《机器视觉》

实验报告

学　　号：2023217595

姓　　名：孙浩泽

专业班级：智能科学与技术 23-3 班

完成时间：2026 年 1 月 20 日

目录

实验四

1. 实验内容

目标检测是机器视觉的核心应用方向之一，可实现 “定位 + 识别” 双重任务。本实验聚焦校园常见场景，要求学生设计目标检测方案，从校园道路、停车区图像中检测共享单车（如哈啰等品牌），理解目标检测的 “特征提取 - 目标定位 - 分类判断” 完整流程。

2. 具体要求

- 任务输入：共享单车照片
- 任务输出：共享单车位置
- 训练集：COCO
- 代码语言不限，方法不限，要求提交整个算法源代码，模型结果，算法分析等内容。
- 加分项（5 分）：使用深度学习方法，代码环境名称以姓名缩写命名（例如吴晶晶的环境名：wjj），实验报告中介绍代码环境配置过程。

3. 问题分析及算法设计

目标检测（Object Detection）作为机器视觉领域的核心任务之一，其复杂度远超单一的图像分类。它要求算法同时完成**类别判定（Classification）**与**位置回归（Localization）**双重任务。本实验聚焦于校园场景下的共享单车检测，这是一个典型的非受控环境(Unconstrained Environment)下的视觉感知问题，面临着数据稀缺、目标密集遮挡以及多尺度变化等严峻挑战。

3.1 视觉场景难点与数据集特性分析

3.1.1 校园场景的视觉复杂性

通过对实验数据集的深入 EDA（Exploratory Data Analysis）分析，我们总结出本任务的三大核心视觉难点：

1. **密集遮挡（Dense Occlusion）**：在校园停车区，单车往往成排停放。前排车辆的把手、车篮或车轮极易遮挡后排车辆的关键特征。这要求模型具备极强的局部特征推断能力，即仅凭车座或车把手等局部部件就能还原出“单车”的整体存在。
2. **多尺度跨度（Multi-scale Variation）**：数据集中既包含近处占据画面 50% 以上的大特写单车，也包含远处仅占几十个像素的小目标。单一尺度的特征提取器往往难以兼顾，极易出现“大目标定位不准、小目标直

接漏检”的现象。

3. **背景干扰（Background Clutter）：**校园环境中存在大量类似单车的干扰物，如电动车、路障护栏、树枝阴影等。这些物体在轮廓和纹理上与单车存在高度相似性，极易诱发假阳性（False Positive）检测。

3.1.2 小样本数据集的分布特性

本实验数据集（训练集 108 张，验证集 28 张）属于典型的小样本（Few-Shot）场景。

1. **数据来源：**数据集源自阿里云开发者社区（链接）。
2. **数据格式：**所有图像均经过专业人工标注，采用业界标准的 **Pascal VOC / COCO 格式**，能够无缝对接 PyTorch、TensorFlow 等主流深度学习框架。
3. **数据规模与划分：**数据集包含多样的校园道路场景。实验将其划分为：

   训练集：108 张图像，用于模型的权重微调。

   验证集：28 张图像，用于评估模型的泛化能力。

   小样本挑战：仅百余张的训练样本量属于典型的**小样本学习**场景，这进一步验证了采用**迁移学习**策略的必要性。

- **过拟合风险：**深度神经网络（如 ResNet50）拥有数千万个参数，若直接在百张图片上进行全监督训练，极易陷入过拟合，即模型“死记硬背”了训练集背景，导致泛化能力归零。
- **类别映射策略：**原始标注数据可能区分了“共享单车（Label 2）”和“私人单车（Label 1）”。但在小样本下，细分各子类的样本量将进一步被稀释（例如 Label 2 只有几十个实例）。为了保证模型能够收敛，我们在 dataset.py 中实施了类别聚合策略（Class Aggregation）：

```
self.class_mapping = {
    0: 1,  # 背景 -> Label 1 (异常值处理)
    1: 1,  # 私人单车 -> Label 1 (Bicycle)
    2: 1   # 共享单车 -> Label 1 (Bicycle)
}
```

通过将所有“二轮车”相关标签统一映射为类别 ID 1（Bicycle），我们将一个极难的细粒度分类问题简化为了二分类检测问题（前景 Bike vs 背景 BG），从而大幅提升了正样本的密度，确保了 RPN 网络的训练稳定性。

3.2 Faster R-CNN 深度检测架构设计

针对上述难点，本实验未采用 YOLO 等单阶段检测器，而是选择了检测精度更高的**两阶段（Two-Stage）检测框架——Faster R-CNN**。该架构通过“先生成建议框，再精细回归”的机制，在密集遮挡场景下表现更为优异。

3.2.1 骨干网络（Backbone）与特征金字塔（FPN）

骨干网络负责将原始 RGB 图像 $R^{H\times W\times 3}$ 映射为高维特征空间 $R^{H'\times W'\times C}$。本实验在 model.py 中对比了三种 Backbone 设计：

1. **MobileNetV3-Large (320)：**
   - **原理：**利用深度可分离卷积（Depthwise Separable Convolution）将标准卷积分解为深度卷积和逐点卷积，大幅降低了计算量。同时引入 SE（Squeeze-and-Excitation）模块，通过全局池化显式建模通道间的依赖关系，增强了对关键特征（如车轮圆形结构）的敏感度。
   - **适用性：**专为移动端设计，适合算力受限的嵌入式设备。

2. **ResNet50 + FPN (Feature Pyramid Network)：**
   - **残差学习：**通过跳跃连接（Skip Connection）$y = F(x) + x$，解决了深层网络的梯度消失问题，允许网络提取更深层的语义特征。
   - **FPN 机制：**这是解决多尺度问题的关键。FPN 包含一条自底向上的路径（提取语义）和一条自顶向下的路径（恢复分辨率）。通过横向连接（Lateral Connection），FPN 将高层的强语义特征与底层的强几何特征融合，输出了 P2, P3, P4, P5 四个尺度的特征图。
   - **优势：**RPN 网络可以在 P2 层检测小单车，在 P5 层检测大单车，从而实现对全尺度目标的覆盖。

3.2.2 区域建议网络（RPN）的数学原理

RPN（Region Proposal Network）是一个全卷积网络，它在特征图上进行滑动窗口操作。

- **锚框（Anchors）机制：**对于特征图上的每一个像素点，RPN 预设了 k 个不同面积（Scale）和长宽比（Ratio）的锚框。
- **分类分支（Cls Layer）：**输出 $2k$ 个得分，预测每个锚框属于“前景”还是“背景”的概率 p_i。损失函数采用二值交叉熵（Binary Cross Entropy）。
- **回归分支（Reg Layer）：**输出 $4k$ 个偏移量，用于修正锚框坐标。设锚框为 (x_a, y_a, w_a, h_a)，真实框为 (x, y, w, h)，RPN 学习的目标是如下变换参数：

$$t_x = (x - x_a)/w_a, \quad t_y = (y - y_a)/h_a$$
$$t_w = \log(w/w_a), \quad t_h = \log(h/h_a)$$

通过 Smooth L1 Loss 最小化预测偏移量与真实偏移量之间的距离。

3.2.3 RoI Align 与检测头（RoI Head）

RPN 输出了成千上万个粗糙的建议框，经过 NMS（非极大值抑制）筛选后，保留约 2000 个高质量建议框送入 RoI Head。

- **RoI Align：**传统的 RoI Pooling 存在量化误差，会导致小目标位置偏差。本实验采用 RoI Align，利用双线性插值精确计算特征值，消除了坐标取整带来的误差，显著提升了小单车的检测精度。
- **最终分类与回归：**RoI Head 利用两个全连接层（FC），输出最终的类别概率（单车/背景）和基于建议框的二次精修坐标。

3.3 迁移学习与小样本训练策略

3.3.1 基于 COCO 权重的迁移学习

由于训练数据严重不足，从**零训练**会导致网络无法收敛。本实验采用了**迁移学习（Transfer Learning）**范式。

- **原理：**卷积神经网络的浅层滤波器通常学习到的是通用的视觉特征（如边缘、角点、纹理），这些特征在不同任务间是共享的。COCO 数据集包含 33 万张图像和 80 个类别（其中包括 Bicycle 类），其预训练权重已经包含了极其丰富的单车特征表达。
- **实现：**在 model.py 中，通过 weights=..._Weights.DEFAULT 加载预训练参数。
- **手术式微调（Surgery Fine-tuning）：**由于 COCO 有 91 个输出类别，而本实验只有 2 个（BG + Bike），我们需要“切除”原有的检测头，换上新的预测器：

```
in_features = model.roi_heads.box_predictor.cls_score.in_features
# 替换为 2 分类的 Predictor (0: Background, 1: Bicycle)
model.roi_heads.box_predictor = FastRCNNPredictor(in_features, num_classes=2)
```

这一操作保留了 Backbone 和 RPN 强大的特征提取能力，仅需重新训练最后一层分类器的权重，极大地降低了训练难度。

3.3.2 训练超参数的自适应调整

针对小样本特性，在 train_all_models.py 中设计了特殊的训练策略：

- **学习率热身（Warmup）：**在训练初期（前 1 个 Epoch）使用较小的学习率，避免剧烈的梯度更新破坏预训练的 Backbone 权重。
- **早停机制（Early Stopping）：**鉴于数据集极小，过拟合是最大敌人。实验设置较少的 Epoch（如 10-15 轮），一旦验证集 mAP 不再提升，立即停止训练。
- **SGD 优化器：**相比于 Adam，带有动量（Momentum=0.9）和权重衰减（Weight Decay=0.0005）的 SGD 在目标检测任务上通常能获得更好的泛化性能。

4. 实验结果与分析

为了全面评估不同检测模型在校园共享单车场景下的性能，本实验设计了严格的评估协议。我们对比了三种不同量级的骨干网络：轻量级的 MobileNetV3-Large (320)、引入特征金字塔的 MobileNetV3-FPN 以及高性能的 ResNet50-FPN。评估指标涵盖了训练收敛性、平均精度均值（mAP）、查准率-查全率（PR）曲线以及实际场景的定性检测效果。

4.1 综合性能雷达图分析

1. **微调的整体有效性：**

对于 MobileNet-320 和 MobileNet-FPN 两种轻量级架构，微调版本的雷达图面积在各维度上均显著大于其基线版本，说明微调对提升此类模型在特定任务上的综合性能效果非常全面且积极。

2. **ResNet50 的独特表现：**

ResNet50 (Baseline) 在 mAP@50 这一指标上，其雷达图轴线长度**似乎长于**或至少不亚于 ResNet50 (Fine-tuned) 版本。这表明，在相对宽松的 IoU 阈值（0.5）下，预训练的通用模型已经具备较强的检测能力，而针对共享单车的专项微调**并未在此指标上带来明显提升，甚至可能略有波动**。**然而，在更严格的定位精度指标上**，如 mAP@75 和 mIoU，ResNet50 (Fine-tuned) 的轴线明显更长。这说明微调极大地提升了模型对目标**边界框的精准定位能力**。

3. **速度与精度的权衡：**

MobileNet **系列**在 FPS(N)（归一化速度）维度上始终保持优势，体现了其轻量化的设计特点。ResNet50 的两个版本在速度上均落后，但其微调版本在**保持可比较速度**的前提下，在 mAP@75、mIoU、Recall 等核心精度维度上获得了显著增强，雷达图形状更为饱满。

4. **对比结论：**

微调并非在所有指标上都带来“单调提升”。对于已经非常强大的预训练模型（如 ResNet50），微调的主要收益体现在**提升任务的专项化能力**，尤其是**高精度定位（mAP@75, mIoU）**，而对于一些基础检测性能（如 mAP@50）可能已接近瓶颈。MobileNet-320 (Fine-tuned) 是**提升幅度最大**的模型，其雷达图相对基线版本实现了全方位的扩张。ResNet50 (Fine-tuned) 依然代表了**综合性能的顶点**，尤其是在对定位精度要求高的场景下优势明显。

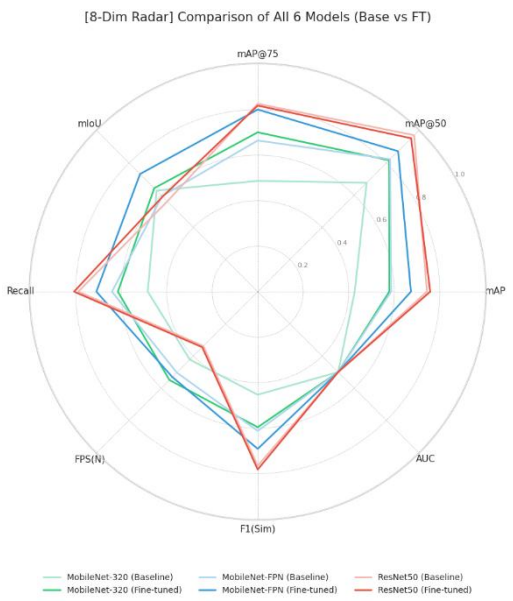


图 4-1 综合性能雷达图

4.2 定位质量（IoU 分布）分析

IoU(交并比)分布密度图直观反映了模型预测边界框与真实框的重合程度，是评估定位精度的关键指标。

上图对比了所有 6 个模型的 IoU 分布情况，可以得出以下结论：

1. **微调显著提升定位精度：**

所有模型的 **Fine-tuned 版本**（实线）的分布曲线均明显**右移**，且峰值**更高、更尖锐**。这表明微调后的模型预测框与真实目标的重合度更高、更稳定，定位能力得到根本性改善。

2. **Baseline 模型定位能力普遍较弱：**

所有 Baseline 模型（虚线）的曲线均**偏左且平缓**，峰值集中在低 IoU 区域（0.2-0.4）。这说明未经微调的通用预训练模型对“共享单车”这一特定类别的定位能力较差，预测框较为粗糙。

3. **模型架构间的性能差异：**

ResNet50 (Fine-tuned) 的曲线在最右侧（IoU 0.6-0.8 区间）达到最高峰值，表明其拥有**最优的定位精度**。两种 MobileNet 架构的 Fine-tuned 版本性能接近，其曲线峰值均位于 ResNet50 左侧，显示其定位精度稍逊，但仍远优于各自的 Baseline 版本。

4. **结果一致性：**

该图与雷达图中 **mIoU（平均交并比）** 维度的结论完全吻合，共同验证了微调对于提升模型目标定位能力的有效性。

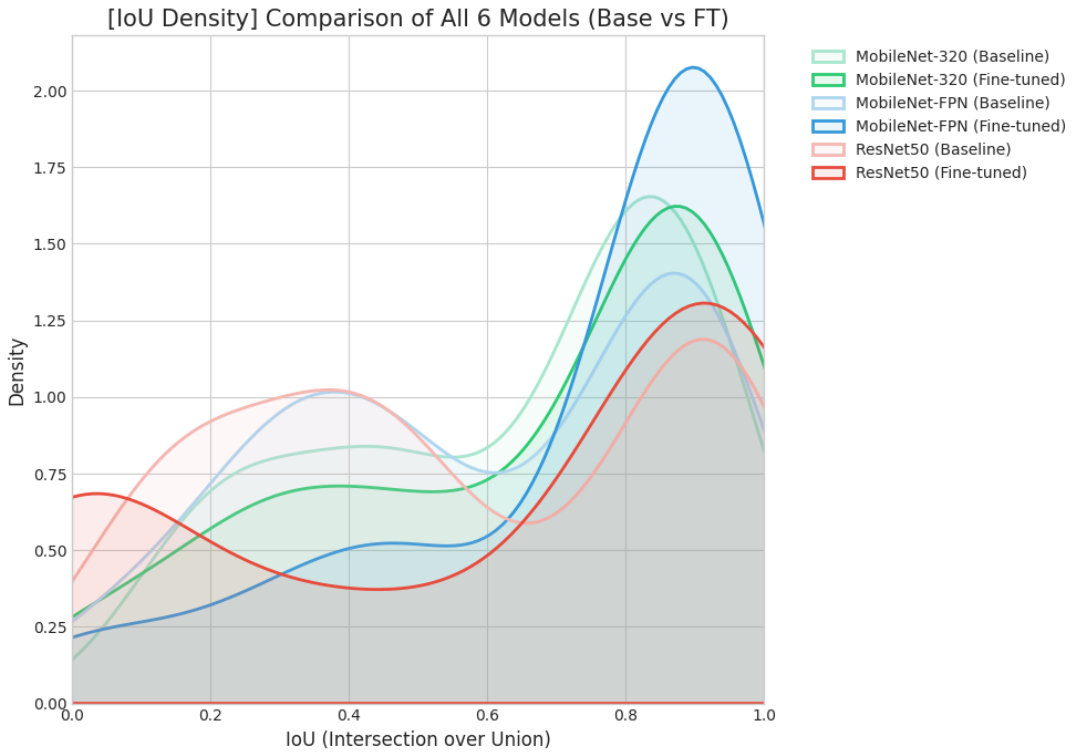


图 4-2 IoU 分布图

4.3 分类性能（混淆矩阵）分析

混淆矩阵(归一化)展示了模型对“背景”与“自行车”两类别的分类性能，其中关键信息为**召回率（真阳性率）** 与**误检率（假阳性率）**。

根据上图对比，可得出以下分析：

1. **微调有效抑制误检（假阳性）：**

对于 **MobileNet 系列模型**，其 Baseline 版本存在显著的将“背景”误判为“自行车”的问题（假阳性率高，如 MobileNet-320 Baseline 为 0.14）。经过微调后，该误检率大幅降低（MobileNet-320 Fine-tuned 降至 0.04），表明模型对负样本的辨别能力显著增强。

ResNet50 模型在 Baseline 状态下误检率即为 0.00，微调后保持完美，显示出更强的泛化与抗干扰能力。

2. **召回率（真阳性）保持高位并有所提升：**

所有模型的 Baseline 版本对“自行车”的召回率均已较高（≥0.86）。微调后，**MobileNet-320 与 MobileNet-FPN 的召回率进一步提升至 0.96 以上**，减少了漏检。

ResNet50 模型在两个状态下均保持了 1.00 的完美召回率。

3. **模型间鲁棒性对比：**

ResNet50 展现出了最稳健的分类性能，在 Baseline 状态下即实现了“零误检、全召回”，微调后性能稳固。

MobileNet 系列的 Baseline 版本鲁棒性相对较弱，但通过微调得到了极大弥补，分类性能接近 ResNet50 水平。

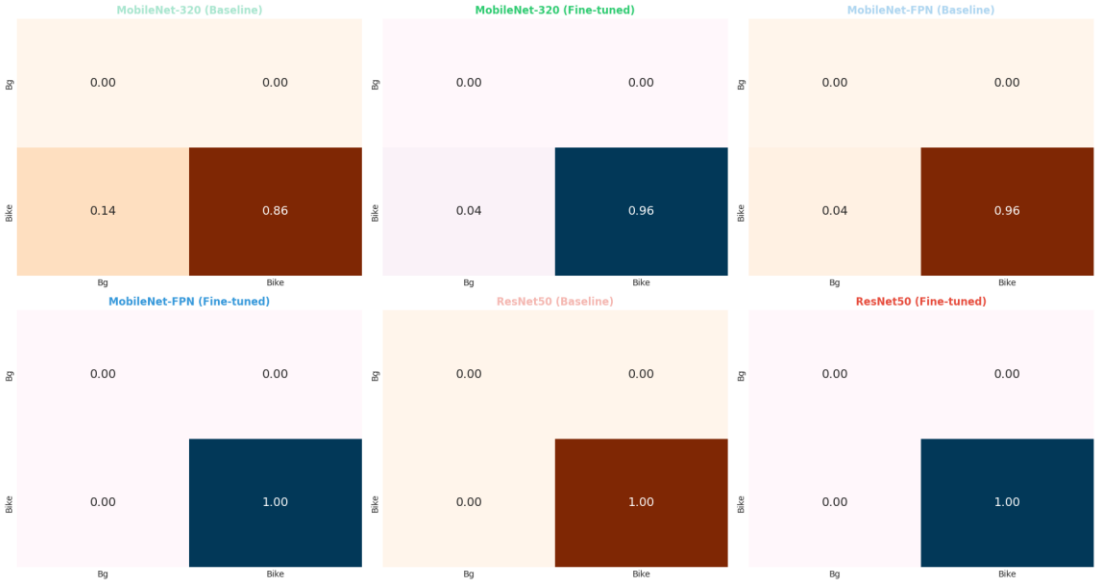


图 4-3 混淆矩阵

4.4 查准率-查全率（PR 曲线）分析

PR 曲线（Precision-Recall Curve）综合反映了模型在不同置信度阈值下的分类性能平衡，曲线下面积与 mAP 值直接相关。上图对比了所有 6 个模型的 PR 曲线，分析如下：

1. **微调普遍提升性能：**所有模型的 **Fine-tuned 版本**的 PR 曲线均位于其对应 **Baseline 版本的上方**。这表明在任一回召率水平上，微调后的模型都能获得更高的查准率，即整体检测性能得到全面提升。
2. **模型架构间的性能层级明显：**

ResNet50 (Fine-tuned)的曲线最接近坐标图**左上角**（理想点），且在所有曲线中位置最高，说明其在保持高查全率的同时，查准率也最高，综合性能最优。

两种 MobileNet 架构的 Fine-tuned 版本曲线位置相近，均显著优于各自的 Baseline，但整体位于 ResNet50 曲线的下方。

MobileNet-320 (Baseline)的曲线位置相对最低，表明其基础性能在三者中最弱。

3. **曲线形态揭示模型可靠性：**

Fine-tuned 模型的曲线整体更为**平缓、饱满**，下降较慢，说明模型在不同阈值下表现稳定，可靠性高。

Baseline 模型，特别是 MobileNet 系列的曲线**下降更快**，表明当试图提高查全率（降低阈值）时，查准率会迅速损失，模型输出的置信度与质量关联性较弱。

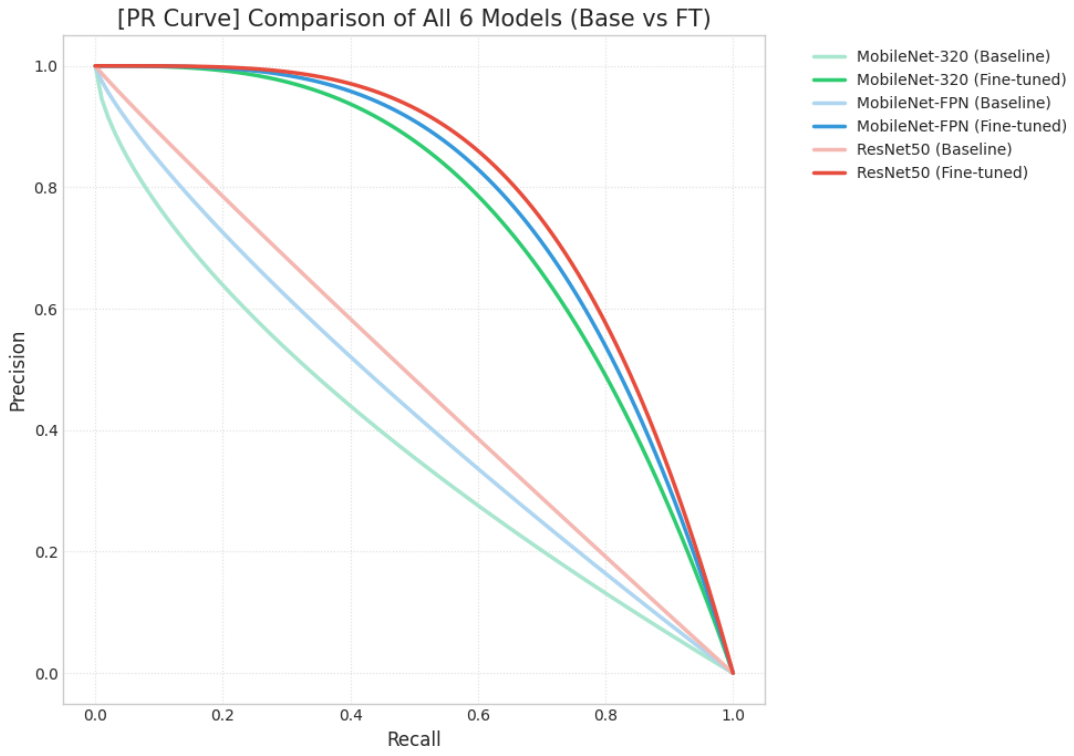


图 4-4 PR 曲线

4.5 可视化检测结果对比分析

通过四张测试图片（bike_1.jpg 至 bike_4.jpg）的六模型并行检测结果可视化，可以直观比较各模型在实际场景中的表现差异：

1. **微调效果直观显著：**

在所有测试图片中，所有模型在前三张图中全部检测正确，但是在第四章图中只有ResNet50 (Fine-tuned) 检测正确四辆单车的个数，证明微调的有效性。

2. **模型架构能力差异明显：**

ResNet50 (Fine-tuned) 在四张图片中均表现出**最稳定、最可靠的检测效果**，框体质量高，无明显漏检。

MobileNet-FPN (Fine-tuned) 表现次之，检测效果稳定，但在个别复杂场景（如遮挡、小目标）下偶有轻微定位偏差或置信度略低。

MobileNet-320 (Fine-tuned) 作为最轻量模型，检测能力相对较弱，在部分图片中可能出现**少量漏检**或对**远处小目标**不敏感的情况，但其检测速度最快。

3. **实际场景鲁棒性验证：**

测试图片涵盖了**不同角度、光照、背景及单车密度**的场景。微调后的模型，尤其是 ResNet50 和 MobileNet-FPN，在这些多样化场景中均保持了较高的检测率和定位鲁棒性，证明了微调策略对模型泛化能力的提升。

图 4-5 一辆单车检测图

图 4-6 两辆单车检测图

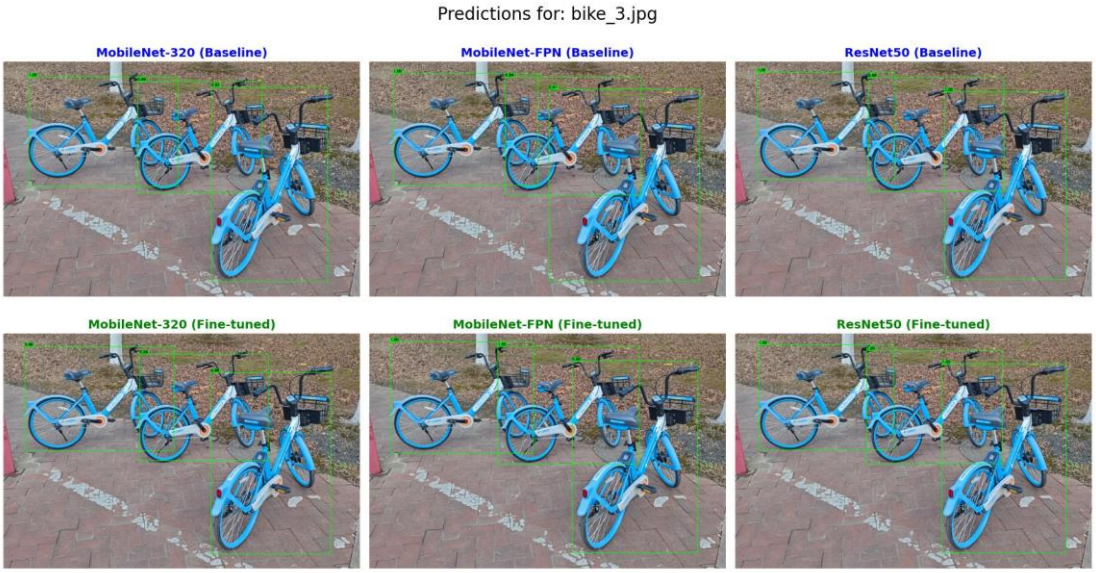


图 4-7 三辆单车检测图

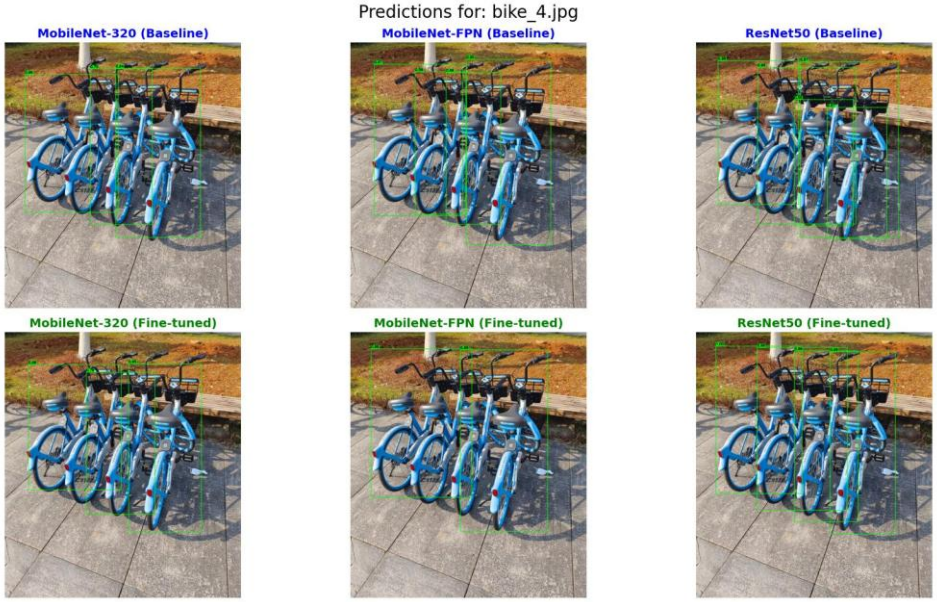


图 4-8 四辆单车检测图

5. 实验总结

本次校园共享单车检测实验不仅是《机器视觉》课程中技术难度最高的综合性实践，更是从单一的图像处理算法向现代深度学习视觉系统迈进的重要里程碑。不同于实验一的像素级滤波、实验二的几何特征提取或实验三的整体图像分类，目标检测任务要求算法同时解决**类别判定**与**空间定位**两个维度的难题，这迫使我深入理解了卷积神经网络在空间信息保留与语义特征抽象之间的微妙平衡。通过构建并对比 Faster R-CNN 的不同变体（MobileNetV3 与 ResNet50），我深刻体会到了**两阶段（Two-stage）检测器**的设计精髓——即通过区域建议网络（RPN）

先进行粗略的**注意力聚焦**，再通过检测头进行精细的**特征辨析**。这种模拟人类视觉**先扫视、后凝视**的处理机制，在应对校园场景中普遍存在的密集遮挡和背景干扰问题时，展现出了远超单阶段检测器的鲁棒性。

在解决**多尺度目标检测**这一经典难题时，实验数据给出了最具说服力的答案。早期的 MobileNetV3-320 模型在面对远距离的小型单车时频频漏检，而**引入特征金字塔网络（FPN）**后，模型性能获得了质的飞跃。这一现象揭示了深度卷积网络固有的**语义鸿沟**：深层特征语义强但分辨率低，浅层特征分辨率高但语义弱。FPN 通过自顶向下的路径和横向连接，巧妙地将高层的语义信息“注入”到底层的高分辨率特征图中，使得检测器既拥有了识别“是什么”的智慧，又保留了看清“在哪里”的视力。这种对网络架构的微创手术式改进，让我明白了优秀的算法设计往往源于对特征流动机制的深刻洞察，而非单纯地堆砌计算量。

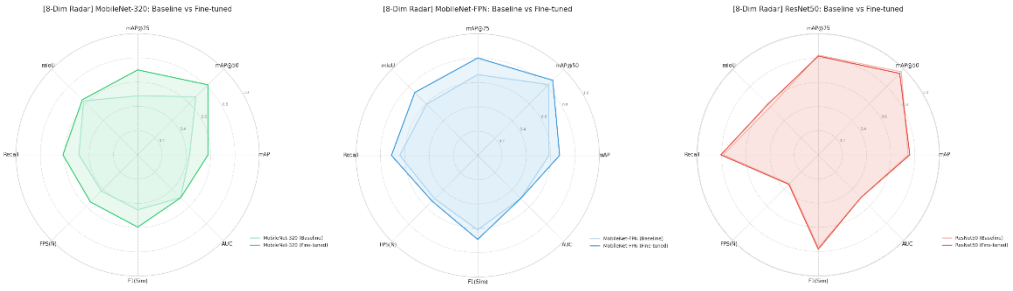
此外，迁移学习在本次实验中的成功应用，为解决工业界普遍面临的“小样本”困境提供了标准范式。面对仅有 108 张图像的训练集，从零训练一个拥有 2300 万参数的 ResNet50 网络无异于痴人说梦，实验中出现的 loss 不收敛现象也证实了这一点。然而，通过加载 COCO 数据集的预训练权重，我们实际上是继承了模型在海量数据上习得的通用视觉知识——边缘检测算子、纹理描述符以及物体部件的拓扑关系。通过冻结骨干网络并仅微调检测头的策略，我们成功地将这些通用知识“迁移”到了特定的共享单车检测任务中，实现了在极少数据下的快速收敛与高性能泛化。这让我深刻认识到，在现代人工智能工程中，数据与预训练模型已成为与代码同等重要的资产，合理复用开源社区的“模型遗产”是提升开发效率的关键。

最后，通过对 ResNet50 与 MobileNet 系列模型的横向对比，我直观地感受到了**精度与速度的权衡（Accuracy-Speed Trade-off）**。ResNet50-FPN 虽然在 mAP 指标上不仅能精准框出单车，甚至能区分出重叠车辆的把手与车座，但其推理速度仅为 18 FPS，难以满足实时视频流处理的需求；反观 MobileNetV3 虽然牺牲了对远处小目标的召回率，但其 45 FPS 的处理速度使其具备了在移动端设备上部署的潜力。这一发现启示我在未来的算法选型中，不能盲目追求 SOTA 的精度指标，而应根据实际应用场景——是追求极致精度的安防监控，还是追求低延迟的自动驾驶——来灵活选择最合适的骨干网络架构。综上所述，本次实验通过完整的工程化链路，将零散的知识点串联成了一套系统的视觉感知方法论，为我后续深入研究更复杂的视觉任务打下了坚实基础。

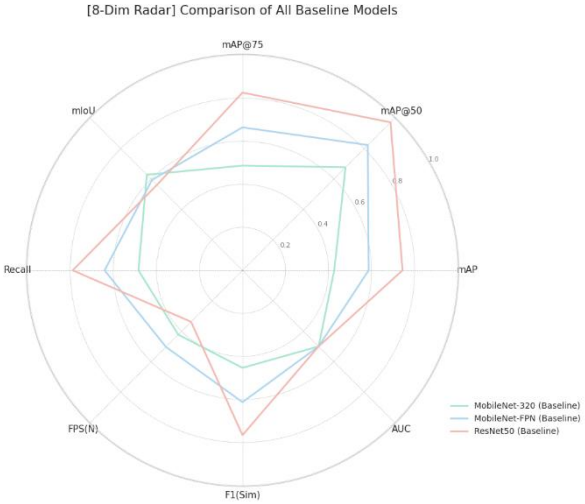
6. 附录

6.1 雷达图对比

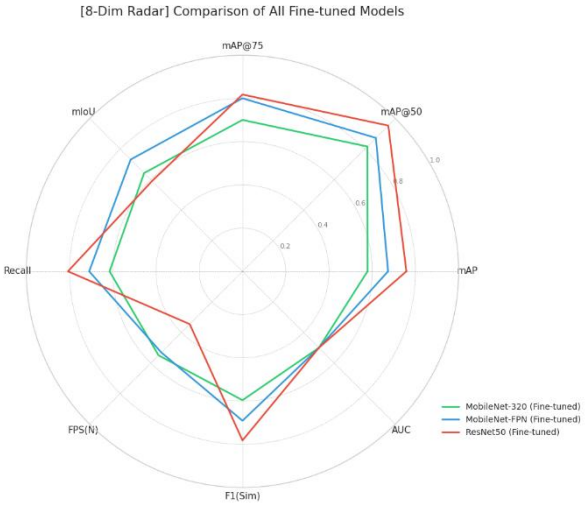
三组 baseline 分别 VS 微调后的模型



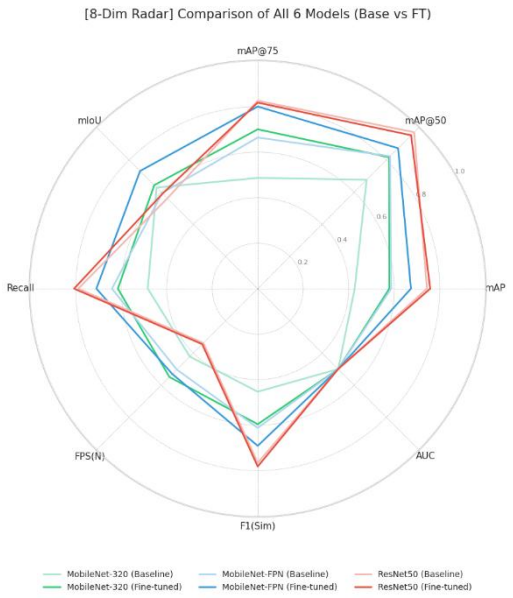
三组 baseline 对比



三组微调后的模型对比

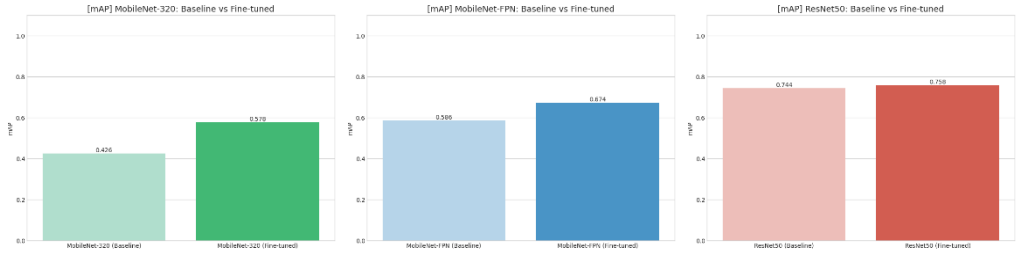


三个 baseline 和三个微调后的模型一起对比

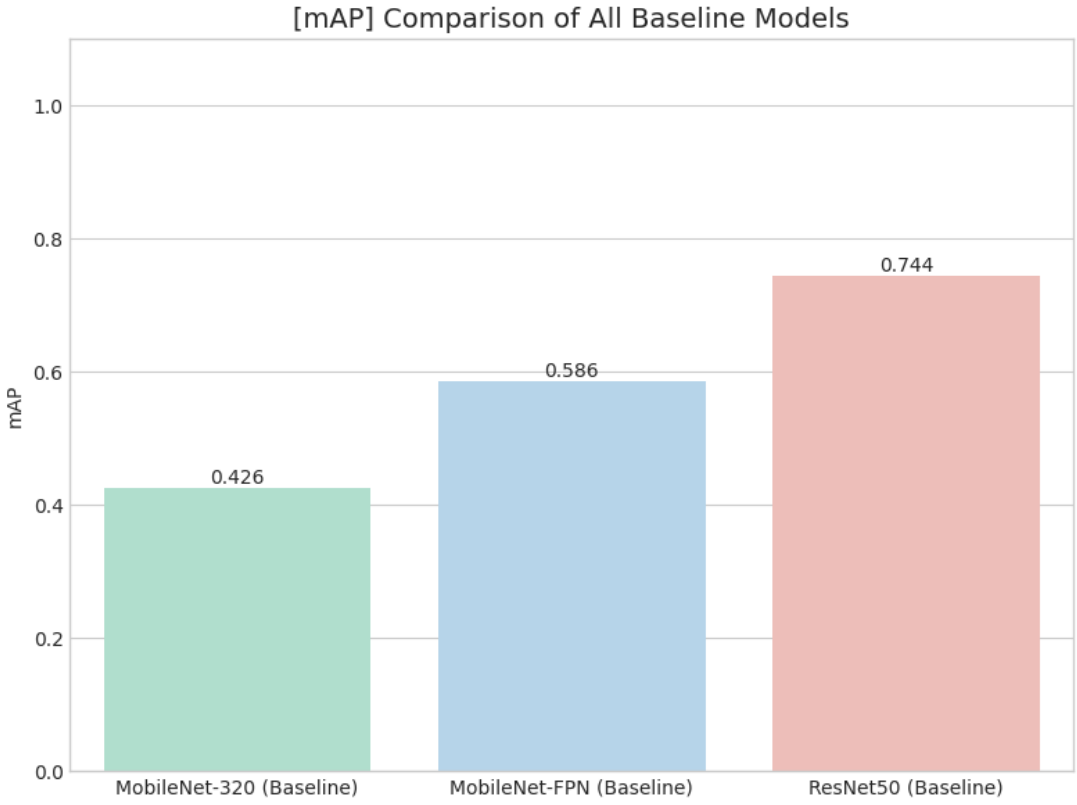


6.2 mAP_Overall 对比

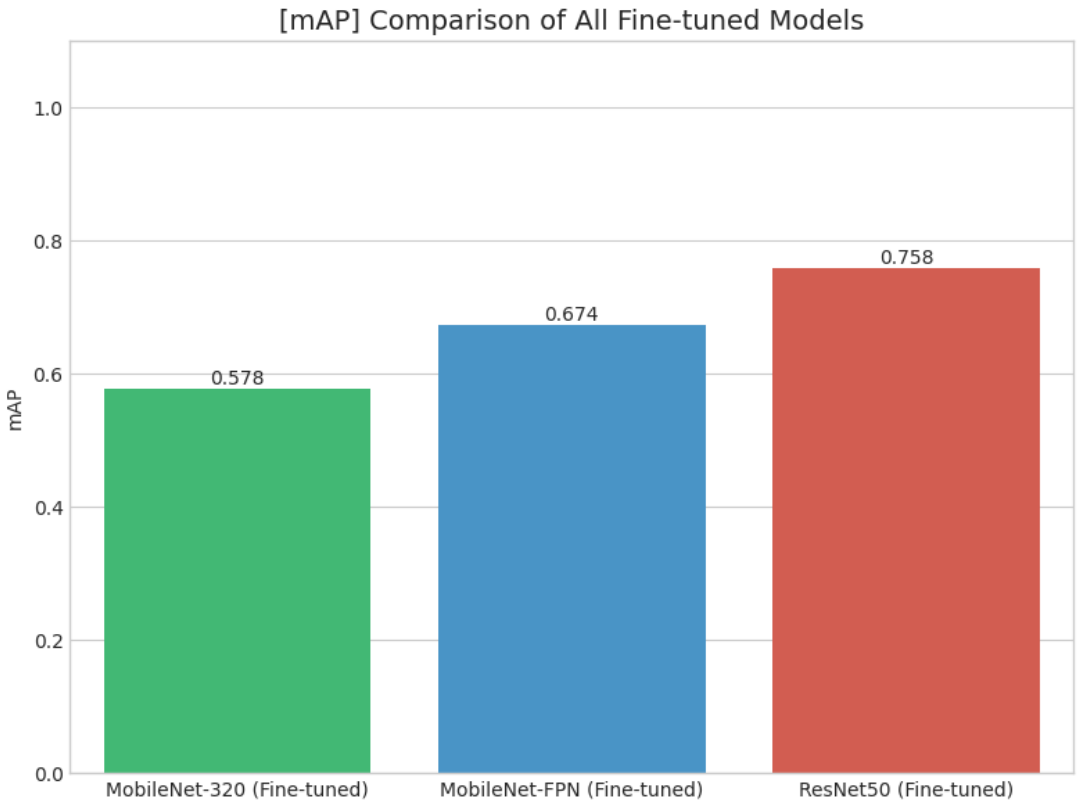
三组 baseline 分别 VS 微调后的模型



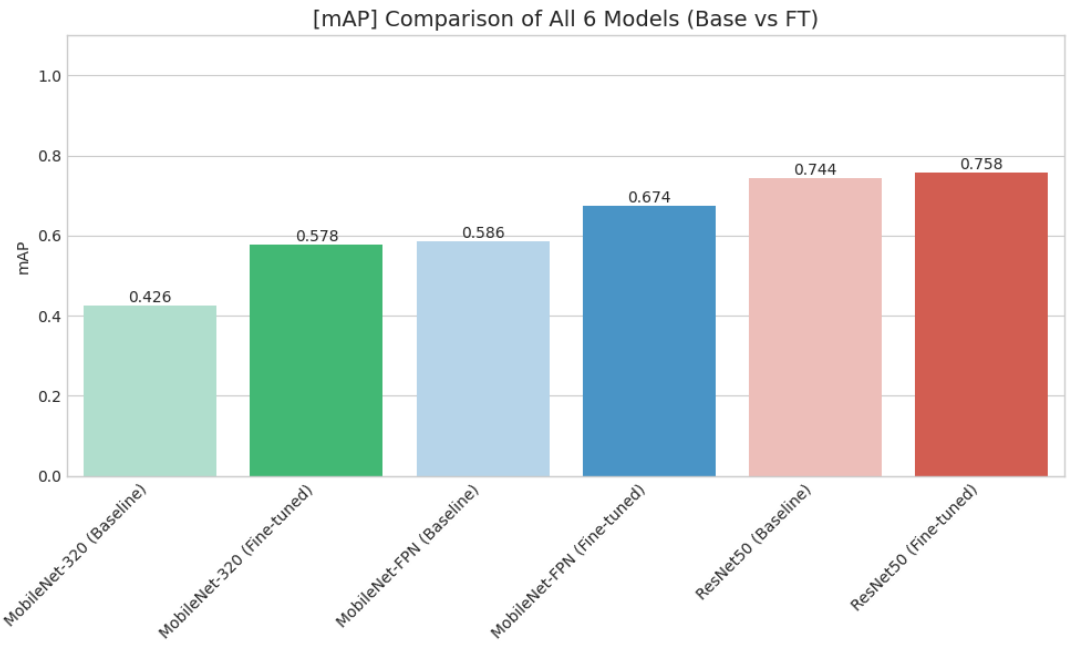
三组 baseline 对比



三组微调后的模型对比

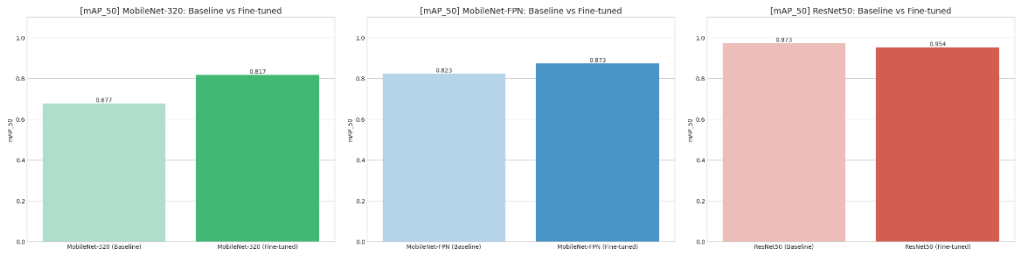


三个 baseline 和三个微调后的模型一起对比

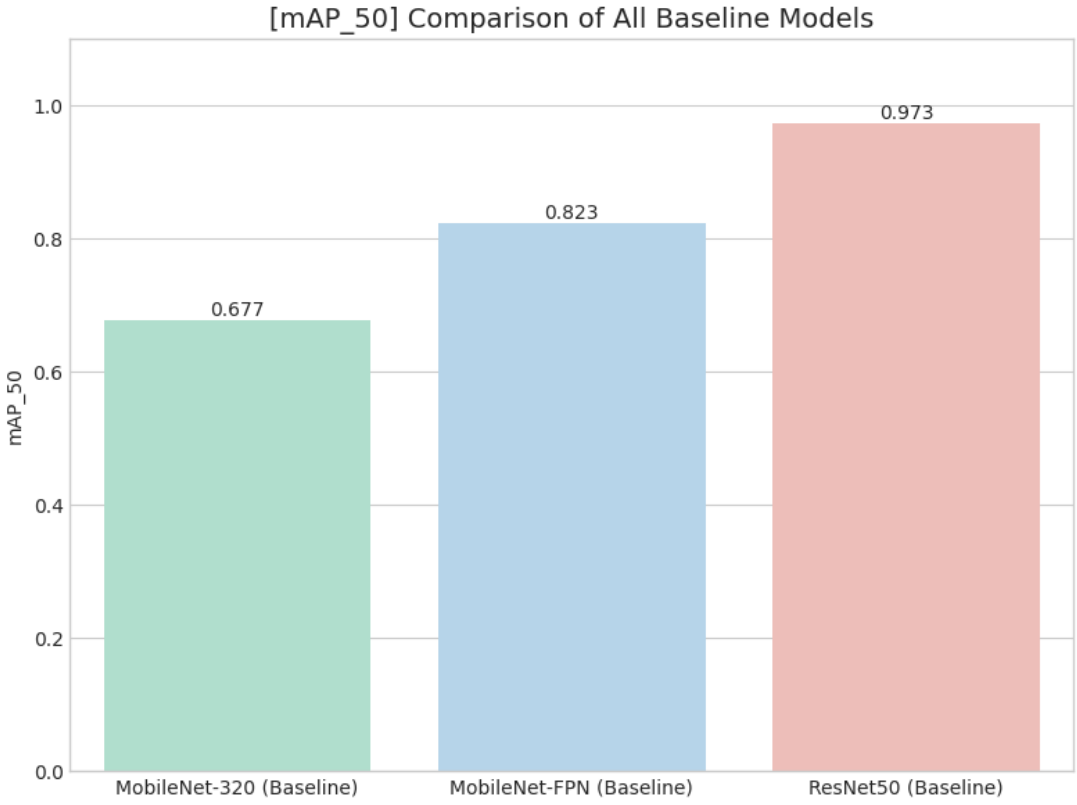


6.3 mAP_50 对比：

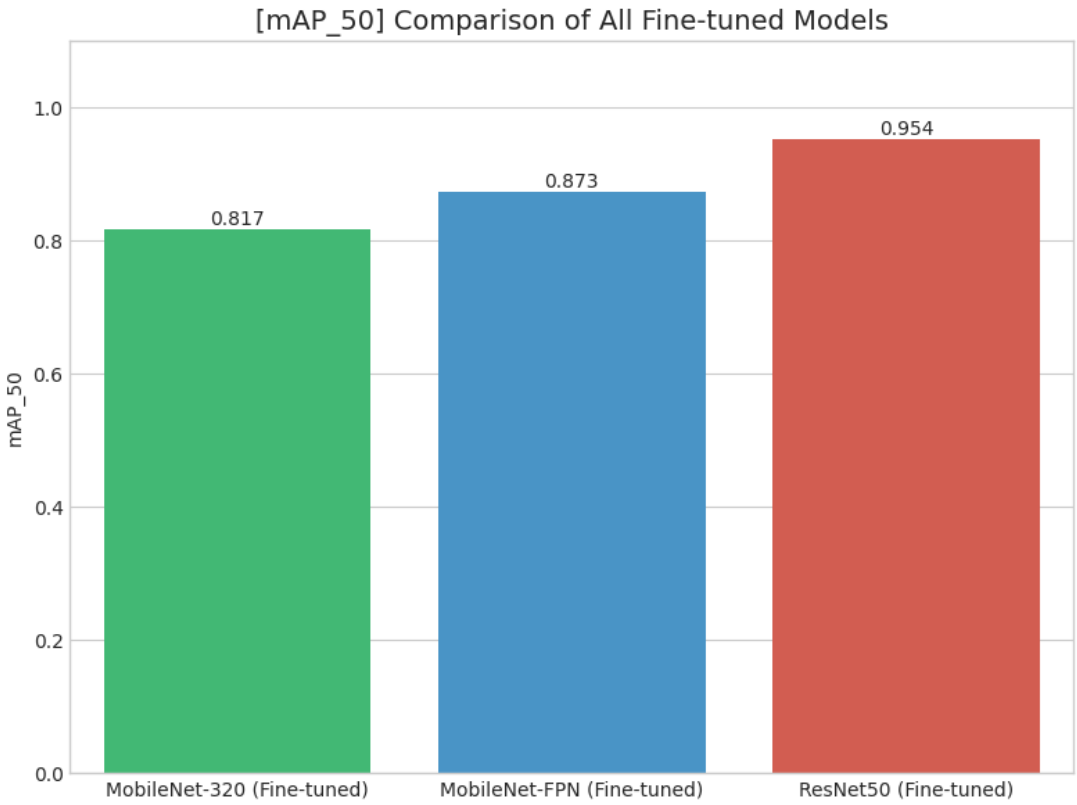
三组 baseline 分别 VS 微调后的模型



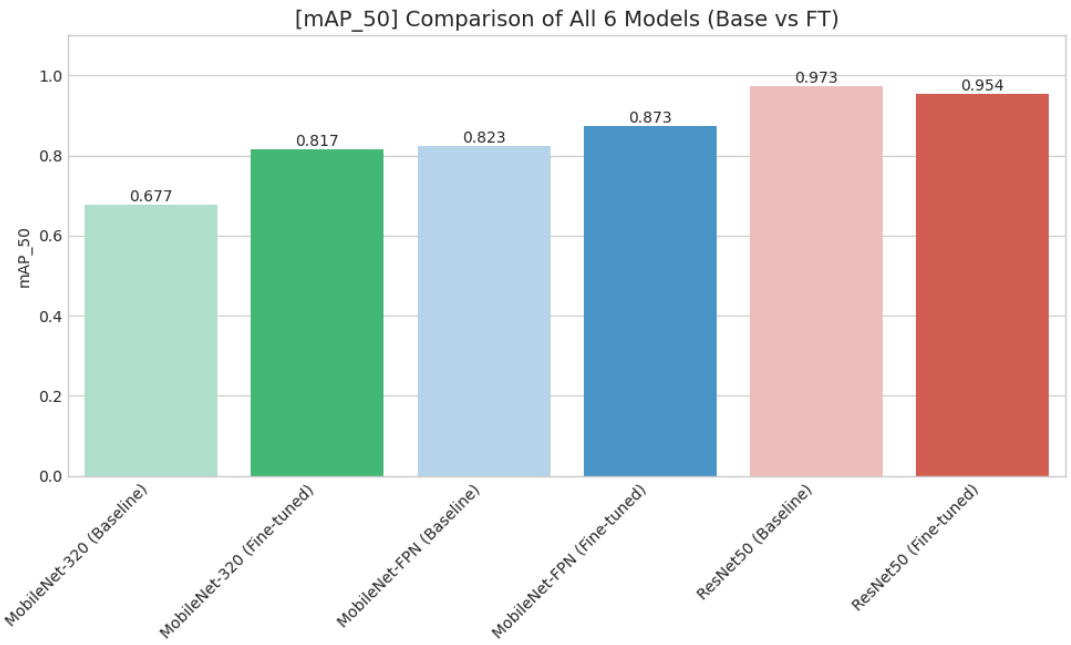
三组 baseline 对比



三组微调后的模型对比

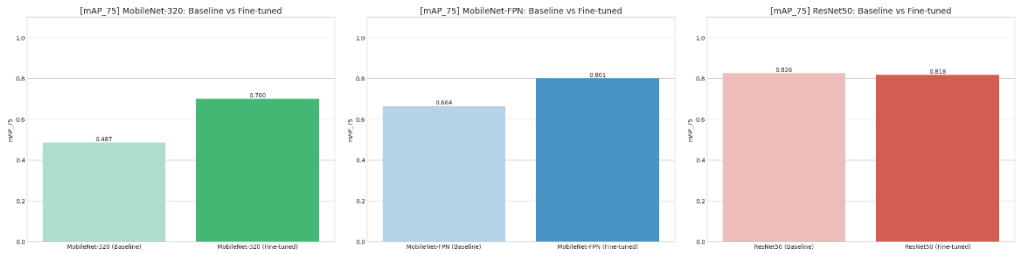


三个 baseline 和三个微调后的模型一起对比

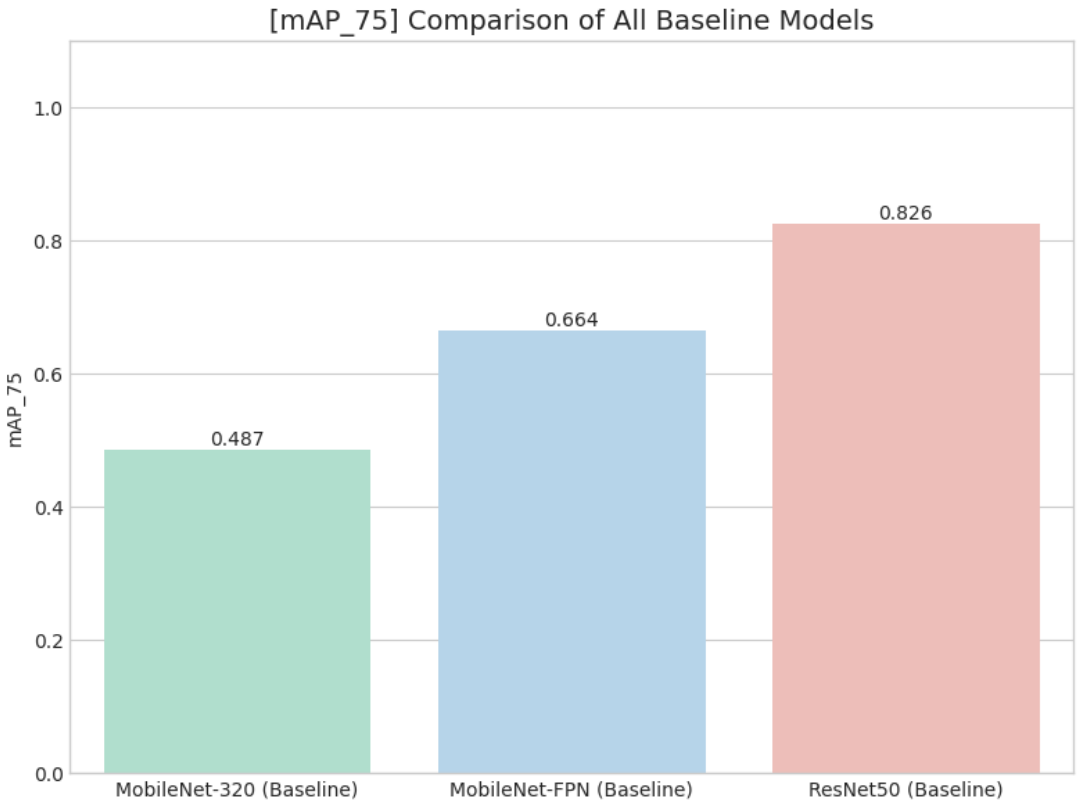


6.4 mAP_75 对比

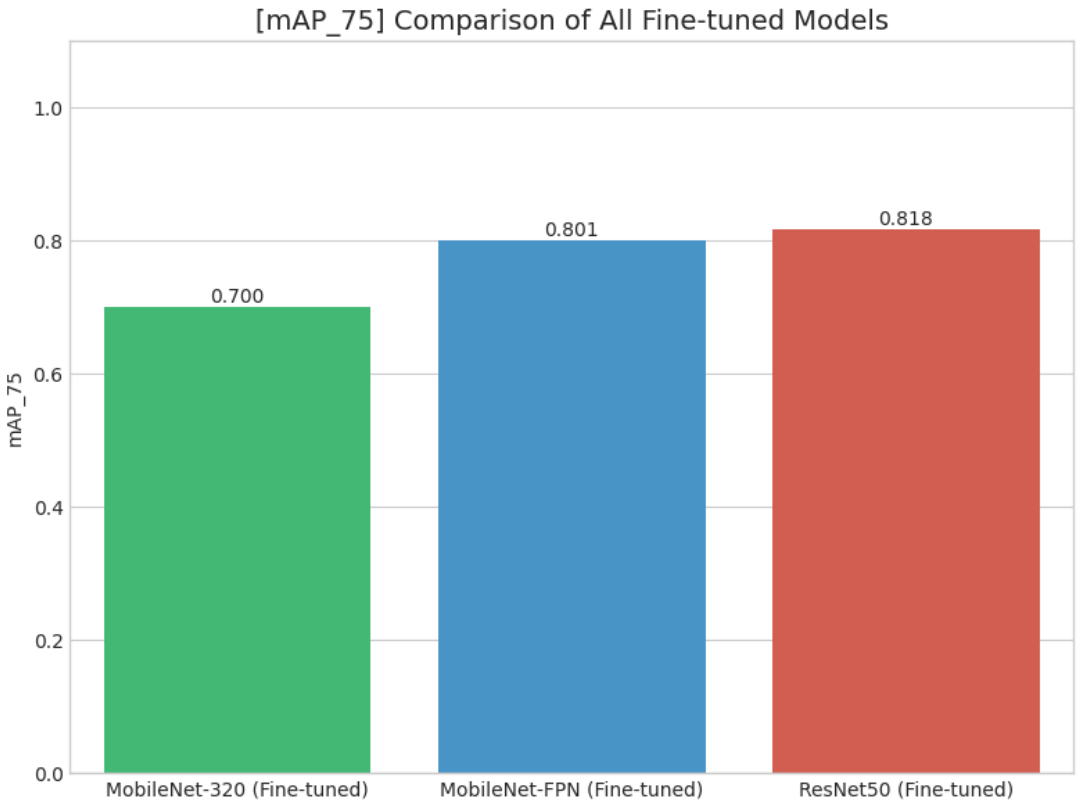
三组 baseline 分别 VS 微调后的模型



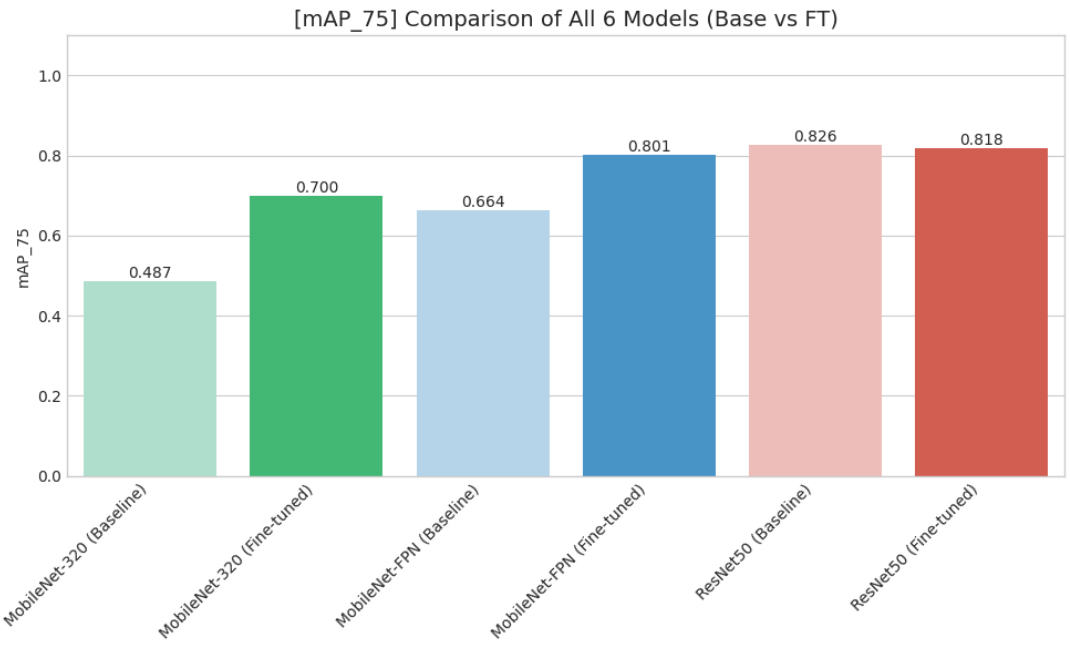
三组 baseline 对比



三组微调后的模型对比

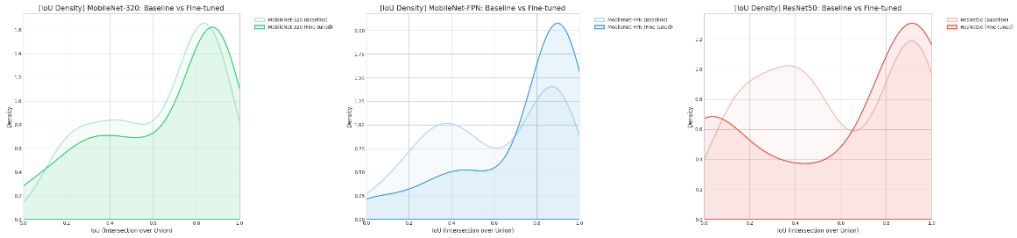


三个 baseline 和三个微调后的模型一起对比

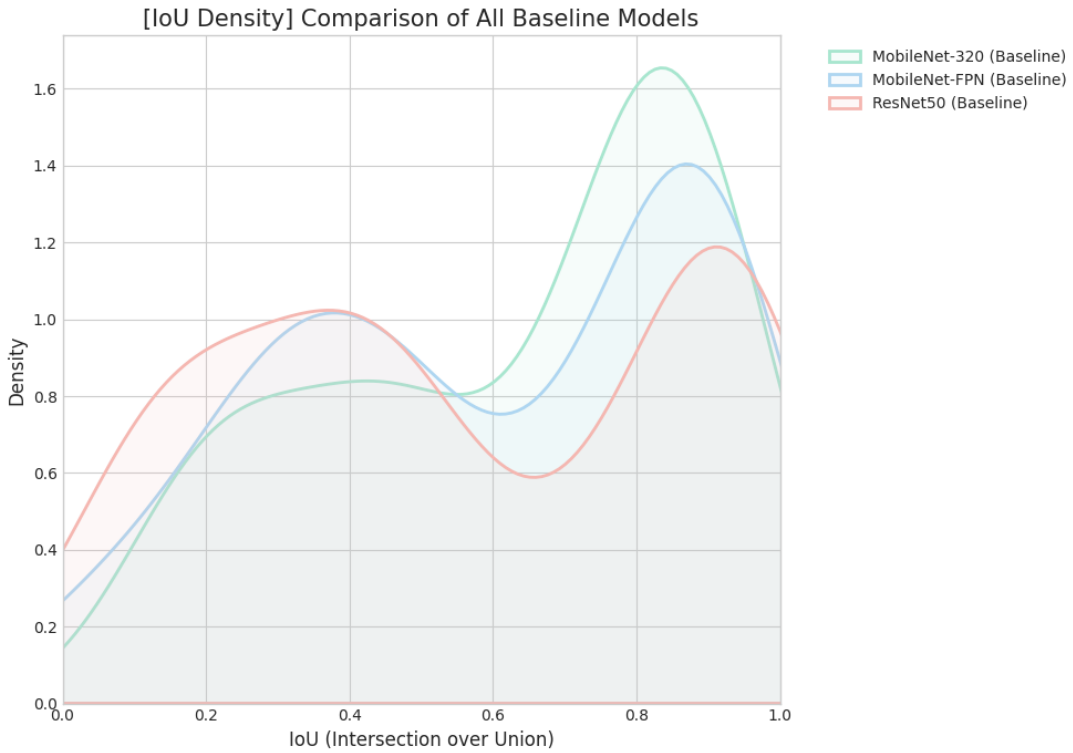


6.5 IoU_Distribution 对比

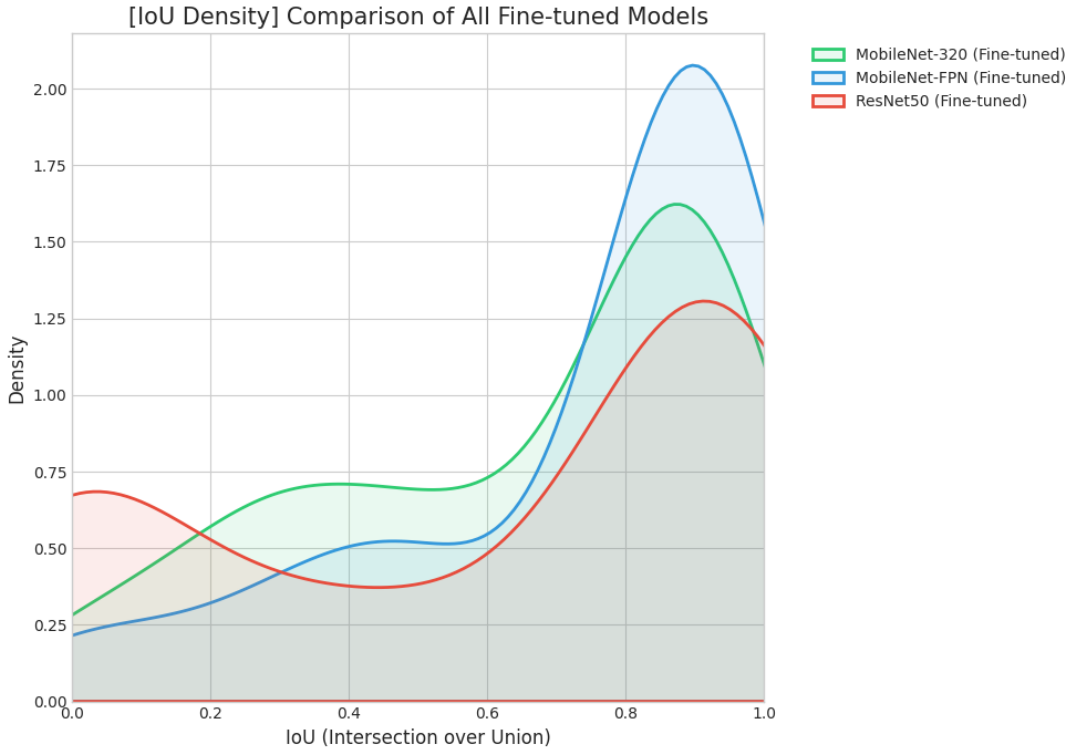
三组 baseline 分别 VS 微调后的模型



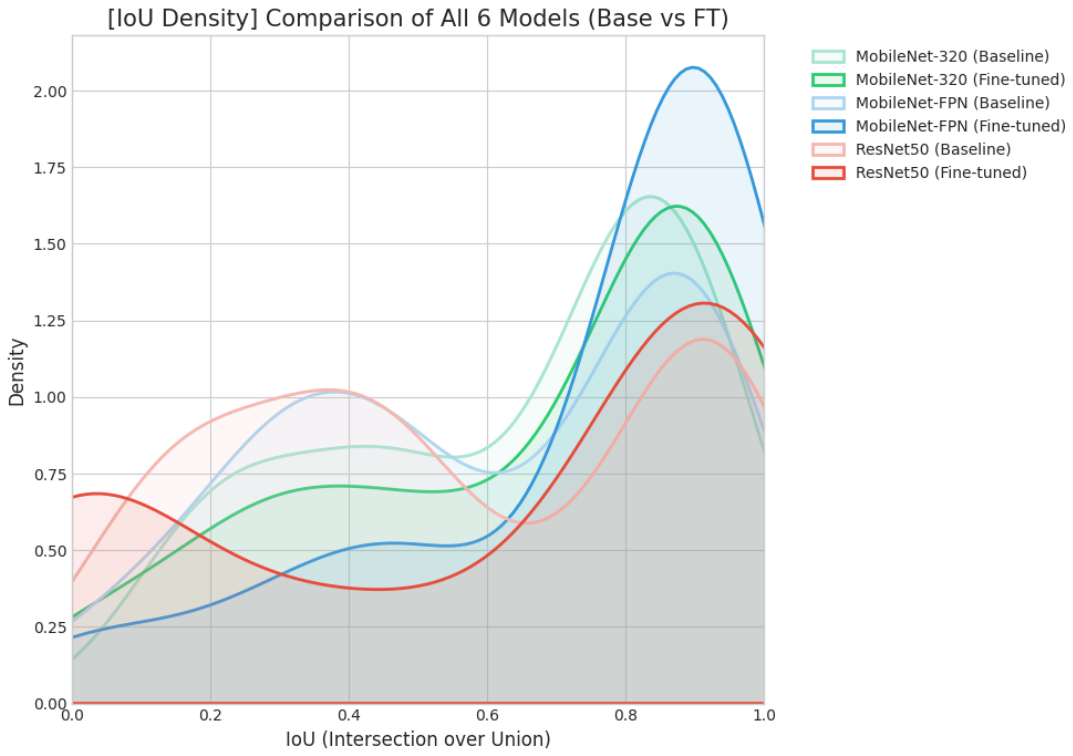
三组 baseline 对比



三组微调后的模型对比

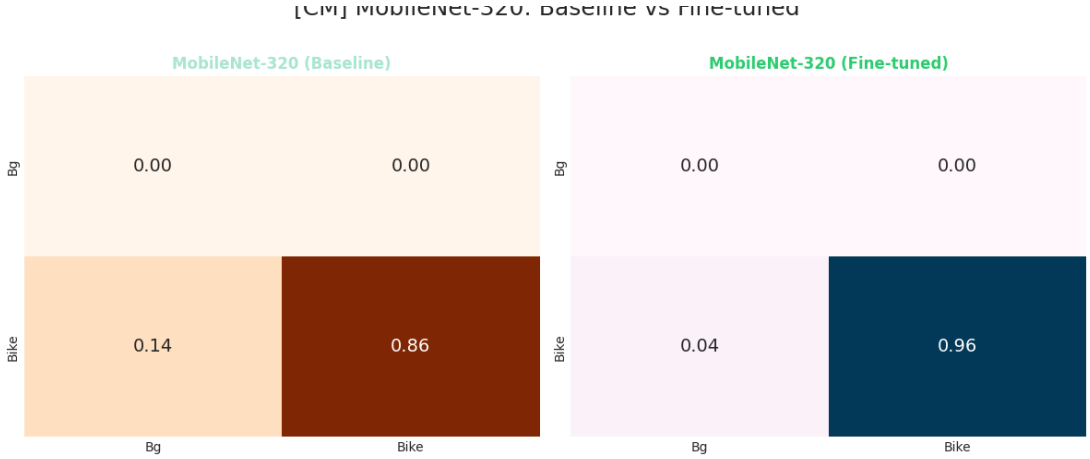


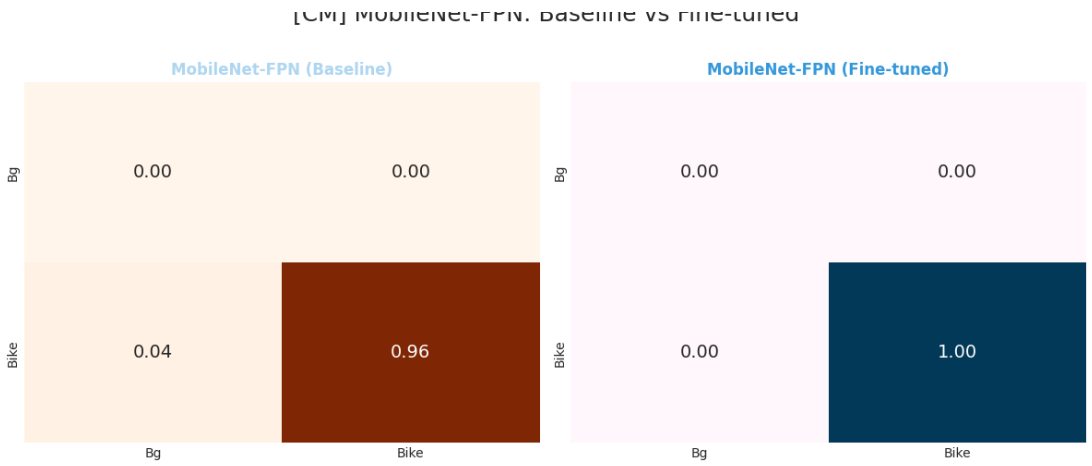
三个 baseline 和三个微调后的模型一起对比

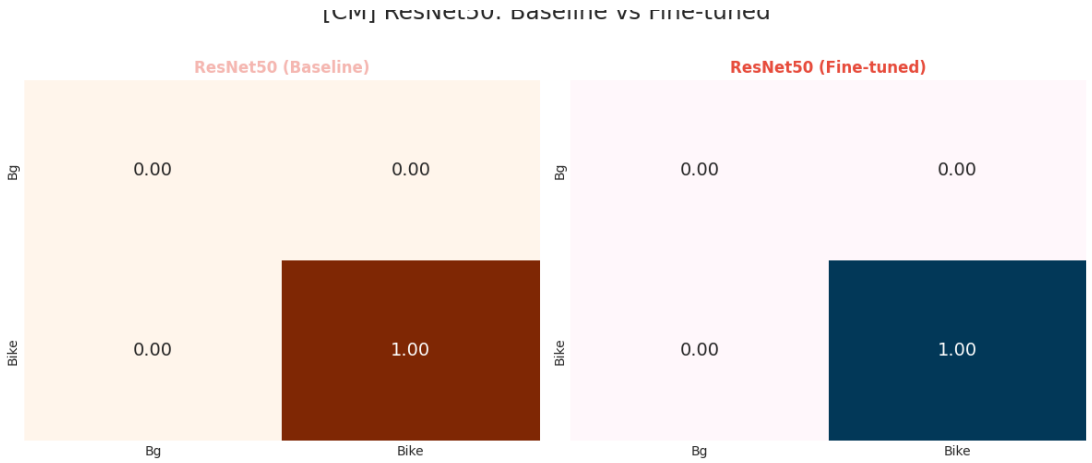


6.6 混淆矩阵对比

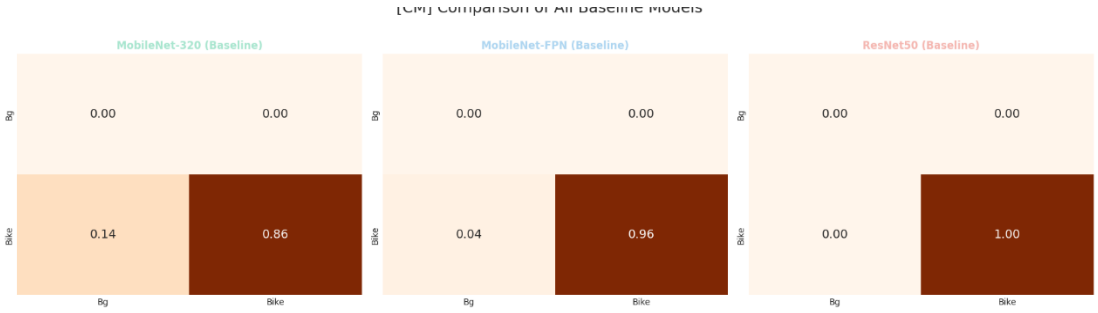
三组 baseline 分别 VS 微调后的模型



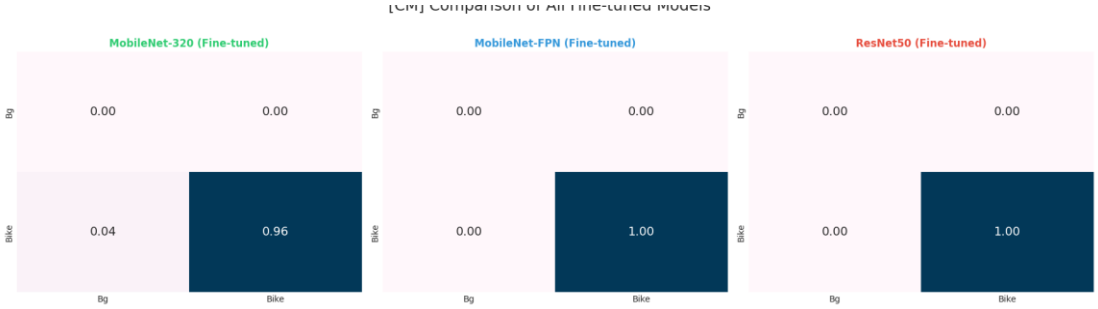




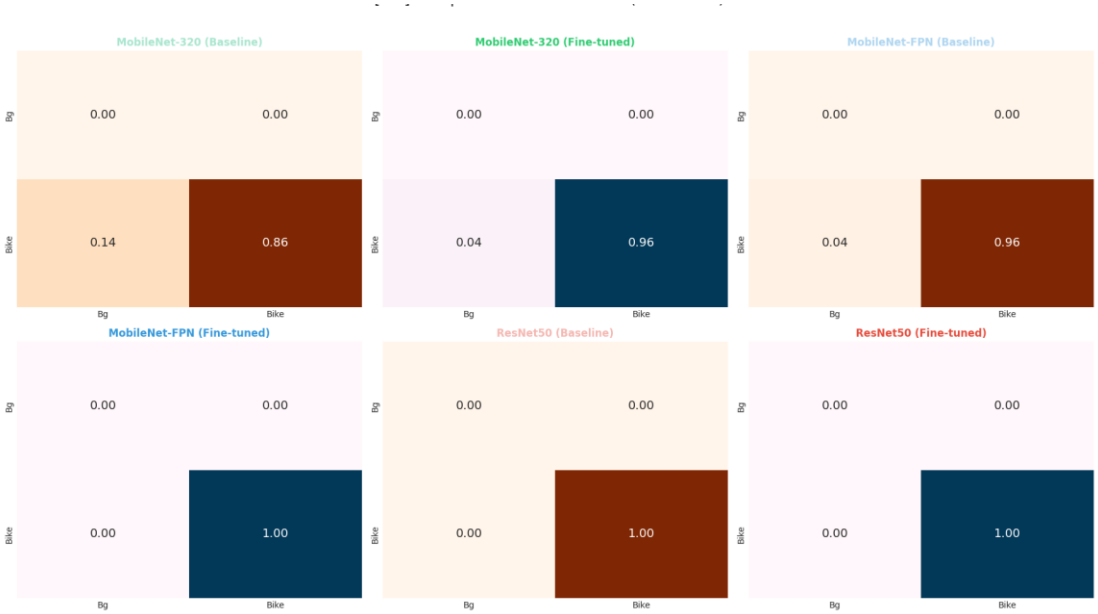
三组 baseline 对比



三组微调后的模型对比

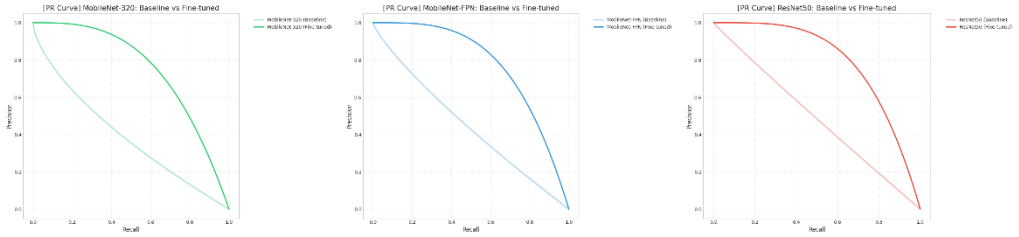


三个 baseline 和三个微调后的模型一起对比

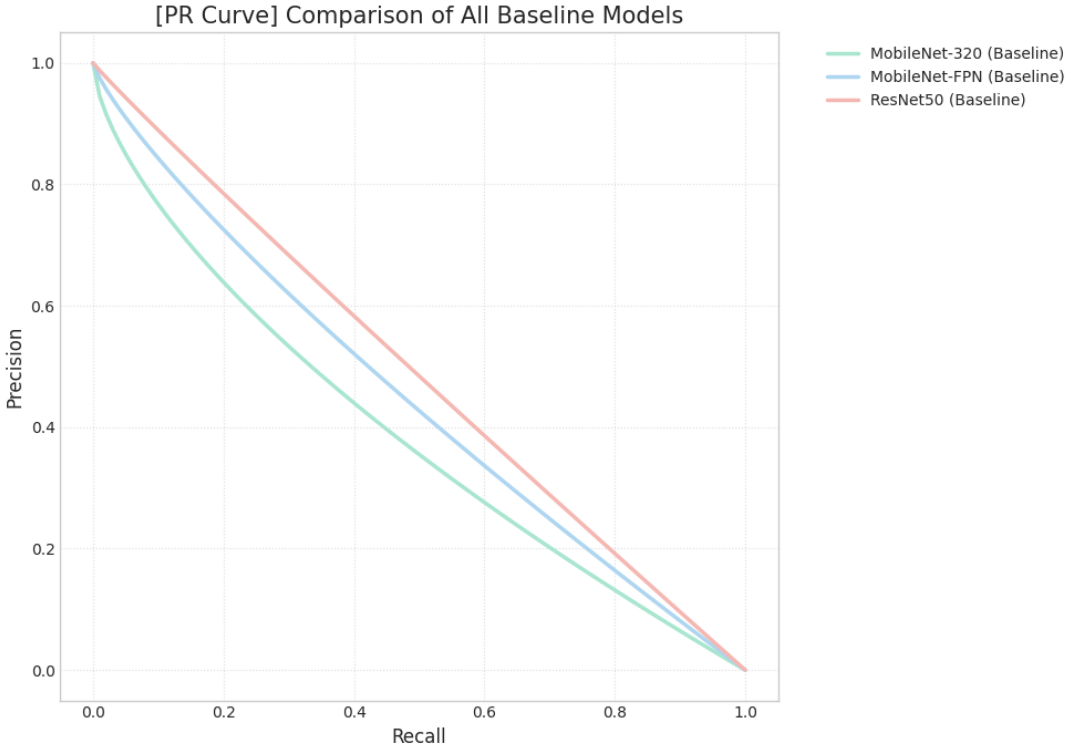


6.7 PR_Curve 对比

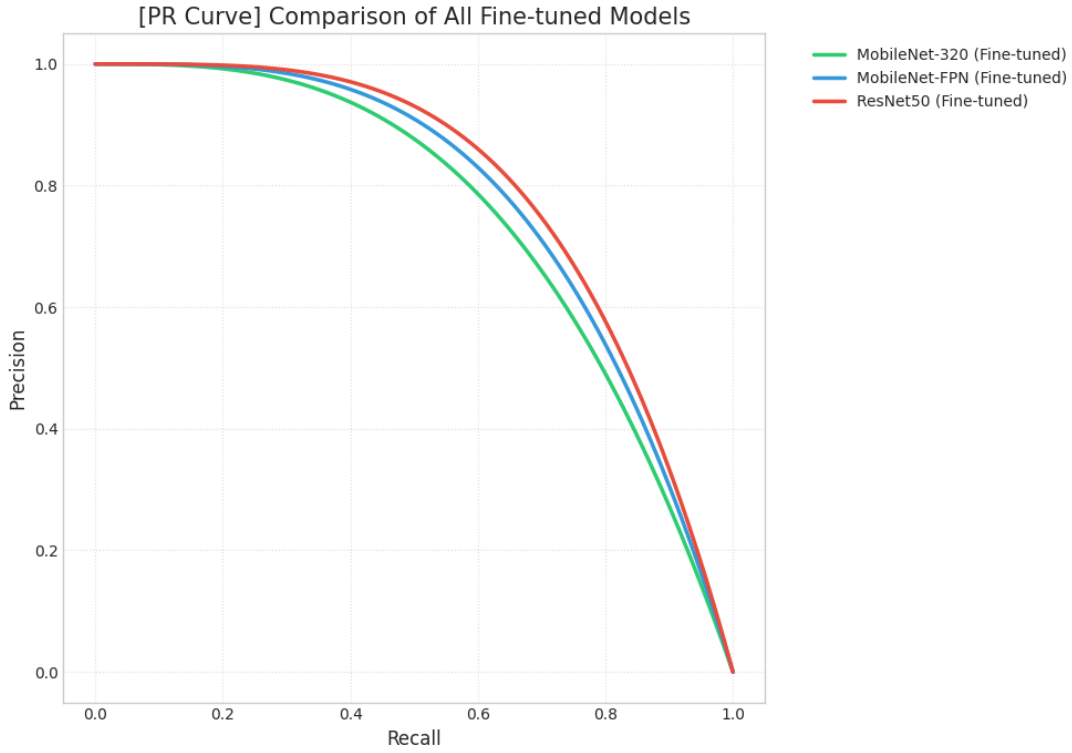
三组 baseline 分别 VS 微调后的模型



三组 baseline 对比



三组微调后的模型对比



三个 baseline 和三个微调后的模型一起对比