以下是一些常见的多模态数据类型，它们在自动驾驶系统中被广泛使用：

1. **图像数据（Visual Data）**：来自车载摄像头的图像数据，可以是单目、双目或多摄像头系统，提供丰富的色彩和纹理信息，用于物体检测、语义分割和场景理解。
2. **激光雷达数据（LiDAR Data）**：激光雷达通过发射激光脉冲并测量反射回来的时间来生成周围环境的精确三维点云，用于高精度的物体检测、自由空间检测和地图构建。
3. **雷达数据（Radar Data）**：雷达通过发射无线电波并接收反射波来检测物体的距离、速度和角度，常用于远距离物体检测和车辆跟踪。
4. **GPS和惯性测量单元（IMU）数据**：提供车辆的位置、速度和方向信息，用于定位和导航。
5. **事件相机数据（Event Camera Data）**：一种新型相机，能够以高时间分辨率记录像素级的亮度变化，适用于动态环境的感知。
6. **音频数据（Audio Data）**：虽然在自动驾驶中使用较少，但音频数据可以提供有关车辆周围环境的额外信息，如紧急车辆的警报声。
7. **车辆通信系统数据（V2X Data）**：包括车辆对车辆（V2V）、车辆对基础设施（V2I）等通信数据，可以提供交通流量、信号灯状态等信息。
8. **高精地图数据（HD Map Data）**：包含道路网络、车道标记、交通标志等详细的地理信息，用于车辆的定位和路径规划。

#### Multi-modal Sensor Fusion for Auto Driving Perception A Survey(2022 arxiv) 综述

多模态自驾的综述：这篇综述论文提供了自动驾驶中多模态传感器**融合技术**的全面概述，包括数据级、特征级和决策级融合方法。介绍了感知模块的重要性以及单模态数据感知的局限性，强调了LIDAR和摄像头数据融合的优势。介绍了一些广泛使用的公开数据集和基准，如KITTI、Waymo和nuScenes。

多模态使用方式：这篇综述论文提供了自动驾驶感知任务中基于多模态方法的文献回顾。它详细分析了超过50篇论文，这些论文利用包括激光雷达（LiDAR）和摄像头在内的感知传感器，尝试解决物体检测和语义分割任务。

创新点：提出了一种新的多模态融合方法分类法，将融合模型分为两大类（强融合和弱融合），四个小类（早融合、深融合、晚融合、非对称融合），这种分类基于融合阶段的观点，更为合理。对当前融合方法进行了深入分析，聚焦于存在的问题，并对未来可能的研究方向进行了讨论。

#### **Multi-Sensor Fusion in Automated Driving: A Survey（2020 IEEE Access）综述**

这篇论文提供了自动化驾驶中**多传感器融合**的综述，涵盖了50多篇利用LiDAR和摄像头的论文，并提出了新的融合方法分类法

多模态使用方式：这篇综述论文主要讨论了近年来自动驾驶中多传感器融合的不同策略。分析了包括雷达、激光雷达、摄像头、超声波、GPS、IMU和V2X等传统传感器的性能和多传感器融合的必要性。

创新点：将融合策略分为四类：基于可辨识单元的融合（FBDU）、基于互补特征的融合（FBCF）、基于多源决策的融合（FBMDM）和基于目标属性的融合（FSBTA）。

强调了在自动驾驶系统中，如何通过融合多传感器数据来提高系统的可靠性和安全性。

#### **PointPainting: Sequential Fusion for 3D Object Detection（2020 CVPR）**

这篇论文介绍了一种新的3D物体检测方法，通过序列融合**LiDAR和相机数据**来提高检测性能

**多模态使用方式**：这篇论文提出了一种名为PointPainting的序列融合方法，用于3D物体检测。该方法通过将激光雷达点投影到仅图像的语义分割网络的输出中，并将类分数附加到每个点上。

**创新点**：PointPainting作为一种通用的融合方法，可以与任何激光雷达检测网络结合使用，提高了检测性能。该方法在KITTI和nuScenes数据集上进行了实验，显示出对三种不同的最先进方法（Point-RCNN、VoxelNet和PointPillars）有显著改进。

#### M2DA: Multi-Modal Fusion Transformer Incorporating Driver Attention for Autonomous Driving（2024 arxiv） (上面好几篇文章这个实验室的)

用于自动驾驶的融合驾驶员注意力（M2DA）的多模式融合变压器和基于激光雷达视觉注意力的融合（LVAFusion）模块。通过结合驾驶员的注意力赋予自动驾驶汽车类似人类的场景理解能力。

#### InfraParis: A multi-modal and multi-task autonomous driving dataset（2024 CVPR）

介绍了一个数据集，RGB、深度、红外数据集，可用来完成多任务——semantic segmentation, object detection, and depth estimation

多模态使用: InfraParis数据集提供了多模态数据，包括RGB图像、深度图和红外图像，这些数据被用于支持自动驾驶中的多个任务，如语义分割、目标检测和深度估计。

创新点: 数据集的独特之处在于它结合了多种数据模态和任务，提供了丰富的注释，并且涵盖了巴黎及其周边地区的多样化场景，包括城市环境、农村景观和高速公路。

#### DV-3DLANE END-TO-END MULTI-MODAL 3D LANE DETECTION WITH DUAL-VIEW REPRESENTATION (2024 ICLR)

透视图和鸟瞰图多模态融合进行车道线检测

多模态使用: 这篇论文提出了一个名为DV-3DLane的端到端多模态3D车道检测框架，它结合了图像和激光雷达（LiDAR）点云数据，通过双视图（透视图和鸟瞰图）来提高车道检测的准确性。

创新点: 引入了双向特征融合策略、统一查询生成方法和3D双视图可变形注意力机制，这些设计有效地结合了来自不同视图的特征，实现了精确的3D车道检测

#### A Survey for Foundation Models in Autonomous Driving（2024 arxiv） 综述

多模态使用: 这篇综述论文探讨了基础模型在自动驾驶中的应用，包括大型语言模型、视觉基础模型和多模态基础模型。这些模型通过整合不同模态的输入数据（如文本、图像、视频等），提高了自动驾驶系统在规划、模拟、3D对象检测和跟踪等方面的能力。

创新点: 提出了一个基于模态和功能对基础模型进行分类的系统化分类法，并探讨了这些模型在自动驾驶中的应用，如用于提高决策制定的透明度和可解释性，以及在模拟和测试中的创新应用。

#### Deep Multi-Scale Multi-Modal Convolutional Dictionary Learning Network (2024 TPAMI)

该工作瞄准跨模态图像间复杂的信息耦合特性，提出了一个基于层级式多尺度可调节字典学习模型的全局主动式可解释神经网络DeepM2CDL。该可解释神经网络为多模态图像复原与融合任务提供了通用的多尺度字典学习框架，在具有良好的网络可解释性的同时，在多模态图像超分辨、去噪、多曝光融合以及多焦点融合等多个任务中取得了卓越的性能。

多模态使用: 这篇论文提出了一个深度多尺度多模态卷积字典学习网络（DeepM2CDL），用于多模态图像处理，包括图像恢复和图像融合任务。

创新点: 该网络通过多层策略执行多尺度多模态卷积字典学习，将不同图像模态以粗到细的方式关联起来，提高了网络的解释性，并且通过展开策略将字典学习整合到深度学习中。

#### Vehicle Dynamics Modeling with a Physics-Informed Neural Network for Autonomous Racing(2023arxiv 引用1)

论文主要介绍了一种名为Deep Dynamics的物理信息神经网络（Physics-Informed Neural Network, PINN），用于自动驾驶赛车的**车辆动力学建模**。

多模态使用: 这篇论文介绍了Deep Dynamics，这是一个物理信息神经网络（PINN），用于模拟自动驾驶赛车的动态。它结合了物理系数估计和动态方程，以在高速下准确预测车辆状态。

创新点: 提出了一个独特的“物理守卫”层，确保内部系数估计保持在其物理意义范围内，以及一个能够准确预测高速赛车动态的PINN模型

#### UniSim-paper（2023 CVPR 引用量100）

UniSim 的多模态能力（相机图像和激光雷达点云）使其能够为自动驾驶系统提供一个丰富的、逼真的虚拟环境，用以测试和评估系统在不同交通场景和配置下的性能。这种多模态传感模拟对于开发和验证自动驾驶车辆的感知和决策算法至关重要，因为它可以提供比单一传感器更全面的数据视图。

#### LimSim: A Long-term Interactive Multi-scenario Traffic Simulator（2023 ITSC）

研究拐角的车辆驾驶预测，和多模态没什么关系，就是看到是pjlab的