

## 一、项目概述

项目地址：/home/xiewei/home/xiewei/250t\_process/ouyangshiyu/semi

### 1.1 项目目标

通过深度学习的方法，利用车头二维码关键点位置信息对铰链车的车头和相机进行角度估计。利用尽量少量的训练图片训练再对各种复杂场景进行测试，包括但不限于黑夜、阴影、大角度偏转、遮挡等场景，示例如图 1-1。



图 1-1 测试示例图片

### 1.2 任务难点及可能解决办法

#### (1) 训练集图片覆盖场景少

- 标注时覆盖大部分角度的图片，其他大部分的情况如黑夜、雨雾天和相机脏污等可以用数据增强来完成。主要的**数据增强**方法 [imgaug](#)、[albumentations](#)，其中 [albumentations](#) 多了模拟相机畸变的情况，本次项目最后选用的是 [imgaug](#) 库。增强后的部分训练集图片如图 1-2 所示。



图 1-2 数据增强在线扩充数据集

- **图像生成**方法可以生成多种场景的图片。经典方法如 [cyclegan](#) 和 [UNIT](#) 等。尝试了 [cyclegan](#) 方法。但是出现了数据训练表现良好，输出结果却不符合预期的情况。如图 1-3 所示，第一列是原图，第二列分别是 [sunny2winter](#) 和 [day2night](#) 的风格转换。[sunny2winter](#) 使用 [cyclegan](#) 的预训练权重；[day2night](#) 使用数据集中的白天黑夜的数据集训练。

- **迁移学习**可以在不进行数据扩充的情况下进行场景迁移，从而适应如训练白天数据迁移到黑夜场景。具体可以看 [weathergan](#) 相关内容。



图 1-3 cyclegan 扩充数据集

- (2) 图片分辨率较大，直接训练会导致内存爆炸。
  - 对图片进行下采样（如 yolo 中的 letterbox 方法）再输入到 backbone，从而减少内存占用，并且提高训练和推理速度。因此导致的量化误差问题可参考 [UDP](#) 等方法。
- (3) 训练数据量少，网络参数过多容易造成过拟合，网络泛化性不高的问题。
  - 使用轻量模型：如 MobileNet、ShuffleNet 等普遍使用的参数量较少的模型或者 Light-HRNet 等基于人体姿态估计设计的 HRNet 的轻量化模型。
  - 本项目部署到车辆中会采用 GPU 进行推理，而 MobileNet、ShuffleNet 等网络会出现[无法完全利用 GPU 资源，而在 CPU 上表现良好](#)的情况，所以可以尝试能在 GPU 上能尽可能利用资源的网络。

## 二、项目整体流程



图 2-1 项目整理流程

- 数据预处理：
  - ◆ 训练数据用 Lableme 工具打标签
  - ◆ Encode 函数将关键点坐标转换为 heatmap 形式
- 模型搭建
  - ◆ Stem 降低输入 backbone 的分辨率
  - ◆ Backbone 提取特征
  - ◆ Head 进行预测
- 模型训练
  - ◆ 选择合适的优化器（AdamW）和学习策略（余弦退火）进行训练
- 模型推理
  - ◆ Python 转换为 onnx 对模型推理进行简化

- ◆ 网络推理完成对输出的热力图进行后处理转换为需要可视化的关键点

### 三、数据集与数据预处理



图 3-1 数据集域数据预处理流程

#### 3.1 数据来源和数据格式

- 数据来源：半挂车运动途中对前车二维码的拍摄视频。
- 数据格式：需要从视频中截取右侧 1280\*720 大小的图像，并且保存为 jpg 格式。

#### 3.2 数据预处理

##### 3.2.1 数据清洗：

因为尽量用少量的带标签数据集进行训练，而识别的一个重点在角度的判定，所以数据集中最好**包括大部分的角度**。最后选择了 172 张图片进行训练（其实还可以更少，因为有角度重叠的部分）。

而不同的环境可以用数据增强手段模拟以及迁移学习的方法找到不同域之间的共同点进行迁移。

##### 3.2.2 标签处理和关键点坐标 encode

利用 Labelme 标注工具进行标注。标注点的类别名称全部命名为 points，如图 3-2 所示。标注点的顺序如图 3-3 所示。

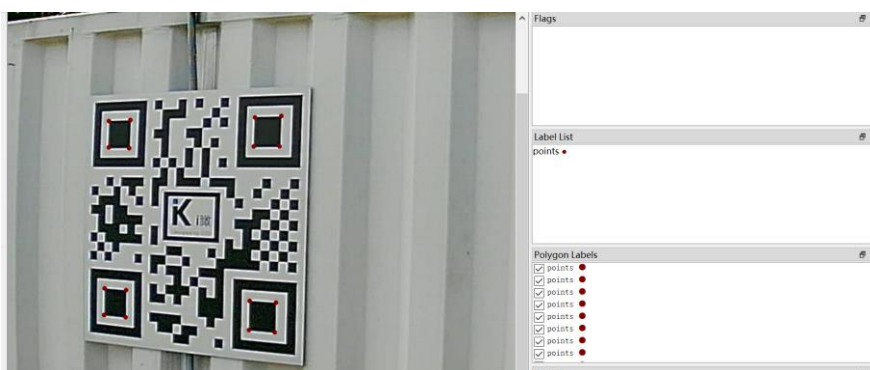


图 3-2 标注示例

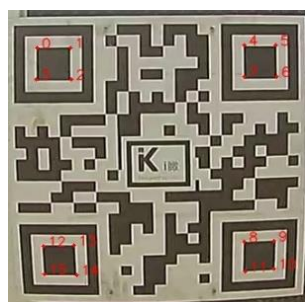


图 3-3 标注数据顺序

获得关键点坐标的 json 文件后，经过 lib.codec.msra\_heatmap.py 中的 encode 函数得到目标热力图。

### 3.2.3 数据增强

利用 imgaug 库中的相关函数采用了多种在线数据增强的方法，如图 3-1 所示。当使用 Fliplr 时，因为反转的不只是图像还有对应的关键点。而 imgug 库中没有对关键点进行翻转，所以本次项目中在封装库中改变了点的对应关系。



图 3-4 使用的数据增强方法

## 四、模型架构

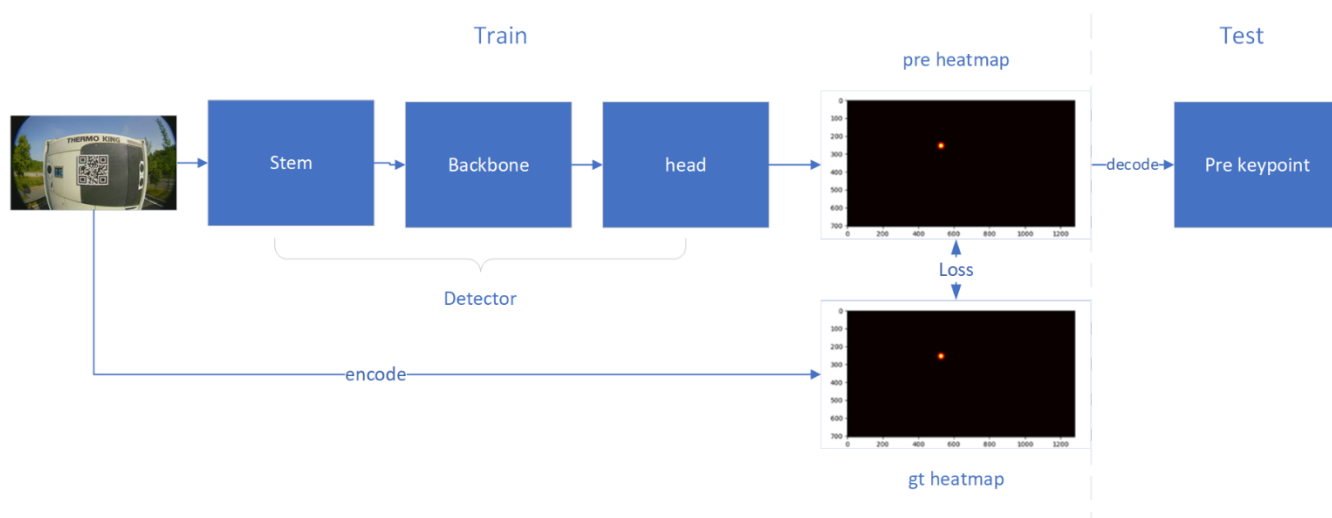
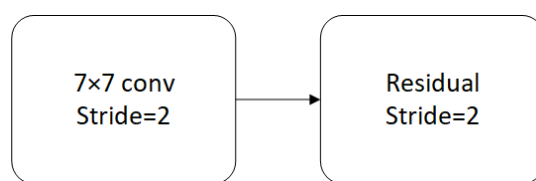


图 4-1 模型架构

### 4.1 Stem

对输入图像进行下采样。Backbone 的参数数量和计算量较多，所以在这之前先减小输入图像的大小。考虑到直接缩小图片会损失图片的信息而这个项目对预测精度要求

较高所以采用增加图片通道减少输入宽度和长度的方法。Stem 结构如图 4-2 所示



4-2 Stem 结构

## 4. 2 Backbone

因为项目代码架构较为简明清晰，所以可以容易集成不同的 backbone 进行对比。在这里设置了超参数 `feature_scale`，来控制中间层的通道数目。

## 4. 3 Head

预测头只有一个分支，用于输出图像关键点的 Heatmap。

# 五、 训练和评估

## 5. 1 训练

训练策略：余弦退火学习率；预热轮数为 10；初始学习率为 0.005；优化器为 AdamW。

可以看出在 100 轮的时候 loss 出现骤降，根据 `bone_loss` 的含义，推断是网络通过关键点之间的约束关系学习到了他们的相对位置关系。也就是说在这个 epoch 之后，主要学习的是怎么减少误差，在这之前学习的是尽量减少出现离群的点。100epoch 之前起主要作用的是 `bone_loss`，之后是 `supervised_loss`。

- `train/bone_loss`

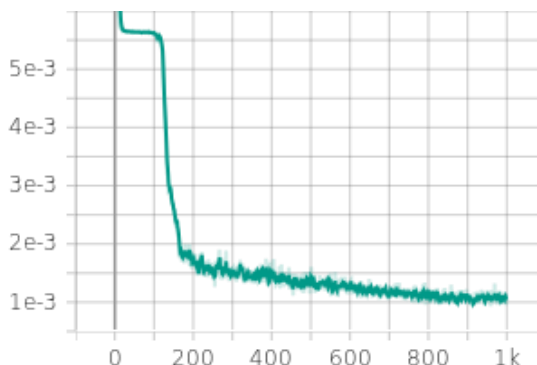


图 5-1 epoch-bone\_loss

- `train/supervised_loss(mse,adaptivewingloss)`

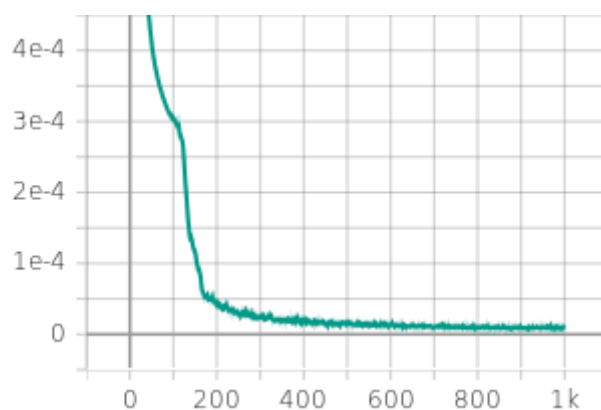


图 5-2 epoch-supervised\_loss

- train/total\_loss

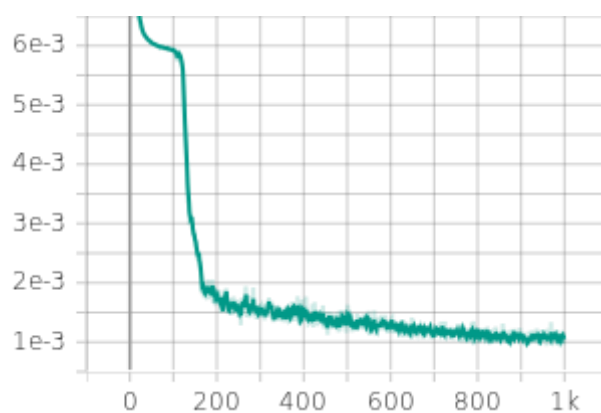


图 5-3 epoch-total\_loss

## 5.2 测试

在测试集上除了冷链车中的畸变非常大的情况。如图 5-4 所示，可能会出现点的标注错误，而这也可以通过增大过滤点的置信度阈值来实现。在其他的大部分的测试集中表现良好，如图 5-5 所示。

在推理过程中，模型的总参数为 1.44M，模型大小为 5.6MiB，输入一张 1280\*720 的图片推理速度为 16.5ms(GTX 1050Ti)



图 5-4 冷链车检测图





图 5-5 测试集中部分检测结果

## 六、 尝试的方向

### 6.1 半监督学习

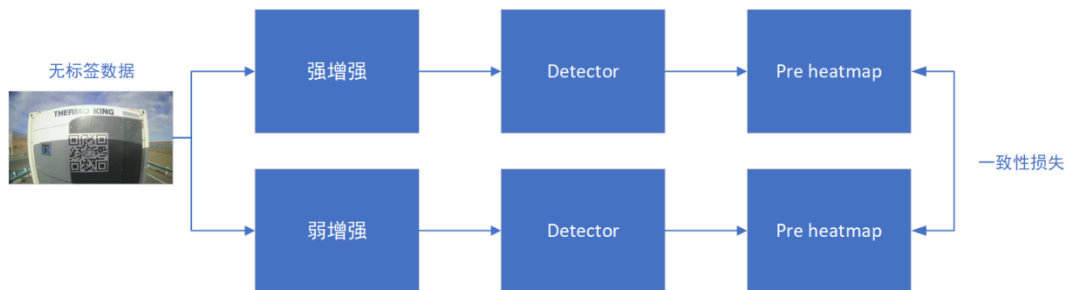


图 6-1 无标签数据半监督学习流程

在监督学习的基础上加上无标签数据的半监督学习，利用了 fixmatch 中的一致性损失的思想。也就是对无标签数据进行两种增强，最后预测的结果应该是一致的。

训练时采用的方法是将 1k 个 epoch 进行训练的权重作为预训练模型，让半监督方法在这个基础上进行训练。最后的结果反而没有一开始好，但是 train\_loss 保持正常，也就是出现了过拟合的特征。

分析是因为 1k 个 epoch 中，网络学到的只是训练集以及训练集经过数据增强后的图像，而不是图像真正的特征。

### 6.2 迁移学习

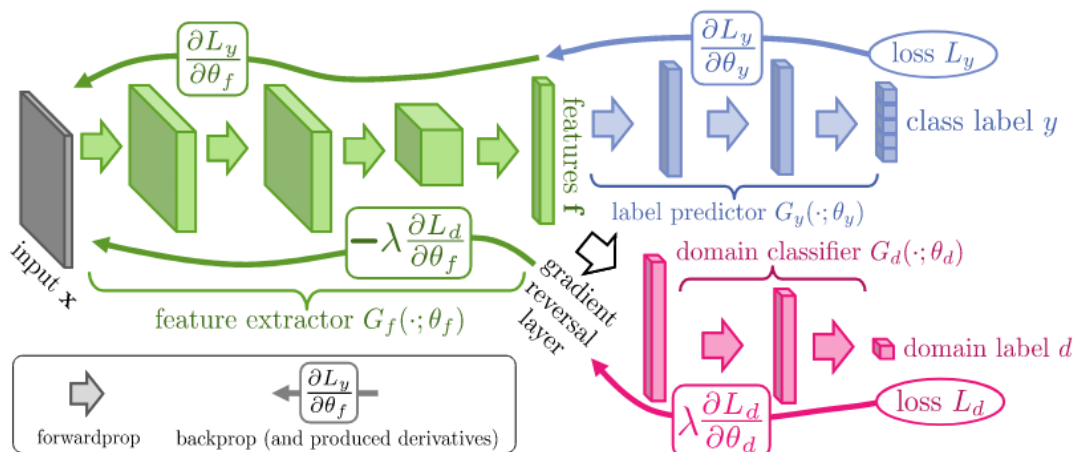


图 6-2 DANN 算法主要思想

比起最开始的全监督的网络，主要是在特征提取曾返回了一个 feature  $f$ ，然后经过一个  $mlp$  用于比较源域和目标域的距离，在算法中量化为  $domain\ loss$ 。而在最后的训练的过程中， $domain\ loss$  不能收敛。分析有以下几个原因：

- 域自适应算法在回归问题上难以奏效。在分类问题上，不同的类之间存在着明确的决策边界。通过领域对抗等方式，目标域上不同类之间的决策边界的距离被扩大，从而有助于模型泛化到目标域。但是在回归问题的输出空间是连续的，因此不存在明确的决策边界。
- 训练集数据太少，因此在使用迁移学习方法时，仍然没有学到图像的根本特征。

## 七、未来发展方向

### 7.1 可优化方向

#### 1. 先框出二维码再对关键点进行检测

可以只将图片的一部分输入网络。用轻量网络对二维码进行检测，再对二维码中的部分进行关键点的提取。也就是用二维码检测时间换取后续关键点提取时间。而且可以提高检测精度。

#### 2. 对 backbone 进行优化

虽然将 UNet 参数量缩小到 1.44M 但是这并非主流的关键点检测网络以及主流的轻量网络，后续可以尝试如 hrnet-light 和 yolo 等网络。

#### 3. 对大模型训练的权重进行微调

本次项目中是直接使用 Unet 从零开始训练，而训练集数据较少，很容易出现过拟合的现象。虽然大量的数据增强缓解了测试集上表现不好的情况，但是从迁移学习的尝试中可以看出还是有过拟合的趋势。

#### 4. 利用无监督模型

MAE 借鉴了 Bert 中对词语进行完形填空的思想，对图像进行遮挡再恢复图像，从而提取特征。基于此，我们也可以先对无标签数据设定一个特定的任务，比如：遮挡然后恢复图像、旋转再推断旋转角度，从而提取大量未标注图像的特征。

### 7.2 关于无标签数据的利用思考

利用无标签数据的有效方法：

- 标记数据 (Labeling)：将无标签数据转化为有标签数据，并进行监督学习。
- 半监督学习 (Semi-Supervised Learning)：利用部分有标签数据和相关的无标签数据，要求它们在目标领域具有相似的特征和标签体系。
- 弱监督学习 (Weak Supervision)：这包括标签数据质量不高（噪声标注）或标签抽象度高的情况，有时也会结合两者。这些数据可以通过自动化处理获得，既可以是人工标注的，也可以是通过网络爬取获得的。
- 自监督学习 (Self-Supervised Learning)：这是一种无监督学习方法，其中数据本身提供训练信号，但信号的设计通常是人工制定的。例如，通过恢复损坏的数据（如 BERT），解决拼图问题 (Jigsaw Puzzle)，或进行图像着色 (Colorization) 等方式进行训练。



此外，这些方法可以用于预训练模型，然后与迁移学习相结合，通过在目标任务上使用有标签数据进行微调，以获得更好的性能。这是当前计算机视觉和自然语言处理领域的常用策略之一。

### **创建监督信号的关键：**

核心思想是在无标签数据中创建有效的监督信号，也就是损失函数。只要有有效的监督信号，就可以从无标签数据中提取信息。通常，我们可以将利用无标签数据的方法分为以下几种：

- 伪标签法（Pseudo-Labeling）
- 对抗训练法（Adversarial Training）
- 基于一致性的正则化增强数据方法
- 对噪声数据进行去噪或恢复

这些方法可以挖掘不同的信号，适用于不同的任务。然而，如何为特定数据集和任务设计监督信号（损失函数）仍然是一个具有挑战性的问题，需要领域专业知识。这也意味着，要充分利用某个数据集的无标签数据，首先需要具备良好的监督学习训练经验，以更好地理解 and 设计相应的算法。

总结起来就是如果需要充分利用未标签数据，需要对数据有足够的理解，建立起 data-centered model，并且熟悉有监督学习的整个流程、调参技巧以及优化方法等。最后根据数据特点制定特定的无标签数据的特征提取任务。