

# 基于PP-PicoDet的轻量级目标检测与LoRA参数高效微调研究

王乐天, 孙仁诚\*

青岛大学计算机科学技术学院, 山东 青岛

收稿日期: 2026年3月9日; 录用日期: 2026年4月17日; 发布日期: 2026年4月28日

## 摘要

针对移动端版面分析中大尺度元素(如复杂表格、整页正文块)识别难度大, 且在小样本约束下轻量级检测器全量微调易过拟合的问题, 本文以PP-PicoDet为基准, 提出一种基于非对称分层LoRA (Low-Rank Adaptation)的参数高效微调方案。该方法通过在主干网络深层与分类检测头中部署高秩适配器( $r = 128$ ), 并在特征融合颈部实施轻量级注入, 以重塑模型对大尺度版面元素的全局结构表征能力。在自建的包含8类大尺度版面元素的小样本数据集(1000张图像)上进行验证, 实验结果表明: 本文方法仅需6.5%的可训练参数量, 即可实现0.581的 $mAP_{0.5:0.95}$ 以及0.847的 $mAP_{0.5}$ , 在推理速度(18.6 FPS)保持不变的前提下, 达到了全量微调94%以上的检测性能。针对 $area = large$ 目标的专项评估显示, 该策略有效解决了全量微调在有限数据下的泛化剧变瓶颈, 实现了精度、稳定性与训练成本的深度折衷(Trade-off)。消融实验进一步证实, 针对特定层级的非对称Rank配置是捕捉版面长程依赖的关键, 为轻量化版面分析模型在边缘侧的快速定制化部署提供了高效路径。

## 关键词

版面分析, 轻量级检测器, 小样本学习, 参数高效微调(PEFT), 低秩自适应(LoRA), PP-PicoDet

## Research on Lightweight Object Detection and LoRA-Based Parameter-Efficient Fine-Tuning for PP-PicoDet

Letian Wang, Rencheng Sun\*

College of Computer Science & Technology, Qingdao University, Qingdao Shandong

Received: March 9, 2026; accepted: April 17, 2026; published: April 28, 2026

\*通讯作者。

## Abstract

To address the difficulty of detecting large-scale elements in mobile document layout analysis, such as complex tables and full-page text blocks, as well as the tendency of full fine-tuning lightweight detectors to overfit under small-sample constraints, this paper takes PP-PicoDet as the baseline and proposes a parameter-efficient fine-tuning scheme based on asymmetric hierarchical LoRA (Low-Rank Adaptation). The proposed method deploys high-rank adapters ( $r = 128$ ) in the deep layers of the backbone and the classification head, while introducing lightweight adaptation into the feature fusion neck, thereby enhancing the model's ability to capture the global structural representations of large-scale layout elements. Experiments conducted on a self-constructed small-sample dataset containing 1000 images and 8 categories of large-scale layout elements show that the proposed method requires only 6.5% of the trainable parameters while achieving 0.561 and 0.820, respectively. Without sacrificing inference speed (18.6 FPS), it attains more than 94% of the detection performance of full fine-tuning. Further evaluation on large objects area = large demonstrates that the proposed strategy effectively alleviates the sharp generalization degradation of full fine-tuning under limited data conditions, achieving a balance among accuracy, stability, and training cost. Ablation experiments further verify that the asymmetric rank configuration for specific network layers is critical for capturing long-range dependencies in document layouts, providing an efficient path for the rapid customized deployment of lightweight layout analysis models on edge devices.

## Keywords

Document Layout Analysis, Lightweight Detector, Few-Shot Learning, Parameter-Efficient Fine-Tuning (PEFT), Low-Rank Adaptation (LoRA), PP-PicoDet

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

版面分析作为文档智能(Document AI)的核心基础任务,旨在对复杂文档中的文字块、大尺度表格及图表等元素进行精确识别与定位,是实现档案自动化处理与数字化转型的关键[1]。与算力充裕的服务器场景不同,移动端扫描及边缘侧办公机器人对模型的参数规模与推理延迟具有极高敏感性。在此背景下,PP-PicoDet 凭借其骨干网络及多尺度特征融合颈部(Neck)的系统性优化,在保持极低计算开销的同时展现了卓越的精度表现,成为轻量化版面检测领域的代表性基准。

现有的版面检测器可分为基于锚框与无锚框两类:两阶段检测器通常基于锚框,通过区域建议网络捕捉复杂的文档元素,并利用特征金字塔网络(FPN)融合多尺度语义特征,以优化对长段落或细碎图表的定位精度;尽管其精度较高,但在移动端文档扫描场景下难以实现实时检测[1]。一阶段检测器在速度与精度间取得了平衡,其中 SSD 通过多尺度检测提升了对公式、注脚等小目标的敏感度[2],而 YOLO 系列则在处理复杂报刊版面时展现出极高的检测效率[3]。无锚框检测器通过消除预设锚框显著提升了对非规则区域的处理效能,如 CornerNet 通过检测角点对定位文本块[4],CenterNet 直接预测目标中心点[5],而 FCOS 则采用逐像素预测并引入中心度分支,有效降低了重叠版面元素的计算开销[6]。后续研究如 ATSS 和 GFL 进一步通过自适应样本选择与质量估计优化了边界框的回归精度[7] [8]。在移动端办公领域,YOLObile、NanoDet 及 YOLOX-Nano 等工作通过轻量化骨干网络与 SimOTA 动态标签分配策略优

化了端侧表现[9];此外,神经架构搜索(NAS)也被应用于优化网络结构,代表性工作如 NAS-FPN 和 MobileDets,在算力受限的情况下实现了版面分析的最优效率-精度折衷[10][11]。PP-PicoDet 作为一种专为移动端与嵌入式设备设计的超轻量化目标检测算法,通过采用增强型 ShuffleNet (ESNet)作为骨干网络,显著提升了在极低计算量下的版面特征提取能力,并引入了改进的路径聚合网络(CSP-PAN)以实现多尺度文档元素的高效融合。该算法采用无锚框(Anchor-free)检测范式,结合 Generalized Focal Loss (GFL)协同优化分类与定位质量,并利用 SimOTA 动态标签分配策略,使其在处理包含文本、图表、页眉页脚等 8 类复杂版面元素时,能够以极小的模型体积实现卓越的定位精度与推理速度。

## 2. 问题分析

在基于 PP-PicoDet 的轻量化版面分析任务中,模型需在小样本约束下实现对大尺度版面元素(如跨页表格、长文本块)的精确识别。针对该任务,骨干网络、特征融合颈部及解耦检测头在处理特定版面特征时存在以下问题:

骨干网络 Stage 3 处于特征分辨率与语义深度结合的关键节点,承载全局布局信息,但由于轻量化算子参数量限制,其对大尺度目标的拓扑结构表征能力不足,导致全量微调在有限数据下难以捕捉长程依赖。Stage 4 作为特征提取的最深层,其高阶语义已高度抽象,但在小样本场景下易受参数过度更新影响,导致特征偏移并削弱模型的泛化稳健性。

在特征融合颈部,由于版面元素具有显著的尺度极差,要求 Neck 模块具备动态调整特征分布的能力。然而,CSP-PAN 在处理复杂嵌套布局时,常表现出多尺度信息聚合效率低下的问题,难以实现底层视觉纹理与高层排版逻辑的精确对齐。

解耦检测头中的分类与回归分支在版面分析中存在不同的参数敏感度。分类分支(Cls Head)面临细粒度语义区分困难的问题,例如文本块与同质化表格的视觉特征高度相似,导致决策边界模糊,难以实现对版面组件类别的准确激活。回归分支(Reg Head)则侧重于空间连续性与几何定位,由于缺乏针对性约束,在小样本环境下易受背景噪声干扰,产生定位随机抖动,从而限制了边界框预测的稳定性。

上述位置存在的表征失效与过拟合风险,构成了轻量化版面分析模型在数据稀缺条件下实现高鲁棒性的核心障碍。在实际工程落地中,版面分析常面临特定领域数据稀缺与标注高昂的约束。在小样本(Few-shot)条件下,传统的全量微调(Full Fine-Tuning)策略暴露出显著缺陷:一方面,由于版面元素具有极强的空间拓扑特征,全量参数更新极易导致模型在有限数据下陷入过拟合,难以在长尾类别上保持泛化稳定性;另一方面,针对不同行业文档维护全量参数模型,极大地增加了存储成本与多任务切换的开销。为此,以低秩自适应(LoRA)为代表的参数高效微调(PEFT)技术为轻量级检测器的域适配提供了新路径[3],其通过在预训练权重旁路嵌入低秩矩阵,实现了参数空间的极简优化。

尽管 LoRA 在大规模预训练模型中取得了显著成效,但将其直接应用于轻量级 CNN 检测器处理大尺度版面元素时,仍面临核心挑战:首先,轻量级模型感受野受限,传统的均匀微调难以弥补模型在大尺度目标(如跨页表格)上的全局语义感知缺陷;其次,卷积算子在提取版面纹理与空间布局时具有高度的结构化特征,如何科学设定 LoRA 的注入位置、秩(Rank)与缩放系数,以匹配大目标的几何特性,尚缺乏系统性研究。基于此,探索一种能够强化大目标表征能力的差异化微调策略,具有重要的学术与工程价值。

基于上述问题,本文以 PP-PicoDet 为研究对象,提出一种面向大目标版面分析的“分层注入、非对称配置”LoRA 微调方案[12]。本文主要贡献概括如下:

1) 提出面向轻量化版面检测器的 LoRA 注入范式:在不改变推理计算图的前提下,通过在主干网络深层与特征融合模块中嵌入低秩适配器,实现了以极小参数代价完成从通用检测域到特定文档域的知识

迁移。

2) 设计了非对称式分层超参数配置策略: 针对大目标对语义连续性的高度需求, 在主干 Stage 3 层与分类头中注入高秩矩阵( $r=128$ ), 显著增强了模型对复杂拓扑结构的表征能力, 有效解决了全量微调在大尺度目标上的识别不稳定性。

3) 构建了小样本版面分析实验基准: 在自建的包含 8 类大尺度版面元素的数据集上进行验证。实验结果表明, 本文方法仅需 6.5% 的可训练参数, 即可达到全量微调 94% 以上的检测精度( $mAP_{0.5} = 0.847$ ), 为轻量化模型处理复杂、大尺度版面任务提供了高效的定制化路径。

### 3. 方法

#### 3.1. 卷积层 LoRA 的参数化表征与重参数化机理

LoRA 的核心思想是将全量参数更新矩阵  $\Delta W$  投影至低秩子空间, 从而极大地压缩搜索空间。对于 PP-PicoDet 中的卷积算子, 其权重张量  $W \in \mathbb{R}^{C_{out} \times C_{in} \times k \times k}$  在微调过程中被冻结, 通过注入并行的低秩支路实现知识迁移:

1) 参数化表征: 引入两个连续的可训练低秩矩阵  $A \in \mathbb{R}^{r \times (C_{in} \cdot k^2)}$  与  $B \in \mathbb{R}^{C_{out} \times r}$ 。其中, A 矩阵通常采用高斯初始化, 而 B 矩阵初始化为零, 以确保微调初期模型输出与预训练模型保持一致, 实现平滑起步。

2) 前向计算与合并:

$$y = \text{Conv}(x; W) + \frac{\alpha}{r}(x * \Delta W), \Delta W = B \times A$$

其中  $\alpha$  为缩放系数, 调节微调权重对原始特征的影响权重。

3) 推理一致性: 微调结束后, 利用线性算子的可加性, 通过  $W_{\text{deployed}} = W + \frac{\alpha}{r}(BA)$  将旁路参数永久性融入原始权重。这一过程确保了在边缘侧推理阶段, 模型无需修改任何底层计算算子, 在保持 18.6 FPS 高速推理的同时, 获得了更强的领域适应性。

#### 3.2. 非对称式注入位置选择策略

针对版面分析中小样本、大目标检测的特殊需求, 本文并未采取传统的全网络均匀注入, 而是提出一种非对称分层注入策略(如图 1), 其设计逻辑如下:

1) 浅层保护机制: 本文有意避开了骨干网络的前两个 Stage。机理分析认为, 浅层特征主要响应图像边缘、线条等通用纹理, 这些特征在版面图像中具有高度的跨领域公用性, 过早注入 LoRA 反而可能引入噪声。

2) 深层语义重构: 在 Backbone Stage 3 与 Stage 4 注入适配器。Stage 3 处于感受野显著扩张的关键节点, 承载了大量的全局布局(Global Layout)信息, 是捕捉大尺度目标的语义核心。

3) 融合颈部动态对齐: 在 CSPPAN 注入 LoRA。由于版面元素常跨越多个特征层级, 通过在融合支路注入适配器, 可以动态修正多尺度特征间的权重分布, 缓解轻量级模型对长程依赖捕捉不足的问题。

#### 3.3. 差异化超参数配置与“按需分配”准则

本文提出一套基于特征敏感度的差异化超参数配置方案(详见表 1), 并从以下维度阐述其设计科学性:

1) 高秩驱动的大目标感知( $r=128, \alpha=145$ ):

在 Stage 3 与 Head-Cls 分支, 本文打破了常规 PEFT 采用极小 Rank (如 4 或 8) 的局限。考虑到分类分支需要对极具挑战性的版面组件(如相似的表格与列表)进行判别, 采用高秩配置能为模型提供更宽广的

假设空间, 从而精确刻画大尺度目标的拓扑边界。

2) 稳健型特征聚合策略( $r = 16\sim 28$ ):

在 Stage 4 与 Neck 层级, 为了在增强语义的同时防止破坏预训练的底层稳定性, 采用了较为克制的 Rank 设定。这种“强分类、稳融合”的设计, 旨在实现精度增益与过拟合风险的动态对齐。

3) 任务驱动的正则化配置:

在大多数支路引入 0.1 的 Dropout 以提升鲁棒性, 唯独在分类头(Cls Head)将其置为 0.0。这一非绝对化设计是基于实验观察: 分类分支在小样本下极易受随机失活的影响, 保持特征流的完整性有助于固定决策边界, 减少预测波动。

Table 1. Hyperparameter settings

表 1. 超参数配置方案

注入层级	匹配模式	秩	缩放系数	Dropout
Backbone Stage 3	backbone.stage3.*conv*	128	145	0.1
Backbone Stage 4	backbone.stage4.*conv*	16	32	0.1
Neck (FPN)	neck.*.conv*	18	36	0.1
Head (Cls)	head.*.cls*	128	145	0.0
Head (Reg)	head.*.reg*	28	56	0.1

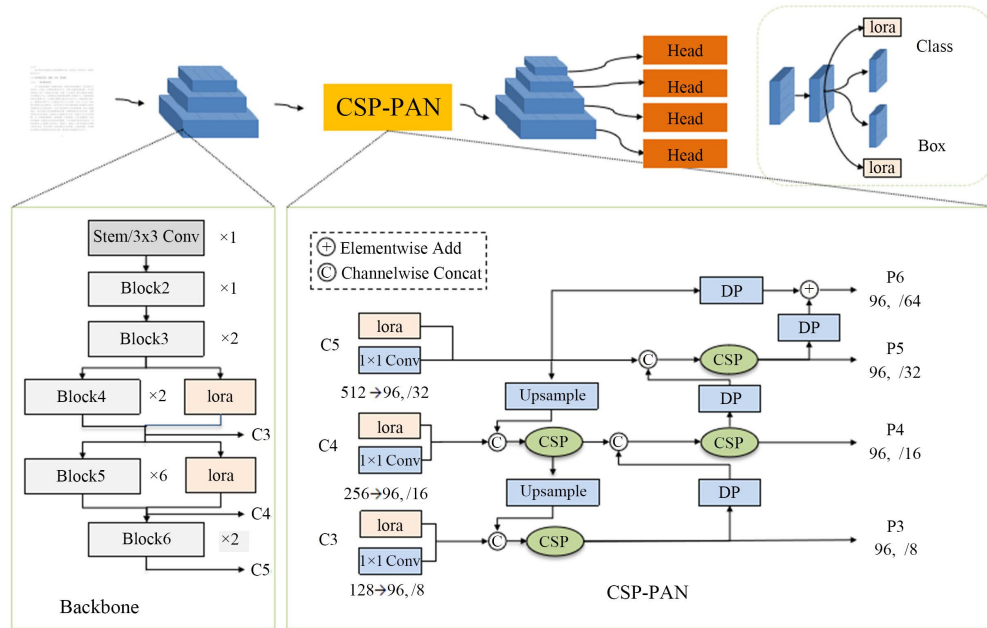


Figure 1. Schematic of LoRA injection positions in the original PP-PicoDet framework

图 1. 基于原 PP-PicoDet 框架图的 LoRA 注入位置示意

综上所述, 本文提出的参数配置方案并非盲目增加 Rank 冗余, 而是基于模块功能特性与过拟合风险的深度对冲。具体而言: 主干网络深层(Stage 3)与分类检测头承担了跨领域语义重塑的核心任务, 故通过高秩(High-rank)适配以扩充模型的表征假设空间; 特征融合颈部与边界框回归分支则更侧重于空间连续性与几何定位的微调, 通过适中秩(Medium-rank)配置在确保多尺度交互稳定性的同时, 有效抑制了背景

噪声对定位精度的干扰；而 Stage 4 则采取低秩(Low-rank)策略以维持预训练模型的高阶语义一致性。这种“因地制宜”的非对称分配策略，旨在以最小的参数扰动(Perturbation)激发模型对版面拓扑结构的最优感知，其有效性将在后续消融实验中得到量化验证。

## 4. 实验设置

### 4.1. 实验数据集与预处理

本文采用自建的政府招标科技项目文档截图数据集进行实验验证，共包含 1000 张高分辨率截图，素材涵盖了公开招标公告、技术投标建议书及政府采购评审报告等标准化公文。该数据集的版面挑战主要体现在：在高度规范化的版面布局中，存在大量视觉特征极度相似的层级条文与正文块，且包含横跨页面的工程进度甘特图、设备采购详细清单等超大尺度组件，对模型的多尺度拓扑建模能力提出了严苛要求。标注过程严格遵循 MS COCO 格式，将 8 类版面元素脱敏标注为  $C_1 \sim C_8$ 。数据划分遵循训练、验证与测试集 7:2:1 的比例(即 700/200/100 张)，并通过分层随机采样确保各类别在子集间的分布均衡性。

### 4.2. 评价指标与损失函数

为全面衡量模型在政务文档版面分析任务中的定位精度与分类效能，本文采用 MS COCO 官方评价准则，选取  $mAP_{0.5:0.95}$  作为综合性能评价的主指标，并辅以  $AP_{50}$  与  $AP_{75}$  以评估模型在不同 IoU 阈值下的鲁棒性。同时，针对招标文档中关键的大尺度组件，重点统计大目标平均召回率( $mAR_{large}$ )，用以定量验证非对称 LoRA 配置在捕捉长程依赖及维护大目标检测完整性方面的有效性。

本文沿用 PP-PicoDet 的多任务损失优化框架，通过分类损失、定位回归损失与概率分布损失的协同作用，实现小样本约束下对大尺度版面元素的识别与定位。总损失函数由 Varifocal Loss (VFL)、Generalized IoU (GIoU) Loss 与 Distribution Focal Loss (DFL) 加权构成[13]，其数学表达为：

$$\text{loss} = \text{loss}_{\text{vfl}} + 2 \cdot \text{loss}_{\text{giou}} + 0.25 \cdot \text{loss}_{\text{dfl}}$$

其中 VFL 分支通过将分类预测与定位质量进行异步耦合，在处理版面正负样本分布不均时，引导模型关注高质量正样本并抑制背景梯度干扰。在回归任务中，GIoU Loss 通过计算非重叠区域的几何关联，解决大尺度目标在初始阶段因 IoU 为零导致的梯度消失问题，从而强化边界预测稳定性。同时，DFL 将边界框偏移量建模为概率分布，使模型聚焦于版面元素边界附近的数值连续性，配合指数移动平均(EMA)策略，提升轻量级检测头在定位组件时的收敛质量[14]。

### 4.3. 训练配置与优化策略

实验基于 PaddleDetection 框架实现。为加速小样本环境下的知识迁移，模型加载了在 ImageNet 与 MS COCO 数据集上充分预训练的权重。训练采用余弦退火学习率衰减策略并配合 Warmup 机制，总周期设为 150 Epoch。考虑到小样本训练易受梯度噪声干扰，本文引入“指数移动平均(EMA)”技术平滑权重更新，以锁定更稳健的分类决策边界。在数据增强方面，本文构建了包含多尺度缩放、随机翻转及色域变换的增强流水线，并谨慎引入 Mosaic 与 Copy-Paste 策略，通过精准控制增强强度，在提升模型对“复杂嵌套布局”感知能力的同时，严格避免破坏政务文档原有的物理拓扑逻辑。此外，结合微小权重衰减与 LoRA 模块特有的差异化 Dropout 设置(见 3.4 节)，进一步抑制过拟合风险。

### 4.4. 对比实验设计

为客观评价非对称 LoRA 策略的适配效能，本文设置了三组对照实验：

1) 线性探测(Head-Only)：完全冻结骨干网络与特征融合颈部，仅微调检测头，用于评估预训练表征

对政务公文版面的原始迁移潜力;

2) 全量微调(Full Fine-tuning): 对模型全网参数进行端到端更新, 作为性能基准以对比小样本约束下全参数微调的泛化稳定性;

3) 非对称 LoRA(本文方法): 在保持基准模型权重冻结的前提下, 仅对 Stage 3 与分类检测头等关键支路实施差异化低秩适配。该方案旨在验证能否以 6.5%的可训练参数代价, 在保持推理速度(18.6 FPS)基本不变的前提下, 实现对招标文档复杂大尺度组件的高效解析与稳定识别。

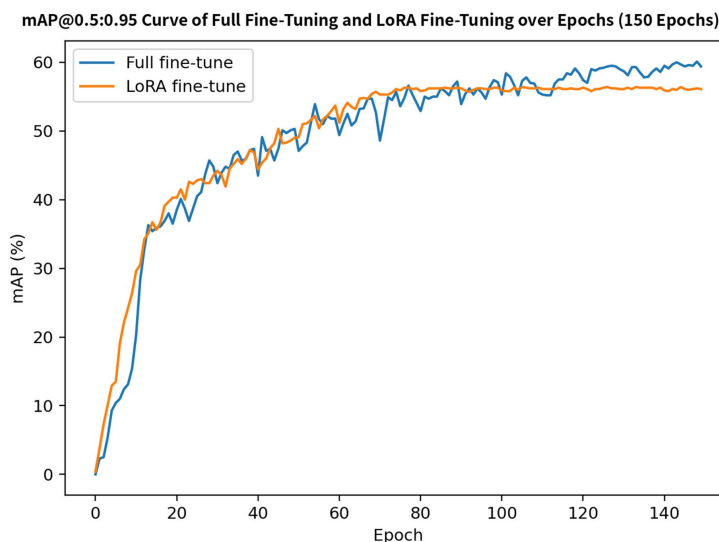
### 5. 实验结果与分析

**Table 2.** Comparison of detection performance of different fine-tuning strategies on the in-house dataset  
**表 2.** 不同微调策略在自建数据集上的检测性能对比

方法	可训练参数占比	AP50	AP75	mAP@0.5:0.95
Original	0%	61.7	30.4	31.6
Head-Only	≈1.2%	80.4	46.9	52.1
Full Fine-tune	100%	86.4	60.6	59.4
LoRA (本文方法)	≈6.5%	84.7	59.4	58.1

表 2 呈现了不同微调策略在自建数据集上的性能指标对比。通过数据分析得出如下结论:

Original 模型在未经过领域微调时, 其 mAP 为 31.6%, 表明预训练权重在处理特定版面布局时存在表征局限。Head-Only 策略以约 1.2%的参数开销将 mAP 提升至 52.1%, 但由于冻结了骨干网络与特征融合层, 模型无法适配目标领域特有的特征分布, 导致性能遭遇瓶颈。



**Figure 2.** mAP@0.5:0.95 curves of full fine-tuning and LoRA fine-tuning over 150 epochs  
**图 2.** 全参微调与 LoRA 微调的 mAP@0.5:0.95 随 Epoch 变化曲线(150 轮)

Full Fine-tune 策略通过更新全部参数设定了精度上限, 其 mAP 达到 59.4%。如图 2、图 3 所示, 两种微调方式均可在小样本版面分析任务中实现有效收敛, 但性能表现存在差异。由 mAP@0.5:0.95 与 AP50 曲线可见, LoRA 微调前期收敛更快、波动更小, 较早进入稳定阶段; 全量微调虽然后期仍能继续提升, 最终精度略高, 但训练过程震荡更明显。结合损失曲线可知, 全量微调在后期训练损失持续下降的同时,

验证损失明显回升, 表现出一定过拟合趋势; 相比之下, LoRA 的训练与验证损失间隔较小, 验证曲线更平稳, 泛化表现更稳定, 如图 4 所示。最终, LoRA 在仅占 6.5% 可训练参数的条件下取得 58.1% 的 mAP, 与全量微调仅相差 1.2%, 说明其在兼顾检测精度、训练稳定性与参数效率方面更具优势。

实验结果证实, 通过在关键支路注入低秩约束, 能够保留预训练模型的通用特征提取能力, 并以低参数扰动实现领域适配。LoRA 的低秩矩阵在微调过程中起到了正则化作用, 在降低计算资源消耗的同时, 缓解了全量微调在数据规模有限时产生的过拟合现象, 实现了检测精度与训练效率的平衡。

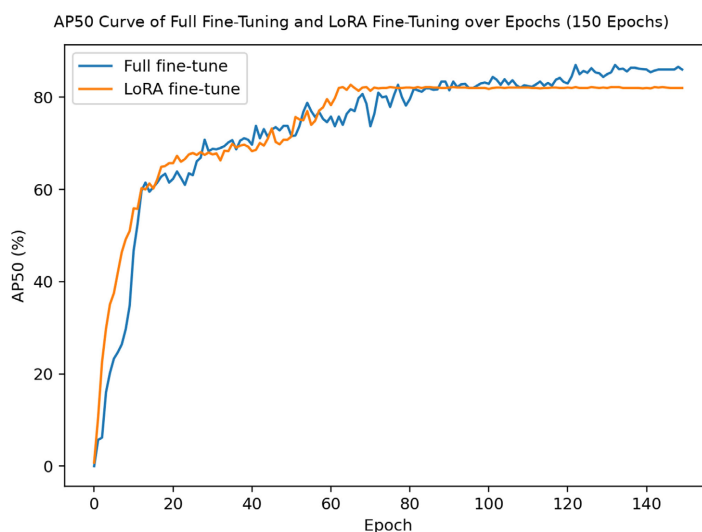


Figure 3. AP50 curves of full fine-tuning and LoRA fine-tuning over 150 epochs

图 3. 全参微调与 LoRA 微调的 AP50 随 Epoch 变化曲线(150 轮)

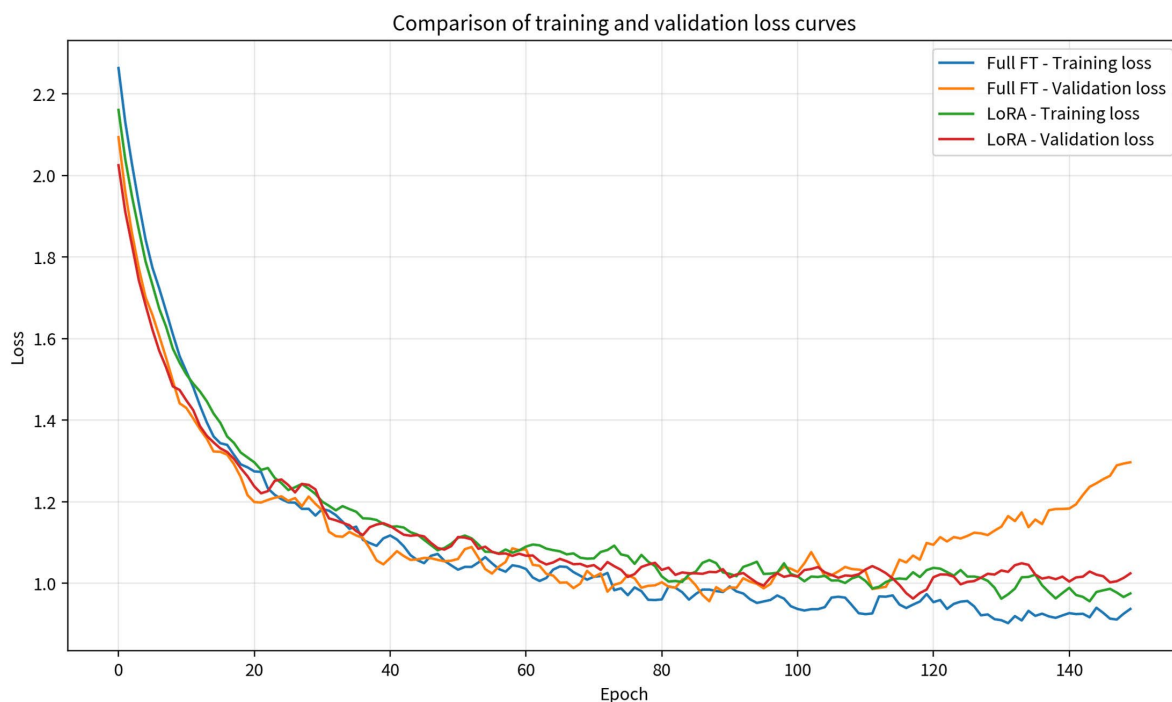


Figure 4. Training and validation loss curves

图 4. 训练集与验证集的损失曲线图

## 5.1. 定位精度与收敛稳定性观察

进一步分析不同 IoU 阈值下的表现可以发现, LoRA 方案在严苛的定位指标下展现出良好的稳健性。相较于 Head-Only 方案, LoRA 在 AP<sub>75</sub> 上提升了 12.5 个百分点, 且与全量微调的差距极小。这说明分层 Rank 配置对回归分支与 Neck 融合层的“针对性增强”, 有助于模型在有限样本下学习到更稳定的边界框预测能力, 减少了定位框的随机抖动。

同时, 通过监控训练曲线发现, LoRA 方案通常能以更少的迭代轮次进入性能平台期, 且其最佳性能点附近的波动幅度显著小于全量微调。这种收敛特征不仅降低了实际生产中的超参调优成本, 也验证了低秩分解在参数空间搜索中的高效性。

## 5.2. 消融实验

为进一步验证本文设计的合理性, 本文从“注入位置”和“超参数配置”两方面进行消融实验。除被消融项外, 其余训练设置与 4.3 保持一致。

### 5.2.1. 注入位置消融

在注入位置的消融实验中, 如表 3 所示, 实验结果呈现出显著的链式协同效应。仅在检测头注入适配器(A1)虽能实现基础的任务迁移, 但由于底层特征未针对文档域校准, 精度难以突破。令人关注的是, 仅在骨干网络注入适配器(A2)时性能剧烈下滑, 其 AP<sub>50</sub> 与 mAP 仅为 14.6% 和 5.2%, 这一断崖式下跌揭示了模型内部的“语义屏蔽”机理: 由于作为特征融合枢纽的 Neck (颈部) 与末端 Head (检测头) 均处于冻结状态, 即便 Backbone 捕捉到了领域增量信息, 这些特征在通过仍处于通用域状态的 Neck 路径时会遭遇严重的语义稀释与路径阻断, 进而与未微调的 Head 产生严重的特征偏移现象, 导致预测逻辑失效。随着注入范围扩张, “主干 + 颈部”的联合适配(A3)使 AP<sub>50</sub> 回升至 73.6%, 有力证明了 Neck 层级在对齐多尺度版面语义中的关键衔接作用。最终, 当实现 Backbone、Neck 与 Head 的全链路协同适配(A4)时, 模型取得了 84.7% 的最优表现, 这充分印证了本文的设计准则: 必须通过跨模块的端到端适配, 消除中层 Neck 的特征瓶颈与末端 Head 的判别鸿沟, 方能构建起从底层视觉感知到高层排版逻辑的完整语义映射链条。

### 5.2.2. Rank 与 Dropout 消融

**Table 3.** Ablation study results of LoRA injection positions

**表 3.** LoRA 注入位置消融实验结果

配置	注入模块	AP50	mAP@0.5:0.95
A1	仅 Head	68.1	44.8
A2	仅 Backbone (stage3~4)	14.6	5.2
A3	Backbone + Neck	73.6	50.4
A4	Backbone + Neck + Head (本文)	84.7	58.1

关于 Rank 维度与 Dropout 概率的消融实验(如表 4 所示)揭露了模型容量与泛化性能之间的微妙平衡。实验结果显示, 当低秩维度  $r$  过小时(如 B1 组配置), 模型假设空间严重受限, 导致其在 1000 张小样本约束下难以准确表征政务文档中复杂的逻辑拓扑关系, AP<sub>50</sub> 仅为 16.6%。而当进一步盲目提升 Rank (如 B3 组)时, 性能并未随参数数量的增加而线性提升, 反而因过高的参数自由度在有限数据下诱发了局部过拟合风险, 导致精度出现明显回落。在正则化策略方面, 实验证实了针对不同任务分支实施差异化 Dropout 的

必要性: B4 组(全支路无 Dropout)相较于本文方案(B2)出现了 14.6 个百分点的 AP<sub>50</sub> 下降, 证明在回归分支与 Neck 结构中引入 0.1 的 Dropout 能够显著平滑梯度波动并提升边界框定位的稳定性。然而, 分类分支(Cls Head)对特征完整度极度敏感, 过强的正则化干预会削弱模型对 Title 与 Header 等关键判别特征的捕捉。综上所述, 本文确定的非对称配置方案——即在分类分支采用高秩( $r = 128$ )且不设 Dropout, 而在其余支路保持低秩并辅以轻量级正则化——在确保决策边界清晰度的同时, 兼顾了空间定位的稳健性, 实现了模型性能的最优平衡。

**Table 4.** Ablation study results of LoRA rank and dropout configurations

**表 4.** LoRA rank 与 dropout 配置消融实验结果

配置	stage3 r	stage4 r	neck r	cls r	reg r	dropout	AP50	mAP@0.5:0.95
B1 (小 rank)	64	8	12	64	16	0.1	16.6	6.2
B2 (本文)	128	16	18	128	28	0.1/0.0	84.7	58.1
B3 (大 rank)	192	32	32	192	48	0.1/0.0	78.7	50.6
B4 (无 dropout)	128	16	18	128	28	0.0	70.1	46.5

## 6. 结论与未来工作

本文面向 1000 张图像、8 类目标的小样本检测场景, 以 PP-PicoDet 为基线, 研究并实现了 LoRA 参数高效微调方案。通过在主干高层、颈部融合与检测头关键卷积中注入低秩可训练模块, 并采用分层 rank 与正则化配置, 本文方法在保持推理结构不变的前提下, 实现了更好的精度 - 参数折中与更稳定的训练收敛。消融实验进一步验证了“Backbone + Neck + Head 协同适配”及“分类分支更强、回归与融合分支更稳”的设计规律。未来工作将结合更复杂的长尾分布与跨场景持续学习需求, 进一步提升方法的可扩展性与工程鲁棒性。

## 参考文献

- [1] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2117-2125. <https://doi.org/10.1109/cvpr.2017.106>
- [2] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [3] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 1-26 July 2017, 7263-7271. <https://doi.org/10.1109/cvpr.2017.690>
- [4] Law, H. and Deng, J. (2018) Cornered: Detecting Objects as Paired Keypoints. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 765-781. [https://doi.org/10.1007/978-3-030-01264-9\\_45](https://doi.org/10.1007/978-3-030-01264-9_45)
- [5] Zhou, X., Wang, D. and Krähenbühl, P. (2019) Objects as Points. arXiv:1904.07850.
- [6] Tian, Z., Shen, C., Chen, H. and He, T. (2019) FCOS: Fully Convolutional One-Stage Object Detection. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October 2019-2 November 2019, 9627-9636. <https://doi.org/10.1109/iccv.2019.00972>
- [7] Zhang, S., Chi, C., Yao, Y., Lei, Z. and Li, S.Z. (2020) Bridging the Gap between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 9759-9768. <https://doi.org/10.1109/cvpr42600.2020.00978>
- [8] Li, X., Wang, W., Wu, L., et al. (2020) Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *Advances in Neural Information Processing Systems*, **33**, 21002-21012.
- [9] Ge, Z., Liu, S., Wang, F., et al. (2021) YOLOX: Exceeding Yolo Series in 2021. arXiv:2107.08430.

- [10] Ghiasi, G., Lin, T. and Le, Q.V. (2019) NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 7036-7045. <https://doi.org/10.1109/cvpr.2019.00720>
- [11] Xiong, Y., Liu, H., Gupta, S., *et al.* (2020) MobileDets: Searching for Object Detection Architectures for Mobile Accelerators. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11958-11967.
- [12] Hu, E.J., Shen, Y., Wallis, P., *et al.* (2021) LoRA: Low-Rank Adaptation of Large Language Models. arXiv:2106.09685.
- [13] Zhang, H., Wang, Y., Dayoub, F. and Sunderhauf, N. (2021) Varifocalnet: An IoU-Aware Dense Object Detector. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 8514-8523. <https://doi.org/10.1109/cvpr46437.2021.00841>
- [14] Yu, G., Chang, Q., Lv, W., *et al.* (2021) PP-PicoDet: A Better Real-Time Object Detector on Mobile Devices. arXiv:2111.00902.