针孔相机和鱼眼相机的异构思路

沈瑞淇

本设计基于的两篇论文分别为:

1) SurroundDepth: Entangling Surrounding Views for Self-Supervised Multi-Camera Depth Estimation (环视针孔相机)

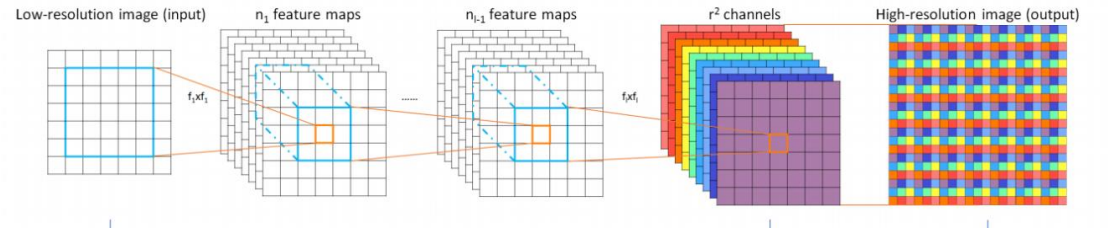
2) OmniDet: Surround View Cameras based Multi-task Visual Perception Network for Autonomous Driving （鱼眼相机）

一．从环视针孔相机的 SurroundDepth 到鱼眼相机的 OmniDet

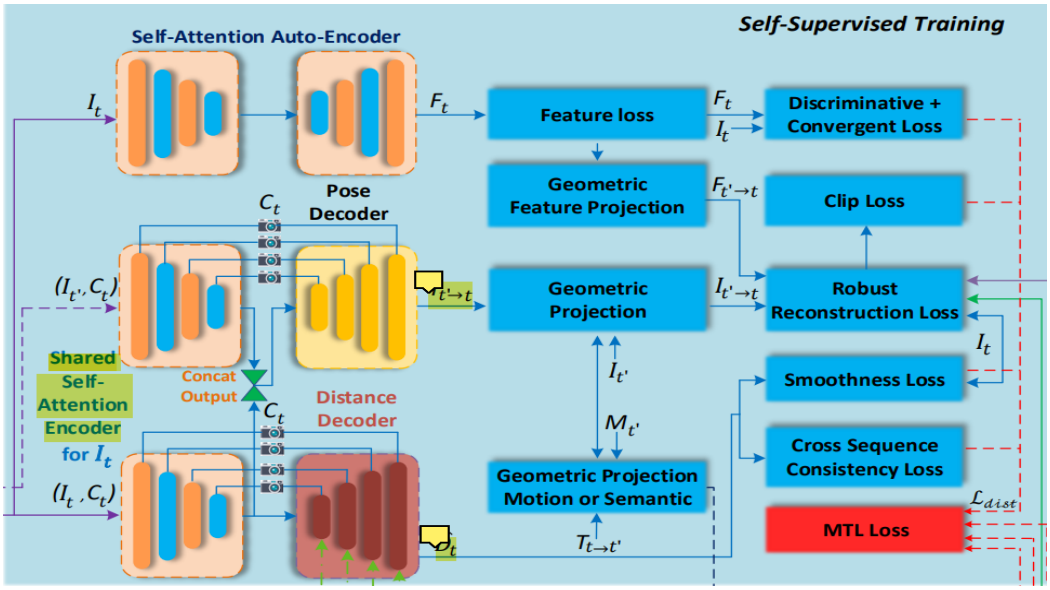
在上述两篇论文中，使用的 encoder+decoder 模型完全一致，都是基于单视角深度估计经典论文 Monodepth2。

以 SurroundDepth 为例，encoder 部分以 ResNet 为主干，每次输入是同一时刻各个视角的图像的整体，卷积层逐个产生不同大小和通道数的特征向量，这些 multi-scale 的特征向量并行进入 CVT 进行特征提取，decoder 部分接收在不同视角间做完 attention 后的特征向量，采取逐层残差连接方式逐步还原大小，输出预测深度图。

OmniDet 则采用 4 个鱼眼相机进行深度估计（与本组自动驾驶车相同），其唯一区别在于 OmniDet 并没有在 encoder 和 decoder 之间对来自各个视角的特征向量做 attention（即 CVT），而是在 decoder 中加入了 pixelshuffle+ICNR，将低分辨率图像转换为高分辨率图像。



Pixelshuffle 像素洗牌，即逐像素卷积，本质上是将原图像的每个像素经过卷积层后增加它的深度，然后再重新排列。例如原图为 224*224*3，卷积后可能得到 224*224*64 的特征图，按照“每个像素的 64 个通道对应高分辨率图像中一个 8*8 子块”的原则去重新排列，从而得到超分辨率图像，ICNR 则负责初始化前面卷积层的网络权重。



上面两幅图分别表示了 OmniDet 中深度估计的原理图，以及四鱼眼视角的方位（前，后，左，右）。

OmniDet 所使用的自监督深度估计的 loss 函数，主要有 4 个，分别是三维重建 RGB 损失，深度图平滑损失，双向序列一致性深度图损失，以及自身特征损失，只有深度图平滑损失与 SurroundDepth 中的完全相同，其余三种损失详解如下：

1）传统的三维重建损失，实质是建立在静态场景与光度一致性的假设之下的，然而实际车辆运动必然产生视野内的物体运动与遮挡的问题，即违反静态假设。这些区域（运动和遮挡）产生的大误差会不可避免降低网络性能。因此，作者引入 clip loss 函数来处理此问题，高于某阈值的误差将被锁死在某一固定值，这些错误将产生零梯度，并不对训练产生影响。

2）双向序列一致性深度图损失（cross-sequence consistency loss）是针对深度图重建的损失。具体实现可参考 *FisheyeDistanceNet: Self-Supervised Scale-Aware Distance Estimation using Monocular Fisheye Camera for Autonomous Driving*　核心原理及公式如下：

$$\mathcal{L}_{dc} = \sum_{t=1}^{N-1} \sum_{t'=t+1}^{N} \left(\sum_{p_t} \mathcal{M}_{t\to t'} \left| D_{t\to t'}(p_t) - \hat{D}_{t\to t'}(p_t) \right| + \sum_{p_{t'}} \mathcal{M}_{t'\to t} \left| D_{t'\to t}(p_{t'}) - \hat{D}_{t'\to t}(p_{t'}) \right| \right)$$

假设我们现在要从 t 时刻的图像（原图 It,深度图 Dt,像素点 pt）来预测 t+1 时刻的 Dt+1

有两种方法来处理：

第一种，逐个像素点处理

（每个二维像素点独立地反投影到三维点，独自变换，再独自投影到新的二维平面）

$$\hat{P}_{t+1} = \pi\left(T_{t\to t+1}\ \pi^{-1}(P_t, D_t)\right),$$

其中 P_t，$\hat{P}_{t+1}$ 均为像素点

第二种，整体像素点处理

（将二维平面的像素点反投影成三维点云，点云整体变换，再整体投影到新的二维平面）

$$P_t = \pi^{-1}(P_t, D_t)$$

$$P_{t+1} = T_{t\to t+1} \times P_t$$

$$\hat{D}_{t+1} = \| P_{t+1} \|$$

理论上，这两种投影方式得到的 t+1 时刻的深度图应该是一致的，双向序列一致性深度图损失计算的正是这两种方法得到的深度图之间的差异。至于“双向”，正向是由 t-1 和 t+1 时刻帧，去预测 t 时刻帧的深度图，反向则是相反顺序（t→t-1/t+1），“双向序列一致性深度图损失”将正向和反向预测都考虑在内。

3）自身特征损失，它来源于鱼眼相机的自身特性。由于“广视角”的特性，鱼眼相机拍摄图像的较大一部分区域的亮度和颜色是相似的，我们称之为“均匀区域”。这些均匀区域在视觉上缺乏细节和纹理，可能导致训练过程中的优化困难或陷入局部最小值。而一旦出现上述情况，网络提取出的特征与目标特征之间，就会出现较大的差异。因此论文针对鱼眼相机，在损失函数中，加入了“自身特征损失”这一项，直接比较输入图像的特征表示和目标特征之间的差异，防止局部最优化的情况发生。

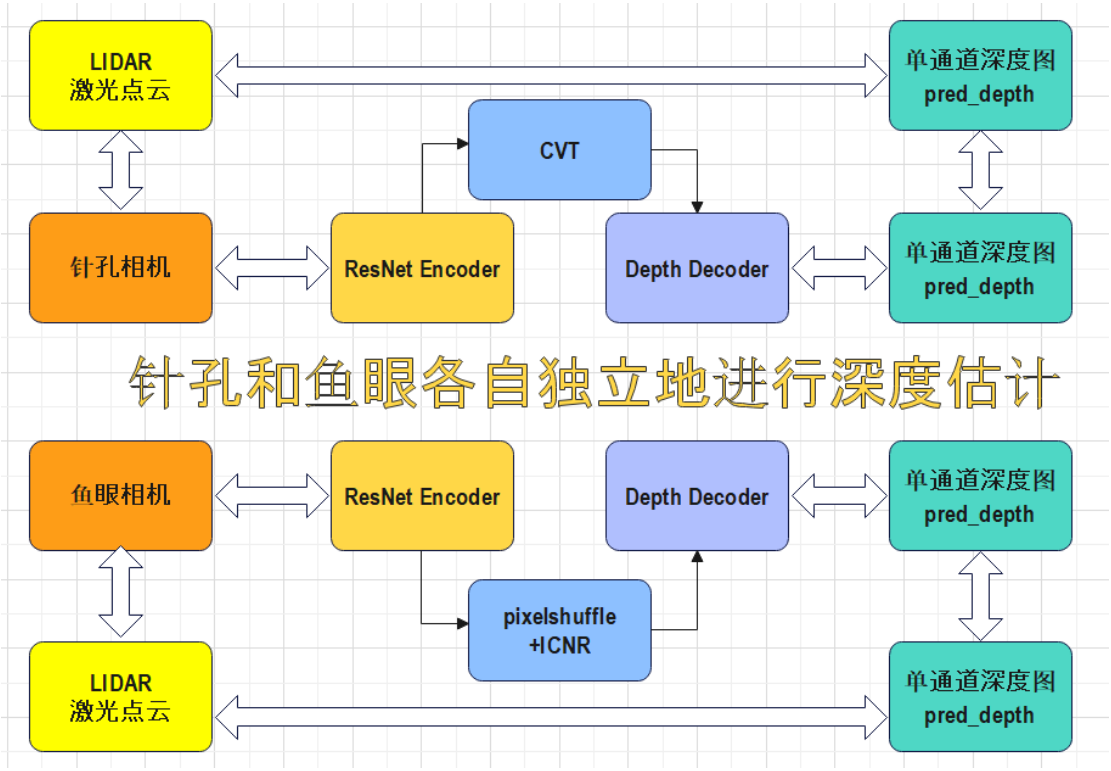
二．针孔相机和鱼眼相机的异构设计

目前设计的异构模型针对的是**一个针孔相机和一个鱼眼相机的异构**，其中鱼眼相机的视野范围要能够涵盖针孔相机（例如鱼眼相机和针孔相机都是前视）。未来有潜力拓展为 6 针孔和 4 鱼眼的联合深度估计模型。

我设计的异构（针孔相机+鱼眼相机）联合深度估计方法，主要分为两阶段训练。

第一阶段是针孔模型和鱼眼模型各自平行且独立地进行训练。

由于之前在 nuscenes 数据集上的监督学习可视化效果明显好于 SfM 自监督学习，因此针孔&鱼眼在此阶段均借助 LIDAR 生成稀疏深度图监督学习的方式来完成训练。示意图如下：



第一阶段：针孔模型和鱼眼模型各自平行且独立地进行训练

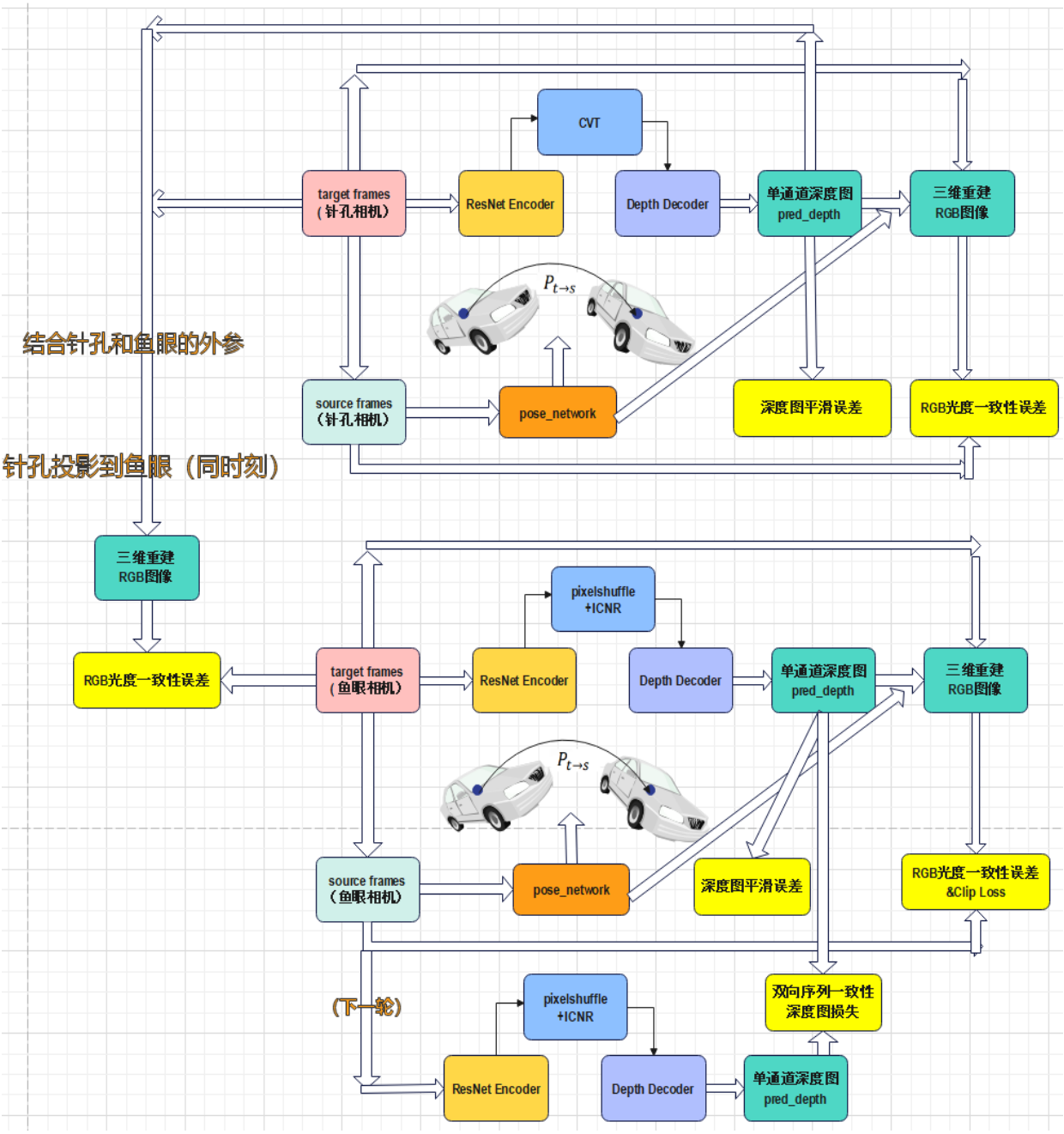
第二阶段是针孔模型和鱼眼模型联合训练。

我还没有做到在二者在特征维度的拼接，其需要严格的数学理论做支撑，也是我未来的研究方向。截至目前，为保证技术的可行性，我着手于 Loss 函数进行异构设计。

具体来说，在 target 时刻，针孔相机模型预测得到的单通道深度图，结合输入针孔相机的 RGB 图像，通过针孔相机和鱼眼相机各自的外参和内参进行投影，投影 RGB 图像到鱼眼相机上，与鱼眼相机此时输入的 RGB 图像作比较，计算出 RGB 光度一致性误差。

除异构投影外，针孔相机模型和鱼眼相机模型，各自还按照 SurroundDepth 和 OmniDet 中的第二阶段训练方式进行正常训练，详细实现已于第一部分讲述。

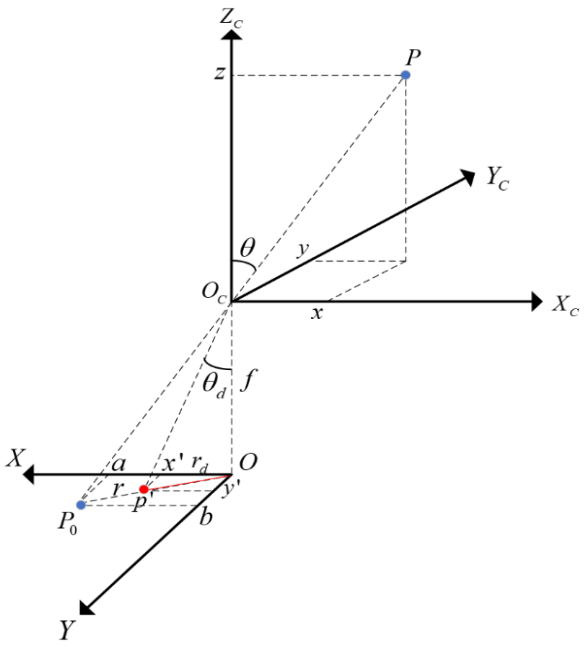
以下是异构深度估计的完整流程图：



第二阶段：针孔模型和鱼眼模型联合训练

附录：鱼眼相机的成像模型（带畸变）

首先，为什么鱼眼相机模型会有畸变？这是由鱼眼的功能决定的。鱼眼相机希望实现广角视野，单独一个透镜做不到这一点，必须由多个透镜相互配合，才能使得广角范围内的所有成像光线，都能经过多次折射，投影在成像平面上。此外，每个透镜并非一定是球面透镜，还可以是非球面透镜，这就使得光线投影过程肯定不能简单地简化为一条直线。人们将这种“光线在鱼眼相机内的投影过程”，描述为“鱼眼相机的畸变”。



鱼眼相机带畸变的成像投影过程

假设鱼眼相机坐标系下物体所在点 P(Xc,Yc,Zc)，假设不存在畸变，那么按照针孔相机投影模型，在不存在畸变的情况下，会投影至 Po(a,b)，此时有如下关系：

$$\frac{a}{f}=\frac{X_c}{Z_c} \qquad \frac{b}{f}=\frac{Y_c}{Z_c} \tag{1}$$

假设 f=1,由于 Pc(Xc,Yc,Zc)已知，所以 a 和 b 可求得。由此，可以确定无畸变针孔模型下的投影点 Po 坐标(a,b)。

由于存在畸变，因此光线发生折射，点 P 投影到 p'点，p'点在图像坐标系下坐标为(x',y')。

$$\frac{x'}{a}=\frac{y'}{b}=\frac{r_d}{r}$$
$$r_d=f\cdot\tan\theta_d=\tan\theta_d\approx\theta_d \tag{2}$$

由于畸变折射后的折射角θd 很小，且 f 假设为 1，因此有 rd=θd，因此上式可写作：

$$\frac{x'}{a}=\frac{y'}{b}=\frac{\theta_d}{r}\quad(r=\sqrt{a^2+b^2}) \tag{3}$$

查阅 opencv 文档可知，发生畸变后的折射角θd 和无畸变时的θ有如下关系：

$$\theta_d=k_0\theta+k_1\theta^3+k_2\theta^5+k_3\theta^7+\cdots$$
假设 $k_0=1$，
则有：
$$\theta_d=\theta\,(1+k_1\theta^2+k_2\theta^4+k_3\theta^6+\cdots) \tag{4}$$

由于θ=arctan(r/f)=arctan(r)，因此θ可求，从而θd 可得。

由此，a,b,r,θd 均已知，由(3)，可求得 x'和 y'，从而知道畸变模型投影点 p'的坐标。