

基于 DATE-FCOS 的空中目标检测研究

陈钊阳, 王玉玫

(华北计算技术研究所, 北京 100083)

摘要: 航空图像中目标的检测是当前研究的热点之一, 高效、准确的探测在军事和民用领域具有很高的价值; 但由于高空环境复杂和空中目标尺度形状多变, 不同用途的飞行器进行改装和涂装使得空中目标在检测中难度较大; 所以提出了一种改进的一阶端到端的空中目标检测算法; 算法采用 DATE-FCOS 为基本框架, 用 CIoU 替代 GIoU 加入到边界框回归损失函数中, 并在此基础上, 利用可形变卷积模块对其骨干网络进行了改进并且在 FPN 结构之后加入 CBAM 模块; 通过实际实验测试, 所提方法在 FGVC aircraft 数据集上提高了检测的平均检测精度, 达到 77.8%, 对比原模型提升 11%, 满足空中目标检测的应用。

关键词: 目标检测; FCOS; 空中目标; CIoU; 可形变卷积; 注意力机制

Research on Air Target Detection Based on DATE-FCOS

CHEN Zhaoyang, WANG Yumei

(North China Institute of Computing Technology, Beijing 100083, China)

Abstract: Target detection in aerial images is one of the current research hotspots, and efficient and accurate detection has high value in military and civilian fields. However, due to the complex high-altitude environment and the variable scale and shape of air targets, the modification and coating of aircraft for different purposes make it difficult to detect air targets. Therefore, an improved first-order end-to-end air target detection algorithm is proposed. The algorithm adopts dual assignment for end-to-end fully convolutional one-stage object detection (DATE-FCOS) as the basic framework, replaces the generalized intersection over union (GIoU) with the complete intersection over union (CIoU), and adds it to the bounding box regression loss function. On this basis, the deformable convolutional module is used to improve its backbone network and add the convolutional block attention module (CBAM) after the FPN structure. Through practical experimental tests, the proposed method improves the average detection accuracy of the fine-grained visual classification of aircraft (FGVC) aircraft dataset by 77.8%, and the average detection accuracy of the proposed method is 11% higher than that of the original model, which meets the application of aerial target detection.

Keywords: target detection; FCOS; air target; CIoU; deformable convolution; attention mechanism

0 引言

近年来, 我国围绕“全面加强军事航天发展: 优先发展战略空间和高新技术, 增强网络空间和电磁空间力量, 全面提高航空航天、防空导弹力量的能力”做出重要部署, 旨在提升国家空天实力, 维护国家安全和领土领空安全, 促进国家经济和科技发展, 并为全球空间治理贡献中国方案。随着国内外无人机相关技术飞速发展, 无人机的应用涉及多个领域, 包括军事应用。高空无人机对航空图像中目标的检测是当前研究的热点之一, 高效、准确地探测在军事和民用领域具有很高的价值。然而, 由于高空环境和光照条件较为特殊, 高空图像采集系统获取的图像存在较多的噪点, 并且存在阴影衰落, 使图像变得粗糙模糊^[1], 因此空中目标检测精度较低, 难度较大。

而空中目标大致可分为 3 类: 1) 空中季节性迁徙的鸟类, 大型的鸟类迁徙高度可达 3 000~6 300 m, 而天鹅等大型种飞行高度可达 9 000 m, 跨越珠穆朗玛峰; 2) 小型飞行器, 主要包括探空气球、空飘气球、轻型和超轻型运动

飞机、轻型直升机、滑翔机、无人机、航模、动力伞、动力三角翼、小型飞艇等; 3) 大型飞行器, 主要有军用和民用的运输机和直升机、客机、轰炸机和侦察机等^[2]。研究的主要重点当然还是各类的大型飞行器, 飞行器, 尤其是飞机有几个方面使目标识别与众不同。首先, 飞机设计跨越了超过一百年, 包括数千种不同的型号、数百种不同的品牌和航空公司; 其次, 飞机的设计因尺寸 (从小型无人机到大型运输机)、目的地 (私人、民用、军用)、用途 (运输机、航母、训练、运动、战斗机等)、推进力 (滑翔机、螺旋桨、喷气式飞机) 和包括技术在内的许多其他因素而异。飞机的结构随着其设计 (机翼、起落架、每个起落架的轮子、发动机等的数量) 而变化; 最后, 任何公开的飞机模型都可以被不同的公司重新使用, 这会导致外观 (涂装) 的进一步变化。根据识别任务的不同, 这些可以被视为噪声或要提取的有用信息。

早期目标检测主要是依靠基于人工特征提取的机器学习方法。这些方法通过色彩、轮廓、纹理等特征来刻画物

收稿日期: 2023-08-16; 修回日期: 2023-09-28。

作者简介: 陈钊阳 (1998-), 男, 硕士研究生。

引用格式: 陈钊阳, 王玉玫. 基于 DATE-FCOS 的空中目标检测研究[J]. 计算机测量与控制, 2024, 32(2): 78-84.

体^[3], 尽管具有方法简单、实时性好等优点, 但时间冗余度高, 基于手工设计的特征鲁棒性、泛化性较差, 仅适用于专门面向的特定任务, 并且严重受到设备的硬件性能和周围环境的影响, 进行高精度的检测任务较为困难, 容易造成大量的检测结果误差, 难以实施于实际的工程应用之中。然而, 随着人工智能的发展和其在计算机视觉领域的研究不断地深入, 深度学习 (DL, deep learning)^[4-6] 的许多方法开始逐渐应用于目标检测领域, 并且由于其优异的表现学术界得到了广泛研究和应用。从目标检测算法的结构上来说可以主要分为两大类: 一类是两阶段检测算法, 主要以 Faster R-CNN^[7]、Mask R-CNN^[8] 等 R-CNN 系列算法为代表, 该算法是第一个提出“先选取候选框, 再通过分类器对候选框内的物体进行分类”的方法。它的主要组成部分包括区域提取网络 (RPN)、特征提取网络 (Backbone)、RoI-Pooling 以及分类和回归网络。两阶段目标检测算法在准确率上有着很大的优势, 在实际应用中也得到了广泛的应用; 另一类是一阶段检测算法, 主要代表有 YOLO (You Only Look Once)^[9], FCOS (Fully Convolutional One-Stage Object Detection)^[10], SSD (Single Shot Detection)^[11], 直接通过神经网络计算输出检测对象的预选框和标签。

王伟臣以 Faster R-CNN 目标检测算法作为研究基础, 并对 Faster R-CNN 提出了两点改进。首先是在原始 Faster R-CNN 算法的主干网络中生成特征金字塔, 充分利用不同层级的特征图来对小尺度目标进行有效的特征提取; 另一方面, 提出了一种改进的 Anchor 匹配规则, 改进的匹配规则能够使所有的真实目标框在 Anchor 匹配过程中匹配到比原始匹配规则更多的 Anchor^[12]。算法有效地提升了检测的精度, 但依旧有着检测速度不理想的问题。薛建伟等人提出了一种利用 YOLOv3 对空中飞行器进行识别追踪的方法^[13], 一阶段的检测算法在检测速度上有着巨大的优势, 该方法提升了对目标的追踪精度和实时性, 但在识别精度方面并无明显提升。王耀领等人提出了利用改进的条件生成网络生成高质量飞行器图像样本^[14], 用以提高模型的泛化能力以此来提升模型的识别精度, 但难以应对现实中空中目标所处的复杂环境。兰旭婷等人提出了以 Resnet50 为基础特征提取的 SSD300 网络模型, 加入注意力机制 CBAM 模块和特征融合 FPN 模块, 采用 Soft-NMS 策略选取最终预测框, 对遥感图像飞机目标更加有效地进行检测^[15], 该方法通过注意力机制提升了模型的性能, 但对于模型的特征提取能力没有提升, 对于飞机目标中相同目标不同角度识别困难的情况效果并不是很理想。

DATE-FCOS 作为在 FCOS 基础上改进的 anchor-free 模型, 不仅有着 FCOS 低计算量的同时保持高准确性的优点, 并且在提升了模型性能的同时简化了模型的训练过程。对于空中目标检测的特点, 本文将 DATE-FCOS (Dual Assignment for End-to-End Fully Convolutional One-Stage Object Detection) 为基准检测框架。并针对空中目标

检测存在的目标尺度变换以及目标检测精度不足的问题, 在原有算法框架的基础上进行改进, 并且进行消融实验论证改进的有效性, 最后与目前常见检测算法进行实验对比。

1 模型介绍

1.1 FCOS 模型

为了弥补 anchor-base 的模型在目标检测领域的几点缺陷:

1) 检测结果对于 anchor 的尺寸、比例和数量等超参数是非常敏感的。不同的 anchor 设置之间存在着较大的 AP 的差距;

2) anchor 的尺寸和比例是固定的, 因此不能很好地应对过大或者过小的目标。

3) 为了获取更高的 recall, anchor-based 检测器通常采取密集采样的方式, 大量的 anchor box 在于 gt box 计算 IoU 时也造成了巨大的算力消耗, 需要较高的算力和时间成本。

因此提出了 anchor-free 的 FCOS。整体框架, 基本上与 RetinaNet 一致, 由 backbone + neck + head 组成。

对于 CNN backbone 的第 i 层特征图上每个像素 (x, y) , 我们可以把它映射到原输入图像。在映射后的该点直接回归目标框, 而不是把这个点看成目标框的中心点生成不同尺寸的 anchor, 从而避免了生成大量的 anchor。FCOS 相当于把每个点看成训练的样本。对于每个 (x, y) , 如果它落在 gt 的框里面, 并且类别和 gt 的框的类别是一样的, 那么就看成正样本 (c^*) , 反之则是负样本 $(c^* = 0)$ 。并且使用点到 gt 框四边的距离作为训练的回归条件。

经典的 anchor-free 检测器 YOLOv1^[16], 由于只检测离中心点近的目标, 导致召回率比较低, 而多级 FPN^[17] 的加入则有效提高了回归率。

为了抑制这些低质量检测到的边界框, 而不引入更多的超参数。FCOS 添加了一个单层分支, 与回归分支并行, 以预测位置的“中心度”。中心度描述了从该位置到该位置负责的对象中心的归一化距离。

1.2 DATE-FCOS 模型

一阶卷积目标检测模型, 如 FCOS 和 RetinaNet^[18], 因其简单性而被广泛熟知并采用。尽管他们取得了成功, 但一对多分配 (o2m) 策略使他们依赖于非最大值抑制 (NMS) 来消除重复的预测。这样的过程使它们对 NMS 的每个参数都很敏感, 并可能导致陷入局部最优解。这个问题促使研究人员消除 NMS, 以实现端到端的完全卷积目标检测。OneNet^[19] 认识到具有分类成本的一对一匹配 (o2o) 策略是端到端检测的关键。具体而言, 通过考虑分类成本, 仅为 GT 分配一个预测, 可以防止其产生冗余预测。在 DATE-FCOS 中重新审视了这两种分配方法, 发现将一对多分配带回到端到端全卷积检测器有助于模型收敛, 提出了双任务分配 (DATE, dual assignment for end-to-end) 的方法: 在 FCOS 模型的基础之上进行了改进, 在训练过程中

构造了两个分支：一对多和一对一分支的分支，在回归子网和分类子网都使用两种分支来提供更多的监督信号来加快一对一指派分支的收敛^[20]，而在进行模型推理时只使用具有一对一分配策略的分支，这不会带来额外的检测开销，并且能显著提高检测性能。DATE-FCOS 模型结构如图 1 所示。

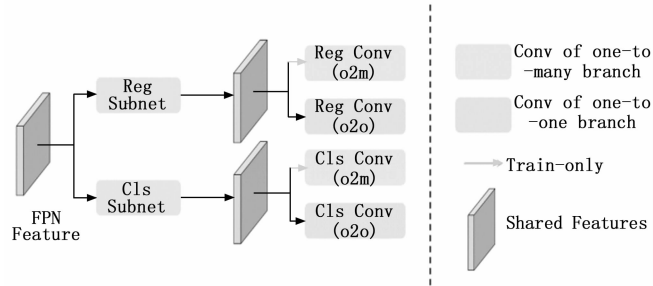


图 1 DATE-FCOS 模型结构

2 模型改进工作

2.1 可变卷积

在计算机视觉领域，同一物体在不同场景、角度中未知的几何变换是检测和识别领域的一大挑战。而对于空中目标这样的情况十分常见，同一目标在不同的角度收集到的图像差异较大，尤其是像飞机这样形状复杂的目标，在不同的角度更容易出现形态的变化，从而提升检测的难度。而通常的处理方法有：

1) 通过充足的数据增强，扩充足够多的样本去增强模型适应尺度变换的能力。

2) 设置一些针对几何变换不变的特征或者算法，比如 SIFT^[21] 和 sliding Windows。

两种方法都有着各自的缺陷。第一种方法因为样本的局限性使得模型的泛化能力比较低，无法泛化到一般场景中；第二种方法则因为手工设计的不变特征和算法对于过于复杂的变换是很难甚至无法设计的^[22]。所以 Deformable Conv (可变形卷积)，被提出以处理这个问题。

它是基于在模块中对空间采样的位置信息做进一步位移调整使采样网格能够自由变形。该位移可以通过模块附加的卷积层从前面得到的特征层中学习，并不需要额外的监督信号。它们为偏移学习只添加了少量的参数和计算，因此属于轻量级的模块^[23]。可变卷积模块可以很便捷地在常见的卷积神经网络中替换原有的卷积核，并且能很容易进行标准反向传播端到端的训练。

传统卷积的输出为：

$$y(p) = \sum_{k=1}^K w_k \cdot x(p + p_k) \quad (1)$$

其中： x 为输入， K 为卷积核的大小， w_k 为权重系数。 p 为输入的位置， p_k 为卷积核中的点相对偏移。而可变卷积的输出为：

$$y(p) = \sum_{k=1}^K w_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (2)$$

我们采用了最新的可形变卷积——DCN v2^[24]，它在前一代 DCN v1 的基础上加入了权重系数 $\Delta m_k \in [0, 1]$ ，以此来降低在改变感受野形状的同时引入的无关背景干扰，并且在模型骨干网络 Resnet50 的 stage3, stage4 和 stage5 的多个 3×3 卷积层使用。 Δp_k 和 Δm_k 的初始值分别为 0 和 1。改进后的网络结构如图 2 所示。

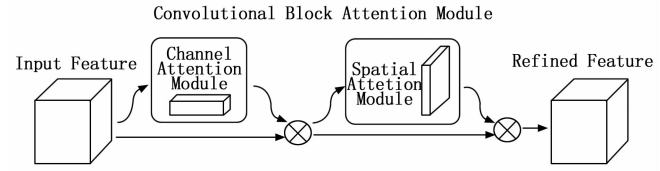


图 2 卷积块注意力模块

2.2 增加通道与空间注意力机制模块

卷积神经网络 (CNNs) 在视觉任务中表现出出色的性能主要是基于其丰富的表示能力。数据经过神经网络中每个不同的卷积核都会产生不同的通道信息，对于经典的神经网络，特征层会默认每个通道上所提取的信息所占的比重是相同的。而对于任何公开的飞机模型都可以被不同的公司使用，并添加更多的设计和进行不同的改装。例如，战斗机外挂弹仓，航空公司对客机进行涂装等等。对于不同的目标类型，不同的特征信息重要程度也是不同的。

因此为了进一步提高细胞神经网络对于目标的表示能力，提升模型对于飞行器目标物体的关注识别。我们在模型中加入了一种简单而高效的卷积块注意力模块 (CBAM, convolutional block attention module)^[25]——通道注意力模块 (Channel Attention Module) 和空间注意力模块 (Spatial Attention Module)，如图 2 所示。本文在进行 FPN 结构之后的特征检测层上分别加入双通道注意力模块 CBAM，通道注意模块可以帮助模型更好地集中检测所需要的信息，而空间注意力模块可以使模型忽略与检测目标无关的边界信息减少干扰，从而提高网络的特征表达能力。

给定其中一个特征图 $F \in R^{C \times H \times W}$ 作为模块的输入，CBAM 依次推断出一个 1 维的通道注意力图 $M_c \in R^{C \times 1 \times 1}$ 和一个 2 维的空间注意力图 $M_s \in R^{1 \times H \times W}$ 。整个注意力过程可以表述为：

$$F' = M_c(F) \otimes F \quad (3)$$

$$F'' = M_s(F') \otimes F' \quad (4)$$

其中： \otimes 为矩阵的逐元素乘法。

对于通道注意力模块：我们通过利用特征的通道间关系来生成通道注意力图，使用它来更加有效地学习目标物体的范围。首先通过使用平均池和最大池运算来聚合特征图中的空间信息，生成两个不同的对于空间上下文的描述： F_{max}^c 和 F_{avg}^c ，它们分别表示平均池特征和最大池特征，用来产生通道注意力映射 $M_c \in R^{C \times 1 \times 1}$ 。最后使用逐元素求和来合并并输出特征向量。通道注意力模块计算具体的表述为：

$$M_c(F) = \sigma(W_1(W_0(F_{max}^c)) + W_1(W_0(F_{avg}^c))) \quad (5)$$

其中： σ 为 sigmoid 函数， W_1 ， W_0 为共享的权重。

对于空间注意力模块, 我们使用特征之间的空间关系生成空间注意力图。相较于通道注意力, 不同的是空间注意力集中在“哪里”的信息部分, 这是对通道注意力的补充。通过使用两个池化操作来聚合特征图的通道信息, 生成两个 2 维的特征图: $F_{avg}^S \in R^{1 \times H \times W}$ 和 $F_{max}^S \in R^{1 \times H \times W}$ 。然后将它们连接起来, 并通过标准卷积层进行卷积, 生成最后 2 维的空间注意力图。空间注意力模块计算具体的表述为:

$$M_S(F) = \sigma\{f^{7 \times 7}[F_{avg}^S; F_{max}^S]\} \quad (6)$$

其中: σ 表示 sigmoid 函数, $[F_{avg}^S; F_{max}^S]$ 为 F_{avg}^S, F_{max}^S 两个特征矩阵的连接, $f^{7 \times 7}$ 表示 7×7 的标准卷积运算。

2.3 加入损失

在原始的 DATE-FCOS 中损失由两个部分组成, 分别是一对一分支损失和一对多分支损失。

一对一分支损失:

$$l_{o2o} = \alpha_{cls} l_{o2o}^{cls} + \alpha_{reg} l_{o2o}^{reg} + \alpha_{iou} l_{o2o}^{iou} \quad (7)$$

其中: $\alpha_{cls}, \alpha_{reg}, \alpha_{iou}$ 是相应的权重, $l_{o2o}^{cls}, l_{o2o}^{reg}, l_{o2o}^{iou}$ 分别是分类, 回归以及 IOU 的损失。

一对多分支损失:

$$l_{o2m} = \beta_{cls} l_{o2m}^{cls} + \beta_{reg} l_{o2m}^{reg} + \beta_{ctr} l_{ctr} \quad (8)$$

其中: $\beta_{cls}, \beta_{reg}, \beta_{ctr}$ 是相应的权重, l_{ctr} 是 FCOS 中 center-ness 分支的损失函数, 其余项含义与一对一分支相同。一对多分支提供足够的监督信号来加速共享多尺度特征提取网络的训练。一对一分配分支的作用类似于接收和调整特征以进行端到端检测。一对多分配策略有助于缓解优化问题, 使一对一分配分支更专注于拟合接收到的表示。因此在训练中一对多分支比重更大。在原始模型中, 一对多分支中使用 $GIoU$ (Generalized Intersection over Union)^[26] 来进行回归损失的计算:

$$GIoU = IoU - \frac{C - (A \cup B)}{C} \quad (9)$$

$$L_{GIoU} = 1 - GIoU$$

$$L_{GIoU} = 1 - IoU + \frac{C - (A \cup B)}{C} \quad (10)$$

先求出两个框的最小凸集 C (即包含 A, B 的最小框), 再计算出 IoU (交并比), 再计算闭包区域中不属于两个框的区域占闭包区域的比重, 最后用 IoU 减去这个比重得到 $GIoU$ 。最后可以得到 $GIoU$ 的损失函数。

$GIoU$ 解决了 IoU 在两个框不相交时损失函数不可导、无法学习的问题, 并且反映出了两个框在相交面积相同时的相交情况, 但依然存在不足。好的目标框回归损失应该考虑 3 个重要的几何因素: 重叠面积, 中心点距离, 长宽比。由此提出了 $CIoU$ (Complete Intersection over Union)^[27], 其收敛的精度更高, 并且考虑到了全部以上 3 个因素。

$$L_{CIoU} = 1 - IoU + \frac{d^2}{L} + \alpha v \quad (11)$$

其中: d 为两个框中心点的欧式距离, L 为最小凸集 C 的对角线长度, αv 是惩罚项, 通过该项因子将宽高比因素考虑到回归损失中, α 是权重函数, 而 v 用来度量长宽比的相似性。

定义为:

$$\alpha = \frac{v}{1 - IoU + v} \quad (12)$$

$$v = 4/\pi^2 \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (13)$$

$CIoU$ 损失函数可以直接最小化两个目标框的距离, 因此比 $GIoU$ 损失函数收敛快得多。并且对于包含两个框在水平方向或垂直方向上这种情况, $CIoU$ 损失可以使回归非常快, 而 $GIoU$ 损失几乎退化为 IoU 损失。

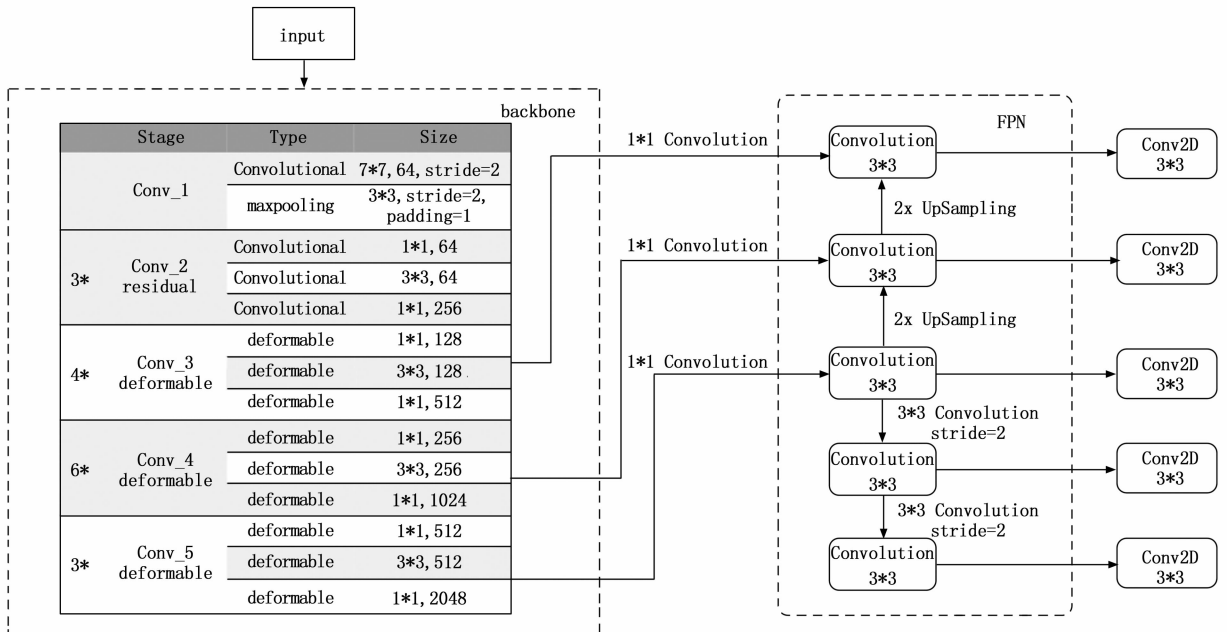


图 3 改进模型网络结构

3 实验结果分析

3.1 实验使用数据集

实验所用数据集为 FGVC-aircraft (Fine-Grained Visual Classification of Aircraft)^[28], 是经典的飞机模型视觉分类的公开基准数据集。

该数据集共有 100 个目标类别, 包括涂装改装后的不同机型数据, 共有 70 多个飞机家族总共 100 个不同型号的飞机。数据集包含了 10 000 张的不同飞机图片。数据集按照 1 : 2 的比例划分, 即训练集 6 666 张, 测试集 3 334 张。

3.2 模型训练

实验运行环境为, CPU 为 Corei7-12 700 K, GPU 为 NVIDIA GeForce RTX-3 060 等。系统环境为 Ubuntu16.04, CUDA11.7, python 3.10, pytorch 1.13.1, 模型框架为基于 pytorch 框架的 mmdetection 目标检测框架^[29]。

本实验在训练过程中采用 AdamW (Adam with Decoupled Weight Decay)^[30] 优化器进行损失优化。初始学习率设为 0.004, 每 500 个 iteration 输出一次日志记录, 为了加快训练学习过程的速度, 设置学习动量为 0.930, 权重衰减系数为 0.001, 以避免数据过拟合的情况出现。Warmup 设置为前 5 个 epoch 线性增长 0.01 当训练 epoch 到 8 次时, 学习率衰减到 10 倍, 当 epoch=12 次, 学习率衰减到 100 倍。最初实验的 batch-size 设为 16, 经过多次调参修改, 最终 batch-size 为 8 时模型的最终损失最小。

模型训练中的损失曲线如图 4 和图 5 所示。

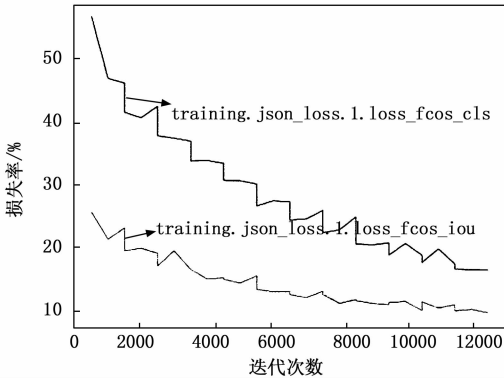


图 4 一对多分支损失

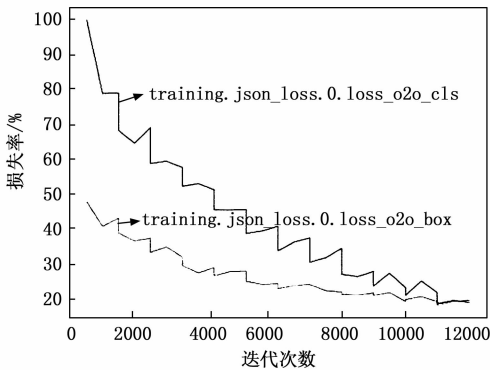


图 5 一对一分支损失

可以看到在训练的损失虽然有波动, 但一直在下降。在 12 000 步后, 模型的两个分支的损失都开始趋于稳定。

在模型训练开始训练后, 我们使用召回率 (Recall, R)、mAP0.5 和 mAP0.5 : 0.95 作为模型性能的评价指标。

其中, mAP0.5 表示 IoU 阈值为 0.5 时所有目标类别的平均检测精度, mAP0.5 : 0.95 表示以步长 0.05, 计算 IoU 从 0.5 到 0.95 总共 10 个 IoU 阈值下的平均检测精度, 通常情况下 IoU 阈值越高对于算法的回归能力要求更高。P、R 以及 mAP 的计算方法如下:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, AP = \int_0^1 P(R) dR, mAP = \frac{1}{n} \sum_{j=1}^n AP_j$$

其中: TP 指被模型预测为正的样本, FP 指被模型预测为正的负样本, FN 指被模型预测为负的正样本, mAP 为所有类别的平均精度, n 为检测目标总类别数。

3.3 消融实验

为进一步论证上述提出的改进检测算法性能, 本文对改进点逐一进行对比分析, 在模块消融实验中, 采用相同的测试集和学习率配置参数, 实验结果如表 1 所示。

表 1 模块消融实验结果

DCN	CIoU	CBAM	mAP/%	R/%	mAP _{0.5} /%
—	—	—	61.8	72.1	63.4
✓	—	—	69.8	75.7	71.5
✓	✓	—	70.3	85.4	71.2
✓	✓	✓	77.8	87.5	79.8

mAP 表示 mAP0.5 : 0.95 精度的数据, mAP_{0.5} 表示 mAP0.5 精度的数据, R 表示模型的回归率。

可以看出可变卷积和 CBAM 两项改进均对准确率有着较大的提升, 两项准确率数据都有着较高的提升; mAP0.5 : 0.95 提升了 11.3% 和 10.6%, R 提升了 4.9% 和 3.6%。

CIoU 的加入虽然对于模型的准确率没有明显的提升, 但加快了训练的回归速率, 并且提升了模型接近 10% 回归率, 增幅达到 12.8%。

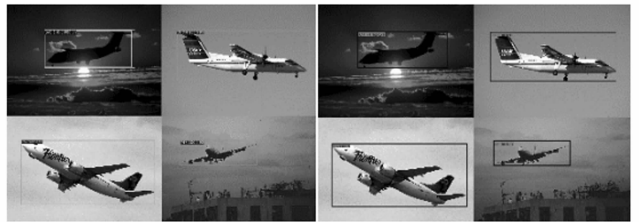


图 6 改进前后检测对比图

如图 6 所示, 将改进后的模型与原模型在 FGVC-aircraft 数据集下的检测结果进行对比, 左侧为原模型检测结果, 右侧为改进后模型检测效果。可以看出改进后的模型对于目标在图像中的标注更加准确, 且准确率更高, 对空

中飞行器的特征检测能力更强。

改进后的模型 0.5: 0.95 达到了 77.8%, 达到了 79.8%, 准确率提升至原模型的 1.11 倍。说明相关改进对于模型对于空中飞行器的检测性能确实具有有效提升。

3.4 对比实验

为综合分析改进后的 DATE-FCOS 模型算法的检测性能, 本文选择与 YOLO、FCOS、RetinaNet 几种经典目标检测算法, 采用相同的测试集进行实验对比以 mAP 、FPS 作为评价指标, 实验结果如表 2 所示。

表 2 对比实验结果

检测算法	$mAP/\%$	$R/\%$	$FPS/\%$
FCOS	66.8	82.7	22.3
RetinaNet	62.4	83.4	22.6
YOLOv3	67.6	85.2	21.4
DATE-FCOS	69.8	78.7	22.1
改进 DATE-FCOS	77.8	87.5	20.8

根据表 2 数据分析可知, 本文算法 mAP 为 77.8%, 比原始网络 DATE-FCOS 提高 9.13 个百分点, 在检测速度上差距并不明显, 这主要是因为一系列的优化改进结果; 而作为经典一阶段目标检测算法 YOLOv3 和 FCOS 则检测能力远不如改进后的 DATE-FCOS 算法, 此外 mAP 也比本文低了超过 10 个百分点, 本文算法在检测速度上略低, 但精度最高, 这主要是加入多层可形变卷积层和 CBAM 模块, 引入 CIoU 回归损失共同作用的结果。

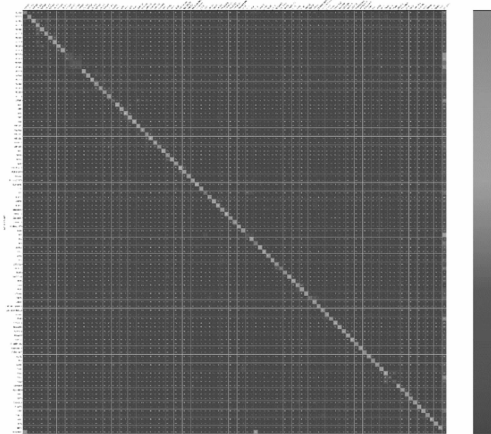


图 7 改进模型混淆矩阵

图 7 为改进后模型的混淆矩阵。图中纵轴代表的是标签, 横轴代表预测结果。由此可知对角线上的方格代表每个类别预测正确, 每个方格内的值代表行为出现的频率, 频率越高颜色越明亮。从混淆矩阵可以看出模型具有较好的准确性和泛化性。

4 结束语

为解决空中目标尺度变换和检测精度低的问题, 本文在 DATE-FCOS 基准框架下, 做出针对性改进。首先在

Resnet 主干网络中加入多个可形变卷积层, 提取更多尺度特征信息, 提升网络的特征提取能力, 之后在多级 FPN 结构后加入 CBAM 模块提升模型的表示能力, 最后修改模型的损失函数为 CIoU, 可提供更加准确的边界框回归方向, 加快训练收敛速度, 提升模型的回归率。

实验结果表明改进 DATE-FCOS 算法在使用相同的空中目标数据集的情况下模型的检测性能更加优秀, 并且从实验得到的客观数据分析看来, 改进后算法在 mAP 方面较原始算法提高显著, 涨幅达到 11%。同时通过对比, 算法性能也优于其他常见检测算法, 改进 DATE-FCOS 算法具有更好的泛化性和鲁棒性, 因此能够满足空中环境中目标的检测。但是仍存在个别特殊类别, 例如完全背光只有轮廓特征的检测目标, 对其识别效果较差, 后续将继续优化网络提高对少数特殊目标的检测精度。

参考文献:

- [1] 朱可汉. 基于 USB 2.0 高分辨率图像采集及实时图像处理算法的研究 [D]. 宁波: 宁波大学, 2013.
- [2] 范少伟, 严振华, 闫彬, 等. 不明空情处置的难点及对策 [J]. 空军雷达学院学报, 2011 (6): 434-436.
- [3] CHEN L, LIU Z, TONG L, et al. Underwater object detection using invert multi-class adaboost with deep learning [C] // 2020 International Joint Conference on Neural Networks (IJCNN), 2020: 1-8.
- [4] HAO X, ZHANG G, MA S. Deep learning [J]. International Journal of Semantic Computing, 2016, 10 (3): 417-439.
- [5] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521 (7553): 436-444.
- [6] SCHMIDHUBER J. Deep learning in neural networks: an overview [J]. Neural Networks, 2015, 61: 85-117.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6): 1137-1149.
- [8] KAIMING H. Mask R-CNN [C] // 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2980-2988.
- [9] REDMON J, ALI F. YOLOv3: An incremental improvement [Z/OL]. ArXiv abs/1804.02767, 2018.
- [10] ZHI T, CHUNHUA S, HAO C, et al. FCOS: Fully convolutional one-stage object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.
- [11] LIU W. SSD: single shot multibox detector [C] // European Conference on Computer Vision, 2015.
- [12] 王伟臣. 基于深度学习的空中目标识别方法研究 [D]. 西安: 西安电子科技大学, 2021.
- [13] 薛建伟, 史庆杰, 周泽强, 等. 基于改进 YOLOv3 和卡尔曼滤波器的飞机检测追踪方法 [J]. 飞控与探测, 2021, 4 (6): 70-77.

- [14] 王耀领, 王宏琦, 许滔. CGAN 样本生成的遥感图像飞机识别 [J]. 中国图像图形学报, 2021, 26 (3): 663-673.
- [15] 兰旭婷, 郭中华, 李昌昊. 基于注意力与特征融合的光学遥感图像飞机目标检测 [J]. 液晶与显示, 2021, 36 (11): 1506-1515.
- [16] JOSEPH R, SANTOSH D, ROSS G, et al. You only look once: unified, real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [17] TSUNG Y, PIOTR D, ROSS G, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2117-2125.
- [18] TSUNG Y, PRIYA G, ROSS G, et al. Focal loss for dense object detection [J]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [19] SUN P, JIANG Y, XIE E, et al. What makes for end-to-end object detection? [C] // International Conference on Machine Learning, 2020.
- [20] CHEN Y Q. DATE: dual assignment for end-to-end fully convolutional object detection [EB/OL]. ArXiv abs/2211.13859, 2022.
- [21] LOWE D G. Object recognition from local scale-invariant features transform [C] // ICCV, International Conference on Computer Vision, 1999.
- [22] LUO W, LI Y, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks [EB/OL]. ArXiv Preprint ArXiv: 1701.04128, 2017.
- [23] DAI T, FENG J. Deformable convolutional networks [C] // 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 764-773.
- [24] WANG R. DCN V2: improved deep & cross network and practical lessons for web-scale learning to rank systems [C] // Proceedings of the Web Conference, 2021.
- [25] SANGHYUN W, JONGCHAN P, JOON-YOUNG L, et al. CBAM: convolutional block attention module [C] // Computer Vision. 15th European Conference (ECCV 2018), 2018.
- [26] REZATOFI GHI, SEYED H. Generalized intersection over union: a metric and a loss for bounding box regression [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 658-666.
- [27] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: faster and better learning for bounding box regression [C] // 34th AAAI Conference on Artificial Intelligence, 2020.
- [28] MAJI S. Fine-grained visual classification of aircraft [EB/OL]. ArXiv abs/1306.5151, 2013.
- [29] CHEN K. MMDetection: open MMLab detection toolbox and benchmark [EB/OL]. ArXiv abs/1906.07155, 2019.
- [30] LOSHCHILOV I, FRANK H. Fixing weight decay regularization in Adam [EB/OL]. ArXiv abs/1711.05101, 2017.
- ***
(上接第 77 页)
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. arXiv: 1804.02767 v1, 2018.
- [7] LIN Q Y, LI S, WANG R G, et al. Research on small target detection technology based on the MPH-SSD algorithm [J]. Computational Intelligence and Neuroscience, 2022: 1-13.
- [8] LIN Q Y, WANG R G, YUANYUAN WANG, et al. Target detection algorithm incorporating visual expansion mechanism and path syndication [J]. IEEE Access, 2023.
- [9] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.
- [10] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks [C] // Proceedings of 2016 Conference Advances in Neural Information Processing Systems. Barcelona, Spain; MIT Press, 2016: 379-387.
- [11] 代小宇. 基于改进 YOLOv5 算法的复杂环境下交通标志的检测与识别 [D]. 银川: 宁夏大学, 2022.
- [12] 尹宋麟, 谭飞, 周晴, 等. 基于改进 YOLOv4 模型的交通标志检测 [J]. 无线电工程, 2022, 52 (11): 2087-2093.
- [13] 赵宏, 冯宇博. 一种基于 CGS-Ghost YOLO 的交通标志检测研究 [J]. 计算机工程, 2023, 49 (12): 194-204.
- [14] DUBEY U, CHAURASIYA R K. Efficient traffic sign recognition using CLAHE-based Image Enhancement and ResNet CNN Architectures [J]. International Journal of Cognitive Informatics and Natural Intelligence (IJCIN), 2022, 15 (4): 1-19.
- [15] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // 2017 IEEE International Conference on Computer Vision (ICCV). Venice; IEEE, 2017: 2980-2988.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas; IEEE, 2016: 770-778.
- [17] RACHMADI R F, KOUTAKI G, OGATA K. Lightweight spatial pyramid convolutional neural network for traffic sign classification [C] // Indonesian Association for Pattern Recognition International Conference, 2018: 23-28.
- [18] 党宏社, 党晨, 张选德. 基于改进 YOLOv5s 的交通标志识别算法 [J]. 实验技术与管理, 2022, 39 (9): 97-102.
- [19] 吕禾丰, 陆华才. 基于 YOLOv5 算法的交通标志识别技术研究 [J]. 电子测量与仪器学报, 2021, 35 (10): 137-144.
- [20] 王林, 张文卓. 一种融合注意力机制与上下文信息的交通标志检测方法 [J]. 计算机测量与控制, 2022, 30 (3): 54-59.