

基于 Improved RetinaNet 的航空图像 定向目标检测

刘克顺, 李爱华, 王长龙, 张玉华

(陆军工程大学石家庄校区, 石家庄 050003)

摘 要: 航空图像中的目标(如车辆、船只、建筑物等)可能以任意方向出现, 目标的尺度差异显著, 传统的不定向边界框难以准确描述目标的方向和形状, 目标在图像中的位置、方向和形状与实际情况难以避免会出现偏差, 导致小目标检测回归损失不确定性较高, 检测精度和效率较低。针对上述问题, 首先应用 Improved RetinaNet 来实现航空图像中定向目标的精确检测, 设计了一个特征交互融合网络来获取定向目标丰富的上下文信息。然后, 引入了一种训练样本动态分配方法, 根据目标的空间位置关系动态地分配训练样本, 以协调训练样本之间的不平衡, 并通过分类子网络预测目标类别。最后, 利用回归子网络, 回归每个定向边界框和邻近的地面真实框之间的偏移量, 并采用长边定义法表示定向边界框, 计算定向边界框和地面真实边界框之间的 KL 散度作为回归损失, 以解决大高宽比定向目标的角度敏感问题。在公开的航空图像数据集上的大量实验证明了 Improved RetinaNet 在检测定向目标方面的有效性和优越性。

关键词: 航空图像; 定向目标检测; 特征交互融合网络; 训练样本动态分配; KL 散度

中图分类号: TP753 **文献标志码:** A **文章编号:** 1671-5497(2026)06-1677-10

DOI: 10.13229/j.cnki.jdxbgxb.20250193

Oriented object detection in aerial images based on improved retinaNet

LIU Ke-shun, LI Ai-hua, WANG Chang-long, ZHANG Yu-hua

(Shijiazhuang Campus of Army Engineering University of PLA, Shijiazhuang 050003, China)

Abstract: Objects in aerial images, such as vehicles, ships, buildings, etc., may appear in any direction with significant differences in their scale. Traditional non oriented bounding boxes are difficult to accurately describe the direction and shape of objects, and the position, direction, and shape of objects in the image may deviate from the actual situation, resulting in high regression loss uncertainty and low detection accuracy and efficiency in small object detection. Therefore, Improved RetinaNet is applied to achieve precise detection of oriented objects in aerial images. Designed a feature interaction fusion network to obtain rich contextual information of oriented objects. A dynamic allocation method for training samples has been introduced, which dynamically allocates training samples based on the spatial position relationship of the

收稿日期: 2025-03-12.

基金项目: 国防科技基础加强计划技术领域基金项目(2019-JCJQ-JJ-015).

作者简介: 刘克顺(1997-), 男, 博士. 研究方向: 计算机视觉, 目标检测. E-mail: llkss123123@126.com

通信作者: 张玉华(1978-), 女, 讲师, 博士. 研究方向: 无人机辅助图像处理. E-mail: zhangyh82022@126.com

object to coordinate the imbalance between training samples and predict object categories through classification sub-network. Based on this, a regression sub-network is used to regress the offset between each oriented bounding box and the adjacent ground truth bounding box, and the long edge definition method is used to represent the oriented bounding box. The KL divergence between the oriented bounding box and the ground truth bounding box is calculated as the regression loss to solve the angle sensitivity problem of large aspect-ratio oriented objects. Extensive experiments on publicly available aerial image datasets have demonstrated the effectiveness and superiority of Improved RetinaNet in detecting oriented objects.

Key words: aerial image; oriented object detection; feature interaction fusion network; training sample dynamic allocation; KL divergence

0 引言

无人机侦察是战争中获取情报的重要手段,其中目标检测是最重要的任务之一。定向目标检测是一个相对新兴但富有挑战性的领域,近年来的一些研究取得了相当大的研究进展^[1-4]。然而,对于航空图像中方向任意、高宽比大的目标,仍然存在着许多挑战,例如它们往往以多种角度分布并且被复杂的背景所包围,只保留了很少的有效特征信息。

目前,主流的定向目标检测器通常利用角度定义法(中心点坐标、宽度、高度、旋转角度)来表示任意方向的目标。虽然这些检测器已经取得了良好的性能,但是当它们在航空图像中检测定向目标时,仍然存在着以下不足:①特征旋转增强的方法^[5-9]精心设计了各种特征增强模块,并在每个位置上设置了不同角度、比例和高宽比的锚框,使得网络变得沉重且复杂,导致了大量的计算开销和内存占用;②样本动态分配的方法^[10-12]设置了一些超参数,如基于锚框的方法的 IoU (Intersection over Union, 交并比) 阈值、无锚框的方法的区域范围等,超参数的不同设置将会产生非常不同的检测结果;③回归损失优化的方法^[13-16]不能根据定向目标的高度和宽度动态调整梯度,难以应对具有挑战性的情况,例如具有大高宽比的定向目标对角度的偏移很敏感,轻微的角度误差就会导致检测精度严重下降。

针对现有方法的不足之处,本文提出了一种称为 Improved RetinaNet 的检测器,用于航空图像中定向目标的精确检测。本文主要贡献如下:

(1) 在主干网络 ResNeXt-101 的基础上设计

特征交互融合网络,在不增加网络复杂度的情况下获取定向目标丰富的上下文信息,同时减少模型参数的数量。

(2) 在分类子网络中引入训练样本动态分配方法,根据目标的空间位置关系动态地分配训练样本,在不设置任何超参数的情况下协调训练样本之间的不平衡。

(3) 在回归子网络中采用长边定义法表示定向边界框,将预测定向边界框和地面真实边界框转换为二维高斯分布,计算二维高斯分布之间的 KL (Kullback-Leibler) 散度作为回归损失,以解决大高宽比定向目标的角度敏感问题。

1 Improved RetinaNet

1.1 总体框架

Improved RetinaNet 是由一个主干网络、一个特征交互融合网络和两个子网络组成的单阶段统一网络。首先,在主干网络 ResNeXt-101^[17] 的基础上设计特征交互融合网络(FIFN),在不增加网络复杂度的情况下获取定向目标丰富的上下文信息。然后,在分类子网络中引入训练样本动态分配(TSDA)方法,根据目标的空间位置关系动态地分配训练样本。最后,在回归子网络中采用长边定义法($\theta \in [-45, 135]$)表示定向边界框,并将 KL 散度作为回归损失。

Improved RetinaNet 的总体框架如图 1 所示。Improved RetinaNet 在每个特征图上设置 A 个锚框。每个锚框分配一个长度为 K 的独热(0-1)向量用于目标分类,其中 K 表示目标类别数量,以及一个长度为 5 的向量用于目标回归,其中 5 表示定向边界框的 5 个参数(x, y, w, h, θ)。

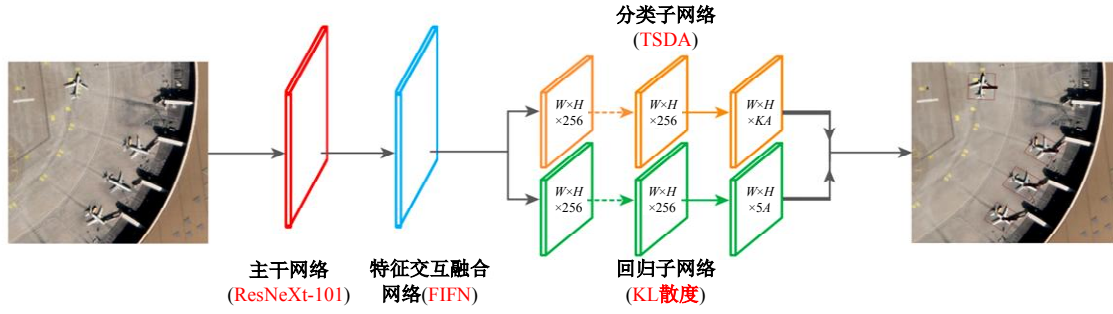


图 1 Improved RetinaNet 的总体框架

Fig. 1 Overall framework of improved RetinaNet

1.2 特征交互融合网络下获取定向目标信息

传统的定向目标检测方法通过设计特征增强模块或构建深层特征提取网络来提高目标检测精度,然而,繁多的参数数量(如特征层数、通道数、卷积核大小等)导致了大量的计算开销和内存占用。因此,特征交互融合网络建立在主干网络 ResNeXt-101 的基础之上,采用聚合变换和分组卷积的方式,可以在不增加网络复杂度的情况下获取定向目标丰富的上下文信息,同时减少模型参数的数量。

与之前的方法不同(直接在通道连接后对相邻特征图进行卷积以输出预测特征图),特征交互融合网络逐步融合了更深层次的特征图。首先,将主干网络的特征图 $f_i (i \in \{1, 2, 3\})$ 与交互特征图 $f_i^{\text{inter}} (i \in \{1, 2, 3\})$ 相乘,以抑制背景干扰,然后通过残差加法来保留原始特征信息。

具体情况如图 2 所示。

由于 f_4 是主干网络的最后一个特征图,直接对 f_4 进行 3×3 卷积来生成 $\hat{f}_4$ 。对于 f_3 ,在 f_4 上执行特征交互来形成 f_3^{inter} 。相似地, f_2^{inter} 和 f_1^{inter} 如下所示:

$$\hat{f}_4 = \text{Conv3}(f_4) \quad (1)$$

$$f_3^{\text{inter}} = \text{Conv3}(\text{Conv1}(f_4 \uparrow)) \quad (2)$$

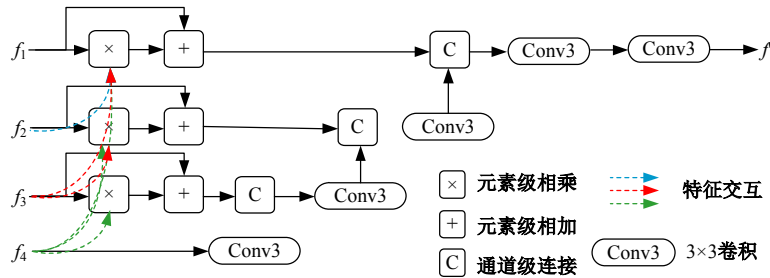


图 2 特征交互融合网络

Fig. 2 Feature interaction fusion network

$$\hat{f}_3 = \text{Conv3}([(f_3^{\text{inter}} \otimes f_3) \oplus f_3; \hat{f}_4]) \quad (3)$$

$$f_2^{\text{inter}} = \text{Conv3}(\text{Conv1}([f_4 \uparrow; f_3 \uparrow])) \quad (4)$$

$$\hat{f}_2 = \text{Conv3}([(f_2^{\text{inter}} \otimes f_2) \oplus f_2; \hat{f}_3]) \quad (5)$$

$$f_1^{\text{inter}} = \text{Conv3}(\text{Conv1}([f_4 \uparrow; f_3 \uparrow; f_2 \uparrow])) \quad (6)$$

$$\hat{f}_1 = \text{Conv3}([(f_1^{\text{inter}} \otimes f_1) \oplus f_1; \hat{f}_2]) \quad (7)$$

$$f' = \text{Conv3}(\hat{f}_1) \quad (8)$$

式中: $\hat{f}_i (i \in \{1, 2, 3, 4\})$ 表示聚合特征图; Conv3 表示 3×3 卷积; Conv1 表示 1×1 卷积; $\uparrow$ 表示上采样; $;$ 表示通道级连接; $\otimes$ 表示元素级相乘; $\oplus$ 表示元素级相加。

最后,将顶部的特征图 f' 作为特征交互融合网络的最终结果,从而使 Improved RetinaNet 在不增加网络复杂度的情况下从输入图像中获取定向目标丰富的上下文信息,同时减少模型参数的数量。

1.3 分类子网络下目标类别预测

在模型训练时,首先需要定义模型的正样本和负样本,然后使用正样本来确定目标的位置。因此,前者是至关重要的。以往的训练样本分配方法设置了一些超参数,如基于锚框的方法的 IoU 阈值和无锚框的方法的尺度范围。对于基于锚框的方法,与锚框形状不同的目标将会产生较

低的IoU,导致缺乏正样本。对于无锚框的方法,固定的尺度范围将会限制高质量正样本的上限。因此,这些超参数的不同设置将会产生非常不同的检测结果。

为解决上述问题,ATSS^[18]提出了自适应训练样本选择方法,通过锚框与地面真实边界框之间的L2距离来选择候选正样本,最后依据目标的统计特征自适应地分配训练样本,在整个过程中只设置了一个超参数。然而,L2距离对目标的大小、方向以及高宽比变化不敏感,这使得ATSS在面对航空图像中方向任意、高宽比大的目标时效果不佳。因此,引入训练样本动态分配方法,根据目标的空间位置关系动态地分配训练样本,并且不设置任何超参数。该方法的步骤如下:

步骤1 对于图像上的每个地面真实边界框 g ,计算所有中心在地面真实边界框 g 中的锚框与地面真实边界框 g 之间的IoU值。

步骤2 将步骤1中的所有IoU值进行求和,并四舍五入为整数 K 。

步骤3 将步骤1中的所有IoU值从大到小依次排序。

步骤4 选择前 K 个IoU值对应的锚框作为最终的正样本,其余的样本则为负样本。

分类子网络的整体训练过程总结如下:首先将特征交互融合网络的融合特征图 f' 作为输入,然后应用训练样本动态分配方法定义模型的正负样本,最后采用Sigmoid激活函数输出二进制目标类别的预测。此外,使用Focal损失作为分类子网络中的分类损失,该损失通过调整焦点因子来控制正负样本对总体损失的影响。

$$L_{\text{cls}} = -\alpha(1 - r_p)^\gamma \log(r_p) \quad (9)$$

$$r_p = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (10)$$

式中: α 为焦点因子,选择 α 为0.25来解决类不平衡的问题; γ 为调制因子,选择 γ 为2来减少大量高置信度负样本的影响; r_p 表示某一类别的分类概率,如果某一类别的分类结果正确,则 $r_p = p$,否则 $r_p = 1 - p$, p 表示某一类别的分类置信度。

1.4 回归子网络下定向边界框偏移回归

回归子网络回归每个定向边界框和邻近的地面真实框之间的偏移量。在回归子网络中,每个定向边界框由5个参数 (x, y, w, h, θ) 表示,其中 x

为定向边界框中心点的横坐标, y 为定向边界框中心点的纵坐标, w 为定向边界框的宽度, h 为定向边界框的高度, θ 为定向边界框的旋转角度。考虑到在 $\theta \in [-90^\circ, 90^\circ)$ 位置上的目标很多,且它们不适合作为边界,因此采用长边定义法($\theta \in [-45^\circ, 135^\circ)$)来表示定向边界框,如图3所示。其中,定向边界框的宽度 w 定义为长边,定向边界框的旋转角度 θ 定义为 x 轴与长边的夹角,顺时针方向为正值。

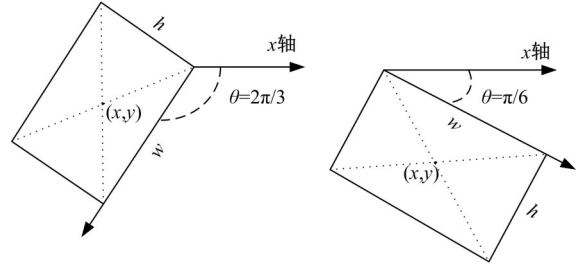


图3 长边定义法($\theta \in [-45^\circ, 135^\circ)$)

Fig. 3 Long edge definition method($\theta \in [-45^\circ, 135^\circ)$)

为了更准确地表示定向边界框,将KL(Kullback-Leibler)散度^[19]作为回归损失。首先,将定向边界框转换为二维高斯分布 $G(\mu, \Sigma)$ 。

$$\Sigma = R\Lambda R^T, \mu = (x, y)^T \quad (11)$$

式中: R 为旋转矩阵; Λ 为特征值的对角线矩阵; μ, Σ 分别为二维高斯分布的均值和方差。

对于定向边界框 $B(x, y, w, h, \theta)$:

$$R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \Lambda = \begin{pmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{pmatrix} \quad (12)$$

两个二维高斯分布之间的KL散度为:

$$D_{\text{kl}}(N_p \| N_i) = \frac{1}{2}(\mu_p - \mu_i)^T \Sigma_i^{-1}(\mu_p - \mu_i) + \frac{1}{2} \text{Tr}(\Sigma_i^{-1} \Sigma_p) + \frac{1}{2} \ln \frac{|\Sigma_i|}{|\Sigma_p|} - 1 \quad (13)$$

$$D_{\text{kl}}(N_i \| N_p) = \frac{1}{2}(\mu_p - \mu_i)^T \Sigma_p^{-1}(\mu_p - \mu_i) + \frac{1}{2} \text{Tr}(\Sigma_p^{-1} \Sigma_i) + \frac{1}{2} \ln \frac{|\Sigma_p|}{|\Sigma_i|} - 1 \quad (14)$$

式中: $N_p(\mu_p, \Sigma_p)$ 、 $N_i(\mu_i, \Sigma_i)$ 表示服从二维高斯分布。

可以看出, $D_{\text{kl}}(N_i \| N_p)$ 中的每一项都由部分参数耦合组成,这使得所有参数形成一个链耦合

关系。在优化过程中,参数相互影响并共同优化,使模型的优化机制是自调制的。

虽然 KL 散度是不对称的,但通过分析不同参数的梯度可以发现,这 2 种形式的优化原理是相似的。以 $D_{kl}(N_p \| N_t)$ 为例,根据式(11)(12),式(13)中的每一项都可以表示为:

$$\begin{aligned} & (\mu_p - \mu_t)^T \Sigma_t^{-1} (\mu_p - \mu_t) = \\ & \frac{4(\Delta x \cos \theta_t + \Delta y \sin \theta_t)^2}{w_t^2} + \frac{4(\Delta y \cos \theta_t - \Delta x \sin \theta_t)^2}{h_t^2} \end{aligned} \quad (15)$$

$$\begin{aligned} Tr(\Sigma_t^{-1} \Sigma_p) &= \frac{h_p^2}{w_t^2} \sin^2 \Delta \theta + \frac{w_p^2}{h_t^2} \sin^2 \Delta \theta + \\ & \frac{h_p^2}{h_t^2} \cos^2 \Delta \theta + \frac{w_p^2}{w_t^2} \cos^2 \Delta \theta \end{aligned} \quad (16)$$

$$\ln \frac{|\Sigma_t|}{|\Sigma_p|} = \ln \frac{h_t^2}{h_p^2} + \ln \frac{w_t^2}{w_p^2} \quad (17)$$

式中: $\Delta x = x_p - x_t$, $\Delta y = y_p - y_t$, $\Delta \theta = \theta_p - \theta_t$ 。

当设置 $\theta_t = 0^\circ$ 时:

$$\frac{\partial D_{kl}(\mu_p)}{\partial \mu_p} = \left(\frac{4}{w_t^2} \Delta x, \frac{4}{h_t^2} \Delta y \right)^T \quad (18)$$

权重 $1/w_t^2$ 和 $1/h_t^2$ 将使模型根据尺度动态调整位置。例如,当尺度较小时,模型会更加关注相应方向偏移的优化。当 $\theta_t \neq 0^\circ$ 时,偏移(Δx 和 Δy)的梯度将根据 θ_t 进行动态调整,以获得更好的优化。

对于 h_p 和 w_p :

$$\frac{\partial D_{kl}(\Sigma_p)}{\partial \ln h_p} = \frac{h_p^2}{h_t^2} \cos^2 \Delta \theta + \frac{h_p^2}{w_t^2} \sin^2 \Delta \theta - 1 \quad (19)$$

$$\frac{\partial D_{kl}(\Sigma_p)}{\partial \ln w_p} = \frac{w_p^2}{w_t^2} \cos^2 \Delta \theta + \frac{w_p^2}{h_t^2} \sin^2 \Delta \theta - 1 \quad (20)$$

一方面, h_p 和 w_p 的优化受到 $\Delta \theta$ 的影响。当 $\Delta \theta = 0^\circ$, $\frac{\partial D_{kl}(\Sigma_p)}{\partial \ln h_p} = \frac{h_p^2}{h_t^2} - 1$, $\frac{\partial D_{kl}(\Sigma_p)}{\partial \ln w_p} = \frac{w_p^2}{w_t^2} - 1$ 时,高度或宽度越小,匹配损失的惩罚越大。这是显而易见的,因为较小的高度或宽度需要较高的匹配精度。另一方面, $\Delta \theta$ 的优化也受到 h_p 和 w_p 的影响:

$$\frac{\partial D_{kl}(\Sigma_p)}{\partial \theta_p} = \left(\frac{h_p^2 - w_p^2}{w_t^2} + \frac{w_p^2 - h_p^2}{h_t^2} \right) \sin 2\Delta \theta \quad (21)$$

式中: $w_p = w_t$, $h_p = h_t$, 然后 $\frac{\partial D_{kl}(\Sigma_p)}{\partial \theta_p} = \left(\frac{h_t^2}{w_t^2} + \right.$

$\left. \frac{w_t^2}{h_t^2} - 2 \right) \sin 2\Delta \theta \geq \sin 2\Delta \theta$, 公式相等的条件是

$h_t = w_t$ 。这表明目标的高宽比越大,就会越关注角度的优化,因为轻微的角度误差会导致大高宽比定向目标的检测精度严重下降。通过以上分析发现,当其中一个参数被优化时,将使用其他参数作为其权重来动态调整优化。换句话说,参数的优化不再是独立的,即优化一个参数也会促进其他参数的优化。这种良性循环的优化是 KL 散度作为一种良好的回归损失的关键。

整体训练过程总结如下:①预测偏移量($t_x^p, t_y^p, t_w^p, t_h^p, t_\theta^p$);②解码预测定向边界框;③将预测定向边界框和地面真实边界框转换为二维高斯分布;④计算两个二维高斯分布的 KL 散度。最终的回归损失 L_{reg} 为:

$$L_{reg} = 1 - \frac{1}{\tau + f(D)}, \tau \geq 1 \quad (22)$$

式中: $f(\cdot)$ 表示一个用来变换距离 D 的非线性函数,使用 $\ln(D+1)$ 作为 $f(\cdot)$ 来使损失更平滑和富有表现力。超参数 τ 用来调节整个损失。

定向目标检测的总体损失为:

$$L = \frac{\lambda_1}{N_{pos}} \sum_{n=1}^{N_{pos}} L_{reg}(b_n, gt_n) + \frac{\lambda_2}{N} \sum_{n=1}^N L_{cls}(p_n, t_n) \quad (23)$$

式中: N 和 N_{pos} 分别为所有锚框的数量和正锚框的数量; b_n 为第 n 个预测定向边界框; gt_n 为第 n 个地面真实边界框; p_n 为用 Sigmoid 激活函数计算出的各种类别的第 n 个概率分布; t_n 为第 n 个目标的类别标签; λ_1, λ_2 为控制整体函数的权衡; 分类损失 L_{cls} 为 Focal 损失。

2 实验验证

2.1 数据集和实施细节

数据集: DOTA-v1.0^[20] 是用于航空图像定向目标检测的大规模数据集之一,包含 15 个常见类别,2 806 张图像和 188 282 个实例。15 个常见类别分别为:港口(HA)、储罐(ST)、大型车辆(LV)、游泳池(SP)、足球场(SBF)、桥梁(BR)、船舶(SH)、飞机(PL)、棒球场(BD)、网球场(TC)、地面跑道(GTF)、篮球场(BC)、小型车辆(SV)、环岛(RA)和直升机(HC)。DOTA-v1.0 数据集的目标在方向、大小和高宽比上变化很大,这增加了目标检测的难度。DroneVehicle^[21] 是一个针

对航空图像中车辆目标检测的数据集,包含 5 个目标类别,31 064 张图像和 441 642 个实例,其中一半是可见光图像。5 个目标类别分别为:公共汽车(Bus)、货车(Fright)、卡车(Truck)、汽车(Car)和面包车(Van)。DroneVehicle 数据集的图像涵盖了从白天到晚上的各种场景,车辆目标具有不同的尺度和角度变化。此外,为准确地检测不同角度的车辆目标,以适应定向目标检测要求,将 DroneVehicle 数据集的数据标注格式从水平标注的 VOC 格式修改为定向标注的 DOTA 格式。

实施细节:所有的实验都在一个具有 8 GB 内存的 NVIDIA RTX 3070Ti GPU (Graphics Processing Unit, 图形处理器)和 MMRotate^[22]平台上进行训练和测试。采用 SGD(Stochastic gradient descent, 随机梯度下降)优化器运行 12 个 epoch,权重衰减和动量参数分别设置为 0.000 1 和 0.9。使用 500 次迭代的学习率预热,初始学习率设置为 0.002 5,并在第 8 个 epoch 和第 11 个 epoch 结束后学习率降低 10 倍。在训练过程中,使用翻转比为 0.25 的水平、垂直和对角线图像翻转作为数据增强的形式。所有的实验都进行了两次以上,并以稳定值作为最终结果。

2.2 消融实验

为验证 Improved RetinaNet 的优越性,进行了一系列消融实验。选择 Rotated-RetinaNet 作为消融实验的基线,采用平均精度均值(Mean average precision, mAP)来评估目标的检测精度,所有的实验条件和参数设置均严格一致。在 DOTA-v1.0 数据集上研究了特征交互融合网络(FIFN)、训练样本动态分配(TSDA)方法和 KL 散度损失对实验结果的影响,如表 1 所示。FIFN 表示是否使用特征交互融合网络,TSDA 表示是否使用训练样本动态分配方法,KL 表示是否使用 KL 散度损失。实验结果表明,每种结构对检测性能都有

表 1 各种结构对实验结果的影响

Table 1 Impact of various structures on experimental results

基线	FIFN	TSDA	KL	mAP/ %	单独 提升	总体 提升
	×	×	×	77.69		
Rotated-	√	×	×	79.87	+2.18	+2.18
RetinaNet	√	√	×	81.01	+1.14	+3.32
	√	√	√	82.37	+1.36	+4.68

不同程度的提高。

表 2 比较了 FIFN 和其他的特征融合网络。与现有的特征融合网络不同,FIFN 能够在不增加网络复杂度的情况下获取定向目标丰富的上下文信息,同时减少模型参数的数量,在模型参数上分别降低了 0.52 M、4.97 M、30.54 M 和 8.03 M,在 mAP 上分别提升了 5.55%、0.65%、0.44% 和 0.19%,取得了最好的定向目标检测结果。

表 2 与不同特征融合网络的比较

Table 2 Comparison with different feature fusion networks

基线	特征融合网络	模型参数/M	mAP/%
	FPN ^[23]	39.23	76.82
	PAFPN ^[24]	43.68	81.72
Rotated-	NAS-FPN ^[25]	69.25	81.93
RetinaNet	FGAFPN ^[26]	46.74	82.18
	FIFN	38.71	82.37

表 3 比较了 TSDA 和其他的训练样本分配方法。与现有的训练样本分配方法不同,TSDA 根据目标的空间位置关系动态地分配训练样本,在 mAP 上分别提升了 15.33%、15.30%、2.48% 和 1.07%,具有更明显的性能提高。

表 3 与不同训练样本分配方法的比较

Table 3 Comparison with different training sample allocation methods

基线	训练样本分配方法	mAP/%
	Max-IoU ^[10]	67.04
	Max-Convex-IoU ^[2]	67.07
Rotated-	SASM ^[11]	79.89
RetinaNet	ATSS ^[18]	81.30
	TSDA	82.37

表 4 比较了 KL 散度损失和其他的回归损失。与现有的回归损失不同,KL 散度损失由于其先进的自调制优化机制,在 mAP 上分别提升了 2.99%、1.46%、1.30% 和 0.69%,取得了最好的定向目标检测结果。

表 4 与不同回归损失的比较

Table 4 Comparison with different regression losses

基线	回归损失	mAP/%
	L1 ^[13]	79.38
	GWD ^[15]	80.91
Rotated-	Smooth L1 ^[5]	81.07
RetinaNet	KRRD ^[16]	81.68
	KL	82.37

2.3 与最先进技术的比较

2.3.1 定量分析

将 Improved RetinaNet 与当前流行的定向目标检测方法在 DOTA-v1.0 数据集上进行了定量分析,如表 5 所示。

由于不同的方法使用不同的图像分辨率、网络结构、训练策略和各种技巧,不能做出绝对公平

的比较。

就定向目标的检测性能而言,Improved RetinaNet 在 15 个目标类别中取得了 10 个目标类别的最高检测精度。就整体性能而言,Improved RetinaNet 的 mAP 达到了 82.37%,取得了最优的定向目标检测结果,比次优的 KRRD 方法的 mAP 增加了 0.69%(从 81.68% 增加到 82.37%)。

表 5 DOTA-v1.0 数据集上不同定向目标检测方法的 mAP

Table 5 mAP of different methods for oriented object detection on DOTA-v1.0 dataset																
方法	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP/%
Gliding Vertex ^[14]	89.64	85.00	52.26	77.34	73.01	73.14	86.82	90.74	79.02	86.81	59.55	70.91	72.94	70.86	57.32	75.02
CSL ^[13]	90.67	86.76	52.21	80.09	72.68	72.62	78.54	90.83	83.99	71.83	67.54	79.38	63.55	82.44	79.91	76.87
R ³ Det ^[5]	90.66	85.52	58.78	83.71	74.72	79.37	79.63	90.82	84.69	73.34	64.27	77.28	72.45	82.33	67.89	77.69
Rotated-RetinaNet ^[10]	90.65	86.04	57.47	84.05	72.02	71.62	78.03	90.77	85.52	72.11	67.86	84.27	63.74	80.68	80.41	77.69
GWD ^[15]	90.67	87.71	56.59	84.02	73.62	76.54	85.11	90.83	83.18	75.47	67.22	83.38	64.84	82.39	81.13	78.84
S ² a-Net ^[6]	88.89	83.60	57.74	81.95	79.94	83.19	89.11	90.78	84.87	87.81	70.30	68.25	78.30	77.01	69.58	79.42
RoI-Transformer ^[7]	88.81	82.48	60.83	83.82	78.34	83.06	88.31	90.87	88.77	87.03	68.65	66.90	79.26	79.71	74.67	80.10
Oriented-RCNN ^[8]	89.84	85.43	61.09	82.82	79.71	82.35	88.82	90.88	86.68	87.73	72.21	70.80	82.42	78.18	74.11	80.87
Rotated-ATSS ^[18]	90.68	87.62	68.77	87.63	73.48	82.34	83.16	90.79	88.68	78.81	70.91	81.76	73.98	84.07	72.24	81.01
SA3Det ^[9]	90.67	87.74	69.52	87.31	73.49	82.52	83.27	90.71	88.77	78.72	71.43	82.02	72.54	82.63	74.50	81.06
DWLA ^[12]	90.65	87.84	70.12	87.44	74.16	82.02	83.55	90.73	88.75	79.28	72.19	81.89	72.32	81.84	76.67	81.30
KRRD ^[16]	90.67	87.78	69.40	87.47	75.25	82.33	83.64	90.74	88.83	80.17	72.42	82.35	72.53	82.25	79.41	81.68
Improved RetinaNet	90.69	87.91	72.18	89.13	73.43	83.41	83.49	90.92	89.14	78.67	74.11	85.94	71.28	81.32	83.05	82.37

将 Improved RetinaNet 与当前流行的定向目标检测方法在 DroneVehicle 数据集上进行了定量分析,如表 6 所示。就定向目标的检测性能而言,Improved RetinaNet 取得了所有目标类别的最高检测精度。就整体性能而言,Improved RetinaNet 的 mAP 达到了 61.92%,取得了最优的定向目标检测结果,比次优的 KRRD 方法的 mAP 增加了 1.26%(从 60.66% 增加到 61.92%)。

表 6 DroneVehicle 数据集上不同定向目标检测方法的 mAP

Table 6 mAP of different methods for oriented object detection on DroneVehicle dataset						
方法	Bus	Fright	Truck	Car	Van	mAP/%
Gliding Vertex ^[14]	84.99	55.05	46.33	83.06	17.22	57.33
CSL ^[13]	85.07	55.32	47.01	83.26	17.42	57.62
R ³ Det ^[5]	84.69	55.51	46.99	84.72	17.39	57.86
Rotated-RetinaNet ^[10]	86.53	55.04	47.24	84.66	17.43	58.18
GWD ^[15]	85.78	56.08	46.45	85.02	17.89	58.24
S ² a-Net ^[6]	85.83	55.92	47.42	85.64	19.43	58.85
RoI-Transformer ^[7]	87.65	56.43	47.81	86.02	18.38	59.26
Oriented-RCNN ^[8]	87.24	55.32	48.23	87.84	19.78	59.68
Rotated-ATSS ^[18]	86.53	57.43	48.98	87.32	21.46	60.34
SA3Det ^[9]	86.97	57.35	49.03	87.52	21.45	60.46
DWLA ^[12]	87.04	57.25	49.06	88.13	21.48	60.59
KRRD ^[16]	87.15	57.11	49.19	88.43	21.44	60.66
Improved RetinaNet	88.38	58.26	51.02	90.53	21.51	61.92

2.3.2 定性分析

将 Improved RetinaNet 和当前流行的定向目标检测方法在 DOTA-v1.0 数据集上进行了定性分析,如图 4 所示。可以观察到,在图 4 的第一列中,Improved RetinaNet 设计了特征交互融合网络来获取目标丰富的上下文信息,不仅可以解决 Rotated-RetinaNet、Oriented-RCNN 和 KRRD 方法将小型车辆(SV)错误检测成大型车辆(LV)

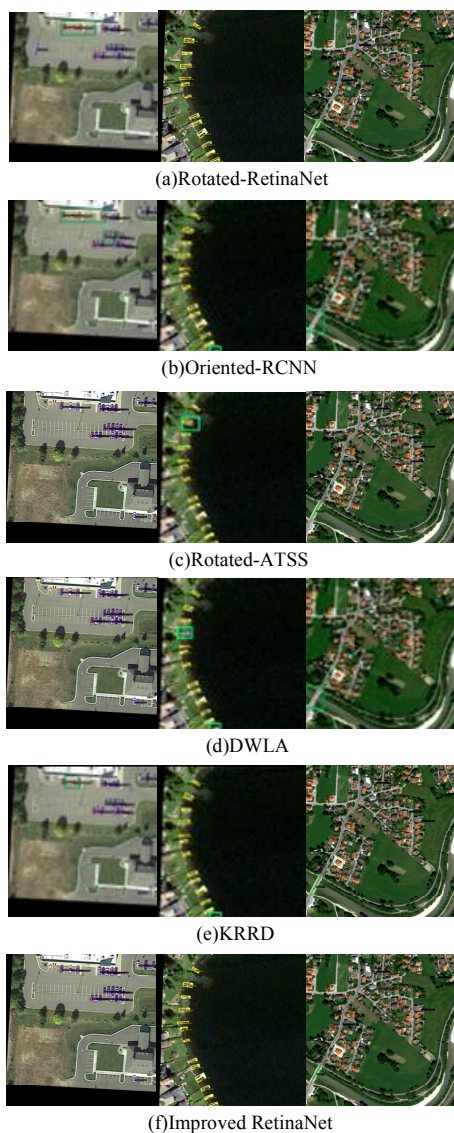


图4 DOTA-v1.0数据集上的可视化比较

Fig. 4 Visual comparison on DOTA-v1.0 dataset

的问题,而且对小型车辆(SV)具有最高的检测精度。在图4的第二列中,Improved RetinaNet引入了训练样本动态分配方法来协调训练样本之间的不平衡,不仅可以检测到Oriented-RCNN、DWLA和KRRD方法遗漏的港口(HA),避免Rotated-ATSS方法对港口(HA)的错误检测问题,而且使港口(HA)的目标边界框更接近其真实形状。在图4的第三列中,Improved RetinaNet采用了KL散度损失的自调制优化机制,不仅可以检测到Oriented-RCNN和DWLA方法遗漏的桥梁(BR),而且对桥梁(BR)具有最高的检测精度。

将Improved RetinaNet和当前流行的定向目标检测方法在DroneVehicle数据集上进行了定性分析,如图5所示。可以观察到,在图5的第一列

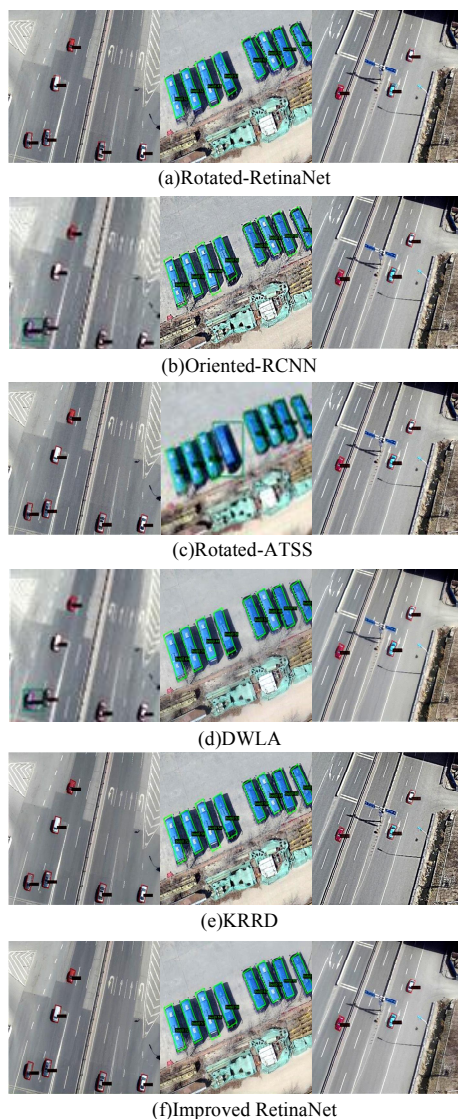


图5 DroneVehicle数据集上的可视化比较

Fig. 5 Visual comparison on DroneVehicle dataset

中,Improved RetinaNet设计了特征交互融合网络来获取目标丰富的上下文信息,不仅可以解决Oriented-RCNN和DWLA方法将汽车(Car)错误检测成货车(Fright)的问题,而且对汽车(Car)具有最高的检测精度。在图5的第二列中,Improved RetinaNet采用了KL散度损失的自调制优化机制,不仅可以检测到Rotated-ATSS方法遗漏的公交车(Bus),而且使公交车(Bus)的目标边界框更匹配其真实形状。在图5的第三列中,Improved RetinaNet对所有的汽车(Car)达到了最高的检测精度,并且生成的目标边界框最匹配汽车(Car)的真实形状。

2.3.3 效率分析

将Improved RetinaNet与当前流行的定向目

标检测方法分别在 DOTA-v1.0 数据集和 DroneVehicle 数据集上进行了检测速度和模型参数的效率分析,采用每秒帧率(Frame per second, FPS)来评估目标的检测速度。需要指出的是,所有的定向目标检测方法都在一个 NVIDIA RTX 3070Ti GPU 上进行测试。表 7、表 8 的实验结果表明,Improved RetinaNet 比大多数定向目标检测方法检测速度更快、模型参数更少,表现出了很高的检测效率。

表 7 DOTA-v1.0 数据集上的效率分析

Table 7 Efficiency analysis on DOTA-v1.0 dataset

方 法	FPS	模型参数/M
Oriented-RCNN ^[8]	12.35	101.53
RoI-Transformer ^[7]	12.66	104.08
GWD ^[15]	15.15	40.56
S ² a-Net ^[6]	15.38	38.62
SA3Det ^[9]	15.63	40.28
DWLA ^[12]	16.13	45.39
R ³ Det ^[5]	16.39	42.03
KRRD ^[16]	16.67	42.17
Rotated-ATSS ^[18]	16.95	40.14
Rotated-RetinaNet ^[10]	17.24	39.23
Improved RetinaNet	16.98	38.71

表 8 DroneVehicle 数据集上的效率分析

Table 8 Efficiency analysis on DroneVehicle dataset

方 法	FPS	模型参数/M
Oriented-RCNN ^[8]	14.53	101.53
RoI-Transformer ^[7]	14.94	104.08
GWD ^[15]	17.32	40.56
S ² a-Net ^[6]	17.57	38.62
SA3Det ^[9]	18.33	40.28
DWLA ^[12]	18.78	45.39
R ³ Det ^[5]	19.47	42.03
KRRD ^[16]	20.31	42.17
Rotated-ATSS ^[18]	20.59	40.14
Rotated-RetinaNet ^[10]	21.26	39.23
Improved RetinaNet	20.89	38.71

3 结束语

提出了一种称为 Improved RetinaNet 的检测器,用于航空图像中定向目标的精确检测。首先,在主干网络 ResNeXt-101 的基础上设计了特征交

互融合网络,在不增加网络复杂度的情况下获取定向目标丰富的上下文信息,同时减少模型参数的数量。其次,在分类子网络中引入了训练样本动态分配方法,根据目标的空间位置关系动态地分配训练样本,在不设置任何超参数的情况下协调训练样本之间的不平衡。最后,在回归子网络中采用了长边定义法($\theta \in [-45^\circ, 135^\circ]$)表示定向边界框,将预测定向边界框和地面真实边界框转换为二维高斯分布,计算二维高斯分布之间的 KL 散度作为回归损失,以解决大高宽比定向目标的角度敏感问题。在具有定向目标的航空图像数据集上进行了大量实验验证,结果表明,Improved RetinaNet 能够有效提升航空图像中定向目标的检测性能。在未来的工作中,将更加关注定向目标检测网络的高效性和轻量化,使其在检测精度、检测速度和模型参数之间取得良好的平衡。

参考文献:

- [1] Han Jia-ming, Ding Jian, Xue Nan, et al. ReDet: a rotation-equivariant detector for aerial object detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, Virtual, 2021: 2786-2795.
- [2] Li Wen-tong, Chen Yi-jie, Hu Kai-xuan, et al. Oriented reppoints for aerial object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, America, 2022: 1829-1838.
- [3] Lu Dong-chen. OSKDet: towards orientation-sensitive keypoint localization for rotated object detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, America, 2022: 1182-1192.
- [4] Doloriel C, Cajote R. Improving the detection of small oriented objects in aerial images[C]//IEEE Workshop on Applications of Computer Vision, Hawaii, America, 2023: 176-185.
- [5] Yang Xue, Yan Jun-chi, Feng Zi-ming, et al. R3Det: refined single-stage detector with feature refinement for rotating object[C]//AAAI Conference on Artificial Intelligence, Virtual, 2021: 3163-3171.
- [6] Han Jia-ming, Ding Jian, Li Jie, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: No. 5602511.
- [7] Ding Jian, Xue Nan, Long Yang, et al. Learning roi transformer for oriented object detection in aerial images[C]//IEEE Conference on Computer Vision and

- Pattern Recognition, Long Beach, America, 2019: 2849-2858.
- [8] Xie Xing-xing, Cheng Gong, Wang Jia-bao, et al. Oriented R-CNN for object detection[C]//IEEE International Conference on Computer Vision, Virtual, 2021: 3520-3529.
- [9] Wang Wen-yong, Cai Yuan-zheng, Luo Zhi-ming, et al. SA3Det: detecting rotated objects via pixel-level attention and adaptive labels assignment[J]. Remote Sensing, 2024, 16: No. 2496.
- [10] Lin, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [11] Hou Li-ping, Lu Ke, Xue Jian, et al. Shape-adaptive selection and measurement for oriented object detection[C]//AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2022: 923-932.
- [12] Zhu Yun-hui, Huang Bu-liao. Dynamic weighting label assignment for oriented object detection[J]. Memetic Computing, 2024, 16: 285-297.
- [13] Yang Xue, Yan Jun-chi. Arbitrary-oriented object detection with circular smooth label[C]//European Conference on Computer Vision, Virtual, 2020: 677-694.
- [14] Xu Yong-chao, Fu Ming-tao, Wang Qi-meng, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(4): 1452-1459.
- [15] Yang Xue, Yan Jun-chi, Ming Qi, et al. Rethinking rotated object detection with gaussian wasserstein distance loss[C]//International Conference on Machine Learning, Cheju-do, Korea, 2021: 11830-11841.
- [16] Zhu Hu-fei, Huang Yong-hui, Xu Ying, et al. Unmanned aerial vehicle(UAV) object detection algorithm based on keypoints representation and rotated distance-IoU loss[J]. Journal of Real-Time Image Processing, 2024, 21: No. 58.
- [17] Xie S, Girshick R, Dollar P, et al. Aggregated residual transformations for deep neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, America, 2017: 5431-5440.
- [18] Zhang Shi-feng, Chi Cheng, Yao Yong-qiang, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Seattle, America, 2020: 9756-9765.
- [19] Kullback S, Leibler R. On information and sufficiency[J]. The Annals of Mathematical Statistics, 1951, 22(1): 79-86.
- [20] Xia Gui-song, Bai Xiang, Ding Jian, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, America, 2018: 3974-3983.
- [21] Sun Yi-ming, Cao Bing, Zhu Peng-fei, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6700-6713.
- [22] Zhou Yue, Yang Xue, Zhang Ge-fan, et al. Mmrotate: a rotated object detection benchmark using pytorch[C]//ACM International Conference on Multimedia, Lisbon, Portugal, 2022: 7331-7334.
- [23] Lin Tsung-yi, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, America, 2017: 2117-2125.
- [24] Liu Shu, Qi Lu, Qin Hai-fang, et al. Path aggregation network for instance segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, America, 2018: 1534-1544.
- [25] Ghiasi G, Lin Tsung-yi, Pang Ruo-ming, et al. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, America, 2019: 7392-7401.
- [26] Zhang Jiang-fan, Zhang Yan, Shi Zhi-guang, et al. Unmanned aerial vehicle object detection based on information-preserving and fine-grained feature aggregation[J]. Remote Sensing, 2024, 16(14): No. 2590.