

基于对齐卷积和椭圆损失函数的遥感图像目标检测算法

郑凌云¹, 李佐勇^{2,3}, 蔡远征^{2,3,4}, 唐郑熠¹

(1.福建工程学院计算机科学与数学学院, 福建 福州 350118;

2.闽江学院计算机控制与工程学院, 福建 福州 350121;

3.福建省信息处理与智能控制重点实验室, 福建 福州 350121;

4.四创科技有限公司, 福建 福州 350100)

✉ 2211308055@smail.fjut.edu.cn; fzulzytdq@126.com; yuanzheng_cai@mju.edu.cn; tangzy84@126.com



摘要:针对遥感图像旋转框定位过程通常会出现分类分数和定位精度不匹配、交并比计算不精确的问题,文章提出了一个鲁棒性高的单阶段对齐检测网络(Single-stage Alignment Network, SAN),首先在网络中采用了对齐卷积,解决了分类分数和定位精度不匹配的问题;其次在网络训练过程中引入椭圆损失函数,将传统采用的定位交并比的计算方式转换为椭圆区域的参数表示计算。该方法在 DOTA 和 HRSC2016 数据集上进行的实验取得了 74.3% 和 89.0% 的平均精度,分别比基线方法高出 13.2% 和 15.5%,优于大部分的主流网络模型。

关键词:目标检测;遥感图像;卷积神经网络

中图分类号:TP311 **文献标志码:**A

Remote Sensing Image Target Detection Algorithm Based on Alignment Convolution and Elliptic Loss Function

ZHENG Lingyun¹, LI Zuoyong^{2,3}, CAI Yuanzheng^{2,3,4}, TANG Zhengyi¹

(1.College of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, China;

2.College of Computer and Control Engineering, Minjiang University, Fuzhou 350121, China;

3.Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Fuzhou 350121, China;

4.Istrong Technology Corporation, Fuzhou 350100, China)

✉ 2211308055@smail.fjut.edu.cn; fzulzytdq@126.com; yuanzheng_cai@mju.edu.cn; tangzy84@126.com

Abstract: Aiming at the common problems of mismatched classification scores and positioning accuracy, as well as inaccurate calculation of Intersection over Union (IoU) in the process of remote sensing image rotation frame positioning, this paper proposes a robust Single-stage Alignment Network (SAN). Firstly, alignment convolution is adopted in the network to solve the problem of mismatched classification scores and positioning accuracy. Secondly, in the process of network training, an elliptic loss function is introduced to convert the traditional calculation method of positioning IoU into parameter representation calculation in the elliptic region. This method achieves average accuracy of 74.3% and 89.0% respectively in experiments on the DOTA and HRSC2016 datasets, which is 13.2% and 15.5% higher than the baseline method, and outperforms most of the state-of-art network models.

Key words: object detection; remote sensing images; convolutional neural network

0 引言(Introduction)

近年来,卷积神经网络(Convolutional Neural Network)已

成功应用于计算机视觉领域的各种任务中^[1-7]。得益于神经网络的蓬勃发展,当前遥感图像识别网络主要围绕经典的检测器

构建,从2017年的旋转视网膜网络(Rotated Retinanet)开始,遥感目标检测的研究逐渐受到关注^[8-9]。遥感目标检测使用主干网络的相同特征进行学习,但是分类分数和定位精度一般不会匹配,对高分类分数的目标不利。目前的损失函数侧重于沿用水平框的面积计算得出交并比(Intersection over Union, IoU),导致在大长宽比的旋转框下,细微角度的偏差会导致IoU偏差巨大,容易让网络过拟合。为了克服上述问题,本文提出一种基于对齐卷积和椭圆损失函数的遥感图像目标检测算法,该网络引入对齐卷积,可以根据尺度、形状和方向生成对齐特征,达到提升精度的目的;使用椭圆框,减少学习过程中IoU损失函数振荡;使用均方误差损失进一步优化回归损失。

1 相关工作(Related work)

随着深度学习的发展,目前的检测器分为两种类别,即单级检测器和两级检测器。单级检测器只需提取一次特征即可实现目标检测,其速度比两级检测器快,但总体精度略低。目前的单级检测器例如YOLO(You Only Look Once)^[10],能直接完成从特征到分类和回归的预测,使用相同的完全连接层实现分类和回归。SSD(Single Shot MultiBox Detector)^[11]网络是一个完全卷积网络,即通过VGG16进行特征提取后,提取 38×38 、 19×19 、 10×10 、 5×5 、 3×3 和 1×1 六层不同尺度的特征,用于SSD分类和回归。两级检测器需要提取多种特征,学习速度慢,但是精度高。常见的两级检测器包括R-CNN^[12]、Fast R-CNN^[13]、Faster R-CNN^[14]、Mask R-CNN^[15]和Feature Pyramid Network(FPN)^[16]。FPN网络构建了一个特征金字塔,可以学习不同尺度的特征和处理大规模的对象变化。综上所述,上述检测模型主要依赖两阶段网络分别进行特征提取和识别。

目前研究人员在遥感图像识别领域取得了如下的研究成果。ReDet^[17]用于提取旋转等变特征,可以准确预测方向并大幅度减小模型尺寸。滑动顶点(Gliding Vertex)^[7]方法没有回归通常矩形的4个顶点,而是在每个矩形框上移动顶点形成旋转框。旋转不变CNN在现有CNN体系结构的基础上,引入和学习一个新的旋转不变层^[18]。LIU等^[19]构建了一个近似封闭的船舶旋转框空间,WANG等^[20]提出了一个中心地图OBB(Oriented Bounding Boxes),通过减少OBB中背景因素进而更好地表示OBB。通过在经典的水平边界框表示法中添加四个滑动偏移变量,可以获得图像的新表示法,检测器也可以学习新的特征。YANG等^[8]提出的RepPoints是一种新的对象检测表示,它由一组表示对象空间范围和语义重要局部区域的点组成。这种表示是通过弱定位监督和矩形锚点隐式识别反馈实现学习的。ZHANG等^[21]开发的CAD-net(Context-Aware Detection Network)由全局上下文网络组成,该网络利用注意力调制特征以及联系全局特征与局部特征检测图像。

2 方法(Method)

2.1 网络架构

本文提出的SAN(Single-stage Alignment Network)主要由4个模块构成,其网络架构如图1所示,主干网络使用ResNet50,将提取的特征传入特征金字塔网络进行多尺度计算。每个金字塔层都使用了9个不同的锚框检测对象,长宽比分别为 $(0.5, 1, 2)$,锚框的大小为 $(1, 2^{\frac{1}{3}}, 2^{\frac{2}{3}})$,比之前只使用3个长宽比的锚框的平均准确率更高^[16]。在每层的特征金字塔网络上增加了两个检测头,每个检测头都有分类和回归的部分。图1中短虚线框的网络为经过改进的视网膜网络模块,相比视网膜模块,增加了对齐检测网络模块,并且使用了椭圆损失函数,分类子网络预测每个锚框对应目标的类别^[22]。W为图像的宽,H为图像的高,分类子网络得出的结果经过对齐检测网络模块之后,计算椭圆旋转框损失。回归网络与分类网络平行,意图回归每个锚框对于真值框位置的偏移量,回归子网络的设计和分类子网络相同,不同的是它为每个空间位置输出36个三线性输出,也就是说每个锚框都有4个输出用于预测锚框的偏移量。图1中中虚线框的网络为旋转检测模块,将前一个网络输入的方向敏感特征和对齐卷积特征进行计算,可以对齐具有不同方向的特征减少参数量,提高计算效率^[23]。

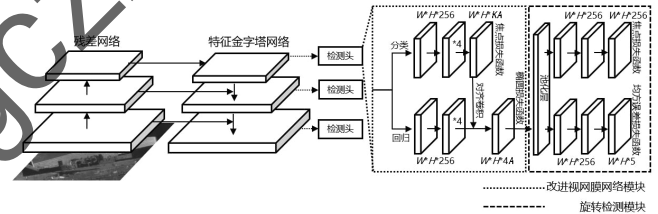


图1 单阶段对齐检测网络框架

Fig. 1 Single-stage alignment detection network framework

2.2 对齐卷积模块

标准的卷积可以用公式(1)表示:

$$Y(p) = \sum_{r \in R} W(r) \cdot X(p+r) \quad (1)$$

其中:W表示卷积核的参数,r为卷积像素相对于当前位置的偏移,p表示当前卷积的空间位置,R为常规的采样位置。

本文研究的网络中,在改进视网膜网络模块中增加了对齐卷积模块(Alignment Convolution Module),与标准卷积不同的是,对齐卷积增加了一个偏移量o:

$$Y(p) = \sum_{r \in R: o \in O} W(r) \cdot X(p+r+o) \quad (2)$$

其中:o为采样到的锚框坐标和真实值之间的偏差值,p中含有如下参数 (x, y, h, θ) 表示p位置的锚框,对于每个规则位置的r,基于锚框的采样位置为

$$L_p^r = \frac{1}{S} (x + \frac{1}{k} (w, h) \cdot r \cdot \mathbf{R}^T(\theta)) \quad (3)$$

其中:k表示卷积核的大小,S为步长,w、h为锚框的长和高,

$\mathbf{R}^T(\theta)$ 是 2×2 的旋转矩阵 $\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}^T$, 当前位置的偏移量 O 为

$$O = \{L_p^r - p - r\}, r \in R \quad (4)$$

图 2 为对齐卷积网络模块的图例, 通过在网络中加入对齐卷积将 $W * H * KA$ 锚框中预测的位置解码成旋转矩形 (x, w, h, θ) , K 为分类的类别, 在 DOTA1.0 数据集的 K 为 15, 本次实验采用 9 个不同的锚框, 所以 A 为 9。与输入特征一起计算对齐卷积, 提取对齐卷积的特征。对齐卷积网络模块由输入通道和输出通道都为 256 的 5 层可变形卷积组成, 经每个卷积层后使用 ReLU 作为激活函数。

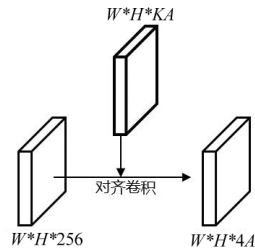


图 2 对齐卷积网络模块

Fig. 2 Aligned convolutional network module

2.3 椭圆损失函数(Elliptic loss function)

计算椭圆 IoU 的流程如图 3 所示, 首先将边界框转换为高斯分布, 将边界框的中心点重合, 其次使用卡尔曼滤波获得重叠区域的分布函数, 最后将得到的分布函数值和旋转框计算 IoU。

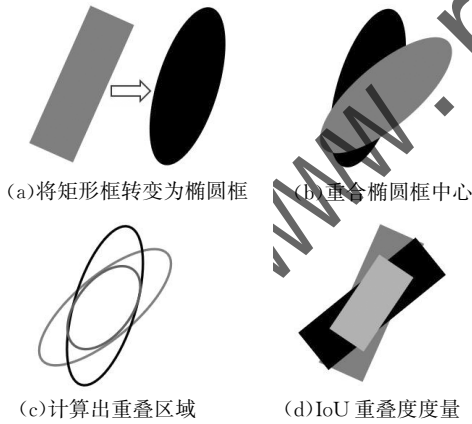


图 3 计算椭圆 IoU 的流程^[23]

Fig. 3 Process of calculating elliptic IoU

2D 旋转框可转换为高斯分布 $G(\mu, \Sigma)$ 形式表示:

$$\Sigma = \mathbf{R} \mathbf{A} \mathbf{R}^T \quad (5)$$

$$\mu = (x, y)^T \quad (6)$$

其中, $\mathbf{R}$ 表示旋转矩阵, $\mathbf{A}$ 表示特征值的对角矩阵。

对于 2D 目标 $B_2(x, y, h, w, \theta)$:

$$\mathbf{R} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (7)$$

$$\mathbf{A} = \begin{pmatrix} \frac{\omega^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{pmatrix} \quad (8)$$

此处, 可基于方差公式计算出高斯分布旋转框的面积 V_B , 由于遥感图像识别检测的是二维平面, 因此此处带入 $n=2$, 并使用协方差累乘, 可得面积 V_B 公式如下:

$$V_B(\Sigma) = 2^n \sqrt{\prod \text{eig}(\Sigma)} = 2^n \cdot |\Sigma|^{\frac{1}{2}} = 2^n \cdot |\Sigma|^{\frac{1}{2}} \quad (9)$$

进一步引入高斯分布计算预测框与真实值之间的重叠部分的面积 G_s , 该过程可以通过由两个高斯分布 G_1 和 G_2 的乘积得到:

$$\alpha G_s(\mu, \Sigma) = G_1(\mu_1, \Sigma_1) G_2(\mu_2, \Sigma_2) \quad (10)$$

其中: α 可以表示为

$$\begin{aligned} \alpha &= G_s(\mu_2, \Sigma_1 + \Sigma_2) \\ &= \frac{1}{\sqrt{\det(2\pi(\Sigma_1 + \Sigma_2))}} e^{-\frac{1}{2}(\mu_1 - \mu_2)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)} \end{aligned} \quad (11)$$

μ, Σ 可表示为

$$\begin{aligned} \mu &= \mu_1 + K(\mu_2 - \mu_1) \\ \Sigma &= \Sigma_1 - K \Sigma_1 \end{aligned} \quad (12)$$

公式(12)中, $K = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$ 。需要注意的是, 公式(12)计算后得到的高斯分布的协方差 Σ , 与协方差 Σ_1 与 Σ_2 有关, 该公式没有体现出位置对于协方差的影响, 其原因在于该计算过程中采用的高斯分布 $\alpha G_s(\mu, \Sigma)$ 并不是标准高斯分布, 不能在不考虑中心点系数 α 的情况下, 直接计算出重叠面积。在计算面积时, 要同时把 α 考虑进去。所以, 引入中心点损失使得两个分布同中心, 这样就不需要考虑中心点系数。椭圆损失函数如下:

$$Loss_E = \frac{V_{B3}(\Sigma)}{V_{B1}(\Sigma_1) + V_{B2}(\Sigma_2) - V_{B3}(\Sigma)} \quad (13)$$

经计算得到 IoU 的取值范围为 $[0, \frac{1}{3}]$, 虽然范围不及正常 IoU 的 $[0, 1]$, 但是同样好用, 性能也几乎是最好的。

接下来, 我们就可以将轴对齐的框的特征转换成任意方向上的特征。

2.4 损失函数

在单阶段图像识别中, 正样本仅占很少的一部分, 更多的负样本对训练没有帮助。本次实验不使用传统的交叉熵损失函数, 改为使用能解决正负样本不平衡问题的 Focal Loss 损失函数。

Focal Loss 损失函数定义如下:

$$Loss_F = -(1 - p_t)^\gamma \log(p_t) \quad (14)$$

其中: γ 是范围在 $[0, 5]$ 的参数, p_t 定义为

$$p_t = \begin{cases} p, & y=1 \\ 1-p, & y=-1 \end{cases} \quad (15)$$

其中: p 为模型属于前景的概率, y 的取值为 1 和 -1,分别代表前景和背景。 $(1-p_t)^{\gamma}$ 作为一个调制因子,当背景容易分辨时, $(1-p_t)^{\gamma}$ 较低,说明对损失的贡献较小,即降低了易区分样本的损失比例。

由于 SmoothL1 损失函数对长宽比完全不敏感,所以本次实验在回归网络中使用均方差损失,公式如下^[24]:

$$Loss_M = \frac{(y - y_0)^2}{2} \quad (16)$$

其中: y 为真实值, y_0 为预测值。

均方差损失的优点在于函数曲线连续、光滑、处处可导,适用于梯度下降算法,并且误差减小时,损失也会减小,有利于收敛。即使使用固定的学习速率,也能很好地收敛到最小值。

在本文的 SAN 网络中,不同的网络模块损失函数的计算方式不同,视网膜模块损失函数的计算方法为

$$Loss_{total} = Loss_F + Loss_E \quad (17)$$

旋转检测模块损失函数的计算方法为

$$Loss'_{total} = Loss_F + Loss_M \quad (18)$$

3 实验(Experiment)

3.1 数据集

DOTA^[25] 是一个用于定向目标检测的大型航空图像数据集,目前一共有 DOTA1.0、DOTA1.5、DOTA2.0 三个版本。DOTA1.0 包含 2 806 张 $800 \times 800 \sim 4\,000 \times 4\,000$ 的图像和 188 282 个实例,包含 15 个常见目标类别:飞机(PL)、棒球场(BD)、桥梁(BR)、田径场(GTF)、小型车辆(SV)、大型车辆(LV)、船舶(SH)、网球场(TC)、篮球场(BC)、储罐(ST)、足球场(SBF)、环岛(RA)、港口(HA)、游泳池(SP)和直升机(HC),其中,数据集的 1.5 版本使用了和 1.0 版本相同的图片,但是额外标注了非常小的目标(小于 10 个像素),并且多了一个类别(集装箱起重机,CC),而 DOTA2.0 包含了 18 个类别,并且有 11 268 张图像和 1 793 658 个实例。与 DOTA v1.5 相比,DOTA2.0 进一步增加了新的图像类别。

DOTA 数据集提供了训练集、验证集和测试集,其中测试集的数据不公开。将训练集和验证集的数据都用于训练。在 DOTA 之后,从原始图像中裁剪了一系列 $1\,024 \times 1\,024$ 的补丁,步幅为 824。为了避免过拟合,没有使用其他技巧。

HRSC2016 是一个带有旋转框的船舶检测数据集,包含 1 061 张船舶的遥感图像,大小范围为 $300 \times 300 \sim 1\,500 \times 900$ 。训练集、验证集和测试集分别包含 436、181 和 444 张图像。使用训练集、验证集进行训练,测试集用于测试。根据原网络的设置不同,把所有图像的大小改为 800×512 或者 512×800 ,但不改变图像的长宽比。训练过程采用随机翻转方式,随机翻转的方向包含水平和竖直方向。

3.2 实验设置

本次实验在两张 NVIDIA RTX A4000 的环境下运行,实验

都使用 ResNet50 初始化。所有网络使用同样的参数设置,学习率为 0.002 5,动量和动量衰减分别为 0.9 和 0.000 1,在 DOTA 数据集下, $epoch$ 为 12,在 HRSC2016 数据集中, $epoch$ 为 36。使用 SGD 梯度衰减。

在 DOTA 数据集中,本次实验采用平均精度均值(mAP)作为评价指标,此评价指标被 PASCAL VOC、COCO、ImageNET、Google Open Image Challenge 等图像检测比赛所使用。在 mAP 的计算中,涉及以下概念。

TP : 实际为正样本,同时被分类器划分为正样本的样本数。

FP : 实际为正样本,但是被分类器划分为负样本的样本数。

FN : 实际为负样本,但是被分类器划分为正样本的样本数。

TN : 实际为负样本,同时被分类器划分为负样本的样本数。

召回率 $Recall$ 的计算方式为

$$Recall = \frac{TP}{TP + FN} \quad (19)$$

准确率 $Precision$ 的计算方式为

$$Precision = \frac{TP}{TP + FP} \quad (20)$$

P-R 曲线是所有 Precision-Recall 点连成的曲线,用于计算 AP 值。通常, P 在纵轴, R 在横轴, P 值越高, R 值越低。 AP 值就是 P-R 曲线围成的面积,由于检测精度需要 IoU 作为度量的边界,使用不同 IoU 度量的 P-R 曲线的 AP 值也不同, mAP 是使用不同 IoU 得出的 AP 值的均值。 $AP50$ 是 IoU 阈值取 0.5 时得出的 AP 值,研究人员在 HRSC2016 数据集中使用此评判方法。

将典型的 Rotate Retina Net(RR)作为基线,视网膜网络作为单级检测器检测速度快,但也会因为添加了模块而增加了计算量。视网膜网络的焦点损失表现较好,在当时精度超过了二级检测器,计算速度超过了单级检测器,是一个很好的挑战目标。

3.3 实验结果

表 1 是本文方法与当前的主流网络的精度对比结果,表 1 中有些网络的名称使用缩写,黑体数字下划线表示类别中最高的精度,黑体数字表示次高的精度,可以看到,本文方法的 mAP 比基线方法提高了 13.2%,比当前主流网络有了较大的提升,并且对难以识别的一些目标(直升机、田径场、大型汽车)也表现出最优的检测效果。在大长宽比的物体识别中,对比其他的网络的 mAP 也有较大的提升,说明网络改进的方向是正确的。

表1 本文方法与各个主流网络的精度比较

Tab.1 Accuracy comparison of the proposed method with each mainstream network	
AP/%	
方法	飞 棒球 桥 田径 小型 大型 船 网球 篮球 储 足球 环 港 游泳 直升 mAP 机 场 梁 场 汽车 汽车 船 场 场 罐 场 岛 口 池 机 %
RR ^[8]	82.4 71.2 36.9 51.8 69.2 48.3 68.9 89.4 70.5 84.0 53.0 62.7 50.6 57.3 20.9 61.1
S ² Anet ^[22]	89.5 83.3 46.2 71.6 78.4 74.5 86.3 90.8 85.5 84.3 58.9 65.4 64.7 69.0 54.1 73.5
KLD ^[26]	88.5 78.3 45.2 62.2 71.5 64.1 76.6 90.8 80.5 78.6 49.1 64.3 62.1 62.2 33.4 67.2
CFA ^[27]	86.8 73.2 47.9 65.3 78.9 73.6 85.7 90.8 78.1 84.6 42.6 60.3 62.2 65.2 17.3 67.5
ConvNet ^[28]	89.4 83.1 42.7 74.3 79.4 71.9 85.4 90.9 87.9 84.8 66.8 64.7 61.4 75.6 54.1 74.2
R3Det ^[24]	88.9 77.8 45.4 65.3 74.7 69.6 78.3 90.8 80.3 83.6 53.4 63.2 60.7 61.1 36.1 68.6
GV ^[29]	89.0 76.0 50.1 66.2 73.7 73.5 85.9 90.8 79.9 84.1 52.1 65.2 65.3 69.3 49.5 71.4
SASM ^[30]	89.4 78.9 48.5 70.2 78.2 73.7 85.1 90.8 83.1 85.0 58.9 66.8 64.5 68.7 50.5 72.8
GWD ^[31]	88.1 78.5 44.2 64.5 72.3 58.0 76.5 90.8 81.9 80.1 55.8 63.8 60.2 64.8 39.8 68.0
本文方法	89.4 81.2 49.7 71.7 78.4 75.0 86.1 90.8 86.6 85.2 62.4 66.5 65.8 69.0 56.7 74.3

对网络进行消融实验,分别对网络是否添加了椭圆损失函数、对齐卷积和均方误差损失做了对比,如表 2 所示,对齐卷积网络模块能让网络学习到角度信息、椭圆损失函数能够更加精确地计算交并比、均方误差损失函数能优化梯度计算。从表 2 中可以看到,当使用了更多的对齐卷积或者卷积函数时,精度会比没使用或者使用更少的方法高,证明每个模块的引用是有意义的。

表2 使用不同算法的结果

Tab.2 Results using different algorithms			
对齐卷积	Loss _E	Loss _M	mAP/%
×	×	×	73.4
×	×	✓	73.9
×	✓	×	73.6
✓	×	×	73.8
×	✓	✓	73.9
✓	✓	×	73.7
✓	×	✓	74.0

替换旋转检测网络模块中的回归损失函数,比较结果如表 3 所示,从表 3 中可以看到,在相同的网络下,均方误差损失在本文方法里发挥的最好,均方误差损失便于梯度下降和利于收敛的性质更契合本文网络,并且模块的增多也会相应地提升精确率。

表3 本文方法与其他损失函数的比较

Tab.3 Comparison of the proposed method with other loss functions			
损失函数	mAP/%	损失函数	mAP/%
BalancedL1Loss	71.9	KLDRepPointsLoss	73.8
GHMRLoss	70.3	L1Loss	74.0
BoundedIoULoss	73.8	MSELoss(本文方法)	74.4

表 4 是本文方法与其他主流方法在 HRSC2016 数据集上的结果。我们用 AP50 作为评判标准。在 HRSC2016 数据集上,本文所提网络的精确率比基线提高了 15.5%,并且优于当前的主流方法。HRSC2016 数据集包含的船舶目标都是任意方向和长宽比的,表明本文所提网络在遥感图像识别领域的表现出色。

表4 各方法在 HRSC2016 数据集上的结果

Tab.4 Results of each method on the HRSC2016 dataset			
方法	mAP/%	方法	mAP/%
Rotated Retina Net	73.5	ATSS ^[33]	86.0
Orient R-CNN ^[32]	79.3	R3Det	82.9
S ² Anet	86.9	GLD	83.7
Rotated Fast R-CNN	75.5	SAN(本文方法)	89.0
KLD	84.9		

图 4 是 DOTA 实例的识别结果。对于相同的图像,图 4(b)和图 4(d)是本文网络的实验结果,图 4(a)和图 4(c)是基线的实验结果。基线网络对于角度的回归不是很稳定,会出现检测框大小正确,但方向有误的情况。图 4(a)和图 4(b)展示了常见的遥感图像,从中可以看出遥感图像的目标更加密集,本文方法对于每个目标的置信度相对于基线方法更高,并且旋转框的定位更加准确,在图像尺寸明显增大的情况下,基线方法有较大的缺点:旋转框的定位不精确,会出现明显的角度偏差,并且由于图像尺寸过大,导致某些目标难以检测。本网络显著提升了定位精度和置信度,也能检测出基线网络检测不出来的实例。

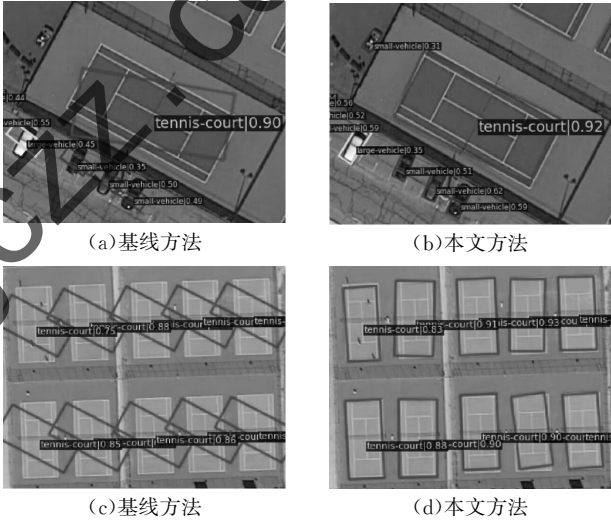


图 4 部分图片的识别结果

Fig. 4 Recognition results of some images

4 结论(Conclusion)

本文研究了遥感目标检测技术的发展现状以及目前存在的问题,提出了一种使用椭圆旋转框和对齐卷积的单阶段检测网络。使用高斯分布将矩形锚框转变为椭圆锚框,有效弥补了水平锚框因为角度偏差出现的 IoU 与实际精度不匹配的缺陷。同时,引入对齐卷积网络模块与协变性特征,解决定位精度和分类分数不高的问题。本文方法和其他主流网络比较,具有明显的优势,在现阶段流行的数据集中取得了优于主流算法的精度。在将来的工作中,要进一步考虑当前算法对于小目标实例召回率不高的问题,通过添加对小目标友好的网络模型子网络分支与显著性模块增强等方式,逐步提升小目标的召回率。

参考文献(References)

[1] KANG X D, DUAN P H, XIANG X L, et al. Detection and correction of mislabeled training samples for hyperspectral image classification[J]. IEEE Transactions on Geoscience

- and Remote Sensing, 2018, 56(10): 5673-5686.
- [2] LI Q P, MOU L C, XU Q Z, et al. R³-Net: a deep network for multi-oriented vehicle detection in aerial images and videos[DB/OL]. (2018-08-16) [2023-04-12]. <https://arxiv.org/abs/1808.05560>.
 - [3] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: a survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.
 - [4] PENG J T, LI L Q, TANG Y Y. Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(6): 1790-1802.
 - [5] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
 - [6] CHENG G, YANG C Y, YAO X W, et al. When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(5): 2811-2821.
 - [7] XU Y C, FU M T, WANG Q M, 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.
 - [8] YANG Z, LIU S H, HU H, et al. RepPoints: point set representation for object detection[C]//IEEE. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 9656-9665.
 - [9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//IEEE. Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2999-3007.
 - [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement[DB/OL]. (2018-04-08) [2023-04-12]. <https://arxiv.org/abs/1804.02767>.
 - [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//LEIBE B, MATAS J, SEBE N, et al. Computer Vision-ECCV 2016. Cham: Springer, 2016: 21-37.
 - [12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 580-587.
 - [13] GIRSHICK R. Fast R-CNN[C]//IEEE. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2015: 1440-1448.
 - [14] 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.
 - [15] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//IEEE. Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2980-2988.
 - [16] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 936-944.
 - [17] HAN J M, DING J, XUE N, et al. ReDet: a rotation-equivariant detector for aerial object detection[C]//IEEE. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 2785-2794.
 - [18] CHENG G, ZHOU P C, HAN J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.
 - [19] LIU Z K, WANG H Z, WENG L B, et al. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(8): 1074-1078.
 - [20] WANG J W, YANG W, LI H C, et al. Learning center probability map for detecting objects in aerial images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(5): 4307-4323.
 - [21] ZHANG G J, LU S J, ZHANG W. CAD-net: a context-aware detection network for objects in remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12): 10015-10024.
 - [22] HAN J M, DING J, LI J, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-11.
 - [23] YANG X, ZHOU Y, ZHANG G F, et al. The KFIoU loss for rotated object detection [DB/OL]. (2022-01-29) [2023-04-12]. <https://arxiv.org/abs/2201.12558>.
 - [24] YANG X, YAN J C, FENG Z M, et al. R3Det: refined single-stage detector with feature refinement for rotating object[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 3163-3171.
 - [25] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//IEEE. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3974-3983.

- [26] YANG X, YANG X J, YANG J R, et al. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence [DB/OL]. (2021-04-18) [2023-03-05]. <https://arxiv.org/abs/2106.01883v3>.
- [27] GUO Z H, LIU C, ZHANG X S, et al. Beyond bounding-box: convex-hull feature adaptation for oriented and densely packed object detection[C]//IEEE. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 8788-8797.
- [28] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[C]//IEEE. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 11966-11976.
- [29] XU Y C, FU M T, WANG Q M, 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.
- [30] HOU L P, LU K, XUE J, et al. Shape-adaptive selection and measurement for oriented object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 923-932.
- [31] YANG X, YAN J C, MING Q, et al. Rethinking rotated object detection with Gaussian Wasserstein distance loss [DB/OL]. (2022-04-18) [2023-01-17]. <https://arxiv.org/abs/2101.11952v1>.
- [32] XIE X X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection[C]//IEEE. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 3500-3509.
- [33] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//IEEE. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 9756-9765.

作者简介:

郑凌云(1999-),男,硕士生。研究领域:目标检测。

李佐勇(1980-),男,博士,教授。研究领域:图像处理,模式识别,深度学习。

蔡远征(1987-),男,博士,讲师。研究领域:目标检测,图像分割,图像检索。本文通信作者。

唐郑熠(1984-),男,博士,副教授。研究领域:区块链技术。本文通信作者。

(上接第40页)



图8 漏报举例

Fig. 8 Misreporting examples

5 结论(Conclusion)

疲劳检测在日常学习、生活的很多方面都有着广泛的应用。例如,随着我国汽车保有量的不断攀升,疲劳驾驶引发的交通事故数量也逐年升高。在疲劳驾驶行为导致交通事故前,驾驶员往往会有明显的先前疲劳特征,因此对疲劳驾驶状态进行检测并做出相应的预警,对于保障交通运输安全至关重要。本文提出的基于Dlib的模型通过人脸检测器、关键点定位器、特征计算器和状态预测器四个模块,能够快速有效地识别出人物打哈欠、瞌睡等疲劳状态。通过实验结果可知,模型能够满足实时性、鲁棒性和准确率的要求。

参考文献(References)

- [1] 李小平,白超. 基于深度学习的司机疲劳驾驶检测方法研究[J]. 铁道学报, 2021, 43(6): 78-87.
- [2] 徐军莉,王平,穆振东. 融合眼动和脑电特征的疲劳驾驶检测研究[J]. 重庆交通大学学报(自然科学版), 2021, 40(12): 7-11.
- [3] 李响,李国正,彭理群,等. 基于语音特征迁移学习的驾驶疲劳检测[J]. 铁道学报, 2020, 42(4): 74-81.

[4] 陆荣秀,张笔豪,莫振龙. 基于脸部特征和头部姿态的疲劳检测方法[J]. 系统仿真学报, 2022, 34(10): 2279-2292.

[5] 姜平,杨欣,胡辑伟,等. 基于边缘计算的疲劳驾驶检测方法[J]. 计算机工程, 2021, 47(7): 13-20, 29.

[6] DINGES D F, GRACE R. PERCLOS: a valid psychophysiological measure of alertness as assessed by psychomotor vigilance[J]. Tech Brief, 1998, 31(18): 19-23.

[7] KNAPIK M, CYGANIEK B. Driver's fatigue recognition based on yawn detection in thermal images[J]. Neurocomputing, 2019, 338: 274-292.

[8] RAMZAN M, KHAN H U, AWAN S M, et al. A survey on state-of-the-art drowsiness detection techniques [J]. IEEE Access, 2019, 7: 61904-61919.

[9] 陈得恩,张建伟,柯文俊. 稳定的视频内头部姿态估计方法[J]. 计算机工程与设计, 2020, 41(12): 3438-3443.

[10] 樊星,刘占文,林杉,等. 融合人眼特征与深度学习的疲劳驾驶检测模型[J]. 计算机工程, 2021, 47(8): 243-250.

作者简介:

宋佳尧(2002-),女,本科生。研究领域:计算机应用,图像处理。

尉斌(1978-),男,硕士,讲师。研究领域:目标跟踪。本文通信作者。

安姝洁(2003-),女,本科生。研究领域:人工智能,图像处理。

张淞嘉(2002-),女,本科生。研究领域:计算机应用,图像处理。

杨玮(2002-),女,本科生。研究领域:电子商务,图像处理。