

基于单目摄像机的行人与自驾车轨迹预测

动机

传统的利用俯瞰图进行轨迹预测的方法只能在静态摄像头情况下使用，在动态情况下会失效。

内容

1. 有两种行为会导致摄像头所拍到的画面发生变化：
 1. 行人的移动，只影响整个画面的一部分。
 2. 摄像头所在的汽车的移动。整个画面都会被影响。
2. 利用自监督训练的范式训练出了一个网络，用于通过画面推断摄像头所在的车辆在做什么运动。基于此，我们可以将行人的运动状态标准化。
3. 行人的运动状态标准化之后，可以用一个简单的线性模型来预测行人的状态

相关工作

1. Monocular Depth estimation：用于检测每个像素相对于单目摄像头的距离
2. Trajectory Prediction from static camera：利用LSTM将人类交互因素和所有轨迹结合起来进行单一的结果预测，缺点是需要一个静态的俯视摄像头。
3. Trajectory Prediction from on-board camera：
 1. 利用CNN预测行人是否会穿过马路
 2. 先用一个RNN预测速度和方位角，再用一个RNN来预测行人的轨迹。缺点是只能预测两个标量，并且需要人工数据标注。
 3. 在2的基础上添加一个RNN用来预测每个行人的意图，但是行人的bounding box时常互相穿插导致方法失效
 4. 利用multi-stream RNN将行人位置，车辆运动和车流预估相结合。缺点是忽略了行人本身的特征，移动仅仅表示为平移和旋转。

具体方法

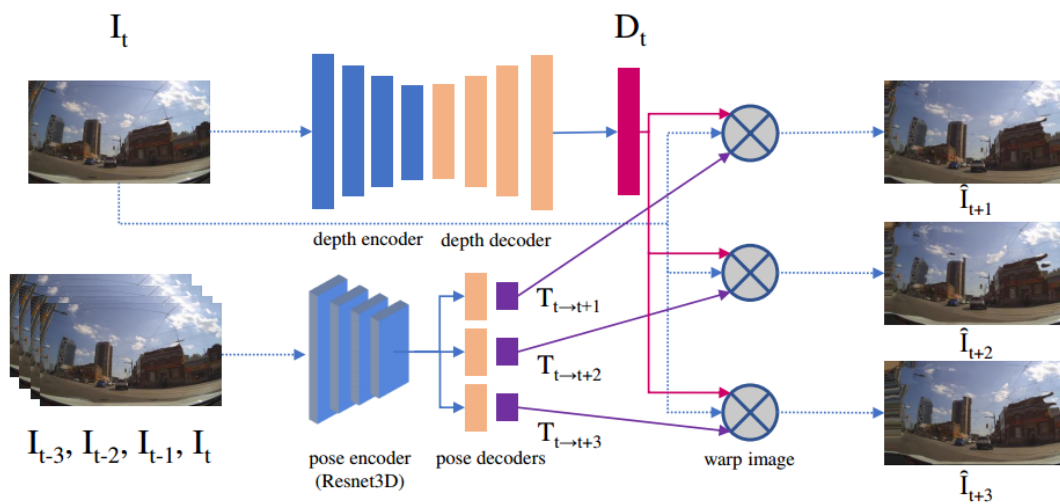
汽车运动判断

已有的范式：

1. 输入：连续的两帧图像 I_{t-1} 和 I_t 。
2. 输出：dense depth map D_t 和pose transformation $T_{t-1 \rightarrow t} \in SE(3)$
3. 方法：用 $T_{t-1 \rightarrow t}$ 和 D_t 把 I_{t-1} 变成 I^* 然后最小化 I^* 和 I_t 的差距。

本文的方法：

1. 输入：一系列图像 I_i
2. 输出：一系列预期动作 $T_{i \rightarrow i+f}$
3. 方法：同上。



标准化和轨迹预测

1. 视角转换：通过矩阵运算进行视角转换。

2. 轨迹预测：

1. 输入： B_t 和 B_{t-l}

2. 输出：运动参数 V 和 A

3. 方法：

先将视角统一，然后就变成了一个统计问题。

相关实现

<https://gitlab.com/lukeN86/pedFutureTracking>