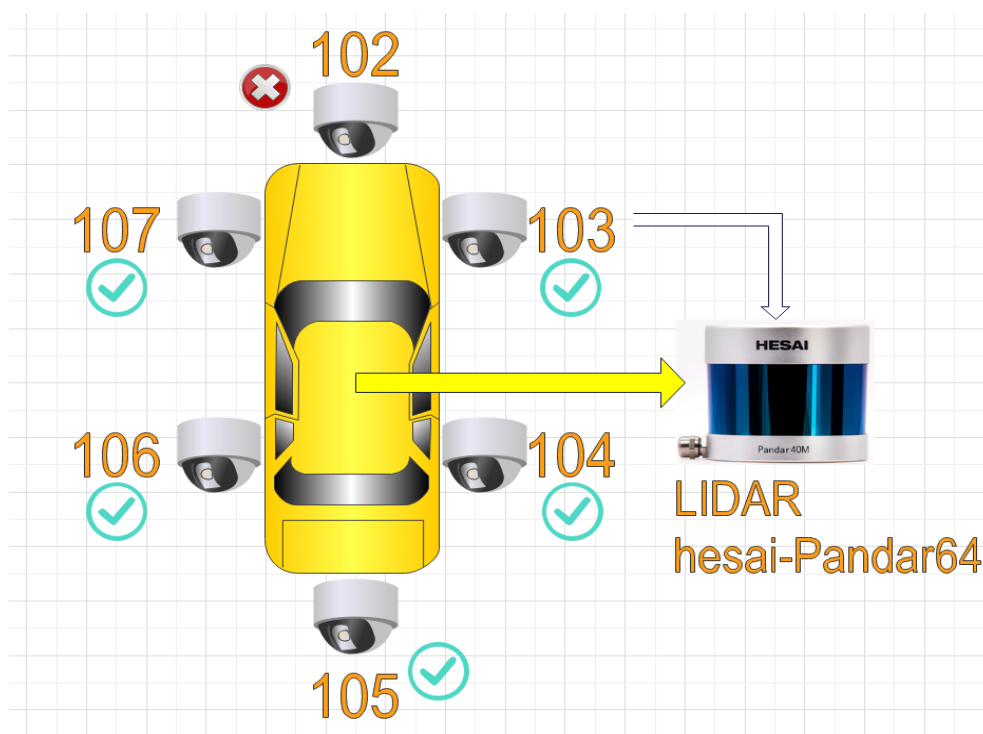


# 基于 CASIA 数据集的 SurroundDepth 模型实现

沈瑞淇

为了进一步验证 SurroundDepth 论文提出的自监督环视相机深度估计模型的应用泛化性，我基于本组自动驾驶汽车，行驶环绕整个自动化所园区，拍摄了数据集。数据集包含 2140 组数据，每组数据包括 6 个环视针孔相机拍摄的 RGB 图像，以及车顶 LIDAR 获取的三维激光点云，示意图如下：



## 一．实验配置与环境

CASIA 数据集将采用 SurroundDepth 中提出的 SfM 进行预训练，特别设定参数如下：

### 1) 图像长宽

针孔相机拍摄图像的长宽由 nusenes 中的 1600\*900，调整为 1920\*1200

### 2) focal\_scale 焦距

nusenes 图像尺寸为 1600\*900，焦距为 500，CASIA 图像尺寸为 1920\*1200，假设相机的光学系统保持不变，为保持相机成像的一致性与几何投影的相似性，将焦距等比例缩放至 800

### 3) min\_depth 和 max\_depth

在 LIDAR 生成的 2140 张稀疏深度图中（只关注有值的像素点），最小为 0.1，最大为 197.8（单位米），因此将 min\_depth 设为 0.1，max\_depth 设为 200

### 4) thr\_dis (threshold\_distance)

在匹配 CASIA 左右相邻两个视角图像的 sift 特征点时，发现计算出的两项指标 distances1 和 distances2，即

```
distances1 = inputs['match_spatial'][j][:, 4]
distances2 = inputs['match_spatial'][j][:, 5]
```

值相对较大，如果维持 nusenes 中关于 thr\_dis 为 10 的设定，那么将过滤掉绝大多数的匹

配点组合，可视化效果不理想，且 loss 容易为 nan，因此经过反复调整，最终将 thr\_dis 设定在 100 最为合适。

## 二．实验代码文件用途解释

代码文件夹路径：`/datab1/ruiqi_shen/cp/SurroundDepth`，conda 环境为 `surrounddepth1`，代码全部为自己编写，重点代码文件解释如下：

**sift\_casia.py** 用来提取 CASIA 针孔图像的 sift 特征点，边缘阈值和对比度阈值维持不变，分别为 8 和 0.01

**match\_casia.py** 匹配 sift 特征点，每个视角的图像和左右两个相邻视角的图像做匹配

**casia\_mono\_dataset.py** 建立在 `mono_dataset` 基础上，建立完整的数据集，包含 RGB 图像信息，LIDAR 生成的深度信息，特征点匹配信息，内外参信息，运动姿态信息等。

**runer\_casia.py** 核心训练和验证函数，规定了训练超参数，数据集和网络模型加载，训练（损失函数类型和定义），验证方法等

**run.py** 主函数，执行即可

**casia\_scale\_pretrain.txt** 第一阶段预训练（sfm）配置参数

**casia\_scale.txt** 第二阶段正式训练配置参数

由于后面第三部分将要提到的原因，实际只使用 103-107 的 5 个环视针孔摄像头作为数据来源（即排除了 102），因此对于网络（`resnet_encoder/CVT/depth_decoder`）也都做了相应修改。

所有的 **CASIA 园区数据**，都在 `/dataa1/ruiqi_shen/real_dataset` 下，包括原始数据集（6 个视角的图像和激光雷达点云），sift 特征点，特征点匹配结果，由 LIDAR 点云生成的稀疏深度图（验证集）等

## 三 实验结果

实验的指标参数如下：

$$\text{Abs Rel: } \frac{1}{|T|} \sum_{d \in T} |d - d^*| / d^*$$

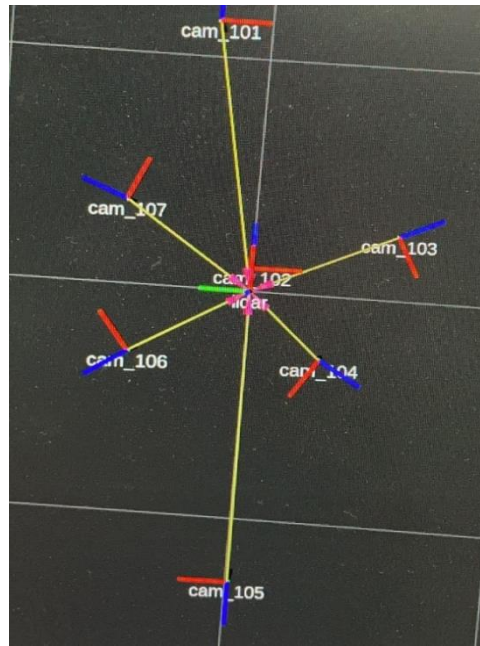
$$\text{Sq Rel: } \frac{1}{|T|} \sum_{d \in T} ||d - d^*||^2 / d^*$$

绝对相对误差 Abs Rel，和平方相对误差 Square Rel，都是越接近 0 越好。

Median 之比：真实深度图的深度值中位数，与预测深度图的深度值中位数之比，越接近 1 越好（说明两幅图相似），表达式如下：

$$\frac{\text{median}(D^*)}{\text{median}(D)}$$

实验起初长时间没有得到期望的效果，后来发现之前的工程师对于 102 摄像头标定存在误差。按照标定好的 6 个摄像头的内参和外参，绘制出的各个摄像头与激光雷达之间的空间位置关系如下：

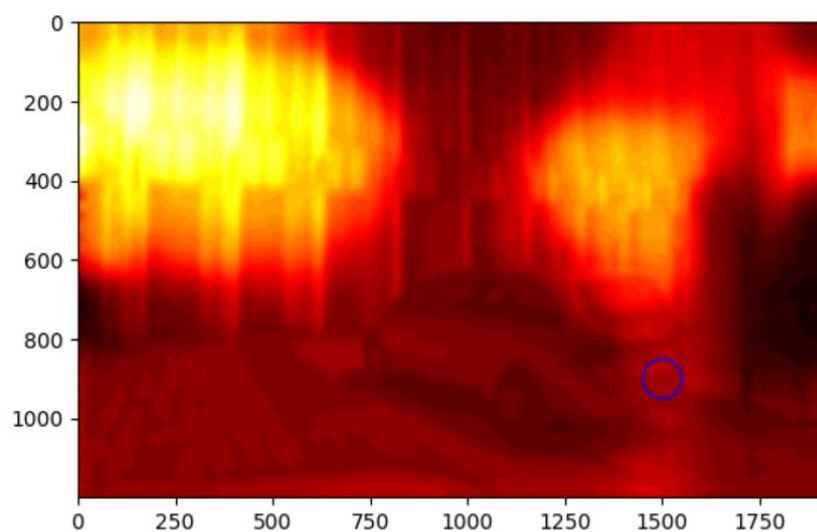


由上图可见, 102 摄像头和车顶激光雷达基本重合, 然而实际上它是置于车前部的摄像头, 与车顶激光雷达之间有一定距离。因此, 我们去除了 102 号摄像头, 只留下剩余 5 个摄像头 (107→103, 逆时针顺序) 做深度估计, 由于 107 号和 103 号摄像头之间没有直接的公共视野区域, 这对于模型的训练会产生一定的影响。

有以下两点需要注意:

1. 101 号摄像头由于是测量远景的, 而其余 6 个摄像头是测量近景的, 因此 101 号摄像头本身就不在考虑范围内。
2. 给到我标定好的外参是由 Lidar 指向相机的, 因此需要取逆, 改变物理意义为相机指向 Lidar, 才能够继续运算。

因为如上所述的原因, 目前网络预测出的深度 (例如对于某一相机), 视野范围内不同像素区域的预测深度在相对深度关系上准确, 但是在绝对深度值上, 仍存在较大误差。



以上图为例, 上图是可视化出来的深度图热力图 (根据深度大小, 将像素的深度值映射到不同的颜色上), 蓝色圈圈所在点距离相机有 17 米。然而实际中肯定不到这个距离, 只有约 5 米左右, 因此深度绝对值估计仍存在问题。然而, 空间不同像素区域之间的相对深度关系是

准确的，因而能清晰地看出深度图描绘的内容。

由于绝对深度估计存在误差，因此在进行第二阶段训练时，效果不甚理想，几个误差指标都很大。因此目前针对 CASIA 数据集，只选用了 5 轮预训练的效果加以展示和可视化，后续会解决此问题，并完成第二阶段的训练。

	median 之比	Abs Rel	Sq Rel
CASIA	0.7944	0.865	9.713

预训练 5 轮后，效果最理想的是第 1 轮，此时指标如上。可视化视频也利用的这个模型。

CASIA 数据集的可视化视频已经放在了附录中，它是由该数据集的 1712 组训练集数据（每组对应 5 个可用视角的 5 张图像），逐张通过训练好的模型做 validation 生成深度图，转为热力图表示，并逐帧拼接而成的视频（每个视角对应一个视频）。

6 个视频分别命名为 casia\_0/1/2/3/4.mp4，其中 0/1/2/3/4 分布对应摄像头序号 107/106/105/104/103，能够看到自动驾驶汽车环绕园区的景象。详情请观看视频。