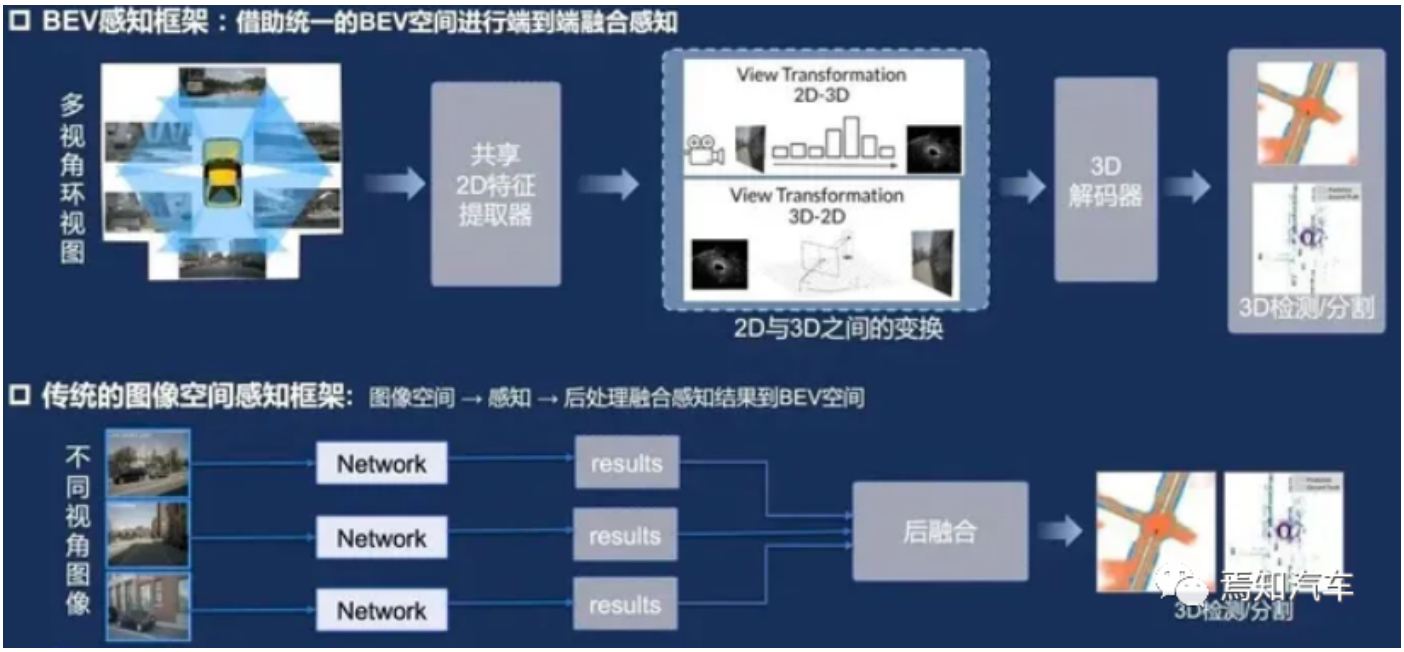
利用Transformer BEV解决自动驾驶Corner Case的技术原理

自动驾驶系统在实际应用中需要面对各种复杂的场景，尤其是Corner Case（极端情况）对自动驾驶的感知和决策能力提出了更高的要求。Corner Case指的是在实际驾驶中可能出现的极端或罕见情况，如交通事故、恶劣天气条件或复杂的道路状况。BEV技术通过提供全局视角来增强自动驾驶系统的感知能力，从而有望在处理这些极端情况时提供更好的支持。本文将探讨BEV（Bird's Eye View，俯视视角）技术如何帮助自动驾驶系统应对Corner Case，提高系统的可靠性和安全性。



Transformer 作为你一种基于自注意力机制的深度学习模型，最早应用于自然语言处理任务。其核心思想是通过自注意力机制捕捉输入序列中的长距离依赖关系，从而提高模型在处理序列数据上的能力。

将以上两者进行有效结合也是在自动驾驶策略中比较热门的一门新兴技术。

BEV的技术优势分析

BEV是一种将三维环境信息投影到二维平面的方法，以俯视视角展示环境中的物体和地形。在自动驾驶领域，BEV 可以帮助系统更好地理解周围环境，提高感知和决策的准确性。在环境感知阶段，BEV 可以将激光雷达、雷达和相机等多模态数据融合在同一平面上。这种方法可以消除数据之间的遮挡和

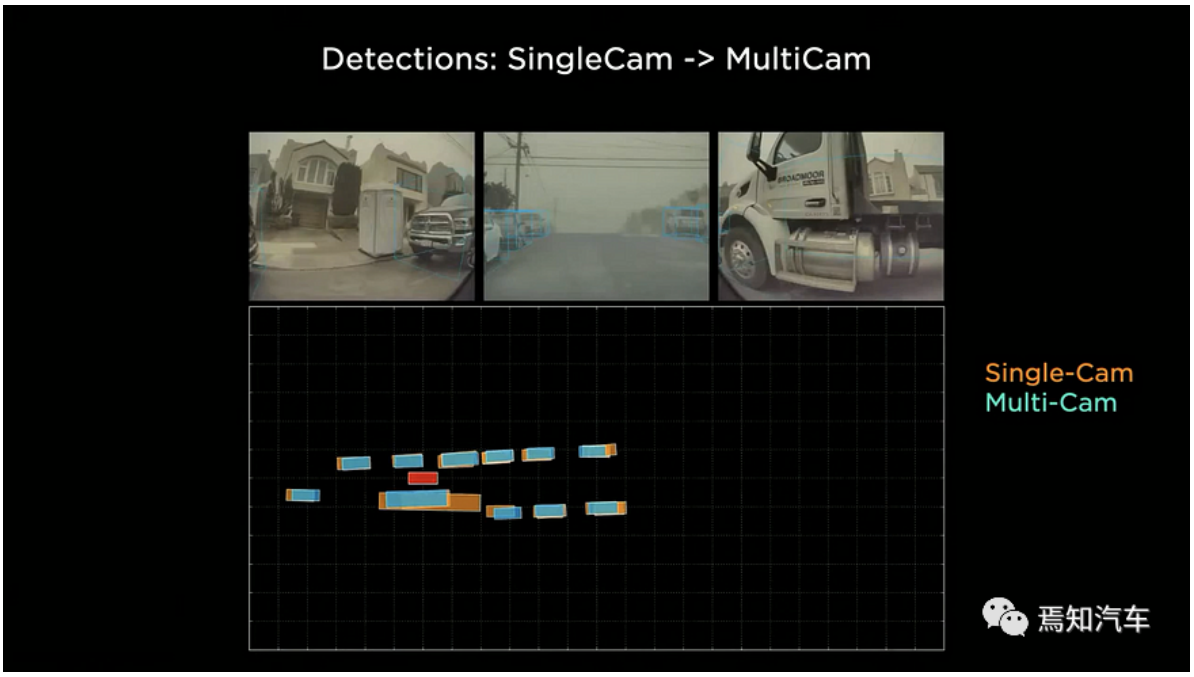
重叠问题，提高物体检测和跟踪的精度。同时，BEV 可以为后续的预测和决策阶段提供清晰的环境表示，有利于提高系统的整体性能。

1、Lidar与BEV技术的比较

首先，BEV技术能提供全局视角的环境感知，有助于提高自动驾驶系统在复杂场景下的表现。然而，激光雷达在距离和空间信息方面具有更高的精度。
其次，BEV技术通过摄像头捕捉图像，可以获取颜色和纹理信息，而激光雷达在这方面的性能较弱。
此外，BEV技术的成本相对较低，适用于大规模商业化部署。

2、BEV技术与传统单视角摄像头的比较

传统单视角摄像头是一种常用的车辆感知设备，可以捕捉车辆周围的环境信息。然而，单视角摄像头在视野和信息获取方面存在一定局限性。BEV技术整合多个摄像头的图像，提供全局视角，可以更全面地了解车辆周围的环境。



BEV技术在复杂场景和恶劣天气条件下，相对于单视角摄像头具有更好的环境感知能力，因为BEV能够融合来自不同角度的图像信息，从而提高系统对环境的感知。
BEV技术可以帮助自动驾驶系统更好地处理Corner Case，如复杂道路状况、狭窄或遮挡的道路等，而单视角摄像头在这些情况下可能表现不佳。
当然在成本和资源占用情况方面，由于BEV需要进行各个视角下的图像感知，重建和拼接，因此是比较耗费算力和存储资源的。虽然BEV技术需要部署多个摄像头，但总体成本仍低于激光雷达，且相对于单视角摄像头在性能上有明显提升。
综上所述，BEV技术在自动驾驶领域与其他感知技术相比具有一定优势。尤其是在处理Corner Case方面，BEV技术可以提供全局视角的环境感知，有助于提高自动驾驶系统在复杂场景下的表现。然而，为了充分发挥BEV技术的优势，仍需要进一步研究和开发，以提高图像处理能力、传感器融合技术以及异常行为预测等方面的性能。同时，结合其他感知技术（如激光雷达）以及深度学习和机器学习算法，可以进一步提升自动驾驶系统在各种场景下的稳定性和安全性。

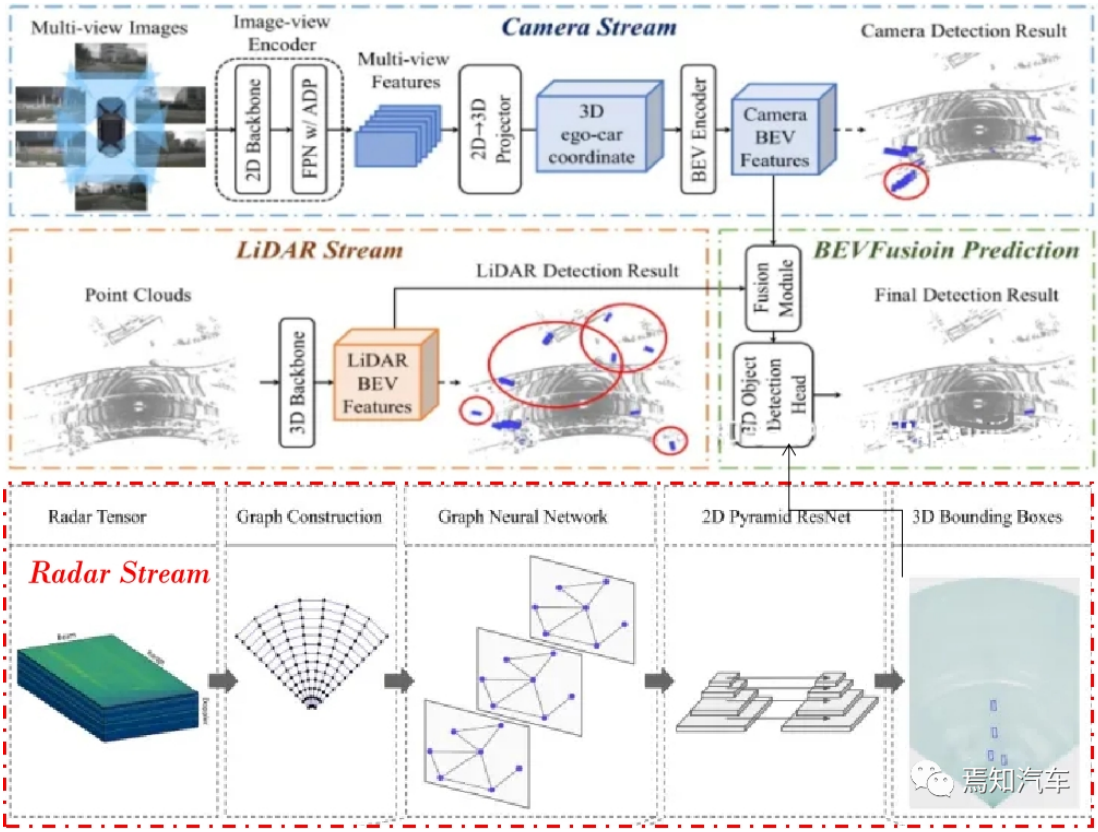
02

基于 Transformer 和 BEV 的自动驾驶系统

与此同时，Bird's Eye View (BEV) 作为一种有效的环境感知方法，在自动驾驶系统中发挥着重要作用。结合 Transformer 和 BEV 的优势，我们可以构建一个端到端的自动驾驶系统，实现高精度的感知、预测和决策。本文也将同时探讨 Transformer 和 BEV 在自动驾驶领域如何进行有效结合和应用，以提高系统性能。
具体步骤如下：

1、数据预处理：

将激光雷达、雷达和相机等多模态数据融合为 BEV 格式，并进行必要的预处理操作，如数据增强、归一化等。
首先，我们需要将激光雷达、雷达和相机等多模态数据转换为 BEV 格式。对于激光雷达点云数据，我们可以将三维点云投影到一个二维平面上，然后对该平面进行栅格化，以生成一个高度图；对于雷达数据，我们可以将距离、角度信息转换为笛卡尔坐标，然后在 BEV 平面上进行栅格化；对于相机数据，我们可以将图像数据投影到 BEV 平面上，生成一个颜色或强度图。



2、感知模块：

在自动驾驶的感知阶段，Transformer 模型可以用于提取多模态数据中的特征，如激光雷达点云、图像、雷达数据等。通过对这些数据进行端到端的训练，Transformer 能够自动学习到这些数据的内在结构和相互关系，从而有效地识别和定位环境中的障碍物。

利用 Transformer 模型对 BEV 数据进行特征提取，实现障碍物的检测和定位。将这些 BEV 格式的数据叠加在一起，形成一个多通道的 BEV 图像。设激光雷达的 BEV 高度图为 H(x, y)，雷达的 BEV 距离图为 R(x, y)，相机的 BEV 强度图为 I(x, y)，则多通道的 BEV 图像可以表示为：B(x, y) = [H(x, y), R(x, y), I(x, y)]

其中 B(x, y) 表示多通道 BEV 图像在坐标 (x, y) 处的像素值，[] 表示通道叠加。

3、预测模块：

基于感知模块的输出，使用 Transformer 模型预测其他交通参与者的未来行为和轨迹。通过学习历史轨迹数据，Transformer 能够捕捉到交通参与者的运动模式和相互影响，从而为自动驾驶系统提供更准确的预测结果。

具体的讲，我们首先使用 Transformer 对多通道 BEV 图像进行特征提取。设输入 BEV 图像为 B(x, y)，我们可以通过多层自注意力机制和位置编码来提取特征 F(x, y)：F(x, y) = Transformer(B(x, y))

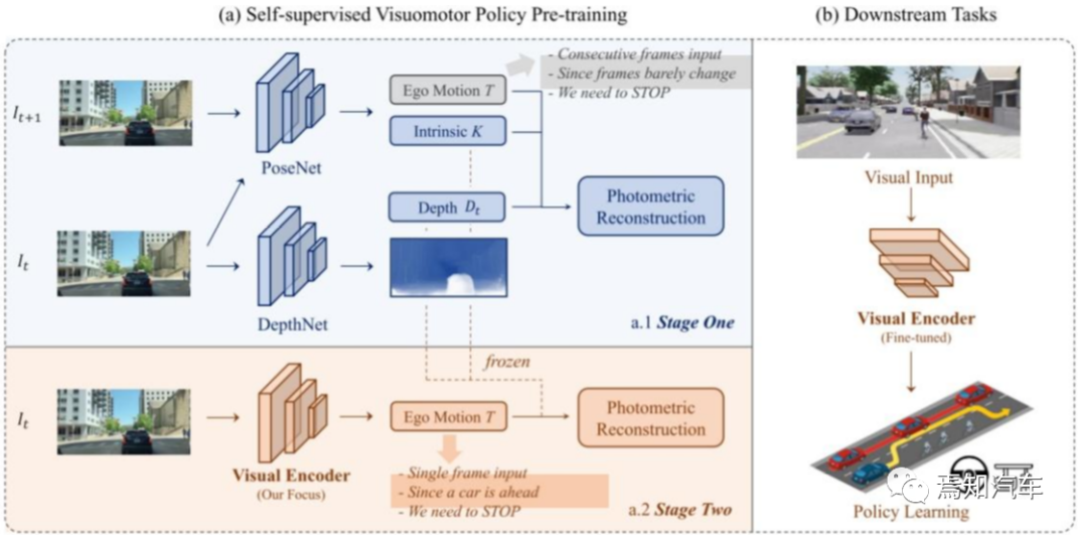
其中 F(x, y) 表示特征图，在坐标 (x, y) 处的特征值。

然后，我们利用提取到的特征 F(x, y) 预测其他交通参与者的行为和轨迹。可以采用 Transformer 的解码器来生成预测结果，如下所示：P(t) = Decoder(F(x, y), t)

其中 P(t) 表示在时间 t 处的预测结果，Decoder 表示 Transformer 解码器。通过以上步骤，我们可以实现基于 Transformer 和 BEV 的数据融合与预测。具体的 Transformer 结构和参数设置可以根据实际应用场景进行调整，以达到最佳性能。

4、决策模块：

根据预测模块的结果，结合交通规则和车辆动力学模型，采用 Transformer 模型生成合适的驾驶策略。



通过将环境信息、交通规则和车辆动力学模型整合到模型中，Transformer 能够学习到高效且安全的驾驶策略。如路径规划、速度规划等。此外，利用 Transformer 的多头自注意力机制，可以有效地平衡不同信息源之间的权重，从而在复杂环境中做出更为合理的决策。

以下是采用该方法的具体步骤：

1、数据收集与预处理：

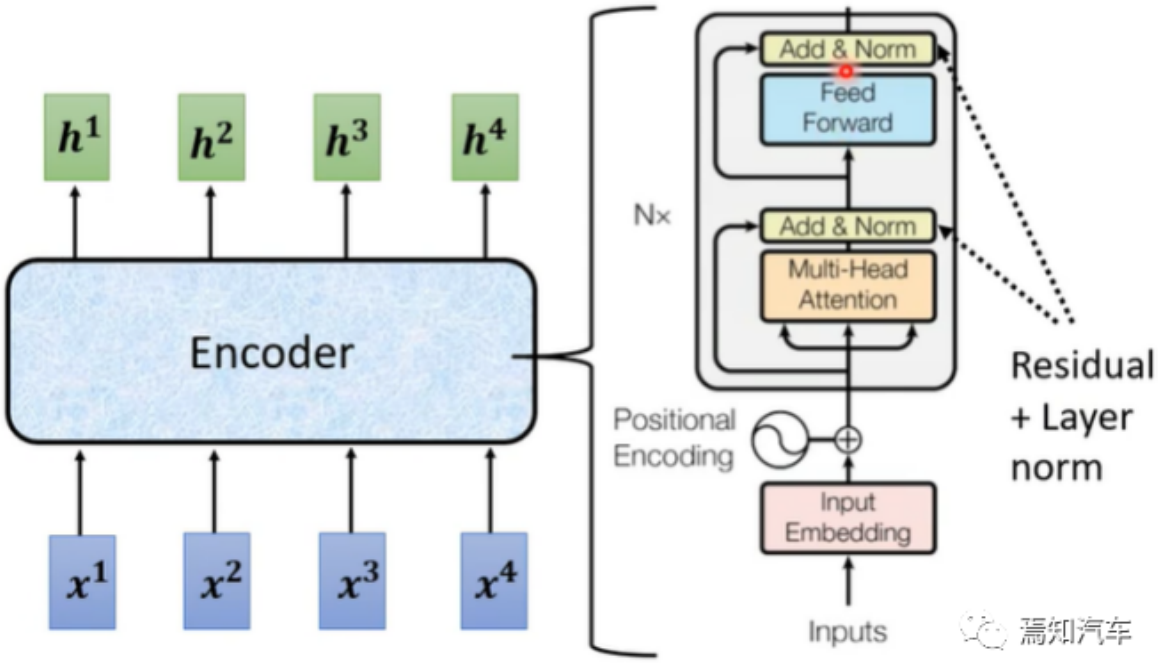
首先，需要收集大量的驾驶数据，包括车辆状态信息（如速度、加速度、方向盘角度等）、路况信息（如道路类型、交通标志、车道线等）、周围环境信息（如其他车辆、行人、自行车等）以及驾驶员采取的操作。对这些数据进行预处理，包括数据清洗、标准化和特征提取。

2、数据编码与序列化：

将收集到的数据编码成适合 Transformer 模型输入的形式。这通常包括将连续的数值数据进行离散化，并将离散化的数据转换成向量形式。同时，需要将数据序列化，以便 Transformer 模型能够处理时序信息。

2.1、Transformer 编码器

Transformer 编码器由多层相同的子层组成，每个子层包含两个部分：多头自注意力（Multi-Head Attention）和前馈神经网络（Feed-Forward Neural Network）。
多头自注意力：首先将输入序列分为 h 个不同的头，分别计算每个头的自注意力，然后将这些头的输出拼接在一起。这样可以捕捉输入序列中不同尺度的依赖关系。



多头自注意力的计算公式为：MHA(X) = Concat(head_1, head_2, ..., head_h) * W_O其中 MHA(X) 表示多头自注意力的输出，head_i 表示第 i 个头的输出，W_O 是输出权重矩阵。
前馈神经网络：接下来，将多头自注意力的输出传递给前馈神经网络。前馈神经网络通常包含两层全连接层和一个激活函数（如 ReLU）。前馈神经网络的计算公式为：
FFN(x) = max(0, xW_1 + b_1) * W_2 + b_2其中 FFN(x) 表示前馈神经网络的输出，W_1 和 W_2 是权重矩阵，b_1 和 b_2 是偏置向量，max(0, x) 表示 ReLU 激活函数。
此外，编码器中的每个子层都包含残差连接和层归一化（Layer Normalization），这有助于提高模型的训练稳定性和收敛速度。

2.2、Transformer 解码器

与编码器类似，Transformer 解码器也由多层相同的子层组成，每个子层包含三个部分：多头自注意力、编码器-解码器注意力（Encoder-Decoder Attention）和前馈神经网络。

多头自注意力：与编码器中的多头自注意力相同，用于计算解码器输入序列中各个元素之间的关联程度。

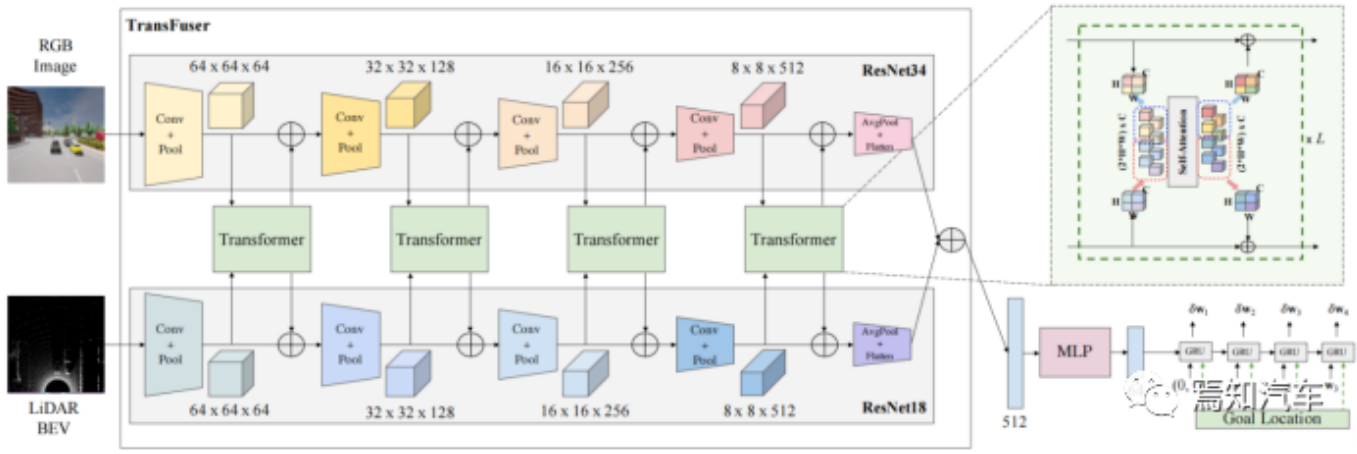
编码器-解码器注意力：用于计算解码器输入序列与编码器输出序列之间的关联程度。其计算方法与自注意力类似，只是查询向量来自解码器输入序列，而键向量和值向量来自编码器输出序列。

前馈神经网络：与编码器中的前馈神经网络相同。解码器中的每个子层同样包含残差连接和层归一化。通过多层编码器和解码器的堆叠，Transformer 能够处理具有复杂依赖关系的序列数据。

3、构建 Transformer 模型：

构建一个适用于自动驾驶场景的 Transformer 模型，包括设置合适的层数、头数和隐藏层大小。此外，还需要根据任务需求对模型进行微调，如使用驾驶策略生成任务的损失函数。

首先将特征向量通过MLP得到低维向量，传递到由GRU实现的自动回归路径点网络，并用其初始化GRU的隐状态。此外当前位置和目标位置也被输入，使网络关注隐状态的相关上下文。



使用单层GRU，用线性层从隐状态预测路径点偏移量

$$\{\Delta w_t\}_{t=1}^{T}$$

，得到预测路径点

$$\{w_t = w_{t-1} + \Delta w_t\}_{t=1}^{T}$$

。GRU的输入是原点。

控制器根据预测路径点，使用两个PID控制器分别进行横向和纵向控制，获得转向、刹车和油门值。将连续帧路径点向量进行加权平均，则纵向控制器的输入为其模长，横向控制器的输入为其朝向。

计算当前帧自车坐标系下的专家轨迹路径点和预测轨迹路径点的L1损失，即

$$L = \sum_{t=1}^{T} |\omega_t - \omega_t^{gt}|$$

4、训练与验证：

使用收集到的数据集对 Transformer 模型进行训练。在训练过程中，需要对模型进行验证以检查其泛化能力。可以将数据集划分为训练集、验证集和测试集，以便对模型进行评估。

5、驾驶策略生成：

在实际应用中，根据当前车辆状态、路况信息和周围环境信息输入预训练的 Transformer 模型。模型将根据这些输入生成驾驶策略，如加速、减速、转向等。

6、驾驶策略执行与优化：

将生成的驾驶策略传递给自动驾驶系统，以控制车辆。同时，收集实际执行过程中的数据，用于模型的进一步优化和迭代。
通过以上步骤，可以采用基于 Transformer 模型的方法在自动驾驶决策阶段生成合适的驾驶策略。需要注意的是，由于自动驾驶领域的安全性要求较高，实际部署时需确保模型在各种场景下的性能和安全性。

03

Transformer+BEV技术解决Corner Case的实例

在本部分中，我们将详细介绍三个BEV技术解决Corner Case的实例，分别涉及复杂道路状况、恶劣天气条件和预测异常行为。如下图分别表示了自动驾驶中的一些Cornercase场景。采用Transformer+BEV的技术可以有效的识别及应对大部分当前所能识别出的边缘场景。

Fig. 2. Examples from CODA. Corner cases are indicated by the bounding boxes, while each color stands for a different object class. CODA contains both *instances of novel classes* (e.g., the dog in the top-left image) and *novel instances of common classes* (e.g., the cyclist in the top-middle image).

1、处理复杂道路状况

在复杂道路状况下，如交通拥堵、复杂的路口或者不规则的路面，Transformer+BEV技术可以提供更全面的环境感知。通过整合车辆周围多个摄像头的图像，BEV生成一个连续的俯视视角，使得自动驾驶系统能够清晰地识别车道线、障碍物、行人和其他交通参与者。例如，在一个复杂的路口，BEV技术能帮助自动驾驶系统准确识别各个交通参与者的位置和行驶方向，从而为路径规划和决策提供可靠依据。

2、应对恶劣天气条件

在恶劣天气条件下，如雨、雪、雾等，传统的摄像头和激光雷达可能会受到影响，降低自动驾驶系统的感知能力。Transformer+BEV技术在这些情况下仍具有一定优势，因为它可以融合来自不同角度的图像信息，从而提高系统对环境的感知。为了进一步增强Transformer+BEV技术在恶劣天气条件下的性能，可以考虑采用红外摄像头或者热成像摄像头等辅助设备，以补充可见光摄像头在这些情况下的不足。

3、预测异常行为

在实际道路环境中，行人、骑行者和其他交通参与者可能会出现异常行为，如突然穿越马路、违反交通规则等。BEV技术可以帮助自动驾驶系统更好地预测这些异常行为。借助全局视角，BEV可以提供完整的环境信息，使得自动驾驶系统能够更准确地跟踪和预测行人和其他交通参与者的动态。此外，结合机器学习和深度学习算法，Transformer+BEV技术可以进一步提高对异常行为的预测准确性，从而使自动驾驶系统在复杂场景中做出更为合理的决策。

4、狭窄或遮挡的道路

在狭窄或遮挡的道路环境中，传统的摄像头和激光雷达可能难以获取足够的信息来进行有效的环境感知。然而，Transformer+BEV技术可以在这些情况下发挥作用，因为它可以整合多个摄像头捕获的图

像，生成一个更全面的视图。这使得自动驾驶系统能够更好地了解车辆周围的环境，识别狭窄通道中的障碍物，从而安全地通过这些场景。

5、并车和交通合流

在高速公路等场景中，自动驾驶系统需要应对并车和交通合流等复杂任务。这些任务对自动驾驶系统的感知能力提出了较高要求，因为系统需要实时评估周围车辆的位置和速度，以确保安全地进行并车和交通合流。借助Transformer+BEV技术，自动驾驶系统可以获得一个全局视角，清晰地了解车辆周围的交通状况。这将有助于自动驾驶系统制定合适的并车策略，确保车辆安全地融入交通流。

6、紧急情况应对

在紧急情况下，如交通事故、道路封闭或突发事件，自动驾驶系统需要快速做出决策以确保行驶安全。在这些情况下，Transformer+BEV技术可以为自动驾驶系统提供实时、全面的环境感知，帮助系统迅速评估当前的道路状况。结合实时数据和先进的路径规划算法，自动驾驶系统可以制定合适的应急策略，避免潜在的风险。

通过这些实例，我们可以看到Transformer+BEV技术在应对Corner Case时具有很大的潜力。然而，为了充分发挥Transformer+BEV技术的优势，仍需要进一步研究和开发，以提高图像处理能力、传感器融合技术以及异常行为预测等方面的性能。

04

结论

本文总结了Transformer和BEV技术在自动驾驶中的原理和应用，特别是如何解决Corner Case问题。通过提供全局视角和准确的环境感知，Transformer+BEV技术有望提高自动驾驶系统在面对极端情况时的可靠性和安全性。然而，当前的技术仍存在一定的局限性，例如在恶劣天气条件下的性能下降。未来的研究应继续关注BEV技术的改进和与其他感知技术的融合，以实现更高水平的自动驾驶安全性。

此文章来源于焉知汽车，作者Jessie

焉知汽车

作者 IJessie

出品 I 焉知