

# 改进 YOLOv3 遥感小目标检测算法<sup>①</sup>

许成林, 黄宇博, 赵 舵

(西南交通大学, 成都 611756)

通信作者: 赵 舵, E-mail: zhaoduo@swjtu.edu.cn



**摘 要:** 针对目前在遥感目标检测领域广泛使用的 YOLOv3 算法存在对小目标物体的特征表达能力不足, 检测效果不好的问题, 本文提出一种改进的 YOLOv3 小目标检测算法. 首先, 引入全局信息注意力机制并改进特征提取网络和特征金字塔结构, 提高模型小目标特征提取能力和检测能力; 其次, 对数据集进行单尺度 Retinex 融合特征增强, 提高模型对小目标特征的学习效果; 最后, 使用自适应锚框优化算法对 anchors 进行优化, 提高 anchors 和目标的匹配程度. 选用遥感数据集 RSOD 进行实验, 本文算法的全类平均精度为 92.5%, 相比经典 YOLOv3 算法, 提高 10.1%, 对遥感小目标的检测效果得到明显提升.

**关键词:** 小目标检测; YOLOv3; 特征增强; 锚框优化; 深度学习

引用格式: 许成林, 黄宇博, 赵舵. 改进 YOLOv3 遥感小目标检测算法. 计算机系统应用, 2023, 32(7): 179-187. <http://www.c-s-a.org.cn/1003-3254/9179.html>

## Improved YOLOv3 Algorithm for Remote Sensing Detection of Small Targets

XU Cheng-Lin, HUANG Yu-Bo, ZHAO Duo

(Southwest Jiaotong University, Chengdu 611756, China)

**Abstract:** As YOLOv3, an algorithm widely used in the field of remote sensing target detection, has insufficient feature expression ability for small targets and a poor detection effect, an improved YOLOv3 algorithm for small target detection is proposed. Firstly, the global context (GC) attention mechanism is introduced, and the feature extraction network and feature pyramid networks (FPN) are improved to enhance the small-target feature extraction ability and detection ability of the model. Secondly, single-scale Retinex (SSR) fusion feature enhancement is applied to the dataset to improve the model's learning effect of small target features. Finally, the adaptive anchor box optimization (AABO) algorithm is adopted to optimize anchors and better match anchors and targets. The experimental results on the remote sensing dataset RSOD show that the mean average precision (*mAP*) of the proposed algorithm is 92.5%, which is improved by 10.1% compared with that of the classic YOLOv3 algorithm, and the detection effect of small remote sensing targets is significantly improved.

**Key words:** small target detection; YOLOv3; feature enhancement; anchor optimization; deep learning

目标检测任务是当前计算机视觉领域研究的核心课题之一, 基本任务是找到感兴趣的目标确定其位置和其所属类别. 传统的目标检测算法主要通过手动设计准确的特征和建立准确的数学模型来实现对各种物体的检测, 如基于滑动窗口的 VJ Det 算法<sup>[1]</sup>和基于行

人特征的 HOG Det 算法<sup>[2]</sup>等. 随着深度学习算法和相关硬件的发展, 目标检测开始进入深度学习的时代. 深度学习下的目标检测主要分为以 YOLO 系列为代表的单阶段 (one-stage) 和以 RCNN 系列为代表的两阶段 (two-stage) 方法<sup>[3]</sup>. 基于深度学习的目标检测算法由于

① 收稿时间: 2023-01-05; 修改时间: 2023-02-03; 采用时间: 2023-02-27; csa 在线出版时间: 2023-04-28

CNKI 网络首发时间: 2023-04-28

其能够对目标特征自动提取而无需手动设计特征,且算法适应性高,能够进行迁移学习,因此被广泛运用于人脸识别、车辆检测、医学影像等领域中.虽然深度学习技术给目标检测带来了突破性的进展,但是在实际任务中,往往需要对小目标进行检测.由于小目标在整幅图像中面积占比小、分辨率低和特征不显著等原因,现有的各种目标检测框架都不能够很好地对小目标物体进行检测<sup>[4]</sup>.许多科研工作者基于已有的各种目标检测框架提出了诸多优化改进方法,以期提高现有检测框架的小目标检测能力.徐思源等人主要通过对小目标数据集进行聚类分析以及改进YOLOv3特征提取结构的方法来提高检测网络对小目标物体的检测能力<sup>[5]</sup>.刘峻文等人将YOLOv3特征提取网络替换为基于ResNet和DenseNet的特征提取网络RD-Net,并引入混合空洞卷积,提高检测框架对小目标的敏感程度<sup>[6]</sup>.王建军等人根据不同尺度特征的重要性对YOLOv3的主干网络进行重构,并引入RFB感受野模块增大特征图的感受野,提高检测框架对小目标的检测效果<sup>[7]</sup>.

本文针对小目标检测任务,对经典YOLOv3检测框架进行改进优化.首先,调整YOLOv3特征提取网络的残差模块数量,在保证不同尺度下网络特征提取能力不变的前提下扩充特征金字塔(feature pyramid networks, FPN)层数,增大输出特征图的感受野,引入全局信息注意力机制(global context, GC),提高上下文特征捕获能力,增强模型对小目标的识别精确度.其次,基于Retinex理论对训练集小目标进行随机数量的单尺度Retinex(single-scale Retinex, SSR)特征融合增强,提高数据集中小目标的特征信息量.最后,结合bounding box在数据集上的分布特性,使用自适应锚框优化算法,对anchors尺度和长宽比例进行优化,提高模型对小目标的匹配程度.

### 1 YOLOv3 算法简介

YOLOv3算法是单阶段目标检测算法的核心代表之一.YOLOv3模型由特征提取网络Darknet-53和特征金字塔FPN组成.其中特征提取网络结构如图1所示<sup>[8]</sup>.假设输入图片尺寸大小为416×416,Darknet-53将会对图片进行8倍、16倍和32倍的下采样,分别得到大小为52×52、26×26和13×13的3层特征图.

将得到的3张特征图经过特征融合模块FPN的处理,每个特征图都充分融合了自身特征信息和前面

位置的特征信息,输出大小为:(13, 13, 512)、(26, 26, 256)、(52, 52, 128)三张特征图,每个特征图经过YOLO探测头,最终得到大小为:(13, 13, 3×(num\_classes))、(26, 26, 3×(num\_classes+5))和(52, 52, 3×(num\_classes+5))的3张预测图.其中3表示每个特征图中一个特征点对应于3个anchor,num\_classes表示数据集要预测的目标种类数量,5表示预测框的置信程度和4个参数(预测框中心坐标x,y和宽高w,h).

Type	Filters	Size	Output	
	卷积层	32	3×3	256×256
	卷积层	64	3×3/2	128×128
1×	卷积层	32	1×1	128×128
	卷积层	64	3×3	
	残差层			
	卷积层	128	3×3/2	64×64
2×	卷积层	64	1×1	64×64
	卷积层	128	3×3	
	残差层			
	卷积层	256	3×3/2	32×32
8×	卷积层	128	1×1	32×32
	卷积层	256	3×3	
	残差层			
	卷积层	512	3×3/2	16×16
8×	卷积层	256	1×1	16×16
	卷积层	512	3×3	
	残差层			
	卷积层	1024	3×3/2	8×8
4×	卷积层	512	1×1	8×8
	卷积层	1024	3×3	
	残差层			

图1 YOLOv3特征提取网络Darknet-53

YOLOv3的损失函数由预测边界框回归损失项、预测置信度损失项和预测类别损失项组成.如式(1)所示:

$$\begin{aligned}
 loss = & \lambda_{\text{coord}} \sum_{i=0}^{S \times S} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_i \times h_i) [(x_i - x^*)^2 + (y_i - y^*)^2] \\
 & + \lambda_{\text{coord}} \sum_{i=0}^{S \times S} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_i \times h_i) [(w_i - w^*)^2 + (h_i - h^*)^2] \\
 & - \sum_{i=0}^{S \times S} \sum_{j=0}^M I_{ij}^{\text{obj}} [C_i^* \log(C_i) + (1 - C_i^*) \log(1 - C_i)] \\
 & - \lambda_{\text{noobj}} \sum_{i=0}^{S \times S} \sum_{j=0}^M I_{ij}^{\text{noobj}} [C_i^* \log(C_i) + (1 - C_i^*) \log(1 - C_i)] \\
 & - \sum_{i=0}^{S \times S} I_{ij}^{\text{obj}} \sum_{c \in \text{class}} [p_i^*(c) \log(p_i(c)) + (1 - p_i^*(c)) \log(1 - p_i(c))]
 \end{aligned} \tag{1}$$

## 2 改进的 YOLOv3 算法

### 2.1 骨干网络的改进与多尺度 FPN 的改进

利用深度学习技术进行小目标检测的一大难点是:小目标(将物体标注框的长宽乘积除以整个图像的长宽乘积,再开根号,如果结果小于3%,则定义为小目标)本身分辨率较低,其像素特征在经过以卷积操作为主的神经网络特征提取器后,基本已经出现较大损失.如果仅仅考虑局部小目标图像的特征提取不难得出,在经过多次卷积后,其特征权重随着卷积次数的增加呈指数下降.特征提取器中越靠后的卷积操作得到的特征图基本已经丧失小目标特征信息,因此无论后续

经过多少次网络学习,算法模型都无法对小目标检测达到一个较好的效果.

基于上述分析,本文对 YOLOv3 网络结构作出如图 2 所示的改进.本文延续 Darknet-53 整体结构特点,首先对残差模块数量和排列作出调整,将 Darknet-53 中最开始的两个残差模块数量分别扩大两倍,得到 Res2、Res4,并将 Res4 的输出作为第 1 个特征图输出,参与后续的特征融合.这样的设计在第 1 次特征引出时,减少了一次残差操作,网络对浅层信息的提取保留水平得到提高,然后将 Res6 模块修改为 Res8 模块,保证模型对大中型目标特征提取的完善程度.

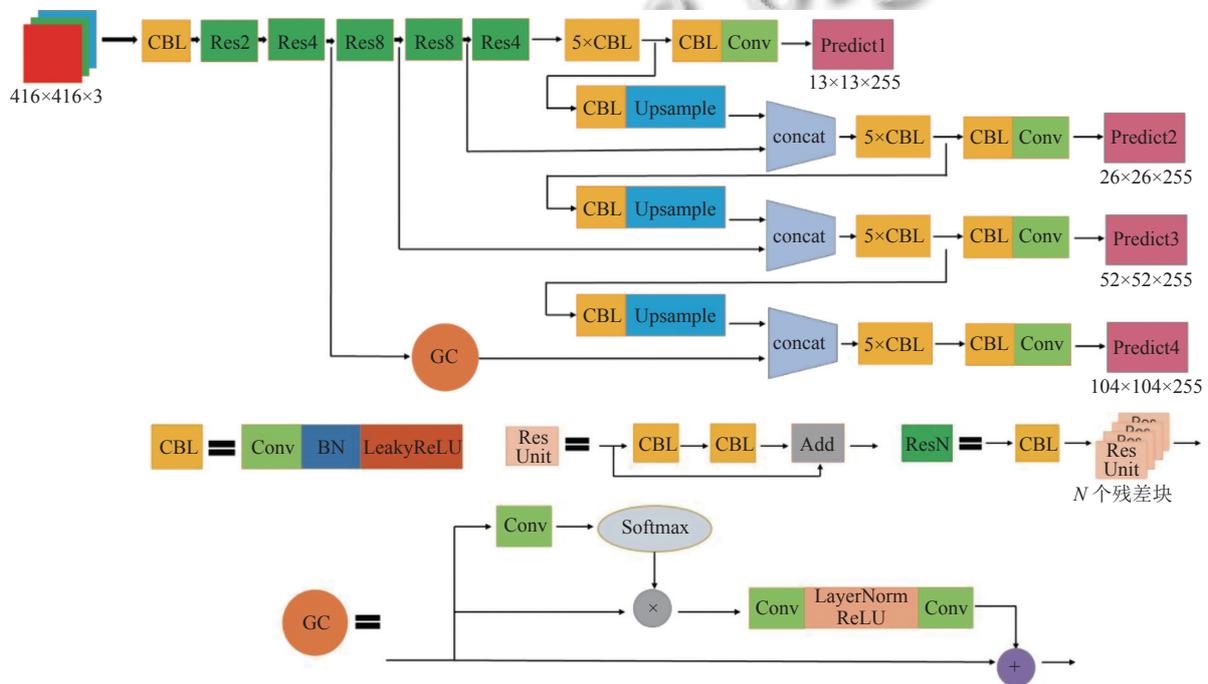


图 2 改进的 YOLOv3 算法模型结构

GC 注意力机制融合 SENet 和 NLNet 两大经典注意力机制的优点,能够充分挖掘上下文信息<sup>[9]</sup>,同时相对于其他注意力机制轻量化程度高,对于模型速度的影响较低.本文算法在第 1 个特征图进行融合操作之前引入 GC 注意力机制,挖掘小目标与背景的差异性特征,提高模型对于小目标检测的识别准确率和定位准确程度.

使用 YOLOv3 模型进行目标检测的输出结果是 FPN 结合 YOLO-head 得到的 3 张尺寸大小分别是 13×13、26×26 和 52×52 的预测张量.以 13×13 预测张量为例,本质是将原始输入的 416×416 尺寸大小的图

像用 13×13 的网格覆盖,每个网格负责预测某个目标.预测张量的尺寸决定了目标检测框架的精细化程度,如果预测张量图越大,那么小目标就更容易被检测出.

因此本文算法提前进行骨干网络特征图输出,得到 4 张特征图,构建 4 层 FPN 结构,模型新输出的 104×104 大小的预测张量能够提高检测网络的精细化程度,进一步提升算法对小目标的检测能力.

### 2.2 小目标数据 SSR 融合特征增强

Retinex 模型<sup>[10]</sup>中的照度分量可以视为无噪声干扰的原始图像在经过周围环境变化、拍摄失真等噪声

干扰产生的一种非线性变换. 这种非线性变换作用于原始图像就得到了最终的实际图像. 因此只需要利用小目标图像所在区域的照度分量, 便能够反解出该区域的反射分量, 即得到小目标的原始图像, 增强数据集中小目标的本质特征信息量.

本文针对小目标检测, 随机对 10% 的训练集图像标注区域进行 SSR 操作<sup>[11]</sup>. 在包含小目标的图像区域内建立 Retinex 模型如下:

$$S(m, n) = R(m, n) \times L(m, n) \quad (2)$$

对式 (2) 两边取对数得到如下公式:

$$\log(S(m, n)) = \log(R(m, n)) + \log(L(m, n)) \quad (3)$$

使用高斯环绕函数对光照分量进行估计, 代入式 (3) 得到:

$$\begin{aligned} r(m, n) &= \log\left(\frac{S(m, n)}{L(m, n)}\right) \\ &= \log(S(m, n)) - \log(S(m, n) \times G(x, y)) \end{aligned} \quad (4)$$

其中, 高斯环绕函数是一个含有超参数 $\sigma$ 的函数:

$$G(x, y)_i = \frac{1}{2\pi\sigma_i^2} e^{-\frac{x^2+y^2}{2\sigma_i^2}} \quad (5)$$

本算法中取 $\sigma$ 为各局部图像的标准差, 对应于 RGB 三通道, 每个通道的高斯环绕滤波均具有各自的 $\sigma_i, i = 1, 2, 3$ .

根据 Retinex 理论, 一幅图像可以视为反射图像和照度图像的合成, 反射图像中包含图像中物体自身的大量特征信息. 照度图像中则包含大量的外界噪声信息. 所以只需要对照度图像进行估计, 就可以得到反射图像, 即得到物体自身的特征信息<sup>[12]</sup>. 最初的 Retinex 算法主要是基于路径比较的算法, 缺点是无法对照射分量进行准确的估计, 得到的反射图像特征损失较大.

SSR 利用高斯环绕函数式 (5) 对照度图像进行估计, 本质是在图像的每个通道用一个符合二维高斯分布的模板去代表照度图像,  $\sigma_i$ 是核参数, 其值越大, 最终得到的反射图像具有的特征信息越多, 但是图像的保真度会降低. SSR 算法相比之前基于路径的 Retinex 算法能够提取图像中更加准确的照射图像, 进而得到更加准确的反射图像. 更加准确的反射图像能够保留更完整的物体特征.

通过上述算法得到局部新图像, 能够充分反映目标的本质特征如轮廓信息等. 结合原始局部图像构建局部图像 SSR 融合特征增强线性加权模型如式 (6).

$$R^*(m, n) = \alpha E(m, n) + (1 - \alpha)R(m, n), \alpha \in [0, 1] \quad (6)$$

其中,  $E(m, n)$ 表示该区域原始图像,  $R(m, n)$ 表示经过 SSR 处理得到的图像. 利用式 (6) 模型对训练集标注框内的目标进行特征增强,  $\alpha$ 称为细节突出因子, 取值范围是 0 到 1, 图 3 展示了 $\alpha$ 从 0.1 到 1 对同一目标图像的处理效果.

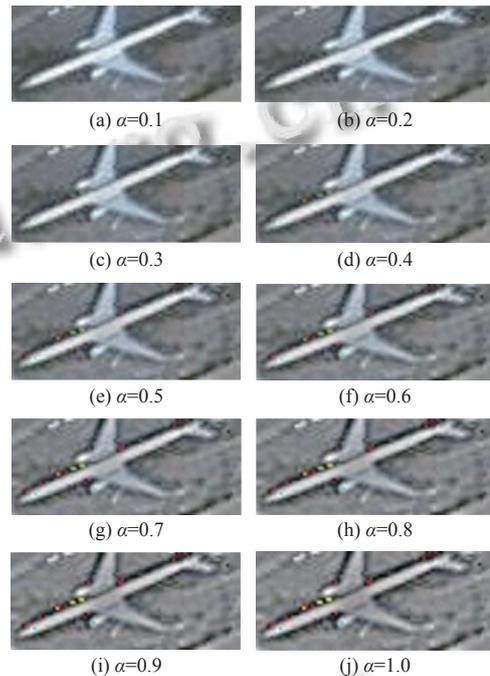


图 3 不同细节突出因子下的飞机图像效果

可以发现, 随着 $\alpha$ 的增大, 飞机轮廓更加明显, 背景噪声得到明显抑制, 飞机的本质属性得到更好的体现, 但同时也损失了较大的环境的噪声信息. 为提高训练过程中神经网络对检测目标本质特征的提取率, 同时为保证模型的鲁棒性, 本文取细节突出因子 $ratio$  (取值的实验过程将在第 3.3 节中详细介绍). 既能够增强目标有效特征的展示, 同时又保持一定程度的环境噪声参与模型训练, 降低网络特征提取时有效特征的丢失率, 保持模型的鲁棒性.

### 2.3 锚框超参数的优化

在经典 YOLOv3 算法中, 使用 K-means 算法得到锚框. 然而文献 [13] 经过实验指出: K-means 算法得到的锚框不够理想并且锚框选取的提升空间是足够大的. 因此, 通过优化改进锚框计算方法进而提高模型表现效果是可行的.

Anchor box 的选取可以用一组超参数来描述, 设

共有  $K$  组锚框, 每个锚框有尺度  $scale$  和长宽比例  $ratio$  两个参数, 则一组锚框组合可以表示为一个向量  $\vec{x} = (scale_1, ratio_1, \dots, scale_K, ratio_K) \in \mathcal{X}$ ,  $\mathcal{X} \subset R^{2K}$  是锚框的搜索空间. 超参数优化的目标函数定义为  $f(\vec{x}) = \vec{x} \rightarrow R$ ,  $f(x)$  是衡量锚框  $\vec{x}$  效果的函数. 考虑到搜索空间的复杂度, 超参数搜索常用的网格搜索、随机搜索不适用于本文场景. 本文采用的自适应锚框优化算法 (adaptive anchor box optimization, AABO)<sup>[13]</sup> 是目前最先进的锚框超参数搜索方法之一, 该算法在基于贝叶斯优化的超参搜索算法 (Bayesian optimization and hyperband, BOHB)<sup>[14]</sup> 的基础上, 针对小目标检测收敛的滞后性对子采样策略进行重写. AABO 算法流程如图 4 所示.

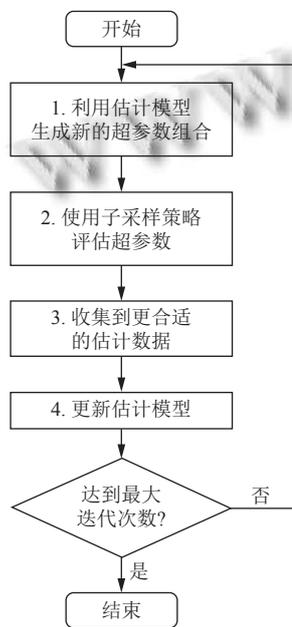


图 4 AABO 算法流程图

图 4 中, 估计模型使用核密度估计函数 (tree Parzen estimator, TPE) 进行建模, TPE 包含两个概率密度函数, 分别是  $l(\vec{x}) = p(y < \alpha | \vec{x}, D)$  和  $g(\vec{x}) = p(y > \alpha | \vec{x}, D)$ ,  $\vec{x}$  是一个超参数组合,  $y = f(\vec{x})$  是超参数组合  $\vec{x}$  的观测结果,  $D = \{(\vec{x}_0, y_0), \dots, (\vec{x}_n, y_n)\}$  为当前的观测数据,  $\alpha = \min\{y_0, \dots, y_n\}$  为当前观测数据中取得的最优结果. 以  $l(\vec{x})$  最小的  $\vec{x}^*$  作为新的超参数组合, 使用子采样策略进行评估, 并利用新获得的数据修正 TPE 模型, 不断重复这个过程, 直至达到最大预算.

子采样策略用于评估一组超参数组合, 其算法步骤见算法 1. 相比基于预算的超参搜索算法 Hyper-

Band<sup>[15]</sup>, 该策略对缓慢收敛的小目标检测有更高的评估准确性.

算法 1. 子采样策略

Input: 超参数组合集合  $\mathcal{I} = \{1, \dots, K\}$ , 阈值  $c_n$ , 预算最小单位  $b$ .

Output: 预算分配结果  $\hat{n}_1, \dots, \hat{n}_N \in \mathcal{I}$ .

1.  $r=1$ , 分别以预算  $b$  评估每一个超参数组合
2. for  $r=2, \dots, N$  do
3. 当前获得最大预算的超参数组合记为  $\zeta^r$ , 称为领先者
4. for  $k \neq \zeta^r$  do
5. 若第  $k$  个超参数组合“优于”领先者, 则给予它额外  $b$  的预算
6. end for
7. 若没有找到“优于”领先者的超参数组合, 则给予领先者额外  $b$  的预算
8. end for

基于式 (7) 和式 (8) 两个条件<sup>[13]</sup> 来评估一个超参数组合是否“优于”领先者, 阈值  $c_n$  使用  $\sqrt{\log n}$ .

$$n_k < n_{k^{cent}} \text{ and } n_k < c_n \tag{7}$$

$$\begin{cases} c_n \leq n_k < n_{k'} \text{ and } \bar{Y}_{1:n_k}^{(k)} \geq \bar{Y}_{j:(j+n_k-1)}^{(k')} \\ \text{for some } 1 \leq j \leq n_{k'} - n_k + 1 \\ \text{where } \bar{Y}_{l:u}^{(k)} = \sum_{v=l}^u Y_v^{(k)} / (u - l + 1) \end{cases} \tag{8}$$

考虑到对一个锚框超参数组合评估的极高时间成本, 设计一个足够紧致但同时不过度损失探索性的搜索空间非常重要. 本文首先对 RSOD 数据集中所有标注框的尺度和高宽比进行统计, 剔除极端分布后作为基础搜索空间. 由于锚框参数的微调对检测效果影响微小, 对尺度和长宽比分别以 0.1 和 0.01 的采样间隔进行离散化, 从而在搜索过程中维持较高的探索性.

目标函数  $f(\vec{x})$  常采用训练过程中验证集的最小损失值. 然而在训练的早期阶段, 大目标的快速收敛主导了验证集的损失值, 因此选取的最优锚框主要是针对大目标达到的收敛最优, 并未对小目标达到最优匹配. 针对这一问题, 本文重新设计了目标函数  $f(\vec{x})$ .

$$f(\vec{x}, epoch) = val\_loss(\vec{x}, epoch) - e^{-\left(\frac{epoch}{max\_budget}\right)} \times mAP_{aircraft}(\vec{x}, epoch) \tag{9}$$

式 (9) 中,  $epoch$  为训练轮次,  $max\_budget$  为 AABO 算法中的最大预算参数, 本文取最大预算参数为 72.  $mAP_{aircraft}$  为模型在 RSOD 数据集 aircraft 类别的平均 AP. 优化后的目标函数可以在训练早期阶段将对小目标适应性强的锚框组合高亮, 使搜索的优化方向更好地集中在小目标检测能力上.

锚框优化效果如图5所示,红色框是真实值,绿色框是第3层FPN对应的锚框,即图2中Predict3.图5(a)采用了YOLOv3中使用的K-means算法,图5(b)采用了AABO算法,迭代次数为27.可以看到图5(a)中的3个锚框都与真实值有较大误差,而经过优化后的图5(b)对小目标的预测具有更高的自适应性.

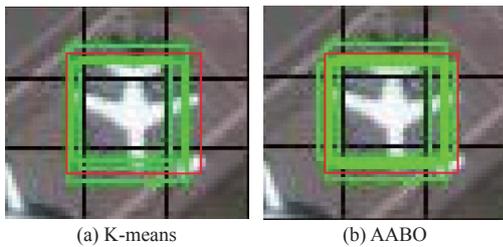


图5 Anchor 优化效果

### 3 算法实验

本文实验在 Windows 10 操作系统下进行,使用 TensorFlow 2.3.0 搭建模型, GPU 使用 NVIDIA Tesla P100-PCIE. 总训练轮次设置为 400, 优化器使用 Adam. 使用余弦退火算法作为学习率的更新方式.

#### 3.1 评价指标

目标检测的效果由预测框的分类精度和定位精度共同决定, 因此目标检测问题既是分类问题, 又是回归问题. 本文采用全类平均精度 (mean average precision, *mAP*) 和 *F1* 分数来评价模型效果.

某个类别的分类准确度用平均准确率 (average precision, *AP*) 衡量, *AP* 计算公式如下:

$$AP = \int_0^1 P(R) dt \quad (10)$$

其几何含义是 *PR* 曲线与坐标轴所围成的图形面积. 对于 *M* 分类的问题, 常使用 *mAP* 值作为分类效果评价指标, 计算公式如下:

$$mAP = \frac{\sum_{i=1}^M AP_i}{M} \quad (11)$$

*F1* 分数是精确率 *P* 和召回率 *R* 的调和均值, 更能客观评价模型的全方位性能, *F1* 计算公式如下:

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (12)$$

#### 3.2 学习率优化调整

理想情况下, 在模型训练过程中, 学习率应由大变

小. 但在实际情况下, 由于模型较大且参数初始化是随机的, 梯度下降方向并不稳定, 因此需要在一段时间内设置较小的学习率<sup>[16]</sup>, 然后再调高学习率, 最后逐渐调小学习率, 以便能够得到更加好的训练效果.

利用余弦退火算法<sup>[16]</sup>对学习率进行更新, 学习率更新公式如下:

$$LR_i = (LR_{max} - LR_{min}) \times \left(\frac{i}{T_1}\right)^2 + LR_{min} \quad (13)$$

式 (13) 是训练开始时的学习率变化公式,  $LR_{max}$  表示训练过程中的最大学习率, 本文取 0.01,  $LR_{min}$  表示训练过程中的最小学习率, 本文取 0.001,  $T_1$  表示学习率上升时间, 本文取 40,  $i$  表示当前处于第  $i$  个 epoch.

$$LR_i = LR_{min} + 0.5 \times (LR_{max} - LR_{min}) \times \left(1 + \cos\left(\frac{i\pi}{T - T_1}\right)\right) \quad (14)$$

式 (14) 是训练到  $T_1$  个 epoch 后的学习率变化公式.  $T$  表示总的训练次数, 本文中  $T = 400$ .  $i$  表示当前处于第  $i$  个 epoch.

将训练过程中的学习率变化可视化如图6所示.

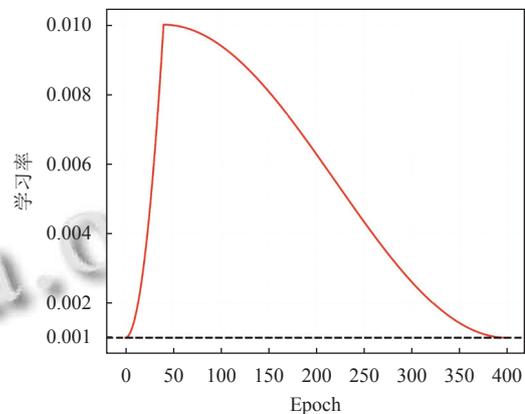


图6 学习率变化曲线

分析学习率变化曲线可以知道, 在训练开始后的一段 epoch 内, 学习率从 0.001 快速增加至 0.01. 此后学习率由 0.01 下降至 0.001, 下降速度由慢变快再变慢.

#### 3.3 不同 $\alpha$ 取值情况下的飞机 *AP* 值对比

由图3可知, 取不同的  $\alpha$ , 对飞机图像特征的展现是不一样的, 因此对模型效果的影响也是不同的, 为了获得最佳的  $\alpha$ , 本文以 0.1 为取值间隔, 从 0 到 1, 分别取  $\alpha=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ ,

1.0 对小目标飞机图像数据集进行特征融合增强, 然后进行模型训练和测试. 得到 10 组  $\alpha$  值下测试集中的飞机 AP 值, 如图 7 所示.

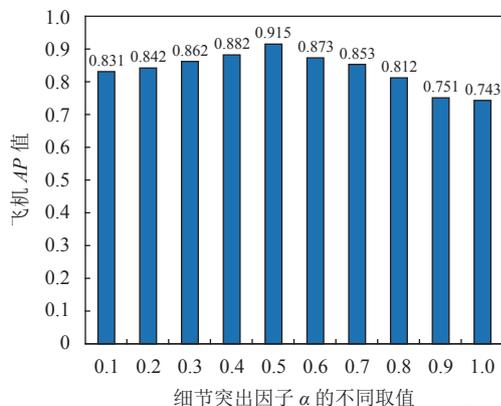


图 7 不同  $\alpha$  取值情况下的飞机 AP 值对比

分析图 7 可知, 随着  $\alpha$  从 0-1 增大, 得到的检测模型对飞机的检测 AP 值先增大后减小, 训练集中的待检测物体特征增强, 背景噪声减弱. 在  $\alpha=0.5$  的时候达到最大的 AP 值 0.915. 因此取  $\alpha=0.5$  能够较好地平衡物体特征信息和背景噪声信息, 使模型能够得到较好的训练效果.

### 3.4 消融实验

本文使用大部分检测物体为小目标的 RSOD 数据集进行模型消融实验. 共设置 5 种消融实验, 分别是经典 YOLOv3 算法、仅改进网络结构的 YOLOv3 算法、改进网络结构并且使用 SSR 增强的 YOLOv3 算法、改进网络结构并且使用 anchors 优化的 YOLOv3 算法和改进网络结构, 同时进行 SSR 增强和 anchors 优化的完整 YOLOv3 改进算法.

将验证集  $mAP@0.5$  作为评价指标, 记录 400 次训练过程中每 10 次结果中的最大  $mAP$  值, 绘制如图 5 所示的  $mAP$  变化曲线.

图 8 表明, 相比于经典 YOLOv3 模型, 经过任意一种文中的优化方法均能够较大幅度提高 YOLOv3 模型的效果.

对比分析经过结构改进的 YOLOv3 模型和经典 YOLOv3 模型的  $mAP$  变化曲线可知: 由于本文重新设置 Darknet-53 的残差模块数量及排列方式同时引入 GC 注意力机制, 提高了模型骨干网络对于 RSOD 数据集中大量小目标图像数据的特征提取能力, 同时能够保证对于中大型目标仍然良好的特征提取能力, 因此

模型的学习效果明显得到提高. 在检测识别过程中, 扩充的 FPN 机制使得模型的感受野得到扩大, 能够对于目标区域物体进行更加精准的识别.

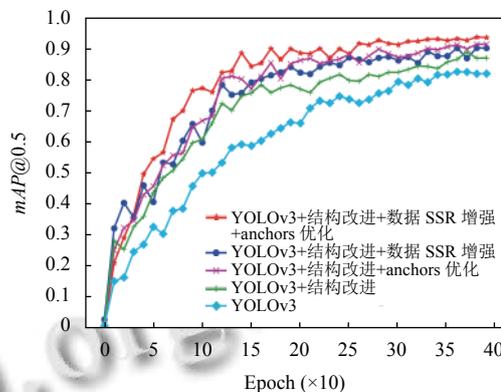


图 8 5 次消融实验的  $mAP@0.5$  变化曲线

对比分析经过结构改进但未进行 SSR 数据增强的模型和经过 SSR 数据增强的模型可知: 经过 SSR 融合增强的数据集能够更好地体现目标物体的特征, 使得模型在训练过程中能够学习到目标物体更多的有效信息. 因此在检测过程中, 相比于仅改变网络结构的模型, 能够得到更好的  $mAP$  值.

对比分析经过结构改进但未进行 anchors 优化的模型和经过 anchors 优化的模型可知: 经过 anchors 优化的 YOLOv3 模型能够更好地对目标检测框进行匹配, 总体上获得更大的  $mAP$  值, 同时模型的收敛速率得到一定提高.

总体来说, 经过本文对 YOLOv3 算法进行的网络结构改进、训练集 SSR 特征融合增强学习和 anchors 优化后, 算法对于小目标的检测效果相比经典 YOLOv3 算法得到明显提高, 最优  $mAP$  值提高 12% 左右.

### 3.5 不同算法对比与检测效果可视化

将本文算法与应用广泛的经典 YOLOv3 算法、YOLOv4 算法、Efficientdet 算法的相关性能指标进行对比, 如表 1 所示, 可以发现本文算法效果明显优于其他常用算法, 在 RSOD 数据集的大部分飞机 (小目标) 检测上具有 0.915 的 AP 值明显高于其他算法.

在不同情况下, 将本文算法与经典 YOLOv3 算法的检测效果进行如下可视化对比分析.

如图 9 所示为复杂背景下经典 YOLOv3 算法与本文算法的检测效果对比. 图 9(a) 为经典 YOLOv3 算法检测效果, 图 9(b) 为本文算法检测效果. 经典 YOLOv3 算法检测到 9 个目标, 本文算法检测到 12 个目标.

表1 不同算法的性能对比表

算法	$mAP@0.5$	$AP$ (飞机)	$F1$ (飞机)
YOLOv3	0.824	0.845	0.843
YOLOv4 <sup>[17]</sup>	0.862	0.864	0.840
Efficientdet <sup>[17]</sup>	0.881	0.603	0.620
本文算法	0.925	0.915	0.900

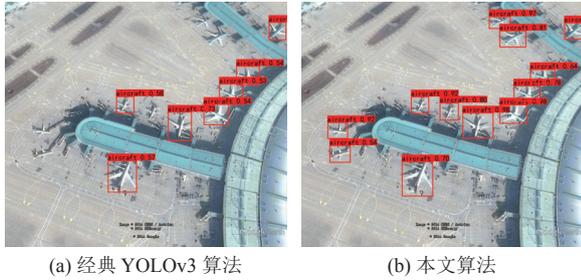


图9 复杂背景下检测效果对比

如图10所示为雾霾环境下经典YOLOv3算法与本文算法的检测效果对比。图10(a)为经典YOLOv3算法检测效果,图10(b)为本文算法检测效果。经典YOLOv3算法检测到11个目标,本文算法检测到24个目标。

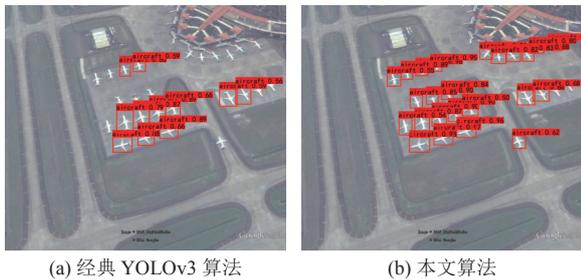


图10 雾霾环境下检测效果对比

如图11所示为多个小目标检测任务下经典YOLOv3算法与本文算法的检测效果对比。图11(a)为经典YOLOv3算法检测效果,图11(b)为本文算法检测效果。经典YOLOv3算法检测到23个目标,本文算法检测到32个目标。可以看出,经过本文优化的YOLOv3算法对于小目标的检测效果得到明显提升。

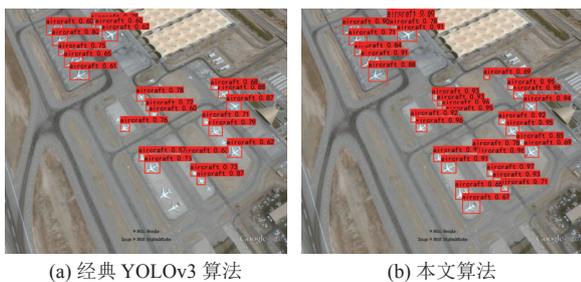


图11 多个小目标任务检测效果对比

如图12所示为集群小目标检测任务下经典YOLOv3算法与本文算法的检测效果对比。图12(a)为经典YOLOv3算法检测效果,图12(b)为本文算法检测效果。经典YOLOv3算法检测到10个目标,本文算法检测到16个目标。待检测的图片左上部分为集群小目标存在处,可以发现经过本文优化的YOLOv3算法对于集群小目标检测效果得到明显提升。

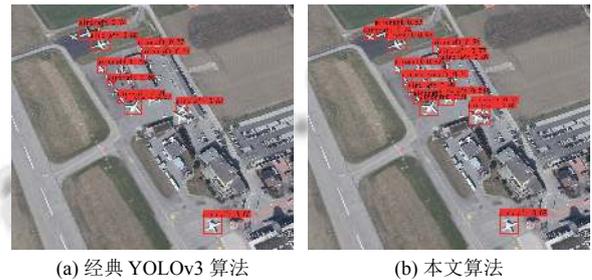


图12 集群小目标任务检测效果对比

#### 4 总结与展望

本文基于经典YOLOv3算法,首先对主干网络Darknet-53进行残差模块数量调整和重新排列组合,保证不同尺度下网络特征提取能力不变,扩充FPN层数,扩大输出特征图的感受野,并引入GC注意力机制,提高模型上下文特征捕获能力,增强了模型对小目标的特征学习和识别能力。其次,基于Retinex理论,建立SSR融合特征增强线性加权模型,取细节突出因子等于0.5,随机对10%训练集目标进行特征增强,提高了训练集小目标特征信息含量,增强了模型训练效果。最后结合bounding box在RSOD数据集上的分布特性,使用AABO算法,对anchors尺度和长宽比例进行优化,提高了模型对小目标的匹配程度。

实验阶段,针对遥感数据集RSOD,首先,设置5种消融实验,对比分析了本文不同优化措施对经典YOLOv3算法效果的影响,最终发现本文算法相比于经典YOLOv3算法,在RSOD数据集上 $mAP@0.5$ 值提高10.1%。然后,将本文算法在RSOD数据集上的表现效果与其他文献中的算法效果进行对比,发现本文算法效果优于其他文献算法效果。最后,针对复杂背景下目标检测、雾霾环境下目标检测、多个小目标检测、集群小目标检测等实际场景下4种典型遥感目标检测任务,将本文算法与经典YOLOv3算法进行效果对比,发现本文算法在4种典型任务中的表现效果均

明显优于经典 YOLOv3 算法。

本文采用不同措施提高 YOLOv3 算法对于遥感小目标的检测效果,毫无疑问将一定程度提高模型的复杂程度,降低模型的实时检测速度。后续研究将考虑对模型进行剪枝操作,得到轻量化模型,以便结合实际硬件对遥感小目标进行实时检测。

### 参考文献

- 1 Viola P, Jones MJ. Robust real-time face detection. *International Journal of Computer Vision*, 2004, 57(2): 137–154. [doi: 10.1023/B:VISI.0000013087.49260.fb]
- 2 Dalal N, Triggs B. Histograms of oriented gradients for human detection. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego: IEEE, 2005. 886–893.
- 3 方路平, 何杭江, 周国民. 目标检测算法研究综述. *计算机工程与应用*, 2018, 54(13): 11–18, 33. [doi: 10.3778/j.issn.1002-8331.1804-0167]
- 4 刘洪江, 王懋, 刘丽华, 等. 基于深度学习的小目标检测综述. *计算机工程与科学*, 2021, 43(8): 1429–1442. [doi: 10.3969/j.issn.1007-130X.2021.08.012]
- 5 徐思源, 储开斌, 张继, 等. 面向小目标检测的改进 YOLOv3 算法. *电光与控制*, 2022, 29(8): 35–39. [doi: 10.3969/j.issn.1671-637X.2022.08.007]
- 6 刘峻文, 张永军, 李智, 等. 基于 RDM-YOLOv3 的头部检测. *激光与光电子学进展*, 2022, 59(8): 0815011.
- 7 王建军, 魏江, 梅少辉, 等. 面向遥感图像小目标检测的改进 YOLOv3 算法. *计算机工程与应用*, 2021, 57(20): 133–141.
- 8 Redmon J, Farhadi A. YOLOv3: An incremental improvement. *arXiv:1804.02767*, 2018.
- 9 Cao Y, Xu JR, Lin S, *et al.* GCNet: Non-local networks meet squeeze-excitation networks and beyond. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop*. Seoul: IEEE, 2019. 1971–1980.
- 10 谢凤英, 汤萌, 张蕊. 基于 Retinex 的图像增强方法综述. *数据采集与处理*, 2019, 34(1): 1–11.
- 11 何艳, 方帅. 一种局部多尺度 Retinex 算法在雾天图像中的应用. *合肥工业大学学报(自然科学版)*, 2015, 38(10): 1333–1338.
- 12 王利娟, 常霞, 张君. 基于 Retinex 的彩色图像增强方法综述. *计算机系统应用*, 2020, 29(6): 13–21. [doi: 10.15888/j.cnki.csa.007430]
- 13 Ma WS, Tian TZ, Xu H, *et al.* AABO: Adaptive anchor box optimization for object detection via Bayesian sub-sampling. *Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020. 560–575.
- 14 Falkner S, Klein A, Hutter F. BOHB: Robust and efficient hyperparameter optimization at scale. *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018. 1437–1446.
- 15 Li LS, Jamieson K, DeSalvo G, *et al.* Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 2017, 18(1): 6765–6816.
- 16 He T, Zhang Z, Zhang H, *et al.* Bag of tricks for image classification with convolutional neural networks. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 558–567.
- 17 冯号, 黄朝兵, 文元桥. 基于改进 YOLOv3 的遥感图像小目标检测. *计算机应用*, 2022, 42(12): 3723–3732.

(校对责编: 孙君艳)