

基于深度学习的骨髓血细胞检测与 识别技术研究

(申请清华大学电子信息硕士专业学位论文)

培 养 单 位： 电子工程系

专 业 领 域： 电子信息

申 请 人： 孙 天 宇

指 导 教 师： 杨 健 教 授

二〇二三年四月

Research on bone marrow blood cell detection and recognition based on deep learning

Thesis submitted to
Tsinghua University
in partial fulfillment of the requirement
for the professional degree of
Master of Electronic and Information Engineering

by
Sun Tianyu
(Electronic and Information Engineering)

Thesis Supervisor: Professor Yang Jian

April, 2023

学位论文指导小组、公开评阅人和答辩委员会名单

指导小组名单

杨健	教授	清华大学
----	----	------

公开评阅人名单

刘 XX	教授	清华大学
杨 XX	研究员	中国 XXXX 科学院 XXXXXXXX 研究所

答辩委员会名单

主席 委员	赵 XX	教授	清华大学
	刘 XX	教授	清华大学
	杨 XX	研究员	中国 XXXX 科学院 XXXXXXX 研究所
	黄 XX	教授	XXXX 大学
秘书	周 XX	副教授	XXXX 大学
	吴 XX	博士	清华大学

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容；（3）按照上级教育主管部门督导、抽查等要求，报送相应的学位论文。

本人保证遵守上述规定。

作者签名： _____

导师签名： _____

日 期： _____

日 期： _____

摘 要

论文的摘要是对论文研究内容和成果的高度概括。摘要应对论文所研究的问题及其研究目的进行描述，对研究方法和过程进行简单介绍，对研究成果和所得结论进行概括。摘要应具有独立性和自明性，其内容应包含与论文全文同等量的主要信息。使读者即使不阅读全文，通过摘要就能了解论文的总体内容和主要成果。

论文摘要的书写应力求精确、简明。切忌写成对论文书写内容进行提要的形式，尤其要避免“第 1 章……；第 2 章……；……”这种或类似的陈述方式。

关键词是为了文献标引工作、用以表示全文主要内容信息的单词或术语。关键词不超过 5 个，每个关键词中间用分号分隔。

关键词：关键词 1；关键词 2；关键词 3；关键词 4；关键词 5

Abstract

An abstract of a dissertation is a summary and extraction of research work and contributions. Included in an abstract should be description of research topic and research objective, brief introduction to methodology and research process, and summary of conclusion and contributions of the research. An abstract should be characterized by independence and clarity and carry identical information with the dissertation. It should be such that the general idea and major contributions of the dissertation are conveyed without reading the dissertation.

An abstract should be concise and to the point. It is a misunderstanding to make an abstract an outline of the dissertation and words “the first chapter”, “the second chapter” and the like should be avoided in the abstract.

Keywords are terms used in a dissertation for indexing, reflecting core information of the dissertation. An abstract may contain a maximum of 5 keywords, with semi-colons used in between to separate one another.

Keywords: keyword 1; keyword 2; keyword 3; keyword 4; keyword 5

目 录

摘 要.....	I
Abstract.....	II
目 录.....	III
插图清单.....	IV
附表清单.....	V
符号和缩略语说明.....	VI
第 1 章 基于改进的 RetianNet 骨髓血细胞检测算法设计与实现	1
1.1 引言	1
1.2 改进的 RetinaNet 骨髓血细胞检测网络	1
1.2.1 基于全局注意的路径聚合网络	2
1.2.2 IOU 预测分支.....	4
1.2.3 训练标签分配策略.....	5
1.2.4 损失函数.....	10
1.3 算法实现与实验结果分析	10
1.3.1 实验环境.....	10
1.3.2 实验结果与分析.....	12
1.4 小结	14
参考文献.....	16
致 谢.....	17
声 明.....	18
个人简历、在学期间完成的相关学术成果.....	19
指导教师评语.....	20
答辩委员会决议书.....	21

插图清单

图 1.1	RetinaNet 基线模型检测错误示例	2
图 1.2	改进的 RetinaNet 网络结构示意图.....	2
图 1.3	路径聚合网络结构示意图	3
图 1.4	全局注意力模块	3
图 1.5	置信度高但交并比低的错误检测示例	5
图 1.6	交并比预测分支结构	5
图 1.7	自适应样本选择阈值计算示意图	7
图 1.8	基于最优运输的标签分配策略	9
图 1.9	骨髓血细胞数据集上的检测结果	14

附表清单

表 1.1	骨髓血细胞检测数据集分布	12
表 1.2	改进 RetinaNet 骨髓血细胞检测结果.....	13
表 1.3	不同目标检测方法在骨髓血细胞测试集上的性能对比	13
表 1.4	骨髓血细胞检测网络模块消融实验	15

符号和缩略语说明

AP	平均精度 (Average Precision)
BP	反向传播 (Back Propagation)
RPN	区域推荐网络 (Region Proposal Network)
CNN	卷积神经网络 (Convolutional Neural Network)
FPN	特征金字塔 (Feature Pyramid Network)
ROI	感兴趣区域 (Region Of Interest)
IOU	交并比 (Intersection Over Union)
ReLU	修正线性单元函数 (Rectified Linear Unit)
CE	交叉熵损失函数 (Cross Entropy)
ONNX	开放式神经网络交换 (Open Neural Network Exchange)
HTML	超文本标记语言 (Hypertext Markup Language)
CSS	层叠样式表 (Cascading Style Sheets)
DOM	文档对象模型 (Document Object Model)
ATSS	自适应样本选择 (Adaptive Training Sample Selection)
PAA	概率标签分配 (Probabilistic Anchor Assignment)
NMS	非极大值抑制 (Non-Maximum Suppresion)
OTA	最优输运标签分配 (Optimal Transport Assignment)
MLP	多层感知机 (Multi-Layer Perceptron)
MSA	多层自注意 (Multi-head Self-Attention)
SA	单头自注意 (Self Attention)
MAP	平均精度均值 (Mean Average Precision)
SGD	随机梯度下降 (Stochastic Gradient Descent)
BMCD	骨髓血细胞数据集 (Bone Marrow Cell Dataset)
ATSS	自适应样本选择 (Adaptive Training Sample Selection)
LN	层标准化 (Layer Normalization)
Faster-RCNN	快速区域卷积神经网络
	矩阵行列式或绝对值

第1章 基于改进的 RetianNet 骨髓血细胞检测算法设计与实现

1.1 引言

第三章中,我们对比了多个单阶段与双阶段检测网络在骨髓血细胞数据上的性能,综合考虑计算量,参数量与检测精度等因素,我们选择将性能优异的 RetinaNet 作为基线模型。此外我们确定了先检测再识别的骨髓血细胞处理流程,即检测网络只需要进行对血细胞进行前景检测与坐标回归。在 RetianNet 基线模型中,尽管模型的检测精度已经很高,但仍然存在着漏检、密集与重叠的血细胞区域边界检测错误等问题,如图 1.1所示。

骨髓血细胞检测主要有如下三个难点:(1)相比外周血红细胞、白细胞、血小板三类血细胞检测,骨髓血细胞种类繁多、形态丰富,尺寸大小不一。(2)在骨髓涂片制作过程中,由于染色剂与光照条件的变化,多个批次的数据存在着色彩差异。此外图像背景复杂,存在较多成熟红细胞的干扰。(3)对于骨髓细胞增生活跃的切片,存在大量血细胞密集堆叠,边缘黏连,易导致漏检、错检等问题。因此精准检测到骨髓血细胞是一项十分具有挑战性的课题。

针对上述难点与基线模型中存在的问题,本章提出了一种改进 RetinaNet 网络。该方法中,我们提出了一种基于全局注意力的自底向上的路径聚合网络结构,缩短了底层与顶层特征之间的信息传递路径,提升网络对位置特征的提取能力。此外探究了不同标签分配策略对检测性能的影响,提出基于最优运输的标签策略用于密集区域的血细胞的标签分配,避免了模糊分配样本的出现,提高网络对血细胞的召回能力。在骨髓血细胞数据集上的实验结果表明,本文提出的改进方法在检测准确率上有较大的提升,达到了较为先进的性能。

1.2 改进的 RetinaNet 骨髓血细胞检测网络

本章提出的改进 RetinaNet 网络结构如图 1.2所示,整体网络结构基于第三章的 RetinaNet 基线模型进行改进。骨干网络为 ResNet50 用于图像特征提取,特征金字塔结构用于多尺度特征提取。锚框的尺寸、数量与分类回归网络结果与基线模型相同。为了提高网络检测的精度,我们在特征金字塔后引入了自底向上的路径聚合模块,该模块基于全局注意力将更浅层的特征与 FPN 深层的进行融合,提升网络定位特征的表达能力,此外我们引入了 IOU 预测分支,将检测框的定位质

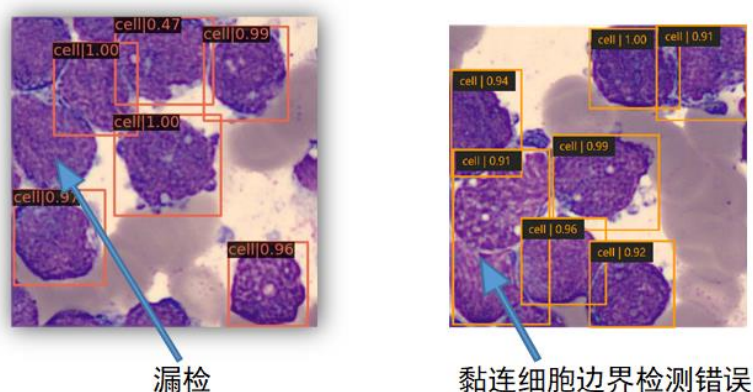


图 1.1 RetinaNet 基线模型检测错误示例

量也纳入到候选框的筛选中。在训练过程中，我们使用基于最优输运的策略进行标签分配。下面各个小节将详细介绍我们的改进点。

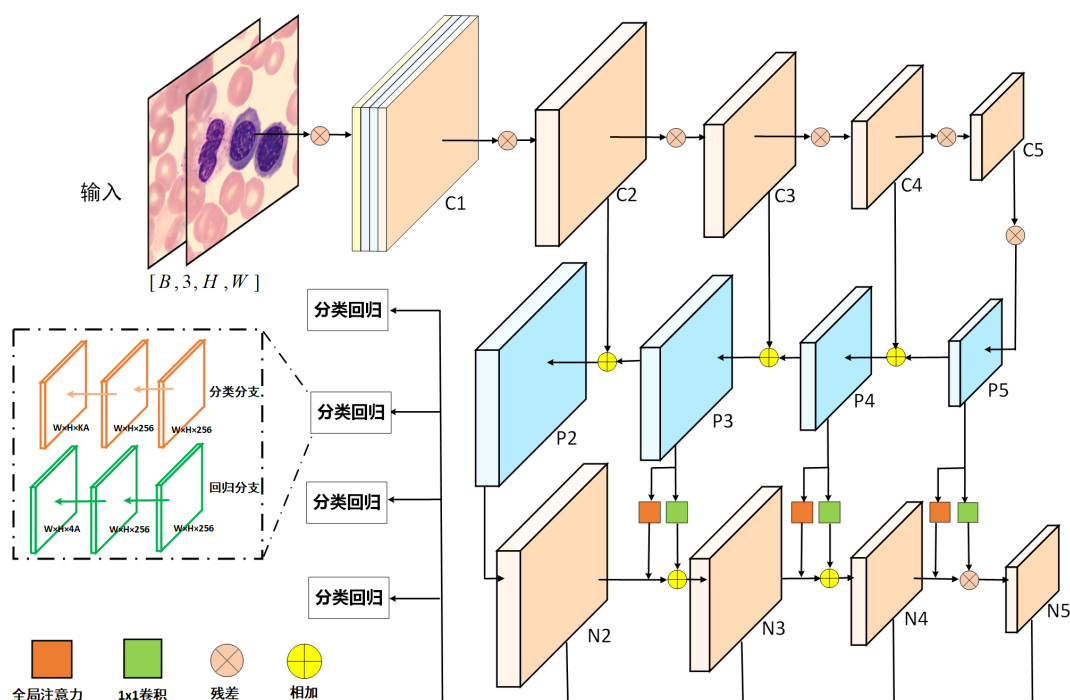


图 1.2 改进的 RetinaNet 网络结构示意图

1.2.1 基于全局注意的路径聚合网络

ResNet50 骨干网络提取了不同层级与尺度的特征图，高层次的特征图表达目标的抽象语义信息，低层次的特征图表达局部的纹理与模式信息，因此引入了特征金字塔网络，增加自上而下的路径来传播语义强的特征，从而增强所有层次特征的分类能力。在骨髓血细胞检测任务中，我们更加关注网络对于血细胞的定位的准确性，这些定位信息主要存在浅层特征图的边缘、纹理信息中。我们构建了自

底向上的路径聚合网络，将浅层的特征与特征金字塔深层的特征图进行融合，通过特征直连缩短了底层到顶层特征之间的信息传递路径。原始结构中底层到顶层需要需要约 50 层网络，如图 1.3 红线所示。自下而上的路径聚合网络引入了特征直连，从底层到顶层的路径只有不到 10 层，如图 1.3 绿线所示，该路径使得浅层的纹理等高分辨定位信息可以更有效的传递到顶层，提升网络定位特征的表达能力。

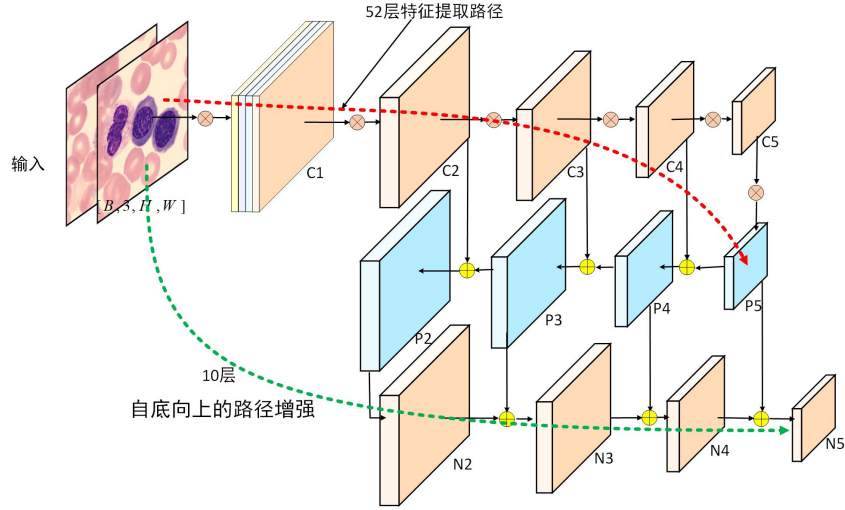


图 1.3 路径聚合网络结构示意图

图中 C_2, C_3, C_4, C_5 为 ResNet50 骨干网络不同阶段生成的特征图， P_2, P_3, P_4, P_5 为特征金字塔生成的不同级别的特征图。自底向上的路径聚合网络从最底层的 P_2 特征图开始，通过步长为 2 的卷积进行下采样并与高层的特征融合后生成新的特征图 N_2, N_3, N_4, N_5 。路径构建模块中使用了全局通道注意力机制^[1]，使用全局上下文信息的高层次特征指导浅层特征的筛选，该结构如图 1.4 所示。

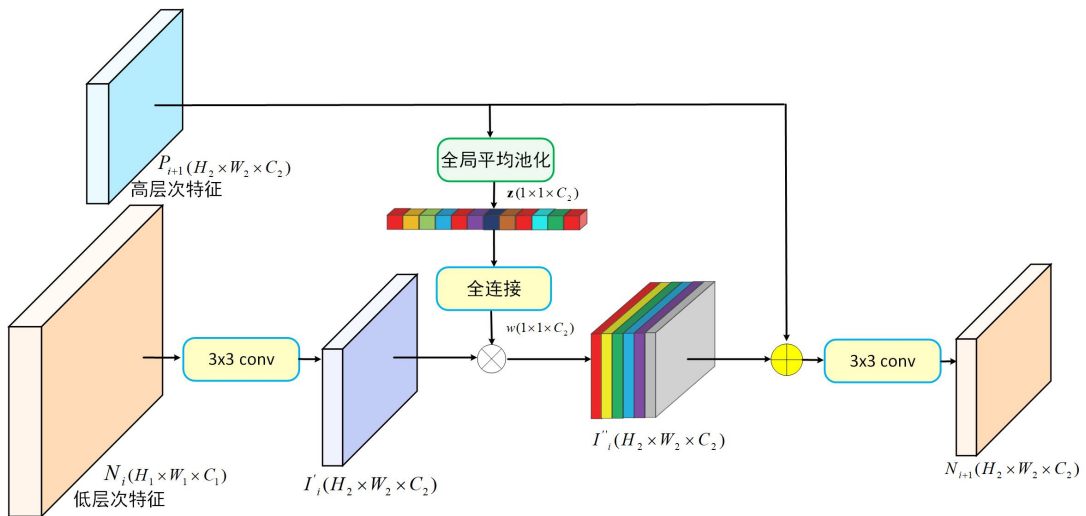


图 1.4 全局注意力模块

全局注意力模块的输入分别是路径聚合网络的低层级的特征图 $N_i(H_1 \times W_1 \times C_1)$ 与特征金字塔的高层级特征图 $P_{i+1}(H_2 \times W_2 \times C_2)$ ，输出新的特征图 N_{i+1} 。首先对特征图 N_i 经过 3×3 、步长为 2 的卷积层降低特征图的尺寸得到 $I'_i(H_2 \times W_2 \times C_2)$ 。对高层级的特征图 P_{i+1} 进行全局平均池化，每个通道压缩为一个值得到 C 维向量 $z(1 \times 1 \times C_2)$ ，如式 1.1 所示：

$$z = F_{sq}(P_{i+1}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W P_{i+1}(i, j) \quad (1.1)$$

然后使用两层全连接结构来全局建模通道之间的依赖关系，第一层全连接的输出使用 ReLU 激活函数，第二层使用 Sigmoid 激活函数，得到权重 $w(1 \times 1 \times C_2)$ ，如式 1.2 所示。

$$w = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (1.2)$$

将权重向量 w 与特征 $I'_i(H_2 \times W_2 \times C_2)$ 按通道点乘得到加权后的特征 $I''_i(H_2 \times W_2 \times C_2)$ ，如式 1.3 所示。

$$I''_i = F_{scale}(I'_i, w) = w_c I'_{ic} \quad (1.3)$$

将 I''_i 与 P_{i+1} 逐元素相加后再经过一个 3×3 卷积层后得到路径聚合网络的下一层特征图 N_{i+1} 。不断迭代上述过程，直到生成 N_5 特征图为止。最终在融合后新的特征图 N_2, N_3, N_4, N_5 上进行区域提取与坐标回归。**anchor** 生成与分类回归结构与 RetianNet 相同，在第三章已进行详细解释。

1.2.2 IOU 预测分支

RetinaNet 在预测阶段会生成密集的检测框，这些检测框会按照置信度高低进行非极大值抑制 (NMS)，去除重复的检测框。上述 NMS 方式默认了一种假设，就是置信度高的锚框定位也会更加精确。但是部分细胞例如杆状核细胞不服从中心分布，因此分类与回归这两种任务不一定严格正相关^[2]，如图 1.5 所示，图中 (a) 无论置信度分数高低，血细胞检测框的坐标都不准确，(b) 中置信度分数更高的检测框右侧与上侧的边界都是不准确的，而置信度较低检测框边界正确。

我们认为需要将检测框的定位质量也纳入到非极大值抑制的考量当中，在挑选分数最大的检测框时同时考虑置信度与交并比。但是在预测阶段，没有目标真实的坐标信息，因此无法使用交并比来判断每个检测框定位质量的好坏。我们在网络的定位部分额外扩展出了一个子分支来预测每一个锚框可能对应真实框的交并比。该分支与定位分支共享特征图信息，使用一个卷积层对于每个锚框输出一个标量值，然后使用 Sigmoid 激活函数进行激活去得到一个零一之间的交并比信

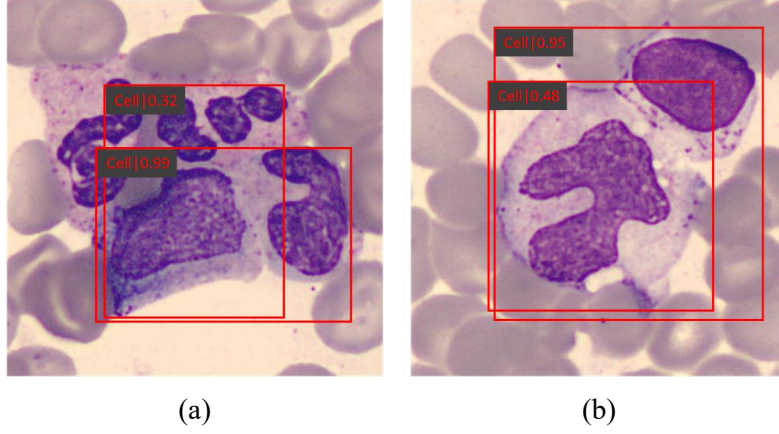


图 1.5 置信度高但交并比低的错误检测示例

息，修改后的网络结构如图 1.6 所示。图中（a）为 RetinaNet 的分类回归分支网络结构。图（b）中在回归分支加入了一个额外的卷积层，去预测交并比信息。

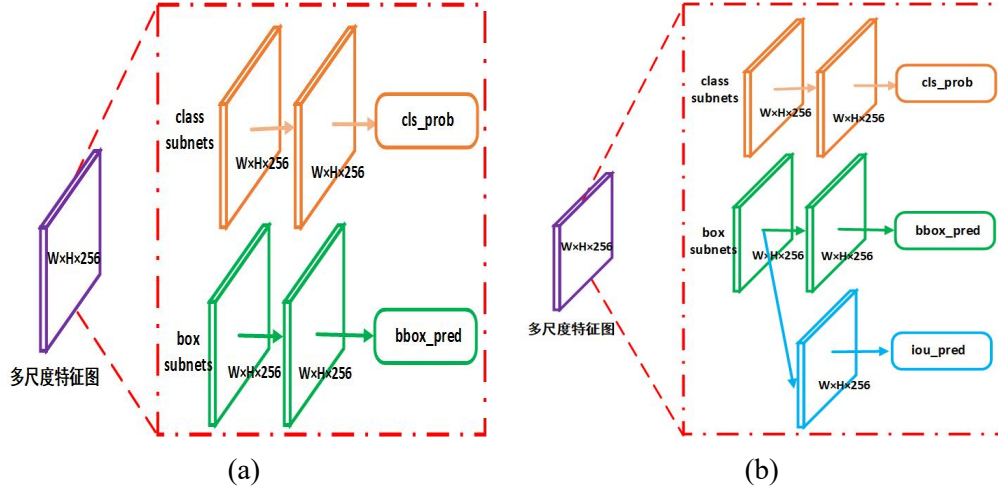


图 1.6 交并比预测分支结构

在训练过程，我们优化的目标有分类，坐标回归与交并比，损失函数如式 1.4 所示，其中 $L_{IoU}(iou_i, iou_i^*)$ 定义为预测 IoU 和真实 IoU 之间的二元交叉熵损失。

$$Loss = \sum_i L_{cls}(p_i, p_i^*) + \lambda_1 \sum_i p_i^* L_{reg}(t_i, t_i^*) + \lambda_2 \sum_i L_{IoU}(iou_i, iou_i^*) \quad (1.4)$$

1.2.3 训练标签分配策略

在目标检测网络训练过程中，标签分配（Label Assignment）是非常重要的一个流程。其目的是将训练样本划分为正负样本，并分配分类与回归的目标，计算其与真值之间的损失来监督训练。标签分配方式为网络训练提供了判别性的监督信号，决定了网络学习与收敛的方向，直接影响模型性能的好坏。本节介绍了我们训练过程中采用的不同标签分配策略，根据样本标签分配的正负权重设计，可

以将这些方法划分为硬标签分配方法与软标签分配方法。

1.2.3.1 硬标签分配策略

硬标签分配策略假设每个锚框非正即负，若 w_{pos}, w_{neg} 分别表示样本属于正负样本的权重，则硬标签划分可以表示为 $w_{pos}, w_{neg} \in \{0, 1\}$ 且 $w_{pos} + w_{neg} = 1$ 。这类方法的核心思想是找到一个最优划分边界，将锚框分割为正样本集合与负样本集合。边界划分规则可以分为静态规则与动态规则这两类。

1.2.3.2 最大交并比

RetinaNet 与 Faster-RCNN 等网络采用的是最常用的基于交并比最大化的标签分配策略（MaxIouAssigner）。该方法主要由以下的几个步骤：（1）初始化，将正样本集合 P 、负样本集合 N 设置为空集，将所有锚框设置为忽略样本；（2）计算多尺度特征图上所有锚框与所有真实框之间的交并比；（3）获取每个锚框交并比最大的 GT 框，如果交并比大于正样本阈值（ pos_thres ），则设置为正样本。如果小于负样本阈值（ neg_thres ），则设置为负样本。（4）如果 GT 框没有被锚框匹配到，则获得与 GT 框 IOU 最大的锚框，如果大于最小的正样本阈值，则将该锚框设置为正样本。

最大交并比标签分配策略是静态预先定义的标签分配方式。骨髓血细胞形状多变，比如杆状核粒细胞，通常呈现出 U 型，这导致目标的中心点通常为背景，并不能代表这个目标，而按照最大交并比的方式会被判定为正样本，导致网络的性能较差。

1.2.3.3 自适应样本选择

自适应样本选择（Adaptive Training Sample Selection, ATSS）方法^[3]基于 L_2 距离与交并比动态计算分割阈值，是一种自适应划分目标正负样本的标签分配策略。具体步骤如下：（1）对于网络输出的多个不同尺度特征图，在每个特征图上计算锚框中心坐标与目标中心坐标的 L_2 距离，选取 K 个 L_2 距离最小的锚框作为候选的正样本，如果有 L 个层级的特征图，那么可以得到 $K \times L$ 个候选正样本。（2）计算每个候选正样本与目标真实框的交并比，得到一组交并比的数据。计算这组交并比的均值 m_g 与标准差 v_g ，将均值与标准差相加，得到交并比的分割阈值 $t_g = m_g + v_g$ （3）在每个层级的特征图的候选正样本中根据阈值，选择真正的正样本加入正样本集合中进行训练。

图 1.7 为 ATSS 自适应计算阈值的示意图，柱状图的横坐标表示不同层级特征图，纵轴为交并比。柱子上的数字表示这个层级特征图上目标与最近 K 个锚框交

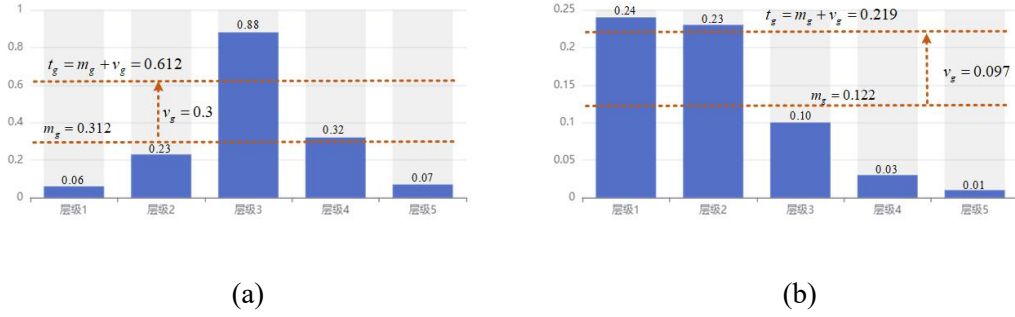


图 1.7 自适应样本选择阈值计算示意图

并比的均值。均值 m_g 代表锚框与目标真实框的匹配程度，如果 m_g 比较大，则候选正样本的质量都很好，可以适当提高阈值来挑选更好正样本。均值低则应当降低分割阈值。标准差 v_g 表示特征层次锚框与目标真实框的匹配程度。如果标准差比较高，则高质量的锚框集中在某一个层级的特征图中，应该提高阈值从最匹配的层级去挑选正样本。标准差比较低，说明所有层级的锚框匹配度都比较高，可以设置一个比较低的阈值广泛的从多个层级的特征图中选取。

1.2.3.4 概率标签分配

概率标签分配 (Probabilistic Anchor Assignment, PAA)^[4] 锚框的得分视为概率分布，并通过最大似然来估计分布参数，然后自适应的计算分割阈值，来选取正负样本。锚框的分数用来衡量其与真实框的相似性，包括了分类得分与定位得分。分类得分是分类分支输出的置信度 $P_i(cls|\theta)$ ，定位得分为预测框与真实框之间的交并比 $\text{IOU}(f(a|\theta), g)$ ，为平衡这两种得分的权重，引入 λ 参数。锚框得分如式 1.5所示：

$$S = P_i(cls|\theta) \times \text{IOU}(f(a|\theta), g)^\lambda \quad (1.5)$$

锚框可以分为正样本、负样本两组，因此可以使用双峰混合高斯分布来建模锚框的分数分布，如式 1.6所示。

$$P(a|g, \theta) = \phi_1 N(a; \mu_1, \sigma_1) + \phi_2 N(a; \mu_2, \sigma_2) \quad (1.6)$$

然后根据最大期望算法 (Expectation-Maximization, EM) 使得似然最大化。首先对参数 ϕ_i, μ_i, σ_i 进行随机初始化，在 E 步中计算 Q 函数如式 1.7所示，式中

x_1, x_2, \dots, x_n 为锚框的得分。

$$Q_i(z^{(i)} = k) = \frac{\phi_k \cdot \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{(x^{(i)} - \mu_k)^2}{2\sigma_k^2}\right]}{\sum_{k=1}^K \phi_k \cdot \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{(x^{(i)} - \mu_k)^2}{2\sigma_k^2}\right]} \quad (1.7)$$

在 M 步中根据式 1.8 计算混合高斯分布的参数 $\phi_1, \phi_2, \mu_1, \sigma_1, \mu_2, \sigma_2$ 。

概率标签分配的具体步骤如下：对于每个真实框，根据网络输出的置信度与坐标计算每个锚框的得分，在每个尺度的特征度选择 K 个得分最高的锚框。采用最大期望算法去估计这组锚框的双峰高斯分布参数，找到两个最高峰对应横坐标的分数 s_1, s_2 ，其中 $s_1 < s_2$ 。（3）将得分低于 s_1 的锚框作为负样本。得分位于 s_1 与 s_2 之间的锚框划分为忽略样本，得分大于 s_2 的锚框作为正样本。最后那些没有参与分配的锚框均视为负样本。

$$\begin{aligned} \mu_k &= \frac{\sum_{i=1}^N Q_k^{(i)} x^{(i)}}{N_k} \\ \sigma_k &= \frac{\sum_{i=1}^N Q_k^{(i)} (x^{(i)} - \mu_k) (x^{(i)} - \mu_k)^T}{N_k} \\ \phi_k &= \frac{\sum_{i=1}^N Q_k^{(i)}}{N} \\ N_k &= \sum Q_k^N Q_k^{(i)} \end{aligned} \quad (1.8)$$

1.2.3.5 最优输运标签分配

最优输运^[5]标签分配将目标检测中的标签分配问题建模为将标签从真实框输运到锚框代价最小的最优输运策略（Optimal Transport）问题，如图 1.8 所示。

最优输运问题的定义如下：假设有 m 个供应商与 n 个需求方，其中第 i 个供应商有 a_i 单元的货物，第 j 个需求方需要 b_j 个单元的货物。每个单元的货物从第 i 个供应商运输到第 j 个的需求方的输运代价是 C_{ij} 。最优输运的目标是找到一种输运计划 $\mathbf{P}^* = \{P_{i,j} \mid i = 1, 2, \dots, m, j = 1, 2, \dots, n\}$ 将供应商的货物全部运输到需求方，并使得运输的代价最小，其中 $P_{i,j}$ 表示第 i 个供应商运输到第 j 个的需求方单元货物的数量。

最优输运的优化目标如式 1.9 所示：

$$\min_{\mathbf{P}} \sum_{i=1}^m \sum_{j=1}^n C_{ij} P_{ij} \quad P_{ij} \geq 0, i = 1, 2, \dots, m, j = 1, 2, \dots, n. \quad (1.9)$$

其中第 i 的供应商输出的数量等于 s_i ，即 $\sum_{j=1}^n P_{ij} = a_i$ 。第 j 个需求方接

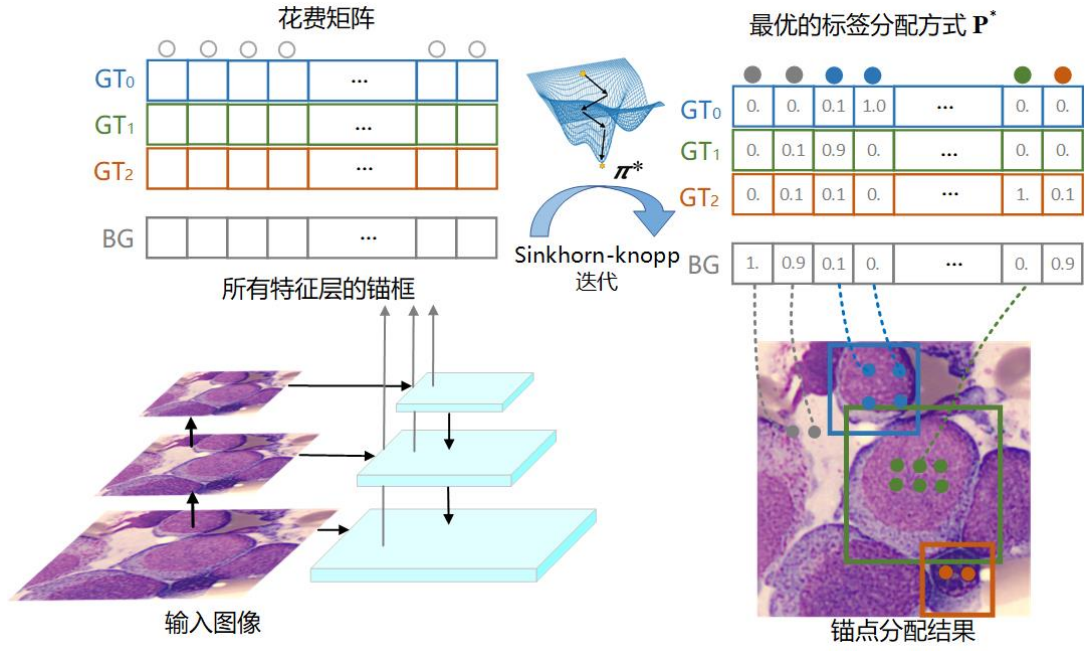


图 1.8 基于最优输运的标签分配策略

收的数量等于 b_j ，即 $\sum_{i=1}^m P_{ij} = b_j$ 。供应商供应的数量等于需求方接收的数量 $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$ 。

最优输运问题可以在多项式的时间复杂度内，通过 Sinkhorn 快速迭代的方法进行求解。该方法的思想源于交叉熵与对偶问题，定义 $\mathbf{H}(\mathbf{P}) \stackrel{\text{def.}}{=} -\sum_{i,j} \mathbf{P}_{i,j} (\log(\mathbf{P}_{i,j}) - 1)$ 上述问题添加了拉格朗日算子和扰动项的形式如式 1.10 所示：

$$E(\mathbf{P}, \mathbf{f}, \mathbf{g}) = \sum_{i=1}^m \sum_{j=1}^n C_{ij} P_{ij} - \epsilon \mathbf{H}(\mathbf{P}) + f_j \left(\sum_{i=1}^m P_{ij} - b_j \right) + g_i \left(\sum_{j=1}^n P_{ij} - a_i \right) \quad (1.10)$$

其中 ϵ 是一个常量超参数，控制熵正则化项的强度，实验中设置为 0.1， $f_j (j = 1, 2, \dots, n)$ 与 $g_i (i = 1, 2, \dots, m)$ 是拉格朗日乘子。通过使得目标函数导数等于零求得最优的输运策略如式 1.11 所示：

$$\frac{\partial \mathcal{E}(\mathbf{P}, \mathbf{f}, \mathbf{g})}{\partial P_{i,j}} = C_{i,j} + \epsilon \log(P_{i,j}) - f_j - g_i = 0 \quad (1.11)$$

因此最优输运策略为 \mathbf{P}^* 如式 1.12 所示：

$$P_{ij}^* = \exp\left(-\frac{f_j}{\epsilon}\right) \exp\left(-\frac{C_{ij}}{\epsilon}\right) \exp\left(-\frac{g_i}{\epsilon}\right) \quad (1.12)$$

令 $s_j = \exp\left(-\frac{f_j}{\epsilon}\right)$ ， $U_{ij} = \exp\left(-\frac{C_{ij}}{\epsilon}\right)$ ， $t_i = \exp\left(-\frac{g_i}{\epsilon}\right)$ 上述变量满足如式 1.13 的

约束关系:

$$\begin{aligned}\sum_i P_{ij} &= s_j \left(\sum_i U_{ij} t_i \right) = b_j \\ \sum_j P_{ij} &= (s_j \sum_i U_{ij}) t_i = a_i\end{aligned}\quad (1.13)$$

上式的约束关系需要同时满足, 因此 t_i 与 s_j 的交替迭代公式如下:

$$s_j^{l+1} = \frac{b_j}{\sum_i U_{ij} t_i^l}, \quad t_i^{l+1} = \frac{a_i}{\sum_j U_{ij} s_j^{l+1}} \quad (1.14)$$

在目标检测的背景下, 假设一张图片中有 m 个真实框, 检测网络所有的 FPN 层总共输出了 n 个锚框。将真实框视为供应商, 可以为 k 个锚框提供正样本的标签, 等价于有 k 个单元的货物 i.e., $a_i = k, i = 1, 2, \dots, m$ 。每个锚框视为需求方, 需要一个标签 i.e., $b_j = 1, j = 1, 2, \dots, n$ 。此外引入一个背景供应商, 来为锚框提供负样本标签, 数量为 $n - m \times k$ 。

将正样本标签从某一个真实框 gt_i 输运到一个锚框 $anchor_j$ 的代价定义为分类与回归的加权损失, 如式 1.15 所示:

$$C_{ij}^{fg} = L_{cls} \left(p_i^*, p_j^{cls}(\theta) \right) + \lambda L_{reg} \left(t_i^*, t_j^{box}(\theta) \right) \quad (1.15)$$

其中 θ 代表模型参数, $p_j^{cls}(\theta)$ 与 $t_j^{box}(\theta)$ 分别代表模型预测的分类得分与坐标回归值。 p_i^* 与 t_i^* 代表了真实类别与坐标。 L_{cls} 与 L_{reg} 为交叉熵损失函数与 GIOU 损失函数。

背景供应商将一个负样本标签传递到锚框的花费为分类损失, 如下式所示:

$$C_j^{bg} = L_{cls} \left(\Phi, p_j^{cls}(\theta) \right) \quad (1.16)$$

在定义好花费矩阵后, 最优输运的策略 \mathbf{P}^* 可以通过 Sinkhorn-Knopp 迭代算法进行求解。每个 anchor 的标签为最大的标签所对应的供应商类别。算法流程如算法 1.1 所示。

1.2.4 损失函数

1.3 算法实现与实验结果分析

1.3.1 实验环境

训练平台使用 linux, 操作系统为 ubuntu20.04, GPU 为 NVIDIA RTX3090, CPU 为 Intel(R) Core(TM) i9-10940X。深度学习框架为 Pytorch1.13.1, 训练的批量大小设置为 16。使用 SGD 随机梯度下降优化器对网络参数进行更新, 动量设置为 0.9, 权重衰减设置为 $5e-4$ 。初始化的学习率为 0.001。网络总共训练 36 个轮次, 在第 16 与第 28 个轮次, 学习率变为原来的 1/10。RetianNet 中 Focal Loss 中正负样本平

算法 1.1 最优输运标签分配**输入:**

- I : 输入图像
- A : 网络输出的一组锚框
- G : 图像中目标真实框标注
- ε : Sinkhorn-Knopp 迭代熵正则化项
- T : Sinkhorn-Knopp 迭代次数
- λ : 式 1.15 中的平衡因子

输出: \mathbf{P}^* 标签最优输运策略

- 1: set $m = |G|, n = |A|$;
- 2: 网络前向计算每个锚框的分数与坐标 P^{cls}, P^{box}
- 3: 动态计算每个 GT 框的 a_i
- 4: 计算得到背景供应商的标签数量 $a_{m+1} = n - \sum_{i=1}^m a_i$
- 5: $b_j (j = 1, 2 \dots n)$ 使用 1 初始化
- 6: 前景分类损失 $C_{ij}^{cls} = FocalLoss(P_j^{cls}, G_i^{cls})$
- 7: 前景回归损失 $C_{ij}^{reg} = IoULoss(P_j^{box}, G_i^{box})$
- 8: 背景花费 $C_{bg}^{cls} = FocalLoss(P_j^{cls}, \Phi)$
- 9: 前景花费 $C_{fg} = C^{cls} + \lambda C^{reg}$
- 10: 计算最终的花费矩阵将 C_{bg} 拼接到 C_{fg} 的最后一行
- 11: 对 s^0, t^0 进行随机初始化
- 12: **for** $i = 1; i < T; i++$ **do**
- 13: 计算 $s^{l+1}, t^{l+1} \leftarrow SinkhornIter(C, s^l, t^l, a, b)$
- 14: **end for**
- 15: 根据式 1.12, 计算并返回最优的标签分配策略 \mathbf{P}^*

衡参数 $\alpha = 0.25$, 难易调整参数 $\gamma = 2$ 。主干网络使用在 coco 数据集上预训练的 ResNet50 网络。

1.3.1.1 数据集介绍

骨髓血细胞图像来自邃蓝智能科技(上海)有限公司合作医院提供, 首先采用第 2.1 小节阐述的主动学习标注策略进行边界框的标注。我们总共标记了 6821 张血细胞图像, 训练集与测试集按照 4:1 的比例进行随机划分, 训练集包含了 5456 张图像, 测试集包含了 1365 张图像。通常每个图像中包含 1 到 10 个有核血细胞, 数据集总共标记了 11352 个血细胞, 训练集有 9065 个血细胞, 测试集有 2287 个血细胞。数据集的分布如表 1.1 所示:

表 1.1 骨髓血细胞检测数据集分布

序号	类别名	类别简写	训练集数量	测试集数量
1	原始细胞	Prim	1856	467
2	淋巴细胞	Lym	996	226
3	单核细胞	Mono	206	52
4	浆细胞	Plas	272	70
5	红细胞	Red	1880	503
6	早幼粒细胞	Promy	357	107
7	嗜中性中幼粒细胞	Myelo	701	150
8	嗜中性晚幼粒细胞	Late	503	144
9	嗜中性杆状核细胞	Rods	998	241
10	嗜中分叶核细胞	Lobu	821	195
11	嗜酸性粒细胞	Eosl	475	132
总计			9065	2287

1.3.2 实验结果与分析

1.3.2.1 评价指标

在??节中，我们提出了先检测、再识别的骨髓血细胞计算处理流程。在本节中，我们只关注血细胞检测的准确性，网络性能评估指标采用准确率、召回率、平均精度均值（mAP）与 PR 曲线，具体定义见 ??节。

1.3.2.2 实验结果

为了验证本节提出的改进 RetinaNet 方法在骨髓血细胞检测上的有效性，我们上节所提的骨髓血细胞数据集上进行测试，在不同交并比阈值条件下，网络的准确率、召回率、f1-score、平均检测精度均值与 FPS 如表 1.2所示，其中准确率、召回率为 f1-score 最大条件下对应的数值，FPS 是网络每秒可计算的图像帧数量。

从表中我们可以看到，本节提出的改进模型效果较好，在交并比为 0.75 的条件下，f1-score 为 0.9474，召回率为 0.950，准确率为 0.9448。随着交并比的逐渐增大，网络的准确率与召回率均开始下降，在交并比为 0.95 时，下降最为明显，平均检测精度只有 0.176。因此，对于高精度的定位（IOU>0.9），模型检测性能仍有待提升，我们认为未来可以在数据量上进行扩充，此外在标注上进行更加精细化的校对。

表 1.2 改进 RetinaNet 骨髓血细胞检测结果

IOU	准确率	召回率	f1-score	mAP	FPS
0.5	0.9624	0.96	0.9612	0.9864	27.2
0.55	0.9594	0.96	0.9597	0.9837	27.2
0.6	0.9594	0.96	0.9597	0.9837	27.2
0.65	0.9579	0.96	0.9589	0.9821	27.2
0.7	0.9559	0.95	0.9529	0.9735	27.2
0.75	0.9448	0.95	0.9474	0.9668	27.2
0.8	0.931	0.92	0.9255	0.9328	27.2
0.85	0.8738	0.86	0.8668	0.8463	27.2
0.9	0.7256	0.72	0.7228	0.6305	27.2
0.95	0.3537	0.33	0.3414	0.1766	27.2

此外,我们对比了 SSD512^[6]、Faster-RCNN^[7]、Cascade-RCNN^[8]、RetinaNet^[9]、YOLOV3^[10]与 DETR^[11]等六种基于深度学习的主流目标检测网络,这些网络均使用在 COCO 数据集上预训练模型进行参数初始化,然后基于迁移学习在骨髓血细胞检测数据集上进行微调。表 1.3为不网络的平均检测精度对比,其中 AP_{50} 表示在交并比为 0.5 条件下的平均检测精度。

表 1.3 不同目标检测方法在骨髓血细胞测试集上的性能对比

方法	主干网络	AP_{50}	AP_{75}	AP_{90}	参数量大小 (MB)	运算次数 (GFLOPs)
SSD512	VGG16	0.9748	0.9178	0.3234	36.04	250.92
Faster-RCNN	ResNet50	0.9819	0.9458	0.4965	41.17	139.25
Cascade-RCNN	ResNet50	0.9806	0.9556	0.5896	69.17	167.24
RetinaNet	ResNet50	0.9801	0.9542	0.5215	36.31	135.73
YOLOV3	DarkNet53	0.9794	0.9508	0.4646	61.95	127.11
DETR	ResNet50	0.9888	0.9559	0.5844	41.30	59.82
本节提出的网络	ResNet50	0.9864	0.9668	0.6305	37.89	140.94

从表中我们可以看到, SSD512 网络检测精度最差,这是由于 VGG16 骨干网络特征表达能力相比与 ResNet 等残差网络更差。在交并比为 0.5 的条件下, DERT 网络的平均精度最高,本节提出的网络略低于 DETR 方法。在交并比为 0.75 与 0.9 的条件下,本节的方法均高于其他对比方法,特别是在交并比为 0.9 的条件下,本节方

法相比与次优的 Cascade-RCNN 方法提升了约 6.93%。该结果说明 Cascade-RCNN 通过级联 RPN 网络的方式，可以有效的提升候选框的定位质量，而本文提出的改进点将 IOU 交并比纳入到检测网络的推理阶段，也可以帮助网络找到定位精度更好的候选框。在网络的计算量与参数量方面，DETR 的计算量最低，SSD512 的参数量最少，本文提出的改进方法相比与基础 RetinaNet 方法计算量增加了 3.8%，参数量增加了 1.58MB。总而言之，本节提出的改进 RetinaNet 方法可以有效提高骨髓血细胞的检测精度。

本节提出的改进模型在骨髓血细胞数据集上检测的可视化结果如图 1.9 所示，图中将置信度阈值大于 0.5 的骨髓血细胞以红色检测框标出，左上角的数字为置信度阈值，从图中我们可以看到，该网络可以有效的对密集与黏连的血细胞区域进行检测。

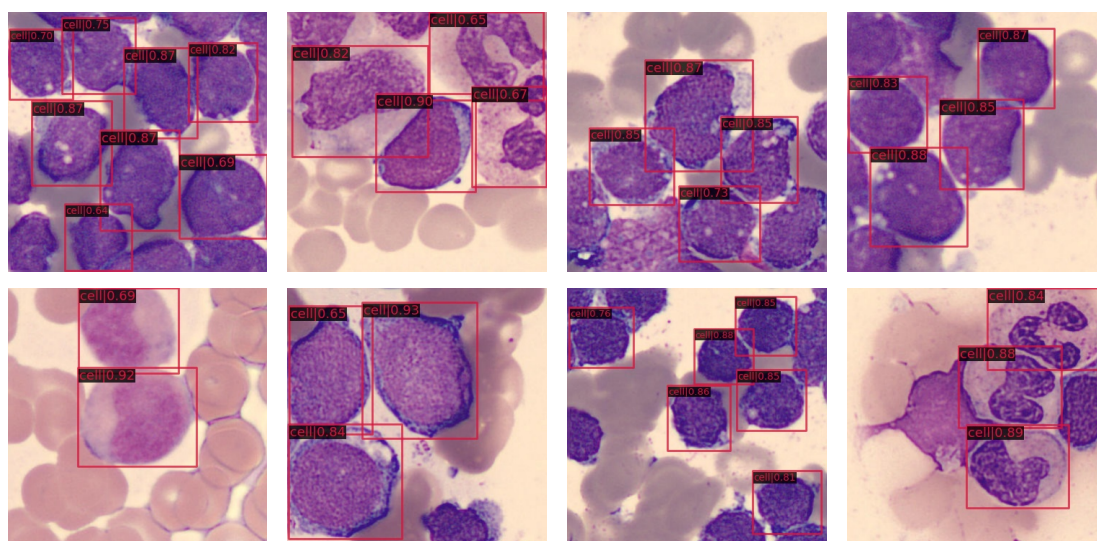


图 1.9 骨髓血细胞数据集上的检测结果

1.3.2.3 改进网络结构

为了验证本节提出了全局注意力路径聚合网络模块与 IOU 预测分支网络模块的有效性，我们分别在基线模型上添加单一模块来验证这两种模块的有效性。不同模型在骨髓血细胞测试集的结果如表 1.4 所示， AP_{50} 、 AP_{75} 与 AP_{90} 分别表示交并比为 0.5、0.75 与 0.9 条件下的平均检测精度。√ 代表网络添加该模块，× 代表网络不包含该模块。

1.3.2.4 标签分配策略

1.4 小结

表 1.4 骨髓血细胞检测网络模块消融实验

主干网络	全局注意力路径聚合模块	IOU 预测模块	AP_{50}	AP_{75}	AP_{90}
RetinaNet	×	×	0.9178	0.3234	36.04
RetinaNet	√	×	0.9458	0.4965	41.17
RetinaNet	×	√	0.9556	0.5896	69.17
RetinaNet	√	√	0.9542	0.5215	36.31

参考文献

- [1] Li H, Xiong P, An J, et al. Pyramid attention network for semantic segmentation[A]. 2018.
- [2] He Y, Zhu C, Wang J, et al. Bounding box regression with uncertainty for accurate object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2888-2897.
- [3] Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9759-9768.
- [4] Kim K, Lee H S. Probabilistic anchor assignment with iou prediction for object detection[C]//Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16. Springer, 2020: 355-371.
- [5] Villani C, et al. Optimal transport: old and new: volume 338[M]. Springer, 2009.
- [6] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, 2016: 21-37.
- [7] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
- [8] Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.
- [9] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
- [10] Redmon J, Farhadi A. Yolov3: An incremental improvement[A]. 2018.
- [11] Zhu X, Su W, Lu L, et al. Deformable detr: Deformable transformers for end-to-end object detection[A]. 2020.

致 谢

衷心感谢导师 ××× 教授和物理系 ×× 副教授对本人的精心指导。他们的言传身教将使我终生受益。

在美国麻省理工学院化学系进行九个月的合作研究期间，承蒙 Robert Field 教授热心指导与帮助，不胜感激。

感谢 ××××× 实验室主任 ××× 教授，以及实验室全体老师和同窗们学的热情帮助和支持！

本课题承蒙国家自然科学基金资助，特此致谢。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

个人简历、在学期间完成的相关学术成果

个人简历

1998 年 08 月 30 日出生于内蒙古自治区赤峰市。

2016 年 9 月考入清华大学电子工程系电子信息科学与技术专业，2020 年 7 月本科毕业并获得工学学士学位。

2020 年 9 月免试进入清华大学电子工程系攻读电子信息专业硕士至今。

在学期间完成的相关学术成果

学术论文：

- [1] 孙天宇, 朱庆涛, 杨健, 曾亮. 基于改进 Vision Transformer 的血细胞图像识别方法研究 [J]. 生物医学工程学杂志, 2022, 39(6):1097-1107(EI, CSCD)

指导教师评语

论文提出了……

答辩委员会决议书

论文提出了……

论文取得的主要创新性成果包括：

1. ……

2. ……

3. ……

论文工作表明作者在 ××××× 具有 ××××× 知识，具有 ×××× 能力，论文 ××××，
答辩 ××××。

答辩委员会表决，（× 票/一致）同意通过论文答辩，并建议授予 ×××（姓名）
×××（门类）学博士/硕士学位。