频链路预测根据已知部分子载波处的 CSI 预测未知的 CSI，在显著节省系统资源的情况下维持通信链路的稳定性。在实际通信系统中，往往面临复杂且突发多变的场景，并且数据收集成本和模型训练开销较大，对传统的基于参数化模型和任务专用 AI 模型的射频链路预测算法提出了挑战。因此，本赛题旨在探索如何利用无线基座模型 WiFo 通过少样本微调快速迁移至该任务，充分挖掘多模态感知数据对智能车射频链路预测任务的增益。

赛题三为无线基座模型 WiFo 赋能的多模态信息智能车定位。即时精准的定位对自动驾驶车辆至关重要，是实现控制决策与路径导航等任务的基础保障。然而，复杂交通场景下，GPS 信号易受干扰，同时现有车辆自主追踪定位算法的可靠性面临严重挑战。因此，需要利用车联网系统，通过多源信息融合与人工智能算法优化实现更加准确鲁棒的智能车定位。赛题要求利用路端设备采集的通信与感知数据实现对环境中车辆目标的位置估计，并提供具有无线信道通用表征与编码能力的基座模型 WiFo，鼓励参赛者充分在其基础上挖掘多模态数据间的内在关联，设计灵活可靠的智能车定位方案。

赛题四为自动驾驶场景交通事故风险预测。面向自动驾驶场景交通事故风险预测，要求参赛队伍根据不同事故类型下的行车环境多模态事故视频理解数据，包括 RGB 图像、文本信息，交通参与者位置信息，实现在固定事故视频长度条件下同时本赛题提供交通参与者的位置信息，可有助于发现事故的风险区域。

赛题五为认知启发的风险驾驶场景关键目标定位。使用风险场景驾驶视觉注意数据集，给定一段风险驾驶场景视频片段，输出与人类驾驶员注视点(gaze fixation)相关的边界框，这些目标被认为是关键目标。特别鼓励参赛者以赛题给定数据集作为基准，并结合多模态大模型（MLLM）进行网络建模。评测将在测试集上进行。即要求参赛者定位出当前正在发生事故视频帧的关键目标，并输出

关键目标的位置(box)。

复杂交通场景通信与感知算法挑战赛将于 10 月 20 日发布赛题任务，10 月 30 日发布赛题任务训练集，11 月 5 日发布赛题任务测试集，并开放测试结果提交，12 月 15 日测试结果提交截止，12 月 20 日公布比赛成绩并颁奖。每支报名参赛的队伍应由 1-10 人组成，可选择赛题中的任意一题或多题报名参赛，请各参赛队伍在提交结果时确认队伍成员信息。

我们热诚欢迎从事无人驾驶技术研发的高校、科研院所与企业积极参与本届比赛，共同为提高无人驾驶智能车研发水平、促进产学研合作、推动国民经济与社会发展做出积极贡献。感谢您一直以来对中国智能车未来挑战赛的关注与支持！

1. 联系方式

邮箱：房老师 fangjianwu@mail.xjtu.edu.cn

张老师 zhangjianan@pku.edu.cn

黄老师 ziweihuang@pku.edu.cn

李老师 670160532lileilei@gmail.com

2. 时间安排

10 月 31 日	算法挑战赛赛题和训练集发布
11 月 10 日	赛题任务测试集发布，测试结果提交开始
12 月 15 日	测试结果提交截止
12 月 19 日	决赛答辩
12 月 20 日	大赛颁奖

二．赛题介绍

赛题一：预训练 LLM4PG 赋能的智能车射频地图构建

本赛题要求选手基于预训练 LLM4PG（Adapting Large Language Model for Pathloss Map Generation）的推理和泛化能力，面向车路协同场景和低空经济场景，利用地面车辆与空中出租车等多车多路端的多模态感知数据，实现城镇与郊区场景下的射频地图构建，为智能车轨迹优化提供支撑。输入数据涵盖不同场景、车流量密度、天气和频段下的感知信息，包括 RGB 图像、深度图像、激光雷达（LiDAR）点云及毫米波雷达点云，如图 1-1 所示。基于此，引入预训练 LLM4PG，融合多模态特征与语义先验知识，利用大模型在跨模态理解与时空特征建模方面的优势，实现复杂环境中智能车射频地图的高精度构建。

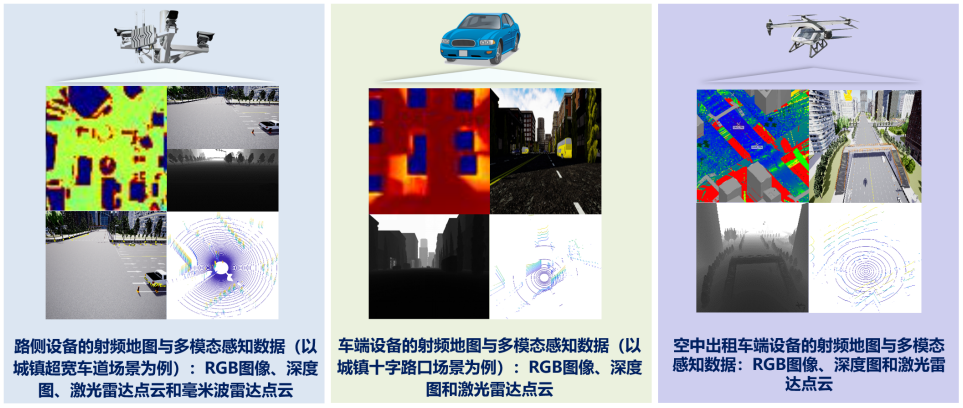


图 1-1 路端/车端设备的射频地图、RGB 图像、深度图像、LiDAR 点云数据和毫米波雷达点云数据

本赛题包含九个车联网场景条件（a-i）和一个低空经济场景条件（j），分别是（a）毫米波晴天高车流量密度城镇超宽车道场景；（b）毫米波晴天低车流量密度城镇超宽车道场景；（c）毫米波晴天高车流量密度郊区分岔路口场景；（d）sub-6 GHz 晴天高车流量密度郊区分岔路口场景；（e）sub-6 GHz 晴天中车流量密度城镇十字路口场景；（f）毫米波晴天中车流量密度城镇十字路口场景；（g）毫米波晴天高车流量密度城镇十字路口场景；（h）毫米波雨天中车流量密度城镇十字路口场景；（i）毫米波雪天中车流量密度城镇十字路口场景；（j）sub-6 GHz 晴

天低车流量密度城镇超宽车道场景。

以下是本赛题任务输入输出的详细介绍：

- 任务输入：每种场景条件下的路端与车端（包含地面车辆和空中出租车）的多模态感知数据，即 RGB 相机的图像文件、深度相机图像文件、LiDAR 点云文件和毫米波雷达的点云文件。
- 任务输出：射频地图信息，以场景条件_车端/路端编号_时刻.txt 文件命名，如场景条件为毫米波晴天高车流量密度城镇超宽车道，则场景条件命名为 mmWave_sunny_highVTD_widelane，生成 Car5 在第 20 个时刻的射频地图生成值，则其生成结果文件命名为 mmWave_sunny_highVTD_widelane_Car5_snapshot20.txt，文件存储的详细格式请到“活动报名”标签页查看。

（1）参赛指南

我们使用均方误差（Mean Squared Error, MSE）指标作为评分标准，评分细则请见下方。

选手应以 txt 格式上传结果，细则请见“活动报名”标签页，请选手严格按照格式提交结果，否则无法参加测评！

请您遵守比赛规则，文明参赛！请在测试结果提交截止时间之前提交您的结果！

（2）评分标准

评估生成的射频地图精确度的指标为 MSE：

$$\mathrm{MSE} = \frac{1}{n}\sum_{m=1}^{n}(\widehat{y_m} - y_m)^2$$

其中，n 为样本数量，为特定场景条件和时刻下的特定设备的射频地图生成值，为特定场景条件和时刻下的特定设备的射频地图真值（ground truth）。

- 低空经济场景条件：$\widehat{y_m}$为特定场景条件和时刻下的特定设备的射频地图生成值（像素值范围 0-255，对应射频地图值 0-255 dB，图像分辨率为71*73），y_m为特定场景条件和时刻下的特定设备的射频地图真值（ground truth）
- 计算不同场景、不同天气、不同车流量密度和不同频段下十个场景条件中的均方误差算数平均值（保留四位有效数字），并由低到高排序

$$\begin{aligned}\text{Score} =& 0.1\text{MSE}_{\text{mmWave,sunny,high,widelane}} + 0.1\text{MSE}_{\text{mmWave,sunny,low,widelane}} \\ &+0.1\text{MSE}_{\text{mmWave,sunny,high,forking}} + 0.1\text{MSE}_{\text{sub6,sunny,high,forking}} \\ &+0.1\text{MSE}_{\text{sub6,sunny,medium,crossroad}} + 0.1\text{MSE}_{\text{mmWave,sunny,medium,crossroad}} \\ &+0.1\text{MSE}_{\text{mmWave,sunny,high,crossroad}} + 0.1\text{MSE}_{\text{mmWave,rainy,medium,crossroad}} \\ &+0.1\text{MSE}_{\text{mmWave,snowy,medium,crossroad}} + 0.1\text{MSE}_{\text{sub6,sunny,low,widelane}}\end{aligned}$$

（3）选手须知

- 所提交结果应与所提交代码运行结果相符，严禁造假，违者将被取消参与排名资格；
- 选手应基于所提供的数据集开发算法，禁止私自增加训练集；
- 禁止使用测试集中的数据参与神经网络训练；
- 应严格按照格式要求提交结果，格式不符者无法参与结果评测；

请在截止日期前提交您的结果，过期提交将被视为无效！

（4）参考文献：为便于选手深入理解赛题以及要求使用的模型，给出以下 2 篇文章供参考

[1] X. Cheng, B. Liu, X. Liu, E. Liu, and Z. Huang, “Foundation Model Empowered Synesthesia of Machines (SoM): AI-Native Intelligent Multi-Modal Sensing-Communication Integration”, *IEEE Transactions on Network Science and Engineering*, early access, 2025.

[2] B. Liu, S. Gao, X. Liu, X. Cheng, and L. Yang, “WiFo: Wireless Foundation Model for Channel Prediction,” *SCIENCE CHINA Information Sciences*, vol. 68, no. 6, pp. 162302, May 2025.

（5） 联系方式

孙铭然 mingransun@stu.pku.edu.cn

赛题二：无线基座模型 WiFo 赋能的多模态感知辅助智能车射频链路预测

本赛题要求选手利用多模态感知信息的环境感知能力和无线基座模型（WiFo）强大的推理和泛化能力，在网联智能车联网/低空经济场景下，实现高精度智能车射频链路预测。输入数据包含车联网场景（分岔路口高车流量场景）和低空经济场景（超宽车道低车流量密度）的多模态感知数据和前 32 个子载波上的 CSI，输出为后 32 个子载波上的 CSI，如图。本项任务将在车联网/低空经济两个场景中分别进行测试，选手的总分为两个场景测试得分的均值。值得注意的是，为了考验选手模型应对实际应用中的感知数据异常情况（比如意外遮挡、数据缺失）的适应能力，车路协同场景和低空经济场景的训练集和测试集中以一定比例向感知数据中引入意外遮挡、数据缺失和 GPS 定位误差三种数据异常处理：

1）数据缺失：每帧中每种感知模态的数据有 10%概率缺失；

2）意外遮挡：RGB 图和深度图约 10%帧数的数据被随机遮盖部分区域，LiDAR 和毫米波雷达点云约 10%帧数的数据随机缺失了部分角度范围的点云；

3）GPS 定位误差：为方便选手处理，车辆的 GPS 信息以二维坐标信息的格式给出，并添加了均值为 0、方差为 5 的高斯噪声，以模拟实际应用场景中 GPS 的定位误差。

以下是本赛题任务输入输出的详细介绍：

任务输入：每种场景条件下的路端的多模态感知数据。对于车联网场景，输入包含 RGB 相机图像文件、深度相机图像文件、LiDAR 点云文件、毫米波雷达点云文件，车辆位置信息文件和前 32 子载波上的 CSI 信息；对于低空经济场景，输入包含 RGB 相机图像文件、深度相机图像文件和 LiDAR 点云文件和前 32 个子载波上的 CSI 信息。

任务输出：每种场景条件下的其余子载波上的 CSI 信息。结果提交的详细格式请到“活动报名”标签页查看。

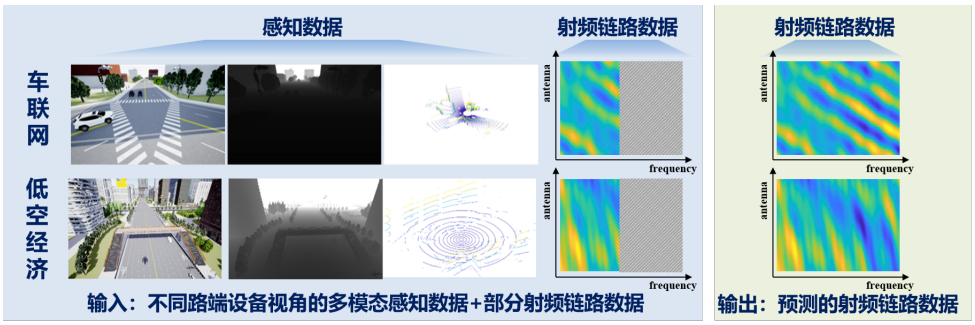


图 2-1 不同路端设备视角的多模态感知数据和预测信道数据

（1）参赛指南

我们使用预测误差 NMSE 指标作为评分标准，评分细则请见下方。

选手应以 csv 格式上传结果，细则请见“活动报名”标签页，请选手严格按照格式提交结果，否则无法参加测评！

请您遵守比赛规则，文明参赛！

请在测试结果提交截止时间之前提交您的结果！

（2）评分标准

车联网/低空经济场景评估预测射频链路精度的指标均为标准化均方误差(NMSE)：

$$NMSE_{\mathrm{V2I}-i,j} = \sum_{n} \frac{\|\hat{H}_n - \bar{H}_n\|_F^2}{\|\bar{H}_n\|_F^2}$$

$$NMSE_{\mathrm{AT2I}-j} = \sum_{n} \frac{\|\hat{H}_n - \bar{H}_n\|_F^2}{\|\bar{H}_n\|_F^2}$$

其中：

$\hat{H}_n$:第 n 个 sample 预测的信道矩阵；

$\overline{H}_n$:第 n 个 sample 真实的信道矩阵；

N：测试集总数据量；i:车联网场景序号；j:链路序号。

为了便于表示，车联网场景与低空经济场景的预测得分将被归一化至 0-1 之间如下：

$$s_{\mathrm{V2I}} = \mathrm{mean}\sum_{i}\sum_{j}\max\{0, 1 - \frac{\log_{10}(NMSE_{\mathrm{V2I}-max})}{\log_{10}(NMSE_{\mathrm{V2I}-i,j})}\}$$

$$s_{\mathrm{AT2I}} = \mathrm{mean}\sum_{j}\max\{0, 1 - \frac{\log_{10}(NMSE_{\mathrm{AT2I}-max})}{\log_{10}(NMSE_{\mathrm{AT2I}-j})}\}$$

其中:

$NMSE_{\mathrm{V2I}-max}$：车联网场景所容许的最大预测误差

$NMSE_{\mathrm{AT2I}-max}$：低空经济场景所容许的最大预测误差

最后取两组场景预测准确度测试得分平均值作为选手在该赛题的得分：

$$s = \frac{1}{2}(s_{\mathrm{V2I}} + s_{\mathrm{AT2I}})$$

（3）选手须知

1）所提交结果应与所提交代码运行结果相符，严禁造假，违者将被取消参与排名资格；

2）选手应基于所提供的数据集开发算法，禁止私自增加训练集；

3）应严格按照格式要求提交结果，格式不符者无法参与结果评测；

4）请在截止日期前提交您的结果，过期提交将被视为无效！

（4）参考文章：为便于选手深入理解赛题以及要求使用的模型，给出以下 2 篇文章供参考

[1] X. Cheng, B. Liu, X. Liu, E. Liu, and Z. Huang, “Foundation Model Empowered Synesthesia of Machines (SoM): AI-Native Intelligent Multi-Modal Sensing-Communication Integration”, *IEEE Transactions on Network Science and Engineering*, early access, 2025.

[2] B. Liu, S. Gao, X. Liu, X. Cheng, and L. Yang, “WiFo: Wireless Foundation Model for Channel Prediction,” *SCIENCE CHINA Information Sciences*, vol. 68, no. 6, pp. 162302, May 2025.

（6） 联系方式

刘伯珣 boxunliu@stu.pku.edu.cn

赛题三：无线基座模型 WiFo 赋能的多模态信息智能车定位

本赛题要求选手基于无线基座模型（WiFo）强大的推理和泛化能力，利用车路射频信道信息与路端感知数据，在网联智能车联网/低空经济复杂场景下，实现复杂场景、多样数据分布条件下高精度、鲁棒定位。输入数据包含车联网场景（十字路口中车流量场景、分岔路口高车流量场景、超宽车道低车流量场景）和低空经济场景（包含空中出租车的超宽车道低车流量场景）的路端设备与目标车辆之间的信道状态信息（CSI）与对应路端的感知数据（RGB 图像），输出为每种场景条件下目标车辆的位置估计结果。值得注意的是，为了考验选手模型应对实际应用场景中的受限情况（如部分时刻可能包含视野遮挡、缺失感知数据等）的适应能力，场景的训练集和测试集中可能包含一定比例的受限情况。

以下是本赛题任务输入输出的详细介绍。

任务输入：本赛题以不同场景为单位，每个场景下提供：

1）该场景的基本信息说明文档，包含场景示意图、全部目标车辆初始时刻的位置、路端设备的位置与朝向。用于选手进行开发的训练数据集额外提供目标车辆每个时刻的位置真值。

2）路端设备的 RGB 相机图像文件以及与对应目标车辆之间无线信道的信道状态信息（CSI）。

任务输出：全部目标车辆全部时刻的位置估计结果，以‘car_目标车编号.txt’文件命名。如目标车 1 的结果文件命名为‘car_1.txt’。文件存储的详细格式规定如下：每个 txt 文件行数为全部时刻帧数 T，每行三个浮点数分别代表该时刻下位置估计结果的 x、y、z 坐标（以米为单位）。

（1）参赛指南

我们使用归一化平移误差***E^T***（百分比）作为评分标准，评分细则请见下

方。

选手应以 txt 格式上传结果，细则请见“活动报名”标签页，请选手严格按照格式提交结果，否则无法参加测评！

请您遵守比赛规则，文明参赛！

请在测试结果提交截止时间之前提交您的结果！

（2）评分标准

车联网/低空经济场景评估位姿估计精确度的指标为归一化平移误差$\boldsymbol{E}$^$\boldsymbol{T}$（百分比）：

$$E_{i,j}^{T}=\frac{1}{N_{i,j}}\sum_{k}^{N_{i,j}}\frac{||\Delta\vec{p}_{est,i,j,k}-\Delta\vec{p}_{gt,i,j,k}||_2}{||\Delta\vec{p}_{gt,i,j,k}||_2}\quad E_{all}^{T}=\frac{1}{\sum_{i=1}^{N}N_i}\sum_{i=1}^{N}\sum_{j=1}^{N_i}E_{i,j}^{T}$$

其中，$E\_$(i,j)^T代表第 i 个场景第 j 个轨迹的归一化平移误差，E_all^T代表所有轨迹归一化平移误差的算术平均值.

选手在本赛题中的最终得分为所有测评场景中$\boldsymbol{E}$^$\boldsymbol{T}$由低到高排序：

$$\mathbf{score}=\boldsymbol{E}_{\boldsymbol{all}}^{\boldsymbol{T}}$$

（3）选手须知

1）所提交结果应与所提交代码运行结果相符，严禁造假，违者将被取消参与排名资格；

2）选手应基于所提供的数据集开发算法，禁止私自增加训练集；

3）应严格按照格式要求提交结果，格式不符者无法参与结果评测；

4）请在截止日期前提交您的结果，过期提交将被视为无效！

（4）参考文章：为便于选手深入理解赛题以及要求使用的模型，给出以下 2 篇文章供参考

[1] X. Cheng, B. Liu, X. Liu, E. Liu, and Z. Huang, “Foundation Model Empowered Synesthesia of Machines (SoM): AI-Native Intelligent Multi-Modal Sensing-Communication Integration”,

IEEE Transactions on Network Science and Engineering, early access, 2025.

[2] B. Liu, S. Gao, X. Liu, X. Cheng, and L. Yang, "WiFo: Wireless Foundation Model for Channel Prediction," *SCIENCE CHINA Information Sciences*, vol. 68, no. 6, pp. 162302, May 2025.

（5） 联系方式

李思江 pkulsj@pku.edu.cn

赛题四：自动驾驶场景的交通事故预测

本赛题要求参赛队伍根据不同事故类型下的行车环境多模态事故视频理解数据，包括 RGB 图像、驾驶员视觉眼动数据，文本信息（事故原因和事故类型文本标注），交通参与者位置信息，特别地，赛题鼓励参赛队伍将视觉眼动数据作为引导条件，实现在固定事故视频长度条件下(帧率为 30，视频时长 5s)的交通事故风险预测。

即要求τ在t_a时刻p_t大于规定的事故风险阈值(通常设置为 0.5)的情况下达到最大,即距离事故起始的时间 TTA(Time-to-Accident)最大。同时本赛题提供交通参与者的位置信息，可有助于发现事故的风险区域。如图 4-1 示例：

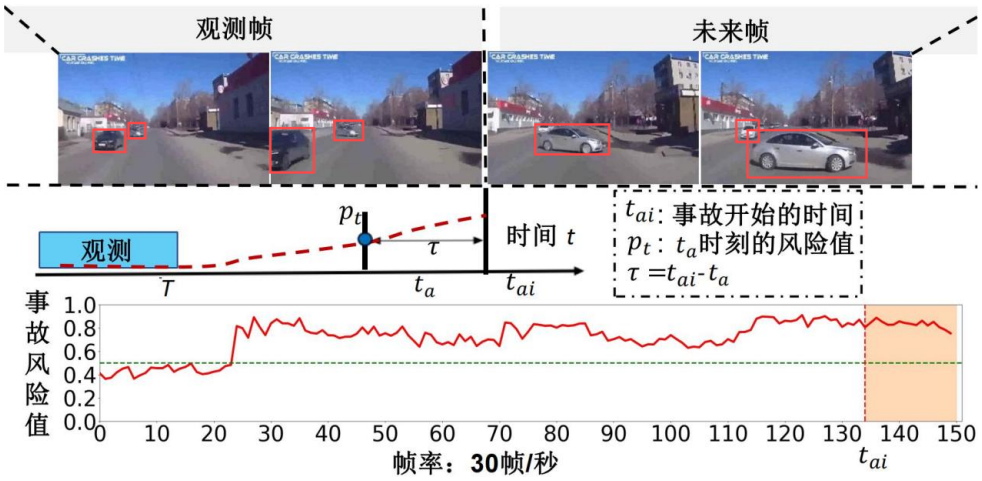


图 4-1 5 秒事故视频（150 帧），自车碰撞其他车辆的事故预测结果

图 4-2 部分事故类型数据

本赛题包含五类行车道路场景：高速、城市、乡村、山区和隧道道路，考虑多种稀有事故情形：特殊天气（雨天、雪天、雾天）与光照（夜晚）。每个场景

中均包含事故时间窗口标注和与事故时间窗口对齐的文本数据。

以下是本赛题任务输入输出的详细介绍。

任务输入：本赛题以不同事故类型视频序列为单位，每个视频下提供

1）包含事故起始时刻到结束时刻的完整视频帧序列和驾驶员视觉眼动数据（分辨率:1280×720,以.jpg 格式保存）；

2） 该视频的基本标注 txt 文件，包含事故标签（发生事故 1/未发生事故 0）、事故类型、事故起始时刻（t_{ai}，其中未发生事故的视频t_{ai}标注为-1）、文本信息，用于选手进行开发的数据额外提供交通参与者位置信息的.json 文件。

任务输出：全部用于测试的 5s（150 帧）事故视频序列的风险预测结果。将每一个测试视频序列的风险预测结果和相应标签存储为字典，如以下方式进行存储：“视频序号”: {“risk”: $risk_1$, $risk_2$, ..., $risk_T$ (规定 T 为 150，即根据观测输出 150 个风险值)], “label”: 1 或者 0，“t_{ai}”：t_{ai}或-1}，风险值以 ndarray 的数组形式进行存储，即字典的键是原始的视频序号，值是预测的风险值、事故标签和事故的开始时刻等标注信息。最后将所有测试结果存储到一个.json 文件里进行逐一读取。

（2）参赛指南

我们使用分类指标（AP、AUC）、事故预测性能指标（TTA_0.5、STTA_0.5）指标作为评分标准，评分细则请见下方。

选手应以 json 文件格式上传结果，细则请见“结果提交”标签页，请选手严格按照格式提交结果，否则无法参加测评！

（3）评分标准

评估模型对事故风险估计精确度的指标为 AP (Average Precision)、AUC (Area under the Curve of ROC)与事故预测风险性能指标 TTA_0.5 (Time to Accident at threshold of 0.5)和 STTA_0.5 (Stabilized time to accident at threshold of 0.5)

精确率(Precision) $= \frac{TP}{TP+FP}$　　召回率(Recall) $= \frac{TP}{TP+FN}$

$$\mathrm{AP} = \sum_{\mathrm{n}} (\mathrm{R_n} - \mathrm{R_{n-1}})\mathrm{P_n}$$

其中正例代表发生事故的测试视频，负例代表未发生事故，正常行驶的视频。TP 表示真正例、FP 表示假正例、FN 表示假负例$\mathrm{P_n}$和 $\mathrm{R_n}$分别是第 n 个阈值的精确率与召回率，AP 表示在不同阈值下精确率与召回率的加权平均值。

$$\mathrm{AUC} = \frac{\sum_{\mathrm{i}\in\text{正样本}} \mathrm{rank_i} - \mathrm{M(M+1)/2}}{\mathrm{M\times N}},$$

其中 M 为正样本数，N 为负样本数，$\sum_{\mathrm{i}\in\text{正样本}} \mathrm{rank_i}$为正样本的序号之和。

$$\mathrm{TTA\_0.5} = \max\{t_{ai} - t_a | p_t > 0.5, 0 \le t_a \le t_{ai}\}$$

其中t_a为事故风险值大于设定阈值 0.5 的时刻。

$$\mathrm{STTA\_0.5} = \max\{t_{ai} - t_{a'} | \forall t \le t_{a'} \le t_{ai}, p_t > 0.5\}$$

其中$t_{a'}$为距离t_{ai}时刻的时间窗口内，对应的事故风险值均大于阈值 0.5 的时刻，STTA_0.5用来评估模型对事故风险感知的稳定输出能力。

选手在本赛题中的最终得分为所有测评事故类型中 AP、AUC 和TTA_0.5的加权平均值：

$$\mathrm{score} = w_{AP}\mathrm{AP} + w_{AUC}\mathrm{AUC} + w_{TTA\_0.5}\mathrm{TTA\_0.5} + w_{STTA\_0.5}\mathrm{STTA\_0.5}$$

依据最终得分从高到低，对全部参赛选手排序。

（4）选手须知

- 所提交结果应与所提交代码运行结果相符，严禁造假，违者将被取消参与排名资格；
- 应严格按照格式要求提交结果，格式不符者无法参与结果评测；
- 请在截止日期前提交您的结果，过期提交将被视为无效！

（5）参考文献

[1] J. Fang, L. Li, J. Zhou, J. Xiao, H. Yu, C. Lv, J. Xue, and T. Chua, “Abductive ego-view accident video understanding for safe driving perception,” in CVPR, 2024, pp. 22 030–22040.

[2] J. Fang, L-L. Li, Z. Zheng, H. Yu, J. Xue, Z. Li, T-S. Chua, "EQ-TAA: Equivariant Traffic Accident Anticipation via Diffusion-Based Accident Video Synthesis," IEEE Transactions on Multimedia, early access, 2025.

（6）联系方式

李磊磊：670160532lileilei@gmail.com

赛题五：认知启发的风险驾驶场景关键目标定位

本赛题要求参赛者使用风险场景驾驶视觉注意数据集，如图 5-1 所示：给定一段风险驾驶场景视频片段，输出与人类驾驶员注视点(gaze fixation)相关的边界框，这些目标被认为是关键目标。特别鼓励参赛者以赛题给定数据集作为基准，并结合多模态大模型（MLLM）进行网络建模。评测将在测试集上进行。即要求参赛者定位出当前正在发生事故视频帧的关键目标，并输出关键目标的位置(box)。如图 5-1 所示，涉事的行人为该场景下的关键目标。

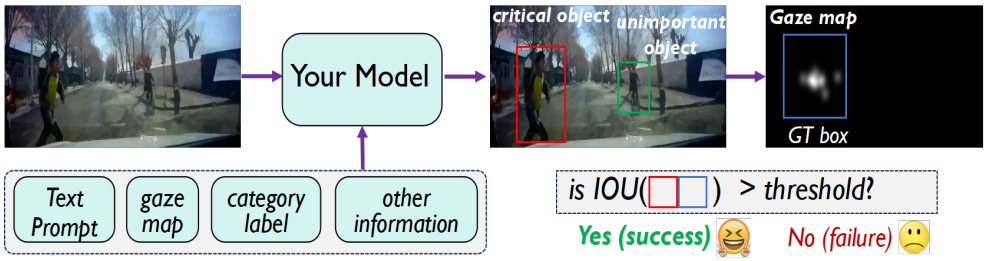


图 5-1 赛题一示意图

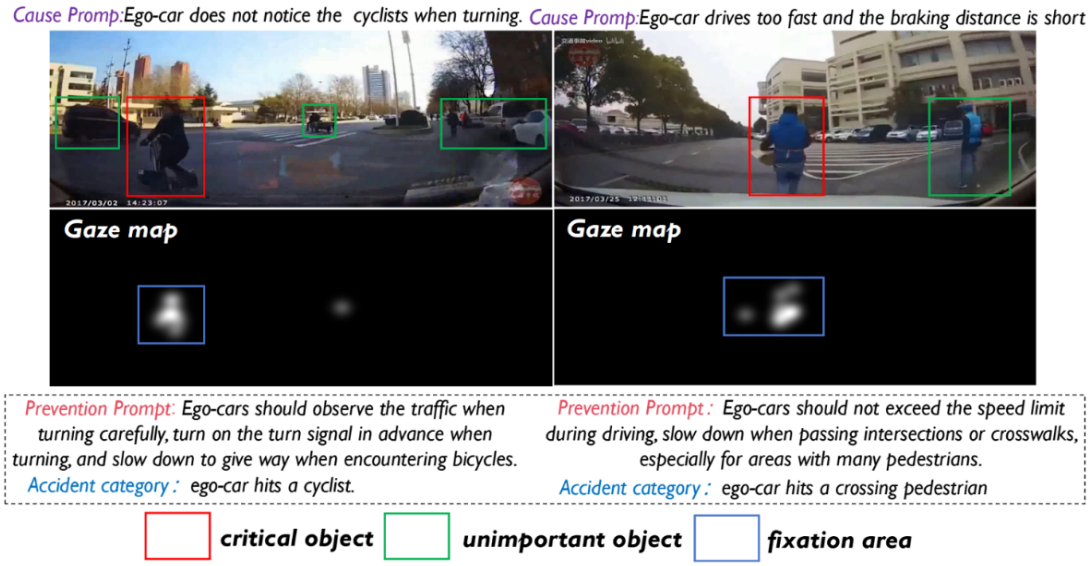


图 5-2 部分标注示例

本赛题使用 DADA-2000 数据集，数据集为 2000 个视频片段，共收集超过 658,476 帧的人类驾驶员注视点标注，如上图所示，事故场景包含大量干扰性（非关键）物体，而驾驶员注视点可聚焦于关键目标。每个场景中均包含驾驶

员的注视点，每个注视图是通过平均累积多个被试者的注视点，并使用(50×50)核大小的高斯滤波器得到的。另外，这些视频分辨率为 1584×660，涵盖 54 种事故场景。包含多种道路线型（高速公路、城市、乡村和隧道）、天气（晴天、雨天和雪天）和光照条件（白天和夜晚）。每个视频都在时间上和手动标注的视频事故原因、预防方案、事故类别和物体框的文本描述对齐。

以下是本赛题任务输入输出的详细介绍。

任务输入：本赛题以不同事故类型视频序列为单位，在训练阶段，①：每个视频下提供包含事故起始时刻到结束时刻的完整视频帧序列（分辨率:1560×660)，该视频帧序列对应的驾驶证注视图（分辨率:1560×660)和原始驾驶员注视点数据；②：该视频的目标框，以 txt 文件保存。请注意，目标框的标注是使用在 COCO 数据集上预训练的 YOLOX 检测器，对原始的 DADA 视频进行初始目标检测，以生成目标边界框标注集。每个.txt 文件包含最多 7 类道路参与者的边界框标注（即汽车、红绿灯、行人、卡车、公交车、骑行者和摩托车）；③：该视频的基本标注 Excel 文件，包含事故标签（发生事故 1/未发生事故 0）、事故类型、事故起始时刻（t_{ai}，其中未发生事故的视频t_{ai}标注为-1）、文本信息等。

任务输出：在测试阶段，要求参赛者输出测试集所有事故视频序列的关键目标定位结果。将每一个测试视频序列的目标定位结果存储为字典，并以如下方式进行存储：“视频序号”: {“bbox”: [x_min,y_min,x_max,y_max] }，即字典的键是原始的视频序号，值是预测的关键目标位置，注意最终提交的文件必须为每一帧包含且仅包含一个关键目标边界框。不允许每帧有多个目标边界框。最后将所有测试结果存储到一个.json 文件里进行逐一读取。

（1）参赛指南

我们使用视频级平均交并比(Video-mIoU)指标作为评分标准，评分细则请见下方。选手应以 json 文件格式上传结果，细则请见“结果提交”标签页，请选手严格按照格式提交结果，否则无法参加测评！

（2）评分标准

评估模型对关键目标定位的指标具体会在四个 IoU（Intersection over Union）阈值下评估性能，分别为 0.0、0.1、0.2 和 0.5，表示预测的边界框与由人类驾驶员注视点（gaze）标注的关键目标真实框之间的重叠程度分别超过 0%、10%、20%和 50%。由于具有挑战性，最终成绩将取这四个阈值下性能的等权平均值作为最终评分结果。评估仅在视频级别进行，旨在强调我们对动态演化中的安全关键驾驶场景理解的关注。测试将在 Drive-Gaze 测试数据集上进行测试。

$$\mathrm{IoU}（\mathrm{P},\mathrm{G}） = \frac{\mathrm{Area}(\mathrm{P} \cap \mathrm{G})}{\mathrm{Area}(\mathrm{P} \cup \mathrm{G})}$$

其中P代表预测出的关键目标框，G代表测试集真实标注的关键目标框，选手在本赛题中的最终得分为所有测评事故类型中IoU在不同阈值下的加权平均值：

$$\mathrm{Video}-\mathrm{mIoU} = [\mathrm{w}(\mathrm{IoU}>0)+ \mathrm{w}(\mathrm{IoU}>0.1)+\mathrm{w}(\mathrm{IoU}>0.2)+ \mathrm{w}(\mathrm{IoU}>0.5)]$$

w取 0.25，最终依据最终 mIoU 得分从高到低，对全部参赛选手排序。

（3）选手须知

所提交结果应与所提交代码运行结果相符，严禁造假，违者将被取消参与排名资格；应严格按照格式要求提交结果，格式不符者无法参与结果评测；请在截止日期前提交您的结果，过期提交将被视为无效！

（5）参考文献

[1] J. Fang, D. Yan, J. Qiao, J. Xue, and H. Yu, “DADA: driver attention prediction in driving accident scenarios,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 4959–4971, 2022.

[2] L.-L. Li, J. Fang, J. Xiao, S. Pang, H. Yu, C. Lv, J. Xue, and T.-S. Chua, “Causal-entity reflected egocentric traffic accident video synthesis,” in ICCV, 2025.

（6）联系方式

李磊磊：670160532lileilei@gmail.com

三、数据集

赛题一：预训练 LLM4PG 赋能的智能车射频地图构建

赛题一中数据的文件夹包含两类场景条件的数据，分别为车联网场景条件和低空经济场景条件，其中车联网场景条件下共包含了 RGB，Depth_map，LiDAR，Radar 和 Pathloss 五个子文件夹，每个文件夹里面包含了九种场景条件下的数据：

(a) mmWave_sunny_highVTD_widelane，

(b) mmWave_sunny_lowVTD_widelane，

(c) mmWave_sunny_highVTD_forking，

(d) sub6_sunny_highVTD_forking，

(e) sub6_sunny_mediumVTD_crossroad，

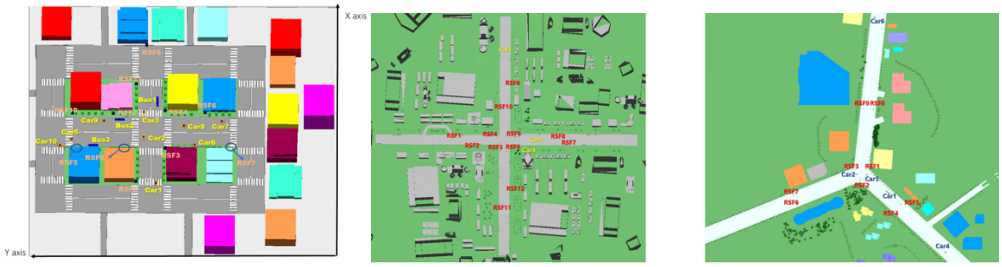
(f) mmWave_sunny_mediumVTD_crossroad，

(g) mmWave_sunny_highVTD_crossroad，

(h) mmWave_rainy_mediumVTD_crossroad，

(i) mmWave_snowy_mediumVTD_crossroad。

低空经济场景条件下共包含 RGB，Depth_map，LiDAR 和 Pathloss 四个子文件夹，每个文件夹里面包含了一种场景条件下的数据：（j）sub6_sunny_lowVTD_widelane。赛题涉及的场景如图 1-1 和图 1-2 所示。



（a）　　（b）　　（c）

图 1-1 赛题一车联网场景车辆和路端设备位置示意图(a)城镇十字路口(b)城镇超宽车道(c)郊区分岔路口

图 1-2 赛题一城镇超宽车道空中出租车场景车辆、空中出租车和路端设备位置示意图

赛题二：无线基座模型 WiFo 赋能的多模态感知辅助智能车射频链路预测

本赛题共包含两大场景：车联网场景与空中出租车场景。其中车联网场景包含高车流量分岔路口场景；空中出租车场景主要为低车流量超宽车道到空中出租车的通信场景。

场景一：车联网场景

赛题二车联网场景下，每个通信链路的每帧包括五种模态的感知数据（包括 RGB 图、深度图、LiDAR 点云、毫米波雷达点云、GPS 信息）和通信链路的 CSI。

值得注意的是，为了考验选手开发的模型对于实际车联网应用中可能出现的感知数据异常情况（比如意外遮挡、数据缺失）的适应能力，并使得感知数据更贴近于真实情况，车联网场景下的训练集和测试集做了如下处理：

数据缺失：所有感知模态随机缺失约 10%帧数的数据；

意外遮挡：RGB 图和深度图约 10%帧数的数据被随机遮盖部分区域；LiDAR 和毫米波雷达点云约 10%帧数的数据随机缺失了部分角度范围的点云；

GPS 定位误差：为方便选手处理，车辆的 GPS 信息以二维坐标信息的格式给出，并添加了方差为 5 的高斯噪声，以模拟实际应用场景中 GPS 的定位误差。

场景二：空中出租车场景

赛题二空中出租车-路场景中，每个通信链路的每帧包括三种模态的感知数据（包括 RGB 图、深度图、LiDAR 点云）和通信链路的 CSI。值得注意的是，该场景同样包含感知数据异常情况（比如意外遮挡、数据缺失），且处理方案与车联网场景相同。

赛题三：无线基座模型 WiFo 赋能的多模态信息智能车定位

文件格式说明

- RGB 图像文件（png 格式）
- 轨迹真值文件（txt 格式）

该文件的内容为多帧的车辆位姿信息，每一行分别代表一帧的车辆位姿信息。每一行的第 1 至 3 列分别为车辆在该帧的 x、y 和 z 坐标，单位为 m；第 4 至 6 列分别为车辆在该帧的 roll、pitch 和 yaw 角度，单位为 rad。

- 发射端和接收端之间的信道状态信息文件（mat 格式）

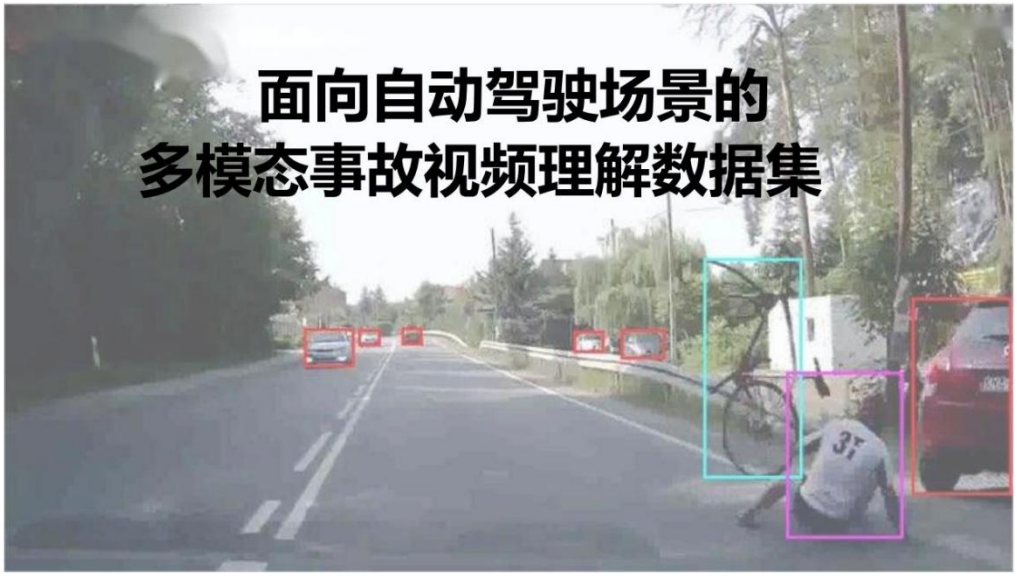
每个 mat 文件中有 1500 个 cell，对应 1500 帧的信道状态信息。每一个 cell

的内容为 32行128 列的信道复增益矩阵，每行依次对应 Tx 端 1-32 号天线， 每列依次对应 Rx 端 1-128 号天线。

- 赛题三的数据集共分为三个场景文件夹，分别是分岔路口高车流量场景、超宽车道低车流量场景和包含空中出租车的超宽车道低车流量场景，每一场景文件夹下分别存有各个车辆以及路端的通信与多模态感知数据，第三个场景（包含空中出租车的超宽车道低车流量场景）还包括了空中出租车的通信与多模态感知数据（在子文件夹 AT 下）。每个场景都可能设置一定的传感器数据缺失，传感器数据残缺具体是指传感器数据会在某些时刻缺失一定的帧数，需要选手在设计算法时需要考虑应对此类问题的鲁棒性。“Documents”文件夹下存有三个场景世界坐标系下的车辆的轨迹以及路端设备的位置。
- 在评测数据集中，将不包含车辆的全部真值轨迹，仅提供车辆世界坐标系下的初始坐标以及路端设备的坐标。评测数据集同时将涵盖数据残缺的特殊情形，用于测试选手定位算法的鲁棒性。

赛题四：自动驾驶场景的交通事故预测

- 数据集展示视频



- 数据集的详细说明：数据集的场景说明、文件夹目录说明、训练集测试集说明等

本届自动驾驶场景交通事故风险预测挑战赛将发布共涵盖 58 种事故类型的多模态事故理解数据集，数据集包含了五种道路场景与四种天气条件供选手使用。

数据集来自公开的驾驶环境数据集和各种视频流网站，原始视频序列请选手在数据集主页进行下载。应下载的文件共有：原始视频序列压缩包、完整的标签 excel 文档（包含所有的事故窗口和文本信息标注）、训练和测试 txt 文件以及每一个视频序列的目标检测数据。

原始视频序列压缩包下载解压后的每一个单位视频文件命名格式应为“事故类型序号/视频序号/images/.jpg”。同时下载解压 train.txt 和 test.txt 标签文件。标签文件中每一行存储的数据格式为“事故类型序号、视频序号、事故标签（1/0）、起始视频帧序号、结束视频帧序号、事故起始帧t_{ai}和视频的事实文本描述”，其中结束视频帧序号和起始视频帧序号的差值始终为 150 帧。请注意文本信息和交通参与者的位置信息不作强制使用，选手可灵活选择。特别注意，如果在训练中使用了文本信息，则必须将 test.txt 中的测试序列视频文本置空为“a video frame

of {}”以模拟实际环境中文本信息的缺失。

选手也可查看完整的 excel 标签文档，根据训练和测试 txt 文件中提供的视频名查找到相应的文本信息进行查看，对数据集进行更加全面的了解。

行车环境多模态事故视频理解数据集将于 10 月 30 日开放下载，敬请期待！

赛题五：认知启发的风险驾驶场景关键目标定位

- 数据集的详细说明

1）训练数据结构

/DADA_Train/：DADA 训练数据集包含以下内容：视觉输入（RGB 图像、注视图和注视点）边界框注释（涵盖一般物体，而不限于关键物体）等。

```
DADA_Train/
└── 1/
    └── 001/
        ├── images/     # RGB frames extracted from the video
        ├── maps/       # Driver gaze maps corresponding to attention
        └── fixation/   # Fixation points per gaze map

└── ...
```

2）目标框标注

DADA Bounding box/每个.txt 文件包含 7 类道路参与者的边界框注释（即汽车、红绿灯、行人、卡车、公交车、自行车和摩托车）。

```
DADA_Dtection/
└── 2/
        └── 002/
                ├── 0001.txt
                ├── 0002.txt
└── ...
```

3）注释元数据格式

训练集中每个视频都附有一个结构化的元数据文件（Excel 格式），用于描述驾驶场景及与事故相关的信息。以下是对该 Excel 文件中元数据字段的 JSON 风格解释：

```
{ "video_hashcode": { "video_name": "1_1","id": "1", "type": "1", "weather": "1","light":
"1","scenes": "4", "linear": "1", "accident occurred": "1", "abnormal_start_frame":
"30","abnormal_end_frame": "115","accident_frame": "63", "total_frames": "440","t_ai":
```

"30","t_co": "63","t_ae": "115", "texts": "a pedestrian crosses the road","causes": "Pedestrian does not notice the coming vehicles when crossing the street","measures": "When passing the zebra crossing, drivers must slow down. When pedestrians or non-motor vehicles cross the zebra crossing, they should stop and give way to other normal running vehicles; When crossing the road, pedestrians must follow the zebra crossing, carefully observe the traffic, and do not cross the road in a hurry."}

| Field Name| Description

| `video_hashcode`| Unique identifier for each video entry

| `video_name`| Video name, composed of accident type and serial ID

| `id`| Video index number

| `type`| Type of accident (refer to accident type list in dataset)

| `weather`| Weather condition: 1 = Sunny, 2 = Rainy, 3 = Snowy, 4 = Foggy

| `light`| Lighting condition: 1 = Day, 2 = Night

| `scenes`| Driving scene: 1 = Highway, 2 = Tunnel, 3 = Mountain, 4 = Urban, 5 = Rural

| `linear`| Road shape: 1 = Arterials, 2 = Curve, 3 = Intersection, 4 = T-junction, 5 = Ramp|

| `accident occurred`| Whether an accident occurred in the video (1 = Yes, 0 = No)

| `abnormal_start_frame` | Start frame index of abnormal event

| `abnormal_end_frame`| End frame index of abnormal event

| `accident_frame`| Frame at which the collision occurs

| `t_ai` | Start of the accident window (same as `abnormal_start_frame`)

| `t_co`| Collision frame (same as `accident_frame`)

| `t_ae` | End of the accident window (same as `abnormal_end_frame`)

| `texts`| Natural language description of the scenario

| `causes`| Cause of the accident

| `measures`| Recommended countermeasures to avoid similar accident

训练数据集将于 10 月 30 日开放下载，敬请期待！

四、结果提交

赛题一：预训练 LLM4PG 赋能的智能车射频地图构建

请选手提交以下内容：

1）提交一个包含详尽的技术路线、预测结果及其说明的技术报告（PDF 文件），文件命名为：队伍名_pathloss_技术报告.pdf

2）提交我们提供的测试数据集中射频地图的生成结果（TXT 文件），以场景条件_时刻_设备编号_pathloss.txt 文件命名。

其中，场景条件的命名规则如下：

(a)超宽车道，晴天毫米波高车流量密度

（widelane_sunny_mmWave_highVTD）

(b)超宽车道，晴天毫米波低车流量密度

（widelane_sunny_mmWave_lowVTD）

(c)分岔路口，晴天毫米波高车流量密度（forking_sunny_mmWave_highVTD）

(d)分岔路口，晴天 sub-6GHz 高车流量密度（forking_sunny_sub6_highVTD）

(e)十字路口，晴天 sub-6GH 中车流量密度

（crossroad_sunny_sub6_mediumVTD）

(f)十字路口，晴天毫米波中车流量密度

（crossroad_sunny_mmWave_mediumVTD）

(g)十字路口，晴天毫米波高车流量密度

（crossroad_sunny_mmWave_highVTD）

(h)十字路口，雨天毫米波中车流量密度

（crossroad_rainy_mmWave_mediumVTD）

(i)十字路口，雪天毫米波中车流量密度

（crossroad_snowy_mmWave_mediumVTD）

(j)超宽车道，晴天 sub-6GHz 低车流量密度（widelane_sunny_sub6_lowVTD）

举例而言，场景为城镇十字路口，场景条件为晴天毫米波高车流量密度（sunny_mmWave_highVTD），在第 20 个时刻针对 car1 的射频地图生成值，则其路径预测结果文件命名为：

crossroad_sunny_mmWave_highVTD_snapshot20_car1_pathloss.txt

生成的射频地图的存储格式要求如下：针对 71*73 的图像，需要先按列从左到右、再按行从上到下的顺序，将射频地图中全部点的数值存入 txt 文件，排成一列，其中第一行为设备名称，举例如下：

name

$value_1$

$value_2$

……

$value_n$

其中，name 为设备名称，$value_n$ 为在图像中第 n 个点的射频地图数值。

3）提供已预训练好的预测模型、代码（包含注释）和 ReadMe 文件（写清运行代码的需求和步骤）

选手最终提交一个压缩包命名为队伍名_pathloss.zip，其中包含 3 个文件夹，分别是技术报告，预测结果，源代码。其中预测结果文件夹内包含 10 个子文件夹，分别是

(a)超宽车道，晴天毫米波高车流量密度

(b)超宽车道，晴天毫米波低车流量密度

(c)分岔路口，晴天毫米波高车流量密度

(d)分岔路口，晴天 sub-6GHz 高车流量密度

(e)十字路口，晴天 sub-6GHz 中车流量密度

(f)十字路口，晴天毫米波中车流量密度

(g)十字路口，晴天毫米波高车流量密度

(h)十字路口，雨天毫米波中车流量密度

(i)十字路口，雪天毫米波中车流量密度

(j)超宽车道，晴天 sub-6GHz 低车流量密度

每个子文件夹内包括所有设备所有时刻的射频地图构建结果的 txt 文件。

请于测试结果提交截止时间之前提交您的结果，逾期提交将无法参加测评！

赛题二：无线基座模型 WiFo 赋能的多模态感知辅助智能车射频链路预测

请选手们提交以下内容：

1） 提交一个包含详尽的技术路线、预测结果及其说明的技术报告（pdf 文件），文件命名为“队伍名_prediction_技术报告.pdf”

2） 提交我们提供的测试数据集的 CSI 预测结果（mat 文件），该结果文件应以“场景_通信链路_CSI.mat”命名，通信链路应以“BS+路端设备序号_Car+车端设备序号”命名。预测结果文件包含大小为“$B \times N \times K$”的 mat 矩阵，其中 B 为该链路的样本个数，N 为天线数目，K=32 为待预测子载波数目。

3） 提供已预训练好的预测模型、代码（包含注释）和 ReadMe 文件（写清运行代码的需求和步骤），将以上内容放在名称为“source”的文件夹中。

选手最终提交一个压缩包命名为队伍名_CSI.zip，其中包含三个文件/文件夹，分别是技术报告文件“队伍名_CSI_技术报告.pdf”，预测结果文件夹“test_CSI”，源代码文件夹“source”。

请于测试结果提交截止时间之前提交您的结果，逾期提交将无法参加测评！

赛题三：无线基座模型 WiFo 赋能的多模态信息智能车定位

请选手们提交以下内容：

1) 提交一个包含详尽的算法方案说明的技术报告（.pdf 文件），文件命名为队伍名_智能车定位.pdf

2) 提交一个测试数据集的智能车定位结果，以测试集的场景为一个单位（一个子文件夹）；每个子文件夹内，每个目标车辆的位置估计结果为一个 txt 文件，如第 X 个场景中编号为 Y 的目标车辆的结果文件命名为 Scene_X/car_Y.txt。

3) 提交对应算法源代码和对应的 ReadMe 文件（要求写清代码文件间的逻辑关系，如何运行）

选手最终提交一个压缩包命名为队伍名_localization.zip，其中包含三个文件夹，分别是技术报告，智能车定位结果，源代码。

请于测试结果提交截止时间之前提交您的结果，逾期提交将无法参加测评！

赛题四：自动驾驶场景交通事故风险预测

请选手们提交以下内容：

- 提交一个包含详尽的技术路线、预测结果及其说明的技术报告（.pdf 文件），

文件命名为队伍名_事故风险预测技术报告.pdf；

- 提交一个测评数据集的风险值预测结果，以测评数据集的每个事故视频为一个单位，以.json 文件的形式，分别存储每一个测试视频 150 帧序列的事故风险值，json 文件命名为队伍名_accident risk.json，每一个单位测试视频的存储示例如: #“视频序号”: {“risk”: $risk_1$, $risk_2$, ..., $risk_T$ (规定 T 为 150，即根据观测输出 150 个风险值)], “label”: 1 或者 0，“t_{ai}”：t_{ai}或-1}；
- 提交事故风险预测算法代码（包含注释）和对应的 ReadMe 文件（写清运行代码的需求和步骤）

选手最终提交一个压缩包命名为队伍名_traffic accident.zip，其中包含三个文件夹，分别是技术报告，风险值预测结果，源代码。其中风险值预测结果为存储所有测试视频序列的 json 文件。

请于测试结果提交截止时间之前提交您的结果，逾期提交将无法参加测评！

赛题五：认知启发的风险驾驶场景关键目标定位

请选手们提交以下内容：

- 提交一个包含详尽的技术路线、预测结果及其说明的技术报告（.pdf 文件），文件命名为队伍名_自动驾驶关键目标定位技术报告.pdf

2）测试阶段提交一个 .zip 压缩文件，其中包含预测的 JSON 文件（命名为 prediction.json）。prediction.json 的结构必须与真实边界框（grounding truth boxes）的格式和键（keys）相匹配。注意准备提交文件时，请严格遵循以下示例格式：

```
. . .
{
  "43_196_0089": [
    {
```

"bbox": [x1, y1, x2, y2] // 格式：左上角和右下角坐标 (x_min, y_min, x_max, y_max)

}

]

}

. . .

- 提交关键目标定位算法代码（包含注释）和对应的 ReadMe 文件（写清运行代码的需求和步骤）

请于测试结果提交截止时间之前提交您的结果，逾期提交将无法参加测评！