同济大学 TONGJI UNIVERSITY

同济大学《深度学习》课程实验报告

基于 UDTIRI 数据集的实例分割任务

任课老师：范睿

班级：10069801

2024 年 1 月

摘要

本实验使用了 UDTIRI 数据集，基于 openmmlab 框架，针对道路坑洞的实例分割这一现实问题，练习搭建与训练 mask-rnn 等实例分割网络，并且基于验证集上的 coco 指标进行调优，最终将测试集的分割结果以 json 文件提交到到在线评估榜，并取得了不错的效果。

关键词：实例分割，openmmlab,Mask R-CNN，mAP

目 录

第一章　实验内容简述

1.1　问题描述

实例分割（Instance Segmentation）是计算机视觉领域的一项任务，其目标是在图像或视频中识别和分割出不同的物体实例，并为每个物体实例分配一个唯一的标识符或编号与语义分割（Semantic Segmentation）不同，实例分割不仅要将图像中的像素分为不同的类别（如人、汽车、狗等），还要将同一类别中的不同物体实例进行区分。本实验旨在使用 UDTIRI 数据集练习实例分割网络的搭建与训练，目标是得到一个对道路坑洞分割效果较好的网络。

1.2　问题背景

在各种 AI 技术深入城市建设之前，城市的建设与维护一直有很多令人头疼的问题，尤其是其中城市基础设施建设与维护中，许多城市道路由于过度使用或是老化，出现了各种缺陷，十分影响通行车辆乘客的舒适度与安全性。如何及时识别并且定位各类道路缺陷，如坑洞，裂痕等，使得城市道路更加安全舒适，同时也使得城市面容更加美观，成为困扰各个城市养路部门的一大难题。

由于如今社会车辆数量的急剧增加，传统人工方式进行道路检修与维护的成本与效率都成为了其“痛点”，效率低下，成本高昂。维护工人们十分辛苦，需要投入大量的精力去找到这些缺陷并填充修复。因此，使用深度学习算法建立出能识别并定位道路缺陷的模型对于解决这一大难题至关重要。

1.3　数据集介绍

本次实验采用的数据集是 UDTIRI（An Open-Source Road Pothole Detection Benchmark Suite）数据集[1]，该数据集是同济大学 MIAS 课题组开源的一个大规模的道路坑洞数据集，包含了 1000 张道路坑洞图像，其中 600 张为训练集，300 张为测试集，100 张为测试集，支持目标检测，实例分割和语义分割等视觉任务，数据集支持 VOC 和 coco 两种格式，本次实验采用了 coco 格式的数据集进行了实例检测的任务，下面是数据集的示意图片：

图 1.1: UDTIRI dataset

1.4　实验环境

1.4.1　基本环境

本次实验在本地进行代码的编写与调试，在 autodl 网站中租用 GPU 进行训练，以下为本实验的基本环境:

项目	内容
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHz
GPU	Nvidia RTX 4090
内存	90GB
OS	Linux（ubuntu20.04）
CUDA	11.3
编程语言	Python 3.8
框架	pytorch 1.10

1.4.2　使用框架

本次实验使用 openmmlab 框架进行网络的搭建与训练。OpenMMLab 是深度学习时代最完整的计算机视觉开源算法体系。自 2018 年开源以来，累计发布超过 20 个算法库，涵盖分类、检测、分割、视频理解等众多算法，本次实验采用其中的 mmdetection 算法库并采用其中的实例分割任务。

第二章　方法描述

2.1　方法简述

本次实验经过多次的对比实验，尝试与调优后，最终采用的方式是使用 Mask R-CNN 网络（backbone 为 resnet50）以 0.02 的初始学习率，采用线性调整学习率的方式，在数据集上训练了 150 轮，并在 112 轮保存了最佳结果，然后以最佳结果为预训练权重，在进行了数据增强的数据集上重新进行了 20 轮微调训练，其中初始学习率为 1e-4，同时在第 15 轮和第 18 轮以 0.25 的倍数调整学习率，最终在第 13 轮保存了最佳权重，但是经过对比，取 20 轮作为最后的权重选择。

2.2　网络选择

本次实验主要尝试了 Mask R-CNN 和 cascade-Mask R-CNN 两种网络，而最终由于 Mask R-CNN 在测试集上表现的更好所以选择了 Mask R-CNN，下面是对两种网络的简要介绍。

2.2.1　Mask R-CNN

Mask R-CNN（Mask Region-based Convolutional Neural Network）[2] 是一种深度学习模型，用于实现目标检测和语义分割任务。它是在 Faster R-CNN 框架的基础上发展而来，通过额外的分支来生成目标的像素级掩码（Mask），从而使模型能够在像素级别上对目标进行识别和分割。Mask R-CNN 是一种端到端的神经网络架构，通过联合优化目标检测、目标定位和目标分割三个任务，能够实现在图像中检测对象并准确地为每个对象生成分割掩码。相较于其他传统的目标检测方法，Mask R-CNN 在精度和效率上取得了更好的性能。其主要特点和组成部分包括：

1. 基于区域的卷积神经网络（R-CNN）：Mask R-CNN 构建在 Faster R-CNN 的基础上，使用了类似的区域提议网络（Region Proposal Network，RPN）来生成候选区域。

2. 目标分类：在候选区域的基础上，Mask R-CNN 使用分类器对每个候选区域进行目标分类。

3. 边界框回归：除了目标分类，它还会回归每个候选区域的边界框，精确定位目标的位置。

4. 掩码生成分支：Mask R-CNN 引入了一个额外的分支，用于生成每个目标的像素级掩码，实现对象的精准语义分割。

5. 特征金字塔网络（Feature Pyramid Network，FPN）：Mask R-CNN 常常与 FPN 结合使用，以便处理不同尺度的目标，提高模型的鲁棒性。

Mask R-CNN 在计算机视觉领域被广泛应用于实例分割（instance segmentation）、物体检测（object detection）、语义分割（semantic segmentation）

等任务，因为它能够同时实现目标检测和目标分割，输出的结果既包括对象的边界框位置，也包括对象的像素级别掩码，对于图像和视频中的对象识别与分割都具有重要意义。

下面是 Mask R-CNN 的网络结构：

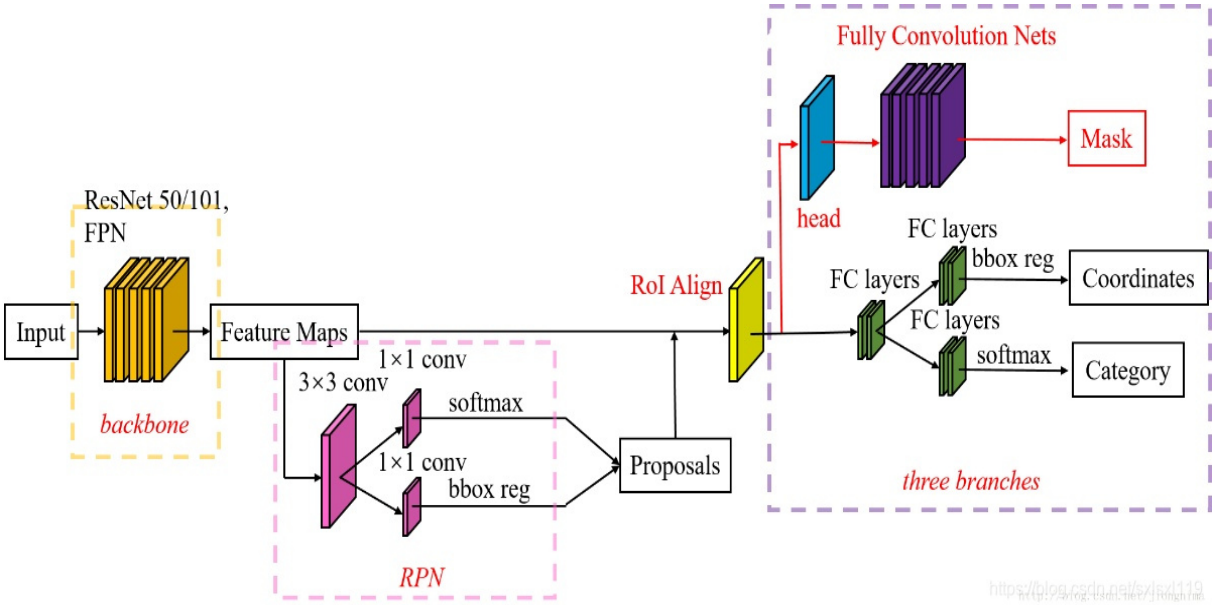

图 2.1: the structure of the Mask R- CNN

2.2.2　cascade Mask R-CNN

Cascade Mask R-CNN 是 Mask R-CNN 的一个改进版本[3]，主要通过级联学习的方式进一步提升实例分割任务的性能。Cascade Mask R-CNN 引入了级联结构，通过级联多个 Mask R-CNN 模型，使得每个级联阶段都能够逐渐提高模型对难以分割目标的性能。具体来说，Cascade Mask R-CNN 的主要特点包括：

1. 级联结构：在 Cascade Mask R-CNN 中，模型被分为多个级联阶段。每个阶段都包含一个目标检测模块和一个目标分割模块。这样，每个级联阶段都有机会在前一阶段的基础上进行学习和优化，以提高难以分割目标的性能。

2. 多阶段的训练：Cascade Mask R-CNN 需要进行多阶段的训练。在每个阶段，模型都会接收前一阶段的输出结果作为输入，然后进行目标检测和目标分割任务。这种逐步学习的方式有助于提升模型对于不同难度目标的适应性。

3.Hard Negative Mining（难负样本挖掘）：在级联学习的过程中，Cascade Mask R-CNN 引入了 Hard Negative Mining 的策略，即对那些在前一阶段难以分割的目标进行更加重视的学习，以便提高对难分割目标的处理能力。

Cascade Mask R-CNN 的这些特点使得它在应对复杂场景和难以分割目标时相对于传统的 Mask R-CNN 有一些性能上的提升。

2.3　backbone 选择

Mask R-CNN 主要支持 resnet50 和 resnet101 这两种 backbone，它们主要的差异是卷积层的数量不同，在本次实验中，resenet101 的表现没有 resnet50 好，个人分析认为是由于该实例分割任务只需要检测坑洞一种 label，相对来说任务比较简单，所以使用较为浅层的 backbone 反而表现好一些。

2.4　数据增强

由于本次实验采用的数据集只有 600 张，个人认为更大的数据量可能会取得更好的效果，同时由于现实生活中道路坑洞出现的情况多种多样，所以对数据集采用随机翻转，光线变化，更改大小等方式有利于提高模型的泛化性。

本人选择了 mmlab 中集成的数据增强方式，选择了以 0.5 的概率进行随机翻转（两个阶段都用了），以及 mmlab 中默认的 AutoAugment 数据增强方式（包含缩放等变换），发现在预训练阶段使用数据增强效果反而发生了下降，而在微调阶段有所提升。

分析原因如下：经过查看训练日志发现在采用了 AutoAugment 后，在部分轮数出现了 loss 为 nan 的情况，经过查询资料后发现，原因可能是由于有些实例过于小，而数据增强中的缩放等操作可能使得其更小，甚至缩为一个点的大小，这样可能会导致计算损失时候出现问题，从而导致影响算法的精度。微调阶段训练轮数较多，遇到 nan 的情况更多可能就会导致负面影响较大，而微调阶段轮数少所以影响较小。

2.5　微调方法

本次实验采取了先以较大学习率进行预训练，然后以小学习率进行微调的方法，现以 0.02 的初始学习率，以线性更新学习率的方式训练了 150 轮；然后以预训练阶段的 best 权重作为预训练模型，继续进行了 20 轮以 1e-4 为初始学习率的微调。

由于选择的保存最佳模型权重的方式是 auto，也就是自动按照 coco 指标的第一项指标进行取优，微调阶段最佳权重保存在了第 12 轮，但是经过观察 log 之后发现第 20 轮其他指标更为优秀（且第一项指标仅仅和第 12 轮差 0.001），尤其是在 map-s 上达到了 0.4 以上，所以个人认为它有可能在测试集上取得更好的效果。

最终在提交到测试榜之后发现在测试集上 coco 各项指标都有所提升，比没有微调提升提升 0.09 左右，且 20 轮的权重比 12 轮的权重提升 0.06 左右，最终选择了第二十轮的权重作为了最佳权重。

第三章　实验结果

3.1　评估指标

本次实验的评估方式是将测试集上推理出来的就送文件提交到在线评估榜单上，榜单上的评估指标是常见的 coco 评估指标，主要包括 map，map50,map75,map-s,map-m 和 map-l，下面是对于该指标的简单说明：

Average Precision（AP）：在精确率-召回率曲线下的面积，用于度量模型在单个类别上的性能。计算方式为对精确率-召回率曲线下的面积进行积分。AP 的计算可以通过不同方法，例如 11-point interpolation 或更精确的积分方法。

mAP（Mean Average Precision）：通过计算所有类别的 AP 并取平均值，得到一个综合评估模型性能的指标。mAP 是目标检测或实例分割任务中常用的评价指标之一，因为它能够全面考虑多个类别的性能表现。而在 coco 指标中 map 一般指 map50-95 即在不同置信度阈值下的平均 map，而 map-s,map-m 和 map-l 指的是模型对于小、中、大模型的检测或分割效果。

3.2　评估结果

下面展示自己在测试集上最好的三次结果：

指标 / 方法	得分	mAP	mAP50	mAP75	mAP-s	mAP-m	mAP-l
仅一阶段训练	0.44498	0.483	0.782	0.526	0.185	0.357	0.539
预训练 + 微调（第 12 轮）	0.44742	0.492	0.809	0.527	0.168	0.359	0.552
预训练 + 微调（第 20 轮）	0.45334	0.492	0.816	0.524	0.184	0.360	0.552

第四章　实验总结

4.1　不足与展望

我认为，本次实验最大的问题在于在小目标上的实例分割效果比较差，迁移泛化能力不足，尽管在验证集上 map-s 可以达到 0.4 以上，但是在测试集上仅有 0.185。这表明模型在捕捉较小的道路坑洞方面能力不足。应该更多地关注这方面的内容，目前考虑到的解决方式有去调整训练时候设置的各类阈值，以保证小型坑洞获得更多的关注；选择对于小型目标分割效果好的网络或者 backbone。此外，本次实验尝试的网络结构比较少，只尝试了 Mask R-CNN 和 cascade-Mask R-CNN 两种算法，应该多尝试其他的一些实例分割网络；近年来 yolo 在实例分割上也表现出色，但是由于 mmlab 对于 yolov8 的支持不是很好，再加上涉及到数据集格式转换的问题，所以本次实验没有尝试 yolov8-seg，未来应该尝试对 coco 数据集进行转换，并且基于 ultralytics 框架进行 yolov8 的算法尝试。除了基于现有的实例分割网络，未来可以尝试修改一些现有实例分割网络或者是自己手动搭建一些实例分割网络。

4.2　实验总结

本次实验主要采用 Mask R-CNN 对道路坑洞进行了实例分割的任务，下面是对实验的总结和个人心得。

实验使用了 Mask R-CNN，这是一种出色的深度学习模型，在道路坑洞的像素级实例分割任务中表现出了良好的性能。通过该模型，成功地对道路图像中的坑洞进行了准确的标注和分割。

模型训练的关键在于数据集的质量和数量。确保数据集中包含多样性、标注准确的道路坑洞样本，有助于提高模型的泛化能力。模型训练过程中需要反复调整参数、学习率和数据增强策略，确保模型能够收敛到最佳状态。

对应一些任务来说，现以较大的学习率进行多轮数的预训练，然后以较低的频率进行小轮数的微调是行之有效的方法。

在实验过程中，评估指标的选择对于理解模型性能至关重要。mAP 是常用的指标之一，但还应考虑其他指标，如 IoU，以全面评估模型的准确性和鲁棒性。

道路坑洞的实例分割对于交通管理和道路维护至关重要，因此，模型的准确性和实用性对于实际应用具有重要意义。

参考文献

[1] S Guo, J Li, S Su, et al. UDTIRI: An Open-Source Road Pothole Detection Benchmark Suite[J]. ArXiv preprint arXiv:2304.08842, 2023.

[2] K He, G Gkioxari, P Dollár, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. [S.l.]: [s.n.], 2017: 2961–2969.

[3] Z Cai, N Vasconcelos. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.]: [s.n.], 2018: 6154–6162.