文章编号：1001-9014(2022)06-1092-10

DOI：10. 11972/j. issn. 1001-9014. 2022. 06. 019

GPNet:轻量型红外图像目标检测算法

李现国[1,2*]，曹明腾[1]，李 滨[1]，刘 意[1,2]，苗长云[1,2]

（1. 天津工业大学 电子与信息工程学院，天津 300387；
2. 天津市光电检测技术与系统重点实验室，天津 300387）

摘要：针对资源受限的红外成像系统准确、实时检测目标的需求，提出了一种轻量型的红外图像目标检测算法GPNet。采用GhostNet优化特征提取网络，使用改进的PANet进行特征融合，利用深度可分离卷积替换特定位置的普通3×3卷积，可以更好地提取多尺度特征并减少参数量。公共数据集上的实验表明，本文算法与YOLOv4、YOLOv5-m相比，参数量分别降低了81%和42%；与YOLOX-m相比，平均精度均值提高了2. 5%，参数量降低了51%；参数量为12. 3 M，检测时间为14 ms，实现了检测准确性和参数量的平衡。

关 键 词：红外图像；目标检测；YOLO；GhostNet；参数量

GPNet：Lightweight infrared image target detection algorithm

LI Xian-Guo[1,2*]，CAO Ming-Teng[1]，LI Bin[1]，LIU Yi[1,2]，MIAO Chang-Yun[1,2]

（1. School of Electronics and Information Engineering，Tiangong University，Tianjin 300387，China；
2. Tianjin Key Laboratory of Optoelectronic Detection Technology and System，Tianjin 300387，China）

Abstract：A lightweight infrared image target detection algorithm GPNet is proposed to address the need for accurate and real-time target detection in resource-constrained infrared imaging systems. The feature extraction network is optimized using GhostNet，feature fusion is performed using an improved PANet，and a depth-separable convolution is used to replace the ordinary 3×3 convolution at specific locations to better extract multi-scale features and reduce the number of parameters. Experiments on public datasets show that the algorithm in this paper reduces the number of parameters by 81% and 42% compared with YOLOv4 and YOLOv5-m，respectively；the average mean accuracy is improved by 2. 5% and the number of parameters is reduced by 51% compared with YOLOX-m；the number of parameters is 12. 3 M and the detection time is 14 ms，which achieves a balance between detection accuracy and number of parameters.

Key words：infrared image，target detection，YOLO，GhostNet，number of parameters

引言

利用红外图像进行目标检测在很多领域具有不可替代的地位，如红外夜视、工业探伤、红外成像制导等[1-2]。目标检测是红外成像系统中最关键也最具挑战性的任务之一。随着深度学习的快速发展，一些基于卷积神经网络（Convolutional Neural Networks，CNN）的红外图像目标检测算法被提出，显著提高了检测准确度。但是这些算法的部署成本比较高——计算复杂度高、参数量大，只有高端图形处理器才能保证其性能。而大多数红外成像系统通常部署在仅配备CPU或中低端GPU的资源受限设备上。因此，研究设计适合红外成像系统的准确、实时检测算法及模型具有重要的意义和实用价值。

红外图像目标检测算法可分为两类：传统算法和基于深度学习的算法[3]。传统算法的主要思想是将图像视为物体、背景和噪声三部分，通过传统的图像处理方法抑制红外图像中的背景和噪声实现目标检测任务。Zhang等人[4]分析了可见光图像和

收稿日期：2022- 05- 25，**修回日期**：2022- 09- 02　　**Received date**：2022- 05- 25，**Revised date**：2022- 09- 02
基金项目：国防科技创新特区项目，天津市重点研发计划科技支撑重点项目（18YFZCGX00930）
Foundation items：Supported by the National Defense Science and Technology Innovation Special Zone Project，Key Projects of Science and Technology Support of Tianjin，China（18YFZCGX00930）
作者简介（Biography）：李现国（1981—），男，博士，山东邹城人，教授，研究领域为智能信息处理、光电检测技术与系统
*通讯作者（Corresponding author）：E-mail：lixianguo@tiangong. edu. cn

红外图像的共享特征,将方向梯度直方图(Histogram of Oriented Gradient,HOG)、AdaBoost以及支持向量机(Support Vector Machine,SVM)引入到红外图像行人检测任务;Ge等人[5]提出将感兴趣区域(Region-of-Interest,RoI)生成、物体分类和跟踪三个模块整合为一个级联,每个模块都利用互补的视觉特征来区分物体和杂乱的背景;Su等人[6]提出一种使用区域估计的帧差法实现车载红外图像行人检测的算法;Zhu等人[7]将双边滤波与横纵多尺度灰度差结合来增强弱目标,抑制背景的同时提高目标强度,并通过自适应局部阈值分割和全局阈值分割提取候选目标;Cai等人[8]提出一种箱粒子标签多伯努利多目标检测算法,通过使用均值滤波对获得的灰度图进行降噪处理,并将所有像素按强度大小进行排序选出强度较大的区域。这类算法计算量小,能一定程度上对背景进行抑制,但参数选择较为复杂,对于复杂的背景检出率较低且鲁棒性较差。

KAIST[9]、FLIR[10]以及CVC-09[11]等红外热成像数据集的公开,促进了基于深度学习的算法在红外成像领域的应用。Ghose等人[12]将Faster R-CNN应用到红外图像上,使用显著性图谱增强红外图像;Devaguptapu等人[13]提出了一个多模型的Faster R-CNN,通过RGB通道获得高级红外特征。这类基于先产生候选框再检测的两阶段目标检测算法虽然准确率较高,但运行速度较慢,且训练成本较高。为了解决两阶段目标检测算法的问题,并便于在资源受限的嵌入式系统上执行,Dai等人[14]提出一种类SSD的红外图像目标检测算法——TIRNet,采用VGG作为特征提取网络并引入残差分支,提高了运行速度;Mate等人[15]将YOLOv3应用在红外图像目标检测中,用于检测恶劣天气下的行人;Song等人[16]将SE模块引入YOLOv3,提高了网络的特征表达能力,在小目标行人检测上取得了更高的精度和更低的误报率;Du等人[17]将可见光数据集迁移到红外数据集,使用YOLOv4进行二次迁移学习,在车辆检测方面取得了良好效果;Wu等人[18]提出了一种基于YOLOv4的行人实时检测算法Rep-YOLO,但存在泛化能力不强的缺点;Li等人[19]基于YOLOv5提出了YOLO-FIRI算法,在红外图像低识别率和高误报率方面有所改善。这类一阶段目标检测算法,实现了端到端的检测,检测速度大幅度提高,但由于红外图像存在波长较长、噪声较大、空间分辨率较差以及对环境温度变化敏感等问题,检测准确度不高。

本文研究并提出了一种基于YOLOv4的轻量型红外图像目标检测算法——GPNet。主要贡献主要有3个方面:第一,以YOLOv4作为基本框架,使用GhostNet替换YOLOv4的主干网络,能够以很低的运算量生成冗余的特征图,提高算法的执行速度;第二,使用深度可分离卷积替换特征提取、特征融合和检测头模块特定位置的普通3×3卷积,可更好地提取深层和浅层的特征并减少参数量;第三,设计了一种改进型的PANet结构,可更好地融合特征,提高检测的准确度。

1　算法的网络结构

YOLOv4和GhostNet在可见光图像目标检测方面取得了良好的性能。与YOLOv3相比,YOLOv4采用CSP(Cross Stage Partial Networks)[20]和PANet(Path Aggregation Network)[21]结构对其进行改进,在检测准确度和计算复杂度方面都更有利于进行目标检测。GhostNet网络主要由Ghost模块组成的步长为1和2两种形式的Ghost Bottle-necks构成,通过Ghost模块代替普通卷积以更低的运算量来生成冗余的特征图,从而降低整个网络的运算量[22]。

本文通过分析YOLOv4和GhostNet这两种网络的结构特点、优化方法等,提出了一种轻量型的红外目标检测算法GPNet,以快速准确地检测图像中的物体。GPNet的整体网络结构如图1所示,主要包含三部分:轻量级的特征提取模块,实现跨阶段的多尺度特征融合模块以及多尺度的检测头模块。首先,特征提取模块对输入的红外图像提取低级空间信息和高级语义特征;然后,多尺度特征融合模块将提取的多层次特征进行融合;最后,多尺度的检测头模块对输入的图片生成密集的边界框并预测类别分数,通过非极大值抑制处理[23],得到最终结果。

在图1中,GPNet的各部分主要由深度可分离卷积块(Depthwise Separable Convolution Module X,DSCMX)、标准卷积(Convolution,Conv)、空间金字塔池化(Spatial Pyramid Pooling,SPP)[23]以及GBX(Ghost Bottleneck X)等四种结构块构成。其中DSCMX有DSCM3和DSCM5两种形式,分别代表深度可分离三次卷积块和深度可分离五次卷积块。GBX有GB1和GB2两种形式,分别为stride=1和stride=2两种步长时的Ghost瓶颈模块。

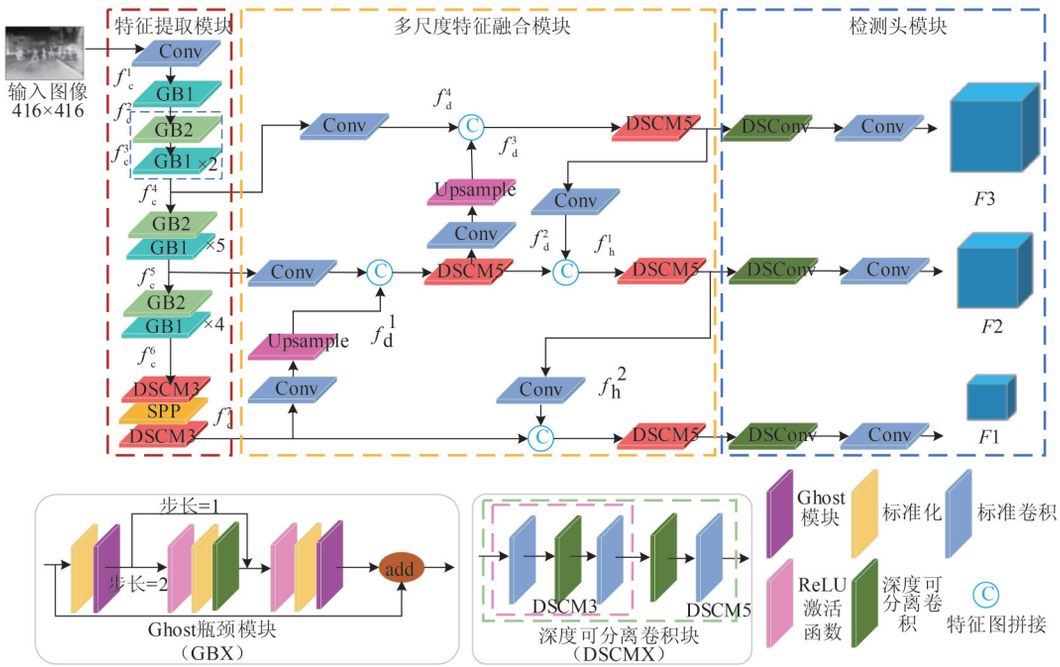


图1 GPNet整体网络结构图

Fig. 1 GPNet overall network structure

1.1 改进的特征提取模块

Ghost模块（如图2所示），最早在GhostNet中被提出。该模块基于一组原始的特征图，应用一系列线性变换，以很小的代价生成许多能从原始特征发掘更多信息的Ghost特征图。该模块即插即用，通过堆叠Ghost模块得到步长为1和2两种步长的Ghost瓶颈模块，从而搭建成了轻量级的神经网络GhostNet。

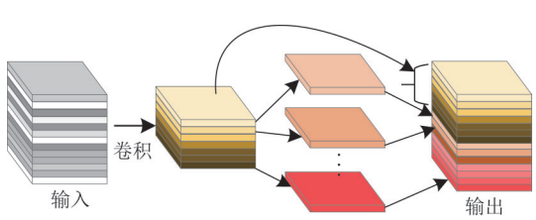


图2 Ghost模块结构图

Fig. 2 Ghost module

深度卷积神经网络[24-26]需要大量的计算成本。尽管MobileNet[27]和ShuffleNet[28]引入了深度可分离卷积和shuffle操作，以较小的卷积核（浮点运算）构建CNN，但其余1×1的卷积层仍占用大量内存和FLOPs。对于普通卷积，给定输入数据 $X \in \mathrm{R}^{c\times h\times w}$，其中 c 代表输入通道数，h 和 w 代表输入数据的高和宽，一个任意的产生 n 个特征图的卷积层的操作可以被表述为

$$Y = X * f + b \quad , \quad (1)$$

其中 $*$ 代表卷积运算，b 代表偏差项，$Y \in \mathrm{R}^{h'\times w'\times n}$ 代表具有 n 个通道的输出特征图，$f \in \mathrm{R}^{c\times k\times k\times n}$ 是这一层中的卷积核，h' 和 w' 分别代表输出数据的高和宽，$k\times k$ 代表卷积核 f 的内核大小。在此卷积过程中，由于卷积核数量 n 和通道数 c 通常很大（例如256和512），所需的FLOPs数量达 $n\cdot h'\cdot w'\cdot c\cdot k\cdot k$ 之多。在GhostNet中，作者指出普通卷积层的输出特征图通常包含很多冗余，并且其中一些彼此相似，无需使用如此大数量的FLOPs和参数来生成这些冗余特征图，可选择用少数原始特征图以更廉价的操作生成这些特征图。这些原始特征图相对较小，并由普通的卷积核生成。具体来说，m 个原始特征图 $Y' \in \mathrm{R}^{h'\times w'\times m}$ 是使用一次卷积生成的，具体计算式为

$$Y' = X * f' + b \quad , \quad (2)$$

其中，$f' \in \mathrm{R}^{c\times k\times k\times m}$ 代表使用的卷积核，$m \le n$，b 代表偏差。为了进一步得到所需的 n 个特征图，文献[22]提出对 Y' 中的每个原始特征图应用一系列廉价的线性运算，以生成 s 个Ghost特征图：

$$y_{ij} = \Phi_{i,j}(y'_i), \ \forall i = 1,\ldots,m, \ j = 1,\ldots,s \quad , \quad (3)$$

其中 y'_i 是 Y' 中第 i 个原始特征图，$\Phi_{i,j}$ 是第 j 个线性运算，用于生成第 j 个 Ghost 特征图 y_{ij}。最终，可以获得 $n=m\cdot s$ 个特征图 $Y=[y_{11},y_{12},\cdots,y_{ms}]$ 作为 Ghost 模块的输出数据。

通过对这些 Ghost 模块堆叠从而组成 Ghost 瓶颈模块，将其简称为 GBX 模块。如图 3 所示，GBX 模块由两个 Ghost 模块堆叠组成，第一个 Ghost 模块用作扩展层，增加通道数。第二个 Ghost 模块用作缩减通道数，以与短路路径匹配，短路路径会用于连接这两个 Ghost 模块的输入和输出。整个 GBX 模块的结构类似于 ResNet[26] 中的基本残差块，并借鉴了 MobileNetV2[29] 除第二个 Ghost 模块仅进行批量归一化（BN：Batch Normalization）[30] 外，其它每一层之后都进行 BN 和 ReLU 非线性激活函数操作。GBX 分为两个支路，当步长为 2 时，会在两个 Ghost 模块中间添加一层深度可分离卷积进行特征图的宽高压缩。

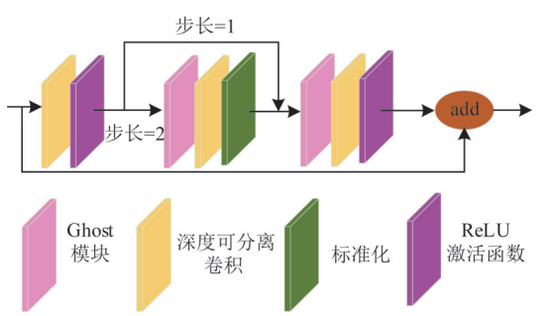


图 3　Ghost瓶颈结构图

Fig. 3　Ghost bottleneck structure

通过对 GBX 模块的堆叠，构建如图 4 所示的 GhostNet 作为本文算法的特征提取网络。

对于输入的红外图像，首先通过一次卷积得到的原始特征图 f_c^1 可表示为

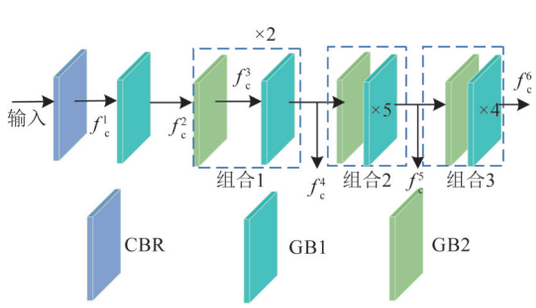


图 4　GhostNet结构图

Fig. 4　GhostNet structure

$$f_c^1=\mathrm{Conv}(x)\quad,\quad(4)$$

其中，Conv 表示卷积操作。这类少量的原始特征图作为 Ghost 模块的输入可以通过简便的操作生成批量的特征图。每经过一个 GB1 模块后得到相应的一个特征图，例如特征图 f_c^2 可表示为

$$F_{\mathrm{Ghost}}(x)=\Phi(\mathrm{Conv}(x))\quad,\quad(5)$$

$$f_c^2=F_{\mathrm{Ghost}}(F_{\mathrm{Ghost}}(f_c^1))+f_c^1\quad,\quad(6)$$

其中，F_{Ghost} 表示经过一个 Ghost 模块的操作，Φ 表示线性运算。每经过一个 GB2 模块后，也得到一个特征图，例如特征图 f_c^3 可表示为

$$f_c^3=F_{\mathrm{Ghost}}(\mathrm{DSConv}(F_{\mathrm{Ghost}}(f_c^2)))+f_c^2\quad,\quad(7)$$

其中，DSConv 表示深度可分离卷积操作。经过组合 1 后得到第一类特征融合模块所需的特征图 f_c^4，再经过组合 2 后得到第二类所需的特征图 f_c^5，最后经过组合 3 后得到第三类所需的特征图 f_c^6。

研究表明[31-32]，相比单纯的使用 $k\times k$ 最大池化的方式，SPP 模块使用 $k=\{1\times1,\ 5\times5,9\times9,13\times13\}$ 的最大池化的方式，即利用四种尺度对特征图进行划分，然后从每个区域中选取一个最大值作为输出，如图 5 所示。再将不同尺度的特征图进行特征图拼接操作，可以更有效地增加主干特征的接收范围，显著地分离上下文特征。当特征图 f_c^6 传入 SPP 前后，都需要经过一个 3 次卷积块。由于深度可分离卷积通过解耦空间和深度信息，可减少模型参数、降低计算量[27]，所以本文将 YOLOv4 中的 3 次卷积块中的 3×3 卷积替换为深度可分离卷积，组成 DSCM3 模块，如图 6 所示。

特征图 f_c^6 在 DSCM3 模块中，首先经过标准卷积得到特征图 f_m^1，然后利用替换的深度可分离卷积进行特征图的提取得到特征图 f_m^2，深度可分离卷积由深度卷积和逐点卷积组成，深度卷积将单个滤波器应用到每一个输入通道，然后，逐点卷积用 1×1 卷积来组合不同深度卷积的输出，大大降低了参数量。深度可分离卷积的参数量为

$$P_{\mathrm{DSConv}}=k\times k\times N+N\times M\quad,\quad(8)$$

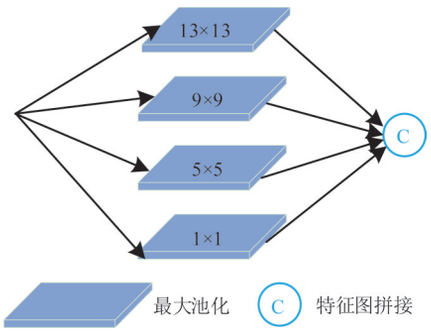


图5 SPP模块结构图

Fig. 5 SPP module structure diagram

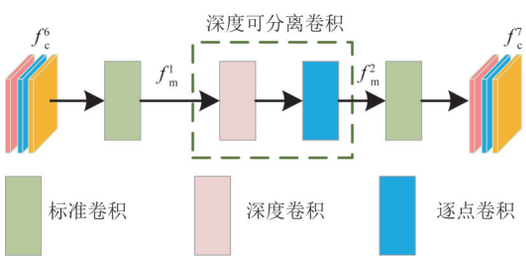


图6 DSCM3模块结构图

Fig. 6 DSCM3 module structure

标准卷积的参数量为

$$P_{\text{Conv}} = k \times k \times N \times M \quad , \quad (9)$$

则参数量的比值为

$$\frac{P_{\text{DSConv}}}{P_{\text{Conv}}} = \frac{1}{M} + \frac{1}{k^2} \quad , \quad (10)$$

其中，k 表示卷积核的尺寸，N 表示输入通道数，M 表示输出通道数。由式(10)可知，当输出特征的通道数很大时，深度可分离卷积的参数量仅为标准卷积的 $1/k^2$，一般 $k>1$。可以证明，本文的DSCM3模块可以显著减少模型的参数量，提升红外图像目标检测的检测速度。

1.2 改进的多尺度特征融合模块

在目标检测领域，为更好地提取融合特征，本文GPNet的特征融合模块中沿用了YOLOv4特征融合中的PANet结构[21]，但进行了一些改进。设计了一种改进型的PANet结构，如图7所示。图中紫色部分为自上而下的FPN层，红色部分为自下而上的特征金字塔层。黄色箭头代表下采样，绿色代表上采样。整个PANet结构主要由FPN层和特征金字塔层融合而成。通过这样融合，FPN层自上而下传达强语义特征，特征金字塔自下而上传达强定位特征，从不同的主干层对不同的检测层进行特征聚合。自上而下的FPN层通过上采样和特征图拼接操作可以得到充足的深层特征图，然后利用DSCM5模块实现对深层特征的充分提取。自下而上的特征金字塔层通过下采样和特征图拼接操作可以得到充足的浅层特征图，之后利用DSCM5模块实现对浅层特征的充分提取。改进型的PANet结构，将原来自上而下的FPN层和特征金字塔层中所用到的5次卷积块中的普通3×3卷积替换为了深度可分离卷积，构成了DSCM5模块。该模块结构与DSCM3模块类似，但不同的是使用两处深度可分离卷积替换普通3×3卷积，如图8所示。

与DSCM3模块作用相同，DSCM5模块进一步减少了网络参数、提高了检测效率。由式(10)可知，每经过一个深度可分离卷积，相比于标准卷积参数量下降为原来的 $1/M+1/k^2$，即，经过一个DSCM5模块可实现两处卷积参数量的下降，由式(8)和式(9)可知，可减少 $2N(k^2M-k^2-M)$ 个参数量。可以证明，本文的DSCM5模块由于模型参数量大大降低，可以有效地提高红外图像目标检测的检测速度。

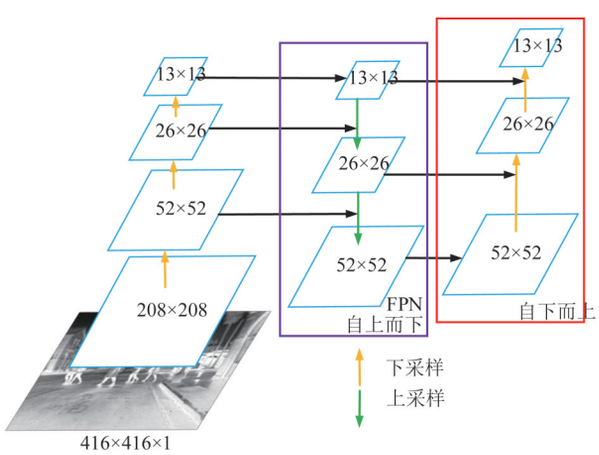


图7 改进的PANet结构图

Fig. 7 Modified PANet structure

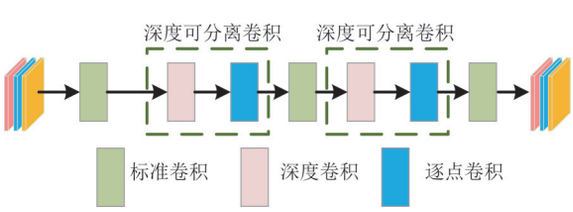


图8 DSCM5模块结构图

Fig. 8 DSCM5 module structure

1.3 改进的检测头模块

得益于深度可分离卷积在特征提取模块和多尺度特征融合模块取得的效果，将深度可分离卷积

同样应用在了检测头模块,将常用的标准的3×3卷积替换为深度可分离卷积。相比采用标准的3×3卷积,深度可分离卷积具有降低参数并提高AP值的优势。

在多尺度特征融合模块中,如图1虚线框所示,首先通过上采样得到特征图 f_{d}^{3} ,然后与特征提取模块中 f_{c}^{4} 卷积操作后进行特征图拼接,最后经过DSCM5模块和两个卷积层进行特征提取得到特征图 $F3$,其特征尺寸为输入图像的 $1/64$,用于检测小目标。同时将特征提取模块中的特征图 f_{c}^{6} 经过卷积层和 f_{d}^{1} 进行特征图拼接,经过DSCM5模块进一步再与下采样得到的 f_{h}^{1} 进行特征图拼接,最后经过DSCM5模块和两个卷积层进行特征提取得到特征图 $F2$,其特征尺寸为输入图像的 $1/256$,用于检测中目标。同理,通过这种方式得到特征图 $F1$,其特征尺寸为输入图像的 $1/1024$,用于检测大目标。特征图 $F1$、$F2$ 和 $F3$ 的具体过程如下所示:

$$F1=\mathrm{Conv}(\mathrm{DSConv}(F_{\mathrm{D5}}(F_{\mathrm{c}}(f_{\mathrm{c}}^{7},f_{\mathrm{h}}^{2})))) \quad , \quad (11)$$

$$f_{\mathrm{d}}^{1}=F_{\mathrm{UP}}(\mathrm{Conv}(f_{\mathrm{c}}^{7})) \quad , \quad (12)$$

$$f_{\mathrm{d}}^{2}=F_{\mathrm{D5}}(F_{\mathrm{c}}(\mathrm{Conv}(f_{\mathrm{c}}^{5}),f_{\mathrm{d}}^{1})) \quad , \quad (13)$$

$$F2=\mathrm{Conv}(\mathrm{DSConv}(F_{\mathrm{D5}}(F_{\mathrm{c}}(f_{\mathrm{d}}^{2},f_{\mathrm{h}}^{1})))) \quad , \quad (14)$$

$$f_{\mathrm{d}}^{3}=F_{\mathrm{UP}}(\mathrm{Conv}(f_{\mathrm{d}}^{2})) \quad , \quad (15)$$

$$f_{\mathrm{d}}^{4}=\mathrm{Conv}(f_{\mathrm{c}}^{4}) \quad , \quad (16)$$

$$F3=\mathrm{Conv}(\mathrm{DSConv}(F_{\mathrm{D5}}(F_{\mathrm{c}}(f_{\mathrm{d}}^{4},f_{\mathrm{d}}^{3})))) \quad , \quad (17)$$

其中,Conv 表示卷积操作,DSConv 表示深度可分离卷积操作,F_{D5} 表示DSCM5模块操作,F_{c} 表示特征图拼接操作,F_{UP} 表示上采样操作。通过多尺度特征融合的方式,将浅层网络中丰富的位置信息和纹理信息更好的与深层网络的语义特征信息相融合,增强模型在小目标下的多尺度特征学习能力,从而提升模型在小目标在复杂场景下的检测能力。

2　实验分析

使用公开的和自制的红外数据集测试本文所提出的红外图像目标检测算法GPNet的性能。首先,从检测精度、速度和参数等方面与SOTA(state-of-the-art)目标检测算法进行对比。然后,进行消融实验,以测试不同方法带来的性能提升。

2.1　检测性能的比较

FLIR的红外数据集是一个经典的公开目标检测数据集,被很多红外图像物体检测算法所评估[19]。采用来自多个短视频的10228张图片,并将其被划分为train和test两个子集,分别包含8862张和1366张图片。

输入图像大小均为416×416,epoch为300,batch size为32,初始学习率为0.001,momentum为0.0005,weight decay为0.937,IoU阈值为0.5,优化器选用SGD,使用mosic数据增强算法扩充样本的多样性。所有实验都是基于Pytorch框架,并利用两块GeForce GTX 1080Ti GPU进行训练。其中GPNet模型训练选用GhostNet在ImageNet数据集上取得73.98%准确率时预训练模型。如图9所示,训练300个epoch后,模型达到收敛。

表1是本文所提出的GPNet和SOTA算法在各项评价指标上的比较。结果以IoU阈值=0.5时的

表1　GPNet和SOTA算法在FLIR红外测试集上的定量比较

Table 1　Quantitative comparison of GPNet and SOTA algorithms on the FLIR IR test set

Model	AP/(%)		mAP50/(%)	Recall/(%)		F1		Params/M	FLOPs/G	Time/ms
	person	car		person	car	person	car			
FasterR-CNN	39.09	61.67	50.38	47.06	69.84	0.43	0.47	136.7	252.7	75
SSD	43.78	58.72	51.25	20.34	42.60	0.33	0.58	23.7	115.7	15
YOLOv3	73.73	85.93	79.83	59.36	77.89	0.70	0.81	61.5	65.5	19
YOLOv4	78.13	84.74	81.44	61.45	73.99	0.72	0.80	63.9	59.8	25
YOLOv5-m	75.24	85.79	80.52	54.25	74.20	0.68	0.81	21.1	21.3	17
YOLOX-m	72.02	80.46	76.24	52.43	68.16	0.66	0.77	25.3	31.1	16
YOLOv4+GhostNet	69.41	86.14	77.77	49.17	76.00	0.63	0.81	39.3	25.6	17
GPNet(本文)	72.65	84.83	78.74	47.32	71.95	0.62	0.79	12.3	7.2	14

AP(平均精度)、mAP50(平均精度均值)、Recall(召回率)和F1(精度与召回率的调和平均)评估算法的准确性,以Params(参数量)、FLOPs(浮点运算数)和检测时间来评估速度。mAP50计算的是person和car两个种类IoU阈值=0.5时的数值。计算Recall和F1时置信度阈值=0.5。

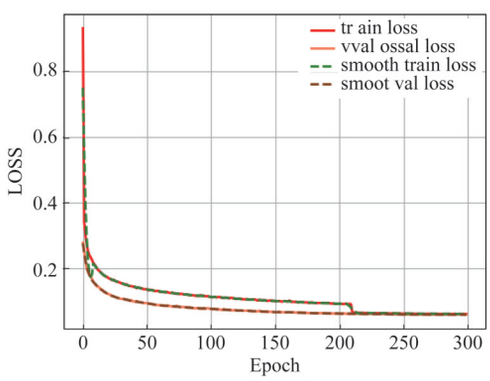


图9 训练损失曲线

Fig. 9 Training loss curve

从表1可以看出,GPNet与原YOLOv4相比,mAP50仅下降3%的情况下参数量降低了约81%,并且在car这个种类上比YOLOv4高0.1%;当种类为car时,该算法的AP值虽然略低于YOLOv3和YOLOv5-m,但该算法的参数量降低明显,分别降低了80%和42%,从这个角度上可以认为该算法在car种类上取得了最优结果。并且,从mAP50方面来看,虽然本文提出的算法分别落后于YOLOv3、YOLOv4和YOLOv5-m结果的1.09%、2.70%和1.78%,但参数量仅为它们的20%、19%和58%。与当前最先进的YOLOX-m相比,本文所提出的算法在AP(person)、AP(car)和mAP50数值上分别高出0.63%、4.37%和2.5%,且参数量仅为其49%。总体而言,该算法虽然在检测精度上有所牺牲,但在参数量上均有大幅度下降。此外,本文的算法在FLOPs(Floating Point Operations)指标上取得最优,该指标可以用来很好地衡量模型的计算复杂度(本文算法的检测时间最短),保证了模型在推理速度和准确度上的平衡。

与仅对YOLOv4替换主干特征提取网络的YOLOv4+GhostNet算法相比,后者相较于YOLOv4在参数量和FLOPs方面分别降低了约38%和57%,检测时间减少了8 ms。而本文GPNet算法,在此基础上参数量和FLOPs又分别降低约69%和72%,检测时间进一步减少了3 ms,同时在准确性方面AP(person)和mAP50数值分别提升了3.24%和0.97%,进一步带来了推理速度和准确度的提升。

为验证算法的鲁棒性,又在KAIST红外数据集的测试集上(set06-set11,10914张图片)进行了实验,结果如表2所示。可看出本文算法在种类person上除Recall指标外均取得了最优结果。

为了进一步验证算法的鲁棒性,采用上文在FLIR数据集训练得到的模型在CVC-09红外数据集(含2884张夜晚图片和707张白天图片)以及自制的校园红外数据集(1103张图片)上进行测试。

表2 GPNet和SOTA算法在KAIST红外数据集上的定量比较

Table 2 Quantitative comparison of GPNet and SOTA algorithms on the KAIST IR test set

Model	Size	AP/(%)	Recall/(%)	F1
Faster R-CNN	416×416	39.52	55.49	0.40
YOLOv4	416×416	50.45	49.49	0.54
YOLOv5-m	416×416	50.69	44.42	0.54
YOLOv5-s	416×416	50.18	44.65	0.53
YOLOX-m	416×416	54.41	48.82	0.56
YOLOX-s	416×416	53.49	47.27	0.55
GPNet(本文)	416×416	**55.04**	47.36	**0.57**

CVC-09红外数据集的测试结果如表3所示。可以看出本文算法除在种类person的Recall指标以外,指标均取得最优,其中AP的数值比YOLOv4分别高出2.76%和7.20%,比YOLOv5-m分别高出0.98%和4.44%。

自制的校园红外数据集图片样例如图10所示,测试结果如表4所示。可看出本文算法在对种类person测试中,AP和F1均取得了最优结果。

综合分析上述4组表格中的数据,本文算法在4种数据集下的多个场景中,性能指标上均有一定优势,验证了本文算法在降低大量参数的同时仍然保持了良好的鲁棒性。

图11为GPNet和SOTA算法在FLIR红外测试集上的检测结果图,第一行是YOLOv4的结果,第二行是YOLOv5-m的结果,第三行是YOLOX-m的结果,第四行是GPNet(本文)的结果。用同样的三张图片来比较检测模型的性能,从图中可以看出,本文提出的GPNet算法相比于其它三种算法,在对小型的车辆和行人检测上不易产生漏检和误检,同时在置信度上也更高。对KAIST、CVC-09和自制数据集也可得到类似的结果,此处不再赘述。

表 3　GPNet 和 SOTA 算法在 CVC-09 测试集上的定量比较

Table 3　Quantitative comparison of GPNet and SOTA algorithms on the CVC-09 IR test set

Model	Size	AP/(%)		mAP50/(%)	Recall/(%)		F1	
		person	car		person	car	person	car
Faster R-CNN	416×416	42.39	67.92	55.15	52.40	75.42	0.40	0.54
YOLOv4	416×416	73.53	79.31	76.42	74.48	70.47	0.70	0.76
YOLOv5-m	416×416	75.31	82.07	78.69	80.89	79.22	0.72	0.76
YOLOv5-s	416×416	76.29	80.53	78.41	71.90	75.56	0.73	0.75
YOLOX-m	416×416	71.86	79.16	75.51	74.41	75.08	0.71	0.75
YOLOX-s	416×416	71.99	75.44	73.72	69.20	70.69	0.70	0.71
GPNet(本文)	416×416	**76.29**	**86.51**	**81.40**	70.59	**85.23**	**0.75**	**0.84**

(a)　(b)　(c)

图 10　自制的校园红外数据集 (a)广场,(b)教学楼,(c)操场

Fig. 10　Self-made campus infrared dataset (a)square, (b)academic Building, (c)playground

表 4　GPNet 和 SOTA 算法在自制校园红外数据集上的定量比较

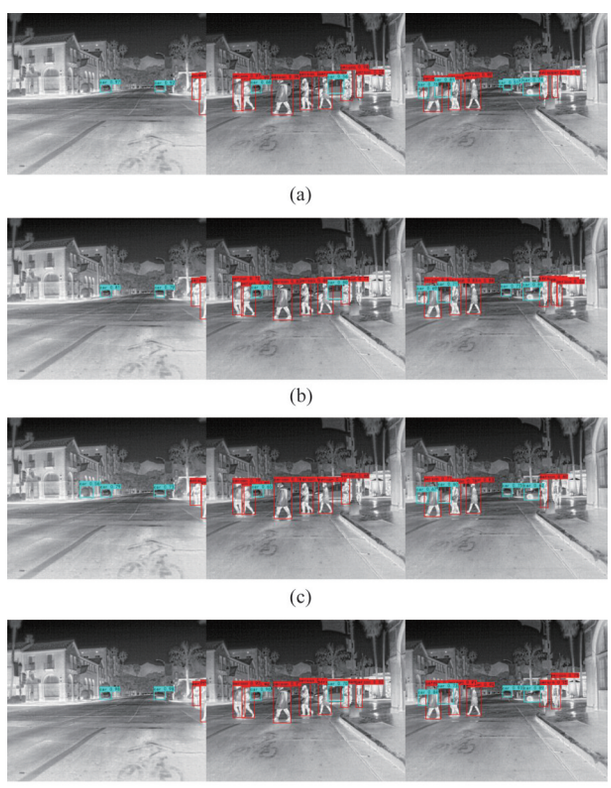
Table 4　Quantitative comparison of GPNet and SOTA algorithms on the self-made campus infrared dataset

Model	Size	AP/(%)	Recall/(%)	F1
Faster R-CNN	416×416	45.28	61.98	0.41
YOLOv4	416×416	81.23	81.11	0.76
YOLOv5-m	416×416	79.63	78.25	0.76
YOLOv5-s	416×416	75.08	62.04	0.73
YOLOX-m	416×416	80.38	72.63	0.77
YOLOX-s	416×416	79.00	69.77	0.76
GPNet (本文)	416×416	**81.46**	78.35	**0.80**

2.2　消融实验

为了更直观地看到不同改进方法对模型性能的影响,进行了消融实验。具体来说,首先将 YOLOv4 的主干网络直接替换为 GhostNet,然后在此基础上逐次利用深度可分离卷积在不同位置进行改进,以观察实验结果并分析其影响。

为了保证消融实验的严谨性,在同一训练平台上设置 300 个 epoch,训练完成后并在 FLIR 测试集上测试,实验数据如表 5 所示。表中 □ 代表该处被改动,Backbone 处有 ▮ 代表被 GhostNet 替换,其它处有 □ 代表普通 3×3 卷积被深度可分离卷积所替换。3-C、5-C 和 Dsample 分别代表三次卷积块、五次卷积块和下采样。AP 代表种类为 person 时的数据。mAP50 计算的是 person 和 car 两个种类 IoU 阈值=0.5 时的数值。



(a)

(b)

(c)

(d)

图 11　GPNet 和 SOTA 算法在 FLIR 红外测试集上的检测结果图 (a) YOLOv4 检测结果,(b)YOLOv5-m 检测结果,(c)YOLOX-m 检测结果,(d)GPNet(本文)检测结果

Fig. 11　Detection comparison of GPNet and SOTA algorithms on the FLIR IR test set(a)YOLOv4 detection results,(b)YOLOv5-m detection results,(c)YOLOX-m detection results,(d)GPNet(ours)detection results

第一组实验数据为在仅替换主干网络时,模型的各项指标。为了验证改进型的 PANet 的有效性,通过对比前两组实验数据可以看出,在对网络特征

表5 在FLIR红外数据集进行消融实验

Table 5 Ablation experiments in FLIR infrared dataset

Backbone	3-C	5-C	Dsample	Head	AP/(%)	mAP50/(%)	Recall/(%)	Params/M	Weight/MB
□					69.41	77.77	46.20	39.3	150.3
□		□			71.21	78.13	46.41	26.2	100.4
□	□	□			69.73	76.72	**49.27**	18.2	68.4
□	□	□		□	**72.65**	**78.74**	47.32	12.7	47.4
□	□	□	□	□	67.37	76.27	44.75	**11.4**	**42.5**

融合模块的PANet中的普通3×3卷积替换为深度可分离卷积后，在参数量下降了13.1M的前提下，AP(person)和mAP分别提升了1.80%和0.36%。为了验证特征融合模块处三次卷积块的有效性，选用前三组实验数据进行对比，结果显示，该实验模型的Recall指标取得最优的49.27%，该指标表示整个数据集中被成功检测出的实例比例，同时参数量下降了8M。为了验证改进检测头的有效性，选用前四组实验数据集进行对比，结果显示，模型的指标再次得到了提升，AP(person)和mAP分别达到了最高的72.65%和78.74%，参数量被进一步降低5.5M。最后一组实验数据表明，虽然该实验模型可以将网络的参数量降到最低，相比与第四组实验数据可以再降低1.3M，但此时的模型各项指标也随之有大幅的下降。综合上述五组实验数据，本文设计的第四组网络模型在检测精度和计算成本上达到了更好的平衡。

3 结论

本文基于YOLOv4和GhostNet提出了一种轻量型红外图像目标检测算法GPNet，设计了其网络结构。将YOLOv4的主干网络的CSP模块替换为了GhostNet，使参数量由原来的63.9 M降低为39.3 M；在网络的特征提取模块、多尺度特征融合模块和检测头模块用深度可分离卷积去替换特定位置的普通3×3卷积，将参数量进一步降低到了12.7 M；优化了PANet结构，更好地融合特征，提高了检测精度。在FLIR红外数据集上对person和car两个种类进行了测试，本文算法在car上的平均精度均值比YOLOv4提高了0.1%，参数量减少了81%；与YOLOX-m相比，平均精度均值提高了2.5%，参数量降低了51%；参数量为12.3M，检测时间为14ms。在KAIST红外数据集上对person种类进行了测试，GPNet相比于YOLOv4取得了最优结果，实现了检测准确性和参数量的平衡；在CVC-09和自制数据集上的测试表明，GPNet的AP和F1指标均有一定的优势，验证了本文提出的算法在红外图像目标检测方面的正确性、有效性和鲁棒性。

References

[1] Han J, Yu Y, Liang K, *et al*. Infraredsmall-target detection under complex background based on subblock-level ratio-difference joint local contrast measure[J]. *Optical Engineering*, 2018, **57**(10):103105.

[2] LI Tong-shun, XI Yong, YIN Jian-Fei. Analysis of the development of key technologies for air-to-air infrared guidance[J]. *Shanghai Aerospace*(李同顺, 奚勇, 印剑飞。对空红外制导关键技术发展分析。**上海航天**), 2021,**38**(3):163-170.

[3] Fang L, Wang X, Wan Y. Adaptable active contour model with applicationsto infrared ship target segmentation [J]. *Journal of Electronic Imaging*, 2016, **25**(4):041010.

[4] Zhang L, Wu B, Nevatia R. Pedestrian detection in infrared images based on local shape features[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007:1-8.

[5] Ge J, Luo Y, Tei G. Real-time pedestrian detection and tracking at nighttime for driver-assistance systems [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2009,**10**(2):283-298.

[6] SU Xiao-Qian, SUN Shao-Yuan, GE Man, *et al*. Pedestrian detection and tracking of vehicle infrared images[J]. *Laser & Infrared*(苏晓倩,孙韶媛,戈曼,等。车载红外图像的行人检测与跟踪技术。**激光与红外**),2012, **42**(8):949-953.

[7] ZHU Han-Lu, ZHANG Xu-Zhong, CHEN Xin, *et al*. Dim small targets detection based on horizontal-vertical multi-scale grayscale difference weighted bilateral filtering[J]. *J. Infrared Millim. Waves*(朱含露,张旭中,陈忻,等。基于横纵多尺度灰度差异加权双边滤波的弱小目标检测。**红外与毫米波学报**),2020, **39**(4):513-522.

[8] CAI Ru-Hua, YANG Biao, WU Sun-Yong, *et al*. Weak Targets Box Particle Labeled Multi-bernoulli Multi-target Detection and Tracking Algorithm [J]. *J. Infrared Millim. Waves*(蔡如华,杨 标,吴孙勇,等。弱目标箱粒子标签多伯努利多目标检测与跟踪算法。**红外与毫米波学报**), 2019,**38**(2):234-244.

[9] Choi Y, Kim N, Hwang S, *et al*. KAIST multi-spectral day/night data set for autonomous and assisted driving[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018,**19**(3):934-948.

[10] 2018. FREE FLIR Thermal Dataset for Algorithm Training. [Online].*Available*: https://www.ir.in/oem/adas/adas-

dataset-form.
[11] Socarrás Y, Ramos S, Vázquez D, *et al*. Adapting pedestrian detection from synthetic to far infrared images[C]// ICCV Workshops. 2013, 3.
[12] Ghose D, Desai S M, Bhattacharya S,*et al*. Pedestrian detection in thermal images using saliency maps[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 1-10.
[13] Devaguptapu C, Akolekar N, Sharma M, *et al*. Borrow from anywhere: Pseudo multi-modal object detection inthermal imagery[C]//Proceedings of theIEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019: 1029-1038.
[14] Dai X, Yuan X, Wei X. TIRNet: Object detection in thermal infrared images for autonomous driving[J]. *Applied Intelligence*, 2021, **51**(3):1244-1261.
[15] Krišto M, Ivasic-Kos M, Pobar M. Thermal object detection in difficult weather conditions using YOLO[J]. *IEEE access*, 2020, **8**:125459-125476.
[16] Song X, Gao S, Chen C. A multispectral feature fusion network for robust pedestrian detection[J]. *Alexandria Engineering Journal*, 2021, **60**(1):73-85.
[17] Du S, Zhang P, Zhang B, *et al*. Weakand occluded vehicle detection in complex infrared environment based on improved YOLOv4[J]. *IEEE Access*, 2021, **9**: 25671-25680.
[18] Wu Z, Wang X, Chen C. Research on light weight infrared pedestrian detection model algorithm for embedded Platform [J]. *Security and Communication Networks*, 2021, **2021**:1549772.
[19] Li S, Li Y, Li Y, *et al*. YOLO-FIRI: Improved YOLOv5 for Infrared ImageObject Detection [J]. *IEEE Access*, 2021, **9**:141861-141875
[20] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J].*arXiv preprint* arXiv:2004.10934, 2020.
[21] Yang J, Fu X, Hu Y, *et al*. PanNet: A deep network architecture for pan-sharpening [C] //Proceedings of the IEEE international conference on computer vision, 2017: 5449-5457.
[22] Han K, Wang Y, Tian Q, *et al*. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:1580-1589.
[23] He K, Zhang X, Ren S, *et al*. Spatialpyramid pooling in deep convolutionalnetworks for visual recognition [J]. *IEEE transactions on pattern analysis andmachine intelligence*, 2015, **37**(9):1904-1916.
[24] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems, 2012: 1097-1105.
[25] Simonyan K, Zisserman A. Very deepconvolutional networks for large-scale image recognition[J]. *arXiv preprint* arXiv:1409.1556, 2014.
[26] He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.
[27] Howard A G, Zhu M, Chen B, *et al*.Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. *arXiv preprint* arXiv:1704.04861, 2017.
[28] Zhang X, Zhou X, Lin M, *et al*. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2018:6848-6856.
[29] Sandler M, Howard A, Zhu M, *et al*.Mobilenetv2: Inverted residuals and linear bottlenecks [C]//Proceedings of theIEEE conference on computer vision and pattern recognition,2018:4510-4520.
[30] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network trainingby reducing internal covariate shift [C]//International conference on machine learning, PMLR, 2015:448-456.
[31] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. *arXiv preprint* arXiv:2004.10934, 2020.
[32] Huang Z, Wang J, Fu X, *et al*. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection [J]. *Information Sciences*, 2020, **522**: 241-258.