

文章编号:1001-5078(2024)09-01469-08

· 图像与信号处理 ·

# 基于改进 CenterNet 的红外小目标检测研究

倪安庆<sup>1</sup>, 李 军<sup>1</sup>, 王耀弘<sup>2</sup>

(1. 重庆交通大学 机电与车辆工程学院, 重庆 400074; 2. 重庆市计量质量检测研究院, 重庆 401121)

**摘要:**随着机器学习技术的不断发展,对目标检测技术的研究也越来越火热。针对目标检测中精度低、实时性差的问题。本文采用了一种单阶段的目标检测算法 CenterNet 完成对目标的快速识别,在算法的主干网络 ResNet50 增加 CBAM 注意力机制,提升了网络对目标的识别精度;在网络的输出模块,采用一种新的 GSConv 卷积模块,在不损失精度的情况下提高了检测速度。改进后的算法在红外数据集上验证其检测的准确性,其检测准确率达到 82.91%。研究表明:改进的 CenterNet 算法,可准确高效的完成对红外小目标的识别。

**关键词:**红外目标检测;CenterNet;注意力机制;GSConv

**中图分类号:**TP181;TN219 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2024.09.019

## Research on infrared small target detection based on improved CenterNet

NI An-qing<sup>1</sup>, LI Jun<sup>1</sup>, WANG Yao-hong<sup>2</sup>

(1. School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074, China;  
2. Chongqing Academy of Metrology and Quality Inspection, Chongqing 401121, China)

**Abstract:** With the continuous development of machine learning technology, the research on object detection technology is becoming increasingly popular. To address the issues of low accuracy and poor real-time performance in target detection, a single stage object detection algorithm CenterNet is adopted to achieve rapid recognition of targets. A CBAM attention mechanism is added to resnet50, the backbone network of the algorithm, to improve the recognition accuracy of the network on the target. In the output module of the network, a new GSConv convolution module is used to improve the detection speed without loss of accuracy. The improved algorithm is validated on the infrared dataset and its detection accuracy reaches 82.91%. The results show that that the improved CenterNet algorithm can accurately and efficiently accomplish the recognition of small infrared targets.

**Keywords:** infrared target detection; CenterNet; attention mechanism; GSConv

### 1 引言

计算机视觉作为人工智能领域的一个方向吸引了很多国内外学者参与研究。在计算机视觉中目标检测是一个重要研究方向,目标检测的任务是完成对目标物体的识别与定位,在识别出物体类别的同

时找到物体在图像中的位置。基于深度学习的目标检测算法在自动驾驶车辆中的行人检测、交通标志识别、道路环境检测等已经得到广泛应用。在低光照度情况下,RGB 图像难以获取外界环境情况,而红外图像受光照条件影响较小,因此在低光照强度

**基金项目:**国家自然科学基金项目(No. 52172381);重庆市研究生联合培养基地项目(No. JDLHPYJD2018003)资助。

**作者简介:**倪安庆(1999-),男,硕士研究生,主要从事自动驾驶车辆环境感知与控制研究,图像识别技术研究。E-mail: naq1999@163.com

**通讯作者:**李 军(1964-),男,教授,主要从事节能与新能源汽车,智能网联汽车技术研究。E-mail: cqleejun@163.com

**收稿日期:**2023-11-29; **修订日期:**2024-01-15

下采用红外图像检测非常有必要<sup>[1]</sup>。

基于深度学习的目标检测算法根据有无锚框,可以分为基于锚框的(Anchor-based)和无锚框的(Anchor-free)。锚框就是一个预先设置目标可能存在的大概位置的先验框,通过这些先验框上能够更加准确的预测到目标的真实边界框,完成对目标的识别分类。基于锚框的算法可分为双阶段(two-stage)目标检测和单阶段(one-stage)目标检测。双阶段目标检测的代表算法有 R-CNN 系列<sup>[2-3]</sup>、SPP-Net<sup>[4]</sup>等,其完成目标检测任务的过程分为两个阶段:先在输入图像中提取感兴趣的候选区域,在筛选过后则通过分类网络对区域内的目标进行分类和定位回归。R-CNN 算法把目标检测问题当成分类问题来处理,其提取特征信息和分类的两个阶段分为四个步骤来进行:首先将图像分割成小区域,找到其中感兴趣的区域;然后通过 CNN 卷积神经网络提取区域中的特征信息;得到特征信息过后,R-CNN 采用了 SVM 分类器来对每一类的卷积神经网络的输出特征进行分类;最后使用回归器来调整输出的候选区域,完成对目标图像的识别分类,其训练过程繁琐、训练时耗时很长无法完成实时检测的任务。由于双阶段目标检测算法的运行速度慢,无法达到要求,因此急需一种简洁高效的算法来实现快速检测的目标。基于此,单阶段目标检测算法应运而生。单阶段目标检测的算法以 YOLO 系列<sup>[5-6]</sup>、SSD<sup>[7]</sup>、CenterNet 等算法为代表。单阶段算法在进行目标检测时,直接在网络中提取特征来预测物体分类和位置。YOLO 系列算法是单阶段目标检测的代表算法,YOLOv1 是单阶段目标检测的开山之作,其创新性在于算法将目标检测直接当成回归问题,只通过一个阶段使用一个网络就完成了目标的类别和位置输出。与 R-CNN 相比其速度更快,并且其基于整幅图像预测的输出泛化能力也更强。YOLOv3 算法在 YOLOv2 的基础上进行了改进,采用了新的残差网络 Darknet-53,通过 FPN 架构实现了多尺度预测,提升了算法整体的性能。此后的 YOLO 系列算法都是不断的在上一代算法的基础上进行优化,提升网络的性能。总的来说,相较于双阶段目标检测,单阶段目标检测耗时更短,检测速度更快,但检测精度通常低于双阶段目标检测算法。

尽管基于锚框的双阶段和单阶段目标检测算法

在提升网络性能方面已经有了很大的进步,但是基于锚框的算法网络结构仍然相对复杂,在实际应用中需要进行更多轻量化操作处理。因此检测速度更快的无锚框目标检测算法被提出,无锚框的目标检测算法不需要设置目标的先验框,简化了网络结构、缩短了检测时间,同时还能通过各种优化方法来提高算法的检测精度。在无锚框的算法中,CornerNet 首次采用了单阶段检测的思想,采用关键点来检测目标,通过检测目标左上、右下的对角点并对对角点进行分类来代替锚框进行识别和定位的作用,这种操作极大的简化了网络的复杂度。CornerNet 通过卷积网络生成目标左上、右下点的热力图,与有锚框的算法相比其只需要目标物体两个方位的信息就能生成预测框,在定位目标时更加容易;同时,其采用了一种新的池化层结构,能够更加准确的找到目标的对角点,保证了算法的检测精度,其检测结果在精度和速度方面都优于 Faster R-CNN 和 YOLOv2。在 CornerNet 的基础上,Zhou X 等在提出了 ExtremeNet<sup>[8]</sup>。ExtremeNet 的整体框架基本与 CornerNet 相同,其改进主要在于关键点检测和关键点分组两部分。在关键点检测上,ExtremeNet 除了对每个目标预测四个顶点,还与目标的中心点进行结合,通过组合预测方式来达到更准确的预测结果。在关键点分组上,ExtremeNet 采用中心分组的方法,通过中心点来判断预测框是否准确,避免四个顶点带来的偏移,同时通过中心点预测的时间也更快。因此 ExtremeNet 在检测精度和速度方面都优于 CornerNet。CenterNet 是在 CornerNet 的基础上提出另一种无锚框的目标检测算法。CenterNet 直接将目标检测问题转换成一个中心关键点的估计问题,其通过目标中心点热力图来对目标进行预测和分类,与 CornerNet 相比其只需要一个关键点检测就能实现对目标的识别定位。相较于有锚框的算法,基于一个关键点检测的方法能够减少错误识别到目标的情况。同时,基于一个关键的检测比 CornerNet 基于两个点和 ExtremeNet 基于四个点的检测速度有了非常大的提升,选取 ResNet 作为主干网络时,其检测速度在相同条件下能够提升 4 到 5 倍。CenterNet 作为一个无锚框的目标检测算法,相较于上述所提到的各种算法在精度和速度方面都具有优势。对于红外图像而言,其受光照环境影响较小,图像更加真实,通过中心点检测的方法能

够更加突出红外图像的优势。

综上,考虑到目标检测算法的实时性和准确性,选择采用无锚框的 CenterNet 作为基础网络,针对进行目标检测的需求进行网络模型改进,优化算法的检测性能,具体贡献如下:(1)在主干网络 ResNet50 中增加注意力模块,能够显著提升网络的计算参数量从而增加网络检测的精度。(2)在网络输出模块处引入卷积改进模块 GConv,能够实现推理速度的加快,减少计算开销,提高计算效率。

## 2 CenterNet 目标检测算法原理

CenterNet 算法由 Law 等<sup>[9]</sup>提出的 CornerNet 算法为基础,Zhou 等<sup>[10]</sup>提出了 CenterNet 目标检测算法,其直接通过目标中心点的特征信息进行目标分类和回归,取得了良好的效果。CenterNet 网络模型通过卷积神经网络提取输入图像的关键信息,然后将得到的特征信息输出为热力图,通过对特征图的预测可以得到目标的关键中心点和预测框的尺寸。其对输入图像预测原理如下:

设  $I, I \in R^{W \times H \times 3}$  是一个高度为  $H$  宽度为  $W$  的输入图像,网络的目标是能够生成一个中心关键点热力图,生成的中心点记为  $\hat{Y} \in [0, 1]_{\frac{H}{R} \times \frac{W}{R} \times C}$ 。其中  $R$  是输出步长,  $C$  是关键点类型的数量。关键点类型代表目标检测中的物体类别。当预测值  $\hat{Y}_{x,y,c} = 1$  时表示检测到该目标,当  $\hat{Y}_{x,y,c} = 0$  时表示为背景。

在训练网络时,对于类别中每个目标其输入图像的真实关键点可表示为  $p \in R^2$ , 计算一个低分辨率下的同等量  $\tilde{p} = \left\lfloor \frac{p}{R} \right\rfloor$ , 即将计算结果取整。然后

将真实关键点利用一个高斯核映射到热力图  $\hat{Y} \in [0, 1]_{\frac{H}{R} \times \frac{W}{R} \times C}$  上, 所用高斯核为  $Y_{xyc} = \exp\left(-\frac{(x - \tilde{p}_x)^2 + (y - \tilde{p}_y)^2}{2\sigma_p^2}\right)$ , 其中  $\sigma_p$  为目标尺寸自适应

应标准差,  $(\tilde{p}_x, \tilde{p}_y)$  为映射到低分辨率特征图上的坐标; 当出现两个同类别目标的高斯核重合时, 规定选取其中较大值的一个。采用 focal loss 的利用惩罚减少像素逻辑回归的损失函数如式(1):

$$L_k = \frac{-1}{N} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^\alpha \log \hat{Y}_{xyc}, Y_{xyc} = 1 \\ (1 - Y_{xyc})^\beta (\hat{Y}_{xyc})^\alpha, Y_{xyc} \neq 1 \\ \log(1 - \hat{Y}_{xyc}), Y_{xyc} \neq 1 \end{cases} \quad (1)$$

其中,  $\alpha$  和  $\beta$  是 focal loss 的超参数;  $Y$  为真实关键点;  $N$  是图像  $I$  中关键点的数量, 通过归一化将所有正的 focal loss 实例归一化为 1。

为了弥补特征提取中下采样输出步幅引起的离散化误差, 对每个中心点的局部偏移量进行了预测。预测时, 所有类别使用同一个偏移预测, 且预测的中心点偏置为  $\hat{O} \in P_{\frac{W}{R} \times \frac{H}{R} \times 2}$ 。利用  $L_1$  损失函数对预测的中心点偏置进行训练, 如式:

$$L_{\text{off}} = \frac{1}{N} \sum_p \left| \hat{O}_p - \frac{p}{R} - \tilde{p} \right| \quad (2)$$

式中,  $\tilde{p}$  为下采样后的关键点;  $p$  为实际标注的关键点, 偏移预测只在关键点处起作用, 所有其他位置均被忽略。

假设  $(x_1^{(k)}, y_1^{(k)}, x_2^{(k)}, y_2^{(k)})$  是类别  $c_k$  的第  $k$  个目标的边界框, 其中心点坐标可表示为:  $p_k = \left(\frac{x_1^{(k)} + x_2^{(k)}}{2}, \frac{y_1^{(k)} + y_2^{(k)}}{2}\right)$ , 利用关键点估计器  $\hat{Y}$  来预测所有的中心点。对每个目标, 回归出其尺寸  $s_k = (x_2^{(k)} - x_1^{(k)}, y_2^{(k)} - y_1^{(k)})$ 。为了减轻计算负担, 对所有对象类别仅使用单一尺寸预测。采用  $L_1$  损失评价尺寸预测, 如下式:

$$L_{\text{size}} = \frac{1}{N} \sum_{k=1}^N |\hat{S}_{pk} - s_k| \quad (3)$$

式中,  $\hat{S}_{pk}$  为每个目标  $k$  的尺寸预测值。  $s_k$  为每个目标  $k$  的回归尺寸。引入两个损失调整系数  $\mu_{\text{size}}, \mu_{\text{off}}$  来缩放损失, 则总的损失函数为:

$$L_{\text{det}} = L_k + \mu_{\text{off}} L_{\text{off}} + \mu_{\text{size}} L_{\text{size}}$$

CenterNet 使用单一网络同时预测目标的关键点  $\hat{Y}$ 、中心点偏置  $\hat{O}$  和尺寸  $\hat{S}$  信息。其中关键点能够得到目标的中心坐标、尺寸信息包含了宽高。因此, 整个网络在每个位置上需要生成  $C + 4$  个预测输出。

在生成推理阶段, 首先要独立地提取每个类别的热力图峰值, 峰值的寻找通过  $3 \times 3$  的矩阵来实现, 每个峰值都要满足大于等于其 8 个相邻区域的值, 然后保留找到的最大的 100 个峰值。设  $\hat{\phi}_c$  为类别  $C$  的检测到的  $n$  个中心点的集合  $\hat{\phi} = \{(\hat{x}_i, \hat{y}_i)\}_{i=1}^n$ , 每个关键点位置由整数坐标  $(x_i, y_i)$  表示, 通过关键点的值  $\hat{Y}_{x_i y_i c}$  作为检测的置信度的度量, 得到输出边界预测框坐标如下式:

$$(\hat{x}_i + \delta \hat{x}_i - \hat{\omega}_i/2, \hat{y}_i + \delta \hat{y}_i - \hat{h}_i/2, \hat{x}_i + \delta \hat{x}_i - \hat{\omega}_i/2, \hat{y}_i + \delta \hat{y}_i - \hat{h}_i/2) \quad (4)$$

式中,  $(\delta \hat{x}_i, \delta \hat{y}_i) = \hat{O}_{x_i, y_i}$  是预测的位置偏移量;  $(\hat{\omega}_i, \hat{h}_i) = \hat{S}_{x_i, y_i}$  是预测的尺寸。

CenterNet 算法的所有输出都是基于对中心点估计产生,不再需要基于 IoU 的非最大值抑制 (NMS) 或其他后处理,通过网络提取的特征能够直接回归预测出目标框的宽和高,因此可以利用  $3 \times 3$  最大池化操作在设备上高效运算<sup>[11]</sup>。CenterNet 算法的网络结构主要分为输入图像前处理模块、主干网络以及检测网络分支。在输入图像后,前处理模块中会将输入图像重新调整为尺寸为  $512 \times 512$  的图像。主干网络主要作用是提取目标关键点的特征信息,原始的 CenterNet 网络模型中提供了 DLA-34、Hourglass-104、ResNet 等结构。其中 ResNet50 作为一种反卷积结构,其网络包含了 50 层卷积神经网络,网络层数在 DLA-34 与 Hourglass-104 之间,能在两种网络中间达到较好的权衡。使用 ResNet50 作为 CenterNet 的主干网络,最能实现检测网络在检测精度与时间上的平衡。在检测网络部分包含有热力图 (Heat map) 分支、中心点偏移预测分支 (Center offset map) 和宽高预测 (Height width map) 分支这三个分支。CenterNet 目标检测网络结构图如图 1 所示。

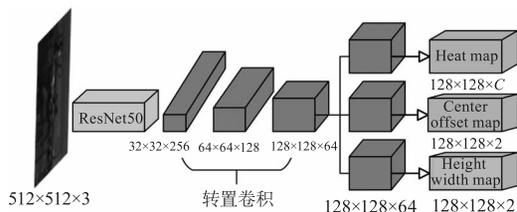


图 1 CenterNet 目标检测网络示意图

Fig. 1 Sketch map schematic diagram of CenterNet object detection network

其中,热力图的输出通道个数对应检测目标的类别个数,中心点偏置用来输出预测  $x, y$  方向上两个维度的中心点位的偏移,宽高图输出一个维度为 2 的特征图用来预测目标框的宽和高。

### 3 CenterNet 目标检测算法改进

#### 3.1 注意力机制改进

CenterNet 目标检测网络基于对目标中心点的估计来输出信息,是否能够准确快速的定位到目标中心点是关键问题。除了通过神经网络对大量的图

像信息进行学习来提高预测精度,如何提升网络本身定位目标中心点的能力也十分重要。要想更准确的定位到目标中心点,就需要更加关注对关键点检测,减少非关键信息对预测结果的影响。因此,可以采用注意力机制来增强网络对关键点的检测。注意力机制模块的设计原理是加强检测器对图片中更加重要的信息的注意,抑制与检测内容不相关的背景信息,提升检测的效果,是一类简单高效的特征提取模块<sup>[12]</sup>。在注意力机制提出之后,各种基于注意力机制的改进方法在计算机视觉领域不断发展,给 CNN 网络带来了显著的性能提升,其中被广泛使用的有,SENet、ECANet、GCNet、CBAM 等。基于注意力机制的方法主要可分为三种:通道注意力、空间注意力以及自注意力。通道注意力主要关注输入图像的通道信息,即不同特征图之间的关系;通过网络学习来获得不同通道特征的受关注程度,利用不同的权重系数来突出重要的特征通道,从而更好的输出图像的特征、提升检测的准确率;但是,由于只关注了通道信息,无法捕捉输入图像的空间信息,通道注意力在通道数很少的情况下对网络的性能提升效果不好。与通道注意力不同,空间注意力机制的核心是突出输入图像的关键特征区域,它利用空间转换模块能够将输入图像的特征信息转换到另一个空间中;同时,也使用不同的权重去强化感兴趣的区域,提高了对特征区域的注意力,这种操作能够与通道注意力达到互补。自注意力受注意力机制的启发,利用图像的内部特征信息来不断交互学习,减少了对外部信息的依赖;因此,自注意力受外部信息的影响较小,特征扰动更低;同时,也由于自注意力只关注图像的内部信息,所以自注意力需要在大量数据上进行学习,在数据量较小的情况下其表现不好。因此,如何充分利用各种注意力方法的特点非常重要。通道注意力或者空间注意力在单独使用时都只获得了输入图像的部分特征信息,都有自己的短板。可以通过将通道注意力和空间注意力的优势结合、互补两种方法的不足来提升使用效果。

基于上述对注意力机制的讨论,通过对几种注意力方法的性能对比,本文采用了一种新颖的轻量且高效的卷积块注意力模块 (CBAM)。CBAM<sup>[13]</sup> 是一个简单而有效的卷积神经网络注意力模块。其核心主要在于两个方面,一是将基于注意力的特征细

化,在寻找特征是能够更加准确,开销也更小;二是将通道注意力和空间注意力两个独立模块相结合,发挥了各自的优势,提高了特征检测的性能。CBAM 结合通道注意力和空间注意力的具体实现方法如下:

对于输入的特征图像,首先通过平均池化和最大池化操作来聚合特征图的空间信息,生成平均池化特征和最大池化特征:  $F_{avg}^c$  和  $F_{max}^c$ 。然后通过一个多层感知器 (MLP) 的共享网络得到通道注意力映射  $M_c \in R^{C \times 1 \times 1}$ 。再通过共享网络使用逐元素求和来合并输出特征向量。通道注意力计算如式:

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

然后,通过卷积操作生成空间注意力映射  $M_s(F) \in R^{H \times W}$ ,使用池化层聚合特征图的通道信息,得到通道中的平均池特征和最大池特征:  $F_{avg}^s \in R^{1 \times H \times W}$  和  $F_{max}^s \in R^{1 \times H \times W}$ 。之后,连接两个特征利用一个卷积操作生成空间注意力图。空间注意力计算如式:

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \\ = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (6)$$

对于一个给定的特征图,该模块能够沿着通道和空间顺序生成注意力图,结合两者的优势。最后在特征细化的操作上,模块通过通道和空间注意力得到的注意力图与输入的特征图相乘完成了特征细化。CBAM 模块结构如图 2 所示。

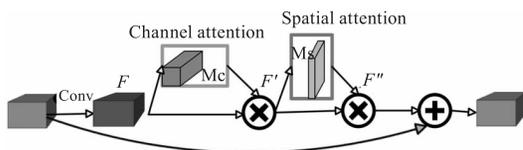


图2 CBAM 模块结构

Fig.2 Module structure of CBAM

在网络中增加注意力机制时需要考虑其带来的计算开销,而 CBAM 本身就是一个轻量级通用模块,在使用时可以很容易的集成到 CenterNet 网络中。选取 ResNet50 作为 CenterNet 的主干网络,ResNet50 采用一种残差结构,包含了 50 个网络卷积层。在应用神经网络时,随着网络的加深,很可能会发生梯度消失和梯度爆炸的问题,通过激活函数可以解决梯度问题,但是随着网络的加深,还可能产生网络的退化问题,即随着网络深度的加深,网络性能反而在下降。残差结构在正常的神经网络中,增加

了一个分支结构,使得网络的输出不再是单纯卷积的输出,而是卷积的输出和前面输入的叠加。通过残差结构可以解决这种退化问题,同时也在一定程度上缓解了梯度消失和梯度爆炸问题,提升网络的性能。ResNet50 相较于 Hourglass-104,其网络结构更加轻量化,因此通过添加 CBAM 注意力机制模块能够很好改善网络的性能。为了保证添加的注意力机制不改变 ResNet50 的网络结构,因此在网络的第一层和卷积层的最后一层加入注意力机制。

### 3.2 卷积模块改进(GSConv)

在实际的工业应用场景中对于目标检测网络的实时检测能力要求很高,CenterNet 目标检测网络有 50 层卷积神经网络,在计算时参数量比较多,因此对网络的轻量化研究也十分重要。大多数构建轻量级网络模型的方法主要集中在采用大量深度可分离卷积,但是这种轻量级模型会损失一部分语义信息,降低预测结果的准确性。因此,需要一种新的方法在不损失模型精度的情况下能够减少计算的参数量。

GSConv 模块<sup>[14]</sup>是一种全新的卷积方法,其模块结构如图 3 所示。

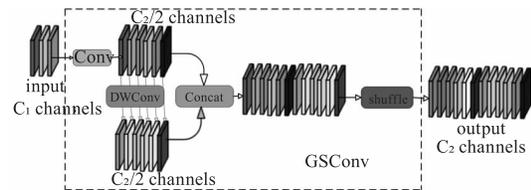


图3 GSConv 模块结构

Fig.3 Module structure of GSConv

图中,Conv 模块包含三种操作:卷积层、批量归一化层和激活层。SC 表示通道密集卷积、DSC 表示通道稀疏卷积,DWConv 模块表示 DSC<sup>[15]</sup>操作。

首先对输入进行一个普通卷积的下采样,然后使用 DWConv 深度卷积,并将两个卷积的结果拼接起来;最后进行 shuffle 操作<sup>[16]</sup>,即将通道数改变一下,让之前两个卷积的对应通道数挨在一起。这些操作让 DSC 的输出更加接近 SC,减少了语义信息的丢失,同时简化了卷积操作使得神经网络的计算量更小。

为了最终加快预测的计算,神经网络中的反馈图像总是要在主干网络中经历类似的转换过程:空间信息一步一步地向通道传输。并且每次特征图的空间(宽度和高度)压缩和通道扩展都会导致语义信息的部分丢失。通道密集卷积计算最大限度地保

留了每个信道之间的隐藏连接,但通道稀疏卷积完全切断了这些连接。CenterNet 目标检测网络在输出检测时,采用转置卷积和标准化操作,压缩了特征图的空间信息。受到 GSConv 卷积模块的启发,在网络的解码过程中引入该卷积模块尽可能的保证解码过程中的语义信息。GSConv 以较低的时间复杂度尽可能多地保留这些连接。GSConv 的优势在轻量级检测器中更为明显,这得益于通过添加 DSC 层和混洗增强的非线性表达能力。但是,如果在模型的所有阶段都使用 GSConv,模型的网络层将更深,这些深层网络将加剧对数据流的阻力,并显著增加推理时间。因此,我们将原始 CenterNet 代码中的上采样卷积模块用 GSConv 模块替换,能够实现推理速度的加快。

## 4 实验设计与结果分析

### 4.1 数据集准备

为了验证改进 CenterNet 网络的有效性,结合目标检测领域公认的数据集,选取了 PASCAL VOC2007 + 12 数据集的组合方式。该数据集检测目标分为 20 类,在训练网络时,在 VOC2007 和 VOC2012 的训练集和验证集(共 16551 张图片)上训练,然后在 VOC2007 的测试集(4952 张图片)上测试。同时,选取了包含 997 张的红外小目标数据集。该数据集收集了大量真实场景下目标飞机的红外图像。训练集和测试集的数量比为 9 : 1,图 4 给出了部分红外图像数据图。

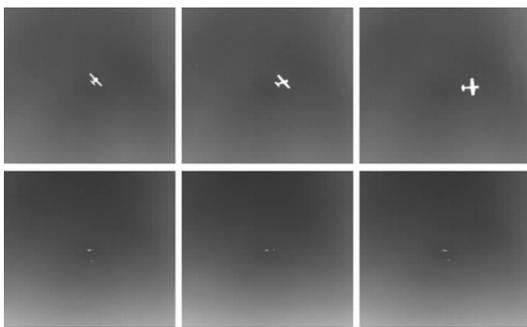


图 4 红外飞机小目标图

Fig. 4 Infrared images of aircraft small target

### 4.2 网络训练

本次实验平台的系统为 Windows 系统,硬件采用 Intel(R) Xeon(R) Platinum 8255C CPU@ 2.50 GHz、16 GB RAM、NVIDIA GeForce RTX2080Ti。使用的深度学习框架为 Pytorch 框架,CUDA 版本为 10.1,编程语言为 Python3.7。利用改进 CenterNet

算法在两类数据集上进行训练,网络共训练 100 轮,选择其中的一轮训练参数用作测试。

### 4.3 评价指标

目标检测算法的评价指标主要分为检测精度和检测速度两大类。评价检测精度时,召回率和精确率是最常用的两个指标<sup>[17]</sup>。精确率是指正确识别为正样本的目标在所有判定为正样本的目标中所占的比例,记为 P;精确率能够衡量一个分类器预测出的真正正样本的概率。召回率是指被正确识别为正样本的目标在所有正样本目标中所占的比例,记为 R;召回率衡量的是一个分类器能把所有的正样本都找出来的能力。在实际运用时需要综合考虑精确率和召回率,因此一般采用 F1 值来作为评价标准,F1 是精确率和召回率的调和平均值。一般来说,模型性能表现越好其 F1 值越高。在实际的应用过程中,还可以使用平均精度 AP 值来衡量模型的性能。通常来说一个越好的分类器,AP 值越高。如果有多类目标,对每类目标的 AP 值求平均,采用 mAP 值来衡量模型的性能。评估检测速度的指标一般采用目标网络每秒检测的图片数即 FPS 来衡量,检测速度越快的网络其 FPS 数越大。

### 4.4 性能对比

首先采用改进 CenterNet 网络模型在 VOC07 + 12 数据集上进行了多次重复实验,针对改进前后算法的具体性能对比如表 1 所示。

可以看出改进后的算法在检测精度和检测时间上都有上有很好的提升,相较于原始算法,其对各个目标的 R 值、AP 值都有显著提升。对于部分目标其精确率相对改进前的算法有一定降低,但是其召回率提升明显,这是因为精确率和召回率成负相关关系。通过对比精确率和召回率的调和平均 F1 值,可以看出改进后的算法并没有降低 F1 值,对大部分目标还提升了 F1 值,对 20 类目标的平均 F1 值提升了 0.011。对于衡量网络性能的平均精度 AP 值,改进后的算法显著提高了对目标的检测精度。其平均 AP 值 mAP 达到 79.51%,相对于改进前的算法提升了 2.44 个百分点。在模型检测速度方面,改进后的算法相较于原始算法有显著提升,改进后的 FPS 达到 93.83,相比于改进前的 91.06 提升了 2.77。实验结果证明了改进 CenterNet 目标检测算法在检测精度和检测速度上都有很好的优势。

表 1 改进前后性能对比

Tab. 1 Performance comparison before and after improvement

目标	改进前 centernet						改进后 centernet					
	F1	AP/%	P/%	R/%	mAP/%	FPS/ (f · s <sup>-1</sup> )	F1	AP/%	P/%	R/%	mAP/%	FPS/ (f · s <sup>-1</sup> )
Car	0.78	87.96	96.22	65.78	77.07	91.06	0.81	90.62	96.85	69.19	79.51	93.83
Horse	0.80	86.91	97.12	67.82			0.80	89.68	97.93	68.10		
Cat	0.81	88.23	94.46	71.51			0.82	89.38	96.27	72.07		
Aeroplane	0.77	84.52	98.89	62.46			0.77	87.86	97.33	63.86		
Motorbike	0.71	86.76	97.86	56.31			0.76	87.84	97.10	61.85		
Bus	0.79	85.32	94.74	67.61			0.79	87.26	94.19	68.54		

#### 4.5 消融实验

上述试验验证了改进后的算法在 VOC07 + 12 数据集上的有效性,为了验证所做出的改进对红外小目标图像检测是否有效,针对红外小目标数据集设计了一组消融实验,对比分析了改进前后算法的性能。在相同实验条件下进行实验,结果如表 2 所示。

表 2 消融实验

Tab. 2 Ablation experiment

网络模型	AP/%	FPS/(f · s <sup>-1</sup> )
CenterNet	77.01	90.03
CenterNet + CBAM	81.49	89.76
CenterNet + GSConv	77.23	94.02
改进 CenterNet	82.91	93.83

通过消融实验可以看出改进后的目标检测网络在红外数据集上性能有很大的提升,采用 CBAM 注意力机制,增加了目标特征,提升了检测精度;采用改进的 GSConv 卷积结构,提高了模型的推理速度,加快了计算速度。针对红外飞机小目标数据集其检测精度和速度都有了显著的提高。

如图 5 所示,改进后的 CenterNet 目标检测网络能够准确的输出红外图像中飞机小目标的中心点热力图,完成对目标的识别。

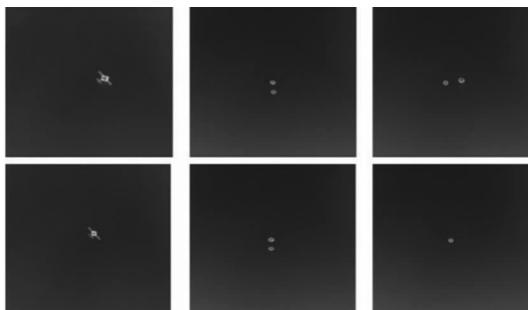


图 5 红外飞机小目标热力图

Fig. 5 Heatmap of infrared aircraft small targets

#### 5 结语

基于改进 CenterNet 的目标检测算法,在主干网络 ResNet50 中融合注意力机制,实现了同时考虑通道和空间信息,提高了网络对信息的提取能力。在特征输出模块引进 GSconv 卷积模块减少了运算时间。从实验结果可以看出改进后的目标检测算法的检测精度和实时性都有显著的提升,在 VOC07 + 12 数据集上其平均检测精度 mAP 提升 2.44 个百分点,检测速度 FPS 提升 2.77,达到了预期效果。在红外飞机小目标数据集上其检测精度 AP 值提升了 5.9 个百分点,检测速度提升了 3.8。实验结果证明本文所提出的改进模型在实际的红外图像目标检测任务上的有效性和优越性,为项目的后续研究采用本文优化改进后算法,针对各种特定复杂环境场景的应用奠定了基础。

#### 参考文献:

- [1] Shen Heng, Gan Zongliang. Forged shadow detection algorithm based on two-stage lightweight YOLO in infrared pedestrian [J]. Laser & Infrared, 2023, 53 (9): 1426 - 1433. (in Chinese)  
沈恒, 干宗良. 基于双阶段轻量 YOLO 的红外行人伪影检测算法 [J]. 激光与红外, 2023, 53 (9): 1426 - 1433.
- [2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2014: 580 - 587.
- [3] Ren S, He K, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks [J]. IEEE Transaction Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137 - 1149.

- [4] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904 – 1916.
- [5] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 779 – 788.
- [6] Redmon J, Farhadi A. Yolov3: an incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
- [7] Liu W, Anguelov D, Erhan D, et al. Ssd: single shot multi-box detector [C] // Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 – 14, 2016: 21 – 37.
- [8] Zhou X, Zhuo J, Krahenbuhl P. Bottom-up object detection by grouping extreme and center points [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 850 – 859.
- [9] Law H, Deng J. Cornernet: detecting objects as paired keypoints [C] // Proceedings of the European conference on computer vision (ECCV), 2018: 734 – 750.
- [10] Zhou X, Wang D, Krähenbühl P. Objects as points [J]. arXiv preprint arXiv:1904.07850, 2019.
- [11] Shi Xianrang, Su Yang, Ti Yan, et al. Improved model based on CenterNet object detection algorithms [J]. Computer Engineering, 2021, 47(9): 240 – 251. (in Chinese) 石先让, 苏洋, 提艳, 等. 基于 CenterNet 目标检测算法的改进模型 [J]. 计算机工程, 2021, 47(9): 240 – 251.
- [12] Shuai Zequn, Li Jun, Zhang Siyi. Crowded pedestrian detection algorithm suitable for vehicle edge computing [J]. Computer Engineering and Applications, 2023, 59(4): 156 – 164. (in Chinese) 帅泽群, 李军, 张世义. 适合车载边缘计算的拥挤行人检测算法 [J]. 计算机工程与应用, 2023, 59(4): 156 – 164.
- [13] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module [C] // Proceedings of the European conference on computer vision (ECCV), 2018: 3 – 19.
- [14] Li H, Li J, Wei H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles [J]. arXiv preprint arXiv:2206.02424, 2022.
- [15] F Chollet. Xception: deep learning with depthwise separable convolutions [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: 1800 – 1807.
- [16] Ma, N., Zhang, X., Zheng, HT., et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [C] // ECCV 2018. Lecture Notes in Computer Science, 2018.
- [17] Zhao Fenkui, Cheng Haifei, Su Sansan, et al. Traffic scene target detection technology based on improved CenterNet [J]. Journal of Chongqing Jiaotong University: Natural Science, 2022, 41(12): 11 – 17. (in Chinese) 赵奉奎, 成海飞, 苏珊珊, 等. 基于改进 CenterNet 的交通场景目标检测技术研究 [J]. 重庆交通大学学报: 自然科学版, 2022, 41(12): 11 – 17.