引用格式：ZHU Zijian, LIU Qi, CHEN Hongfen, et al. Infrared Small Vehicle Detection Based on Parallel Fusion Network[J]. Acta Photonica Sinica, 2022, 51(2):0210001
朱子健,刘琪,陈红芬,等.基于并行融合网络的航拍红外车辆小目标检测方法[J].光子学报,2022,51(2):0210001

基于并行融合网络的航拍红外车辆小目标检测方法

朱子健[1],刘琪[2],陈红芬[1],张贵阳[3],王福宽[4],霍炬[2]
(1 哈尔滨工业大学 航天学院，哈尔滨 150001)
(2 哈尔滨工业大学 电气工程及自动化学院，哈尔滨 150001)
(3 常熟理工学院 机械工程学院，江苏 苏州 215500)
(4 广西大学 机械工程学院，南宁 530004)

摘　要：提出了一种基于并行融合网络的航拍红外车辆小目标检测方法。以并行残差块搭建的网络作为主干网络，完成对目标高精度且强鲁棒性的识别和分类。在此网络的基础上进行特征提取与特征融合，提出了基于跨层连接的改进YOLOv3算法，充分利用底层的信息完成对红外车辆小目标的高精度检测与定位。最后，采用soft-NMS替代NMS来缓解目标重叠问题。实验结果表明，本文方法能够准确检测复杂运动背景下的红外车辆小目标，并在误报率较低的情况下达到较高的检测精度。其中虚警率仅有0.01%且漏检率仅有1.36%。

关键词：模式识别；红外图像；车辆检测；小目标检测；并行融合
中图分类号：TP391.4　　**文献标识码**：A　　**doi**：10.3788/gzxb20225102.0210001

0 引言

红外小目标检测在许多重要任务中起着至关重要的作用，例如现代防御、红外监视和空中交通管制等[1-3]。近年来，陆地车辆侦察技术是战场把控与监视能力建设的关键研究方向。随着低空空域的逐步开放，无人机、四旋翼等小型飞行器，通过搭载红外摄像头，可以隐秘的完成对地面车辆的采集与跟踪。然而，由于观察距离长，目标通常非常小弱，几乎没有或完全没有形状和纹理信息[4]。另一方面，由于地面环境的复杂性，车辆目标可能存在遮挡，重叠，模糊等特点。

因此快速而又准确检测地面车辆目标极具挑战性。此外，由飞行器的运动引起的背景运动也使得路面复杂场景中弱小目标的检测变得更加困难。在过去的几年里，许多针对无人机航拍图像中的车辆目标检测的算法逐步涌现出来。随着人工智能与神经网络的兴起，卷积神经网络被广泛运用于目标检测之中。针对路面车辆检测，JOSEPH Redmon提出的YOLO(You Only Look Once)[5]和HE Kaiming提出的Faster-RCNN[6]都能通过训练COCO数据集完成对明显车辆目标的高精度快速检测，但是在对微小车辆目标的检测任务中检测能力较为一般。基于特征融合的单发多盒检测器(Feature Fusion Single Shot Multibox Detector, FSSD)[7]的作者LI Zuoxin等指出，特征融合技术是解决微小目标识别的关键，他通过将不同尺度的不同层级联在一起形成特征层，运用连续下采样来生成新的特征金字塔，最后将此馈送到多盒检测器中预测最终的检测结果。通过这一思想，JOSEPH Redmon通过上采样构建特征金字塔，得到YOLOv3模型[8]，使得YOLO系列方法得到进一步优化发展。LIANG Xi等提出了FS-SSD(Feature Fusion and Scaling-based Single Shot Detector)[9]，他们使用平均池化操作添加反卷积模块的额外缩放分支以形成特征

基金项目：国家自然科学基金(No. 61473100)
第一作者：朱子健(1998—)，男，硕士研究生，主要研究方向为计算机视觉与深度学习。Email: hit_zzj@163.com
导师(通讯作者)：霍炬(1977—)，男，教授，博士，主要研究方向为光电信息处理。Email: torch@hit.edu.cn
收稿日期：2021-07-21；**录用日期**：2021-09-06

金字塔，并融入上下文分析来提高微小车辆检测精度。上述的方法在针对彩色图像且少遮挡的车辆小目标检测中能发挥出优异的特性，但是面对红外灰度图像，低分辨率、弱光源、强遮挡等等不利因素仍然会弱化检测准确率，造成虚警、漏检等等状况。并且对于众多的通用目标检测算法，都无法很好的移植到红外车辆小目标检测中，如Faster-RCNN等两步法在效率上不占优势，YOLO或是SSD系列等方法在未引入特征融合以及FPN技术前甚至无法胜任红外小目标的检测。

另一方面，飞行器捕获的车辆图像通常具有帧间连续的特点，因此对于红外小目标检测衍生出两类方法：单帧检测和序列帧检测。对于前者，PHIPIP Chen[10]考虑到红外图像中目标与背景的巨大差异，结合人类视觉注意力机制，提出了局部对比度测量（Local Contrast Method，LCM）。在此基础上也衍生出许多改进策略，如多尺度局部对比度测量（Relative Local Contrast Measure，RLCM）[11]、基于多尺度补丁的对比度测量（Multiscale Patch-based Contrast Measure，MPCM）[12]、双邻域梯度法（Double-Neighborhood Gradient Method，DNGM）[13]等等。然而这些方法容易受到杂波与噪声的干扰，从而导致检测精度不高。对于序列帧检测，通常利用相邻图像间的强相关性来进行检测。SLC（Spatial Local Contrast）利用空间对比度增强目标，TLC（Temporal Local Contrast）利用时间对比度增强目标，STLCF（Spatial-temporal local contrast filter）[14]则结合二者形成时空对比度滤波器。这些方法通过序列图像生成轨迹使得精度提升，但是他们忽略了运动的背景，时刻大幅度震动的无人机平台会导致背景运动加剧，从而产生大量虚警。在基于深度学习的红外小目标识别方面有许多学者给出一定的研究，WANG Huaichao[15]等考虑到了这一问题，采用深度学习策略在空间上初步定位目标信息，接着运用图匹配以及序列图像的光流图完成对虚警的剔除，增强了识别的可信度。但是，对于漏检的情况他们并没有做出优化。文献[16]介绍了一种基于CNN的用于红外图像船舶检测的目标检测器，它设计了一个TNet来生成合成目标，并将检测任务分为两个步骤，包括候选目标提取和候选识别。船舶检测的过程比较复杂，无法端到端的实现。文献[17]将红外小目标检测作为分割任务，设计了一个去噪和自动编码器网络，称为CDAE。CDAE可以端到端地检测红外小目标，但是存在误报的问题。文献[18]为了实现端到端检测，设计了一种高效的基于CNN的目标检测器ISTDet。上述方法大多针对于单一的红外空中目标，并且运用各种技术手段来减少虚警。然而对于地面目标，存在多目标且交错的场景，导致漏检的概率远远大于虚警。通常来说，过分的剔除虚警无疑会产生漏检，而通过强特征学习后能对漏检做出弥补，却又会导致虚警的产生。因此针对无人机航拍地面车辆小目标的检测，如何在漏检与虚警之间找到平衡，是提高检测鲁棒性与准确率的关键所在。

综合考虑上述问题，考虑到实际对车辆的拍摄情况是实时进行的，无法从后续图像来推测当前图像信息。因此本文采用单帧检测办法，提出一种端到端的红外车辆目标检测的算法，即基于并行融合网络的航拍红外车辆小目标检测方法。首先利用基于并行残差块的网络（Parallel Residual Net，PaRNet）。通过对并行残差块的重复堆叠完成主干网络PaRNet的搭建，并且在特征融合上充分抽取底层特征，从而实现对红外车辆小目标的高精度目标检测。

1 目标检测算法介绍

1.1 主干网络介绍

主干网络，作为目标检测框架中特征提取的模块，在可信度与鲁棒性上起着不可或缺的作用。通常来说，主干网络都是由用于分类的深度卷积神经网络演变而来[19]。如今，比较流行的深度卷积神经网络包括VGG[20]，ResNet[21]，GoogLeNet[22]和DenseNet[23]。这些网络都是基于模块化的思想，以卷积构建块为基础，我们所需要做的是找到最优的局部构造并在空间上重复它。

ResNet是深度神经网络领域重要的神经网络模型类型。众所周知，在加深神经网络层时，由于训练精度的降低，很难学习从输入到输出的直接映射。然而ResNet可以很好的克服这个问题[24]。ResNet引入了一种全新的网络结构，称为残差网络，它代表了一个前向反馈快捷网络[25]。将x定义为神经网络模型的输入。传统的神经网络模型倾向于学习训练数据的输入输出映射$H(x)$，而残差网络模型倾向于学习表示为$F(x)=H(x)-x$的映射函数。这样就避免了学习从x到$H(x)$的直接映射，同时学习了它们之间的差异。此差值与零更为相近，更能收敛到预期的效果。残差块模型的结构如图1所示。对于浅层的网络采用

BasicBlock的结构形式(由两层3×3卷积核组成),对于深层的网络则更适合采用BottleBlock的形式(由一层3×3卷积核以及两层1×1卷积核组成)。

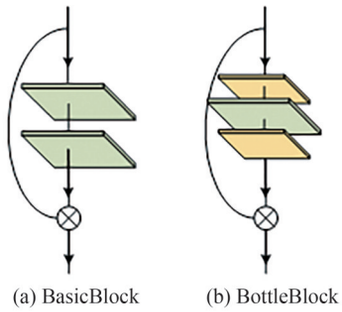

图1 残差块示意图
Fig.1 Residual block

ResNet引入了残差块residual,在检测精度上产生了飞跃,为网络深度奠定了基础。GoogLeNet引入了Inception块,运用并行运算提高模型效率,但是随着网络宽度和深度的不断增加,大量的3×3卷积核与5×5卷积核易于导致模型计算量骤增,产生梯度爆炸,在小数据集上有时会起到适得其反的效果[26-28]。本文在ResNet的基础上,提出了一种基于并行连接的卷积神经网络,PaRNet。它主要由连续卷积与一些并行的残差块穿插而成,在较为轻量级的网络上提升了模型泛化性能与准确率。

1.2 目标检测算法的介绍

在目标检测算法中,YOLO系列因快速性与准确性而广泛被使用。YOLOv3采用Darknet作为主干特征提取网络。在特征融合部分,YOLOv3提取多特征层进行目标检测,一共提取三个特征层。三个特征层位于主干部分Darknet53的中间层,中下层,底层。在获得三个有效特征层后,利用这三个有效特征层进行特征金字塔FPN层的构建,构建方式如图2所示。

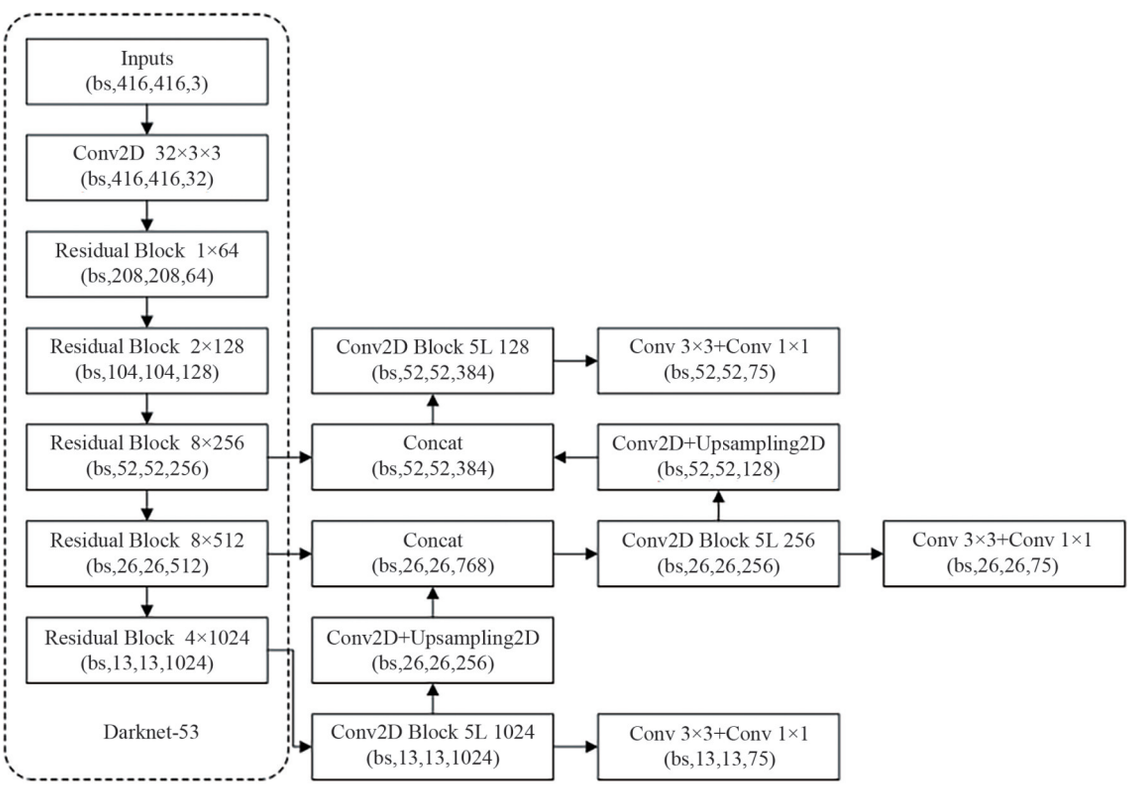

图2 YOLOv3主干网络与特征提取示意图
Fig.2 The backbone and feature map of YOLOv3

2 基于并行融合网络的目标检测算法

2.1 PaRNet主干网络的搭建

PaRNet模型如图3所示，首先采用多层连续卷积，这种连续的卷积结构是典型的VGG中的网络结构，用连续少量的接收器域（5×5和3×3窗口）以增强特征学习能力，使得特征图中单个元素的感受野变得广阔。接着采用局部响应规范化方法，来增强模型的泛化能力。然后采用基于并行的残差块来堆叠形成主干部分，这也是PaRNet中至关重要的一部分。基于并行的残差块结构如图4所示，输入部分被赋予两条路径，主干是通过传统的残差块块进行图像尺寸的压缩。分支则是通过3×3的最大池化层与1×1的卷积层对图像特征进行充分提取后，与主干进行通道维数的融合，从而获取通道维数上的递增。此结构包含以下两点贡献。

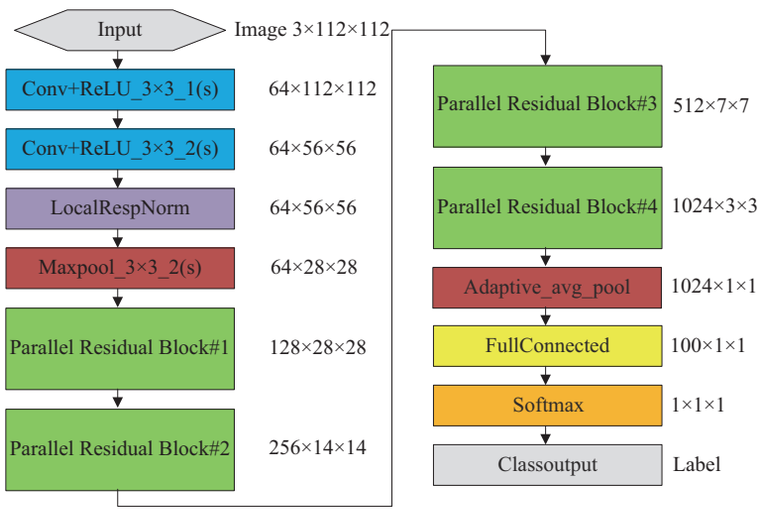

图3 PaRNet网络结构示意图

Fig.3 PaRNet structure diagram

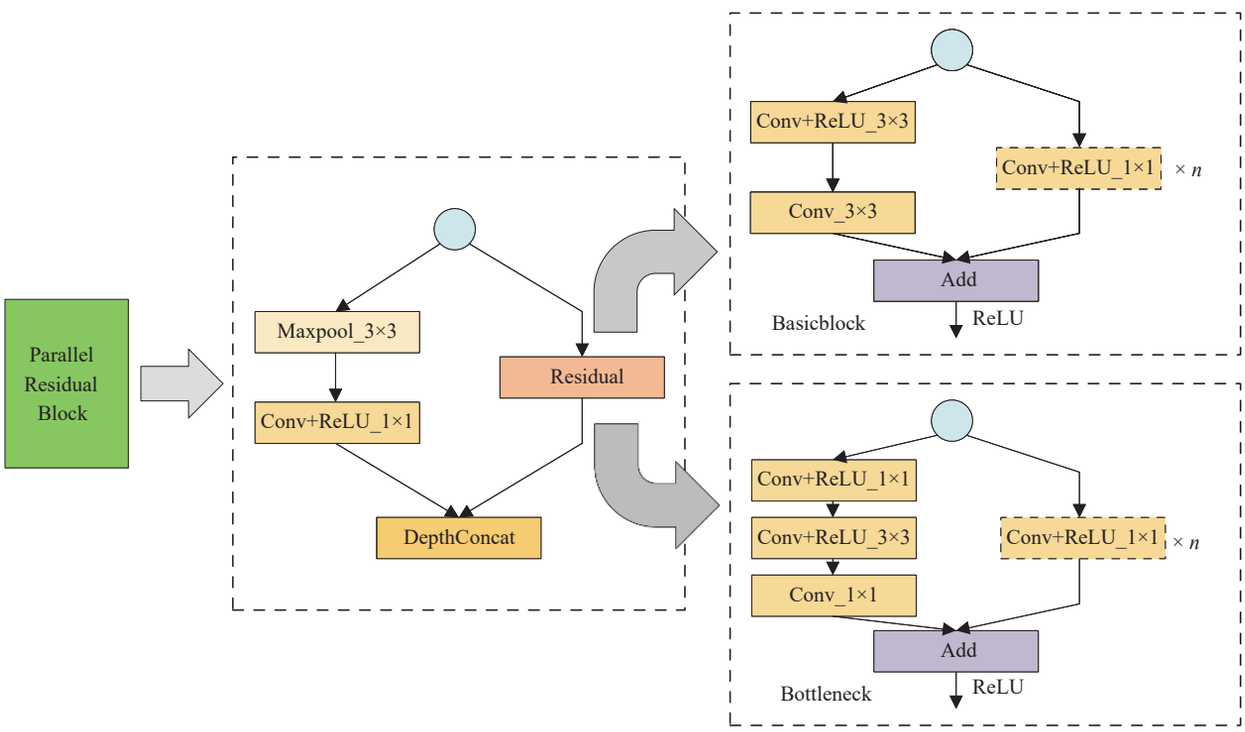

图4 并行残差块设计示意图

Fig.4 Parallel residual block designing diagram

2.1.1 通道融合策略

本结构中,有效深度的增加体现在图像尺寸的减小,通道的扩张则是由通道融合实现的。因此,某个特定尺度下图像的多方位特征可以得到充分提取。通常,随着网络尺寸的不断增加,过拟合的现象也越发严重,ResNet虽然采用了估计残差的形式,但是对于深度网络仍会出现梯度消失的状况。而并行的架构会将网络整体稀疏化,在优化过程中能极大程度的缓解梯度爆炸和梯度消失的问题。再则,并行架构提升了计算资源的利用率,浅层的网络可能可以达到与深层一样的效果,起到了轻量化的作用。

通道融合策略完成了底层特征与顶层的融合。ResNet通常在若干个残差块的最后一块中进行图像尺寸的对半压缩,而前面的所有部分均为同尺度的维数扩张以及感受野的扩大。如图5所示,由于采用3×3的卷积核,为了保持尺寸不变,必须设置填充。因此下层图像的每一块像素都会获取上层图像中2×2的感受野。以3块残差块为例,最下层中的每一像素获取到中间层2×2的感受野以及最顶层3×3的感受野。在获取顶层感受野时,其中的9个像素并不是同等待遇获取的,内圈的像素会得到多次的利用而四周的区域得不到充分使用。这种做法难免会丢失掉少量像素信息,而对于目标本身就不占太多像素的红外灰度图像,会导致训练准确度的下降。然后本文吸取GoogleNet的部分思想,在此模型基础上并联网络,实现顶层特征与底层特征直接融合。在此支路中,所有像素均同等考虑,从上层获取的特征一部分经少量网络后直接与底层特征相融合。因此丰富了图像的理解与表达,在针对小目标或是传统数据集时算法效能都有一定提升。

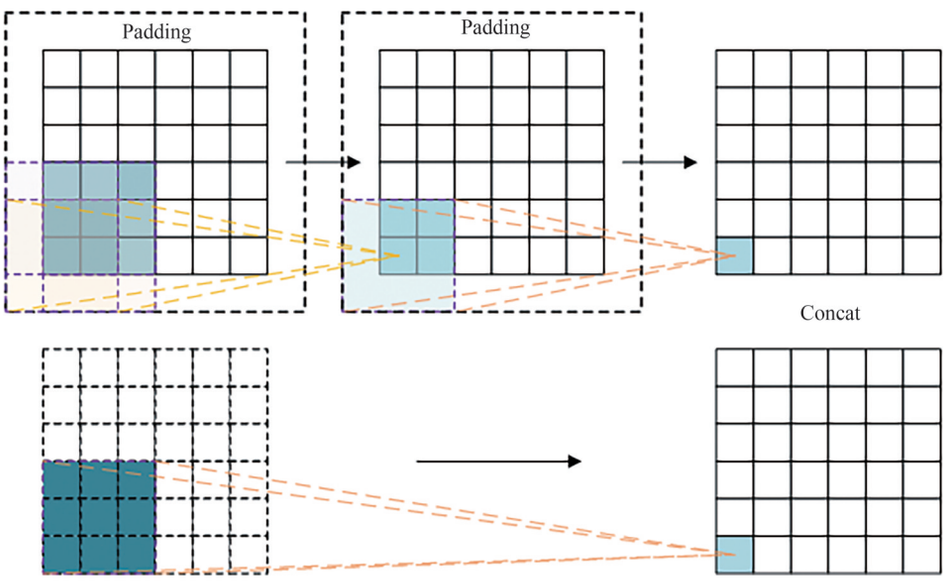

图5 通道融合策略感受野示意图

Fig.5 Schematic diagram of the receptive field of channel fusion strategy

2.1.2 3×3最大池化与1×1卷积

3×3最大池化用于缓解卷积层对位置的过度敏感性。它使模型可以抽取更广范围的特征并在一定程度上避免过拟合的发生。1×1卷积核,可以改变通道维数,起到不同通道间的信息交互,并且可以在保持特征层尺度输出的前提下大幅增加非线性特性。值得注意的是,3×3卷积,5×5卷积或者并联更多的分支网络,效果都适得其反。诸如3×3,5×5等大容量卷积核会耗费大量的参数,与残差块并联后会增大网络整体的负载。为了适应网络,学习率务必下调,这样反而破坏了ResNet原有的思想。

根据ResNet的网络搭建方式,我们将若干个基于并行的残差块进行串联,得到PaRNet-35和PaRNet-51,如图6所示,当选择“Basicblock”时为PaRNet-35,选择“Bottleneck”时为PaRNet-51。它们都是由4块并行残差块组成,不同的是,PaRNet-51网络层数更深,因此在搭建残差块时采用BottleNeck作为基础块。在通道上,采用[64, 128, 256, 512, 1024]递增的形式来构造。

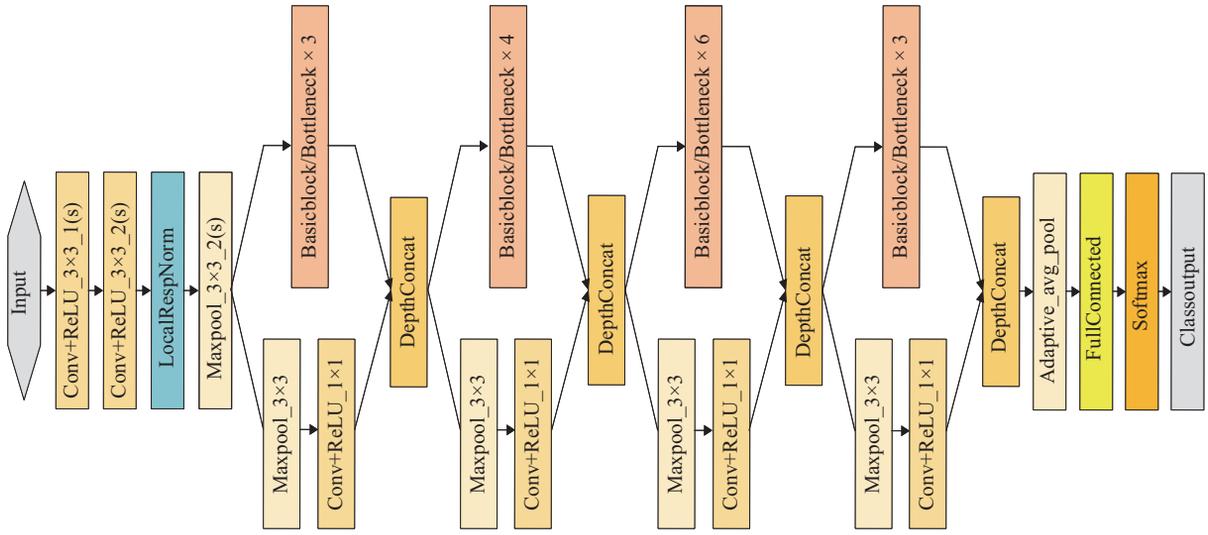

图6 PaRNet-35与PaRNet-51网络示意图

Fig.6 Architecture of PaRNet-35 and PaRNet-51

2.2 改进特征融合的YOLOv3算法

将PaRNet-51作为目标检测的主干网络，提出基于跨层连接的YOLOv3来搭建特征金字塔实现高精度微小目标检测。如第1节所述，已经有很多算法试图观察并充分利用金字塔特征。一般而言，针对微小目标的识别，通常需要更深层次的网络以及更加细腻的特征。因此，我们分别提取第2，3，4块并行残差块的输出特征，通过自下而上扩大图像尺寸的方式来进行特征层的融合。在传统的YOLOv3基础上，我们将上采样操作全部替换为反卷积，从而更加充分的还原微小目标的细节特征。

反卷积也被称为转置卷积，在生成对抗网络中第一次被提出[29]。反卷积其实就是卷积的逆过程。将底层的图像通过卷积操作还原到尺寸更大或是原始尺寸的图像，以此来反应图像中像素级别的特征[30-32]。较单纯的上采样相比，反卷积操作能挖掘到每个像素特征，还原更为精细的局部特征，如图7所示，因此更适用于对小目标的检测工作。然而反卷积很容易有不均匀的重叠，特别是当核尺寸不能被步长整除的时候，会产生不均匀的重叠从而导致棋盘格效应[33]。因此，我们在最后一层反卷积上设置步长为1来缓解此状况。

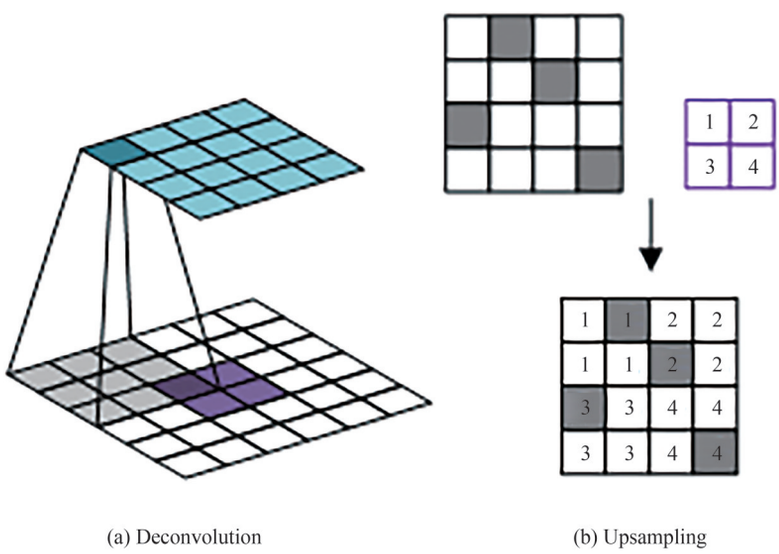

图7 反卷积与上采样

Fig.7 Deconvolution and upsampling

与此同时，最底层的输出特征，即通道数为2 048的特征层，经过重复卷积后，被融合到网络的每一层中，而不是只针对前一层。我们同样设置多层反卷积并在通道维数上进行逐层与跨层的连接。更进一步，

第一块并行残差块的输出特征通过卷积层与最大池化层作用后，同样与其它特征层进行有效融合，以保证特征的完备性与全面性。具体结构与如图8所示。

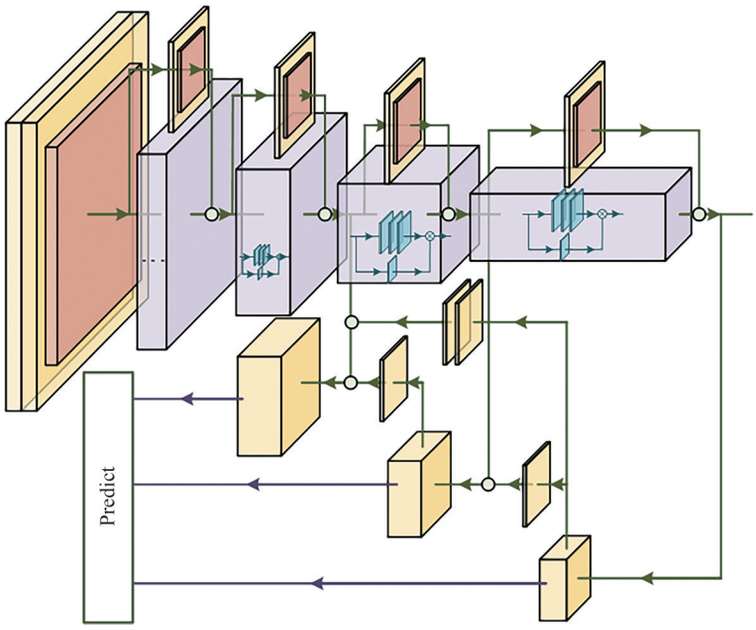

图8　基于跨层连接的特征融合示意图

Fig.8　Diagram of feature fusion based on cross-layer connection

反卷积模块由2个步幅为1的2×2反卷积层组成，每一层都由整流线性单元（ReLU）激活，然后进行批量处理归一化操作。与传统的双线性上采样方法相比，反卷积层在提高特征层分辨率基础上，更加注重图像细粒度的识别，有助于提高网络特征的代表性。输入的微小变化会导致损失函数的较大变化，从而使得梯度变大并缓解梯度消失的问题。反卷积模块将特征图放大后与浅层特征融合，通过连续卷积得到最终预测层。

3　实验与分析

3.1　PaRNet模型验证与评估

将PaRNet运用于公认的CIFAR-10数据集和CIFAR-100数据集。并运用通用标准去评估我们的网络，如TOP-1错误率，TOP-5错误率，准确率等等。最后通过与ResNet等流行网络的比较，验证本文方法的有效性。

3.1.1　CIFAR-10数据集

CIFAR-10数据集[34]是由50 000个训练图像和10 000个测试图像组成的10分类图像集。我们在训练集中进行实验，并在测试集中进行评估。在训练上，采用小批量梯度下降的方式进行训练，学习率设置为0.1，动量设置为0.1，批量设置为16，一共训练1 000个周期。实验环境配置为GTX2070 + Ubuntu 16.04 + PyTorch 1.6.1 + CUDA 10.1 + CUDNN 7.6.5 + Python 3.6.9。将训练结果与相似层数的ResNet和GoogLeNet进行了对比。

表1展示了PaRNet-35和PaRNet-51的模型参数Params、浮点计算量FLOPs、TOP-1错误率以及TOP-5错误率。我们将其与传统网络，如ResNet-50，ResNet-101以及GoogLeNet进行了对比。从参数量与计算

表1　不同网络在CIFAR-10上的表现

Table 1　Performance of different networks on CIFAR-10

	Params/(×10^6)	FLOPs/(×10^8)	TOP-1/%	TOP-5/%
GoogLeNet	6.6	3.9	13.01	0.72
ResNet-50	25.6	11	18.32	1.29
ResNet-101	44.5	20	16.71	1.04
PaRNet-35	22.8	4.4	14.19	0.67
PaRNet-51	31.7	10	13.38	0.64

量上不难看出，相较于相似层数的网络而言，PaRNet拥有更多的参数但却拥有更少的计算量，这得益于并行的优势，反映出与GoogLeNet网络类似的稀疏性。随着网络层数的加深，ResNet在准确率上会得到提升，而我们的网络吸取了这一优点。在前1 000次的训练中，无论是PaRNet-35还是PaRNet-51网络，都能在低网络负载的情况下得到更为精确的分类结果。

图9是上述网络在CIFAR-10上正确率随迭代周期的动态变化趋势。无论是在前期下降段速率以及震荡情况，还是后期的平稳程度与准确度，PaRNet都要优于传统的网络。一方面，由于加入了并行模块，网络得以分流从而导致在前期获得较大的下降梯度，因此PaRNet网络分类误差在整体上拥有更大的速降率。另一方面，PaRNet网络模型更具有平稳性与鲁棒性。在周期达到500之后时，PaRNet曲线更为平滑。

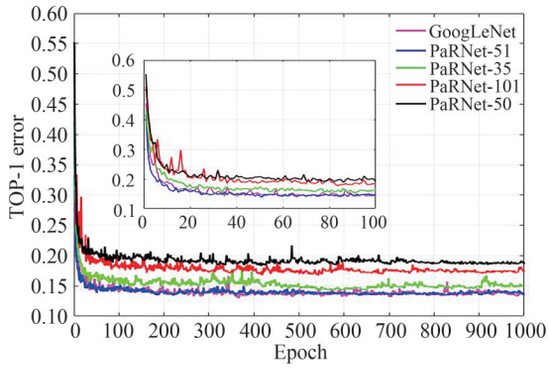

图9 CIFAR-10数据集上不同网络中TOP-1值随训练周期的变化图

Fig.9 The change of TOP-1 value in different networks with the training epochs on CIFAR-10

3.1.2 CIFAR-100数据集

CIFAR-100数据集[34]是由50 000个训练图像和10 000个测试图像组成的100分类图像集，是一个更具挑战性的数据集。运用和上述相同的方式进行训练与测试。

表2展示了在复杂分类任务的数据集上，PaRNet-35和PaRNet-51的TOP-1误差以及TOP-5误差。针对复杂的学习任务，PaRNet能表现出更高的准确率。

图10展示了上述网络在CIFAR-100上正确率随迭代周期的动态变化趋势。PaRNet综合了ResNet与

表2 不同网络在CIFAR-100上的表现

Table 2 Performance of different networks on CIFAR-100

	Params/($\times10^6$)	FLOPs/($\times10^8$)	TOP-1/%	TOP-5/%
GoogLeNet	6.6	3.9	41.97	17.27
ResNet-50	25.6	11	48.50	22.81
ResNet-101	44.5	20	48.09	22.00
PaRNet-35	22.8	4.4	41.08	17.19
PaRNet-51	31.7	10	39.67	15.86

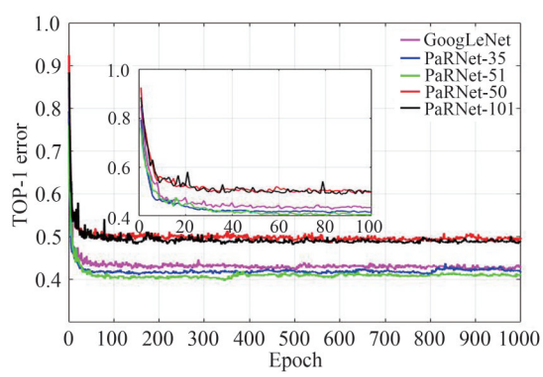

图10 CIFAR-100数据集上不同网络TOP-1值随训练周期的变化图

Fig.10 The change of TOP-1 value in different networks with the training epochs on CIFAR-100

GoogLeNet的优势，一方面，随着网络加深，获得的准确率收益也增大；另一方面，PaRNet也表现出了更快的收敛率与可信度。

3.2 目标检测算法的验证与评估

本数据集采用的红外数据为无人机对地拍摄的序列图像[35]。单帧图像分辨率为640×480、1个波段、8 bit位深。包括遮挡，错车，平台移动等等特殊复杂连续场景。每个典型场景对应1个数据段、共计64个数据段、每个数据段250张图像。如图11与图12所示为列举出的几种典型场景中的图片。

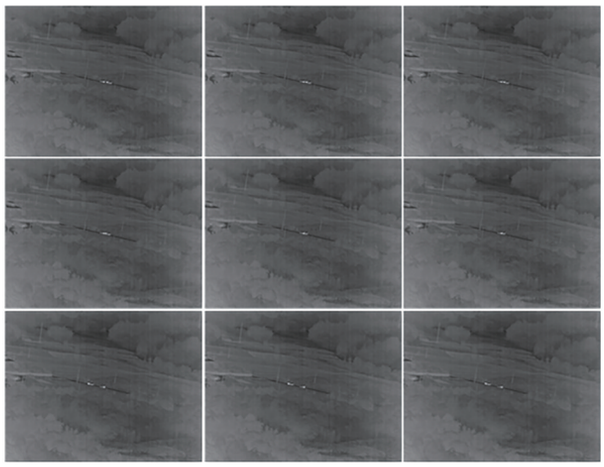

图11 错车场景部分帧图像

Fig.11 Part of the frame image of the meeting

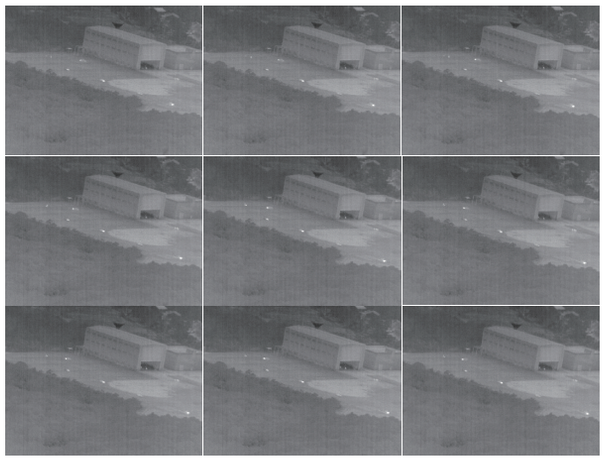

图12 多目标模糊场景部分帧图像

Fig.12 Part of the frame image of the multiple fuzzy targets

将48个文件夹，共计12 000张图像作为训练集，其余的4 000张图像作为测试集。分别运用经典的YOLOv3以及本文改进的方法进行训练与测试。在与上述相同的配置环境下进行实验。运用Adam（Adaptive Moment Estimation）算法进行优化，设计初始学习率为0.001，使用StepLR机制来动态减小学习率。由于网络结构的更改导致缺少对应的预训练权重，因此我们在运用YOLOv3算法训练时也去掉预训练权重。在训练方式上采用冻结训练，防止训练初期权值被破坏的同时加快训练速度。设置冻结训练世代为50次，非冻结训练世代为100次。对于计算检测框中传统的非极大抑制NMS（Non-Maximum Suppression）算法，我们将其改为文献[36]中的soft-NMS算法以解决车辆重叠程度大而导致的漏检问题。由于对于红外车辆目标而言，我们关注的是是否能正确无误检测车辆，而忽略给车辆分类的正确性问题，因此运用传统指标mAP对结果进行评价意义不大。虚警率（False Alarm Rate，FAR）和漏检率（Missed Detection Rate，MDR）由式（1）进行计算得到。

$$\begin{aligned} \mathrm{FAR} &= \mathrm{AlarmObj}/\mathrm{TotalObj} \\ \mathrm{MDR} &= \mathrm{MissedObj}/\mathrm{TotalObj} \end{aligned} \tag{1}$$

式中，AlarmObj为存在虚警的目标数量；MissedObj为存在漏检的目标数量；TotalObj为图像目标总数量。

兼顾检测的准确性与实时性，得到5个典型场景中的虚警率，漏检率以及FPS（Frames Per Second）三个指标如表3所示。由于通过测试，Fast-RCNN，SSD以及FSSD等传统算法由于特征融合不够深入或是对底层特征不敏感等劣势，在红外车辆小目标数据集上，几乎无法完成对目标的正常的检测。而YOLOv4，YOLOv5等改进只是锦上添花而并未针对主体网络架构进行大幅度改动。在此将本文方法与YOLOv3以及文献[15]中提出的方法（在此称为FCR-G）进行对比。

由表3不难看出，针对于红外拍摄的车辆小目标而言，本文的改进算法在精度上要优于传统的YOLOv3算法。首先，在时间上两者差异并不大，可以说都具有高效性。在精度上，本文的方法总体上达到了1.36%的虚警率以及0.01%的漏检率，优于传统的YOLOv3网络。而相比于FCR-G算法，两者在漏检与虚警上的差异并不大，但是由于无需后续的图匹配以及光流处理，本文的算法显得更为高效。

将本文的算法与现有的针对红外小目标检测的算法进行对比。如引言所说，将这类方法分为两类，第一类为未引入深度学习的一些基于局部特征对比的方法，另一类则为基于深度学习的方法，其中也包括一

表3 不同序列场景上评估指标值的计算

Table 3 Calculation of evaluation index values in different sequence scenarios

Scene sequence	FAR/%			MDR/%			FPS		
	YOLOv3	FCR-G	Ours	YOLOv3	FCR-G	Ours	YOLOv3	FCR-G	Ours
Seq 1:Platform movement	0.02	0.01	**0.01**	0.21	1.17	**0.15**	**33.9**	1.1	33.1
Seq 2:Target occlusion	0.02	**<0.01**	0.02	0.80	0.64	**0.20**	33.3	1.0	**33.5**
Seq 3:Overlapping targets	0.01	<0.01	**0**	1.36	**0.24**	1.20	**33.1**	1.2	32.9
Seq 4:Bad conditions	0	**<0.01**	0.01	2.64	2.24	**2.13**	**33.2**	1.1	32.8
Seq 5:Multiple objects	0	<0.01	**0**	3.15	**2.56**	3.12	**34.1**	1.1	33.2
Total test dataset	0.01	<0.01	**<0.01**	1.63	1.37	**1.36**	**33.5**	1.1	33.1

注:表3中加粗标记为更具优势者

些通用的目标检测的方法。表4为与基于局部特征对比的一些方法的比较。表5则为与基于深度学习方法的比较。在此处的漏检率与虚警率取所有场景的平均值。由表4得到,未采用深度学习的方法无论在效率还是精确度上都达不到要求。而从表5中可以看到,本文提出的算法兼顾了精度与效率,在漏检率、虚警率以及检测周期上都有一定提升。

表4 不同算法能力评估对比(未采用深度学习)

Table 4 Comparison of different algorithm (without deep learning)

Detector	FAR/%	MDR/%	FPS
MPCM	38.4	36.2	1.45
RLCM	4.60	46.2	1.44
DNGM	3.89	45.3	2.46
STLCF	3.76	19.4	1.52
Ours	<0.01	1.36	33.1

表5 不同算法能力评估对比(融合深度学习)

Table 5 Comparison of different algorithm (with deep learning)

Detector	FAR/%	MDR/%	FPS
YOLOv3	0.01	1.63	33.5
RefineDet	0.01	1.56	1.44
ISTNet	<0.01	1.35	32.4
FCR-G	<0.01	1.37	1.11
Ours	<0.01	1.36	33.1

综上所述,本文提出的PaRNet主干网络具有较高的准确率以及收敛速率,更重要的是其具有一定的鲁棒性,对于复杂环境下的模糊目标也具备一定的识别能力。另一方面,本文改进后的算法加入了跨层的连接与更深程度的特征融合,使得网络对底层小目标的理解更为充分。最后引入的soft-NMS算法也在一定程度上解决了车辆因错车而引起的重叠问题。但是目标在多目标模糊场景下的高漏检率还是不容忽视的。最后值得注意的是,对于航拍的红外车辆小目标检测而言,虚警的情况是比较少见的,而漏检却处处存在,如何进一步优化网络特征融合的模型是提高检测精度的关键所在。图13展示了检测效果图。

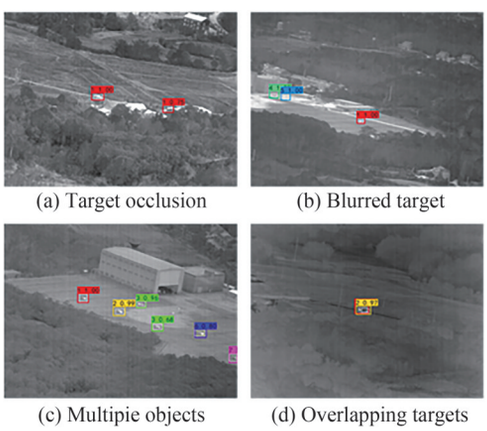

图13 检测可视化

Fig.13 Detection visualization

4 结论

在处理低空航拍的红外车辆小目标检测时,由于背景复杂且包含运动,传统检测方法无法得到准确的检测结果。本文提出了一种基于并行融合网络模型的检测方法,包括主干网络PaRNet的搭建设计以及对特征层的跨层连接。实验结果表明,在多个复杂的运动背景测试序列中,该方法能够以较低的虚警率准确检测目标,其中整体虚警率仅为0.01%,漏检率仅为1.36%。未来,可以在网络结构与训练中加入上下文信息,引入时序信息,通过在相邻帧之间引入卷积来进一步提升算法性能。

参考文献

[1] LI Li, LI Hui, LI Tian, et al. Infrared small target detection in compressive domain[J]. Electronics Letters, 2014, 50(7): 510-512.

[2] LIU Hongkang, ZHANG Lei, HUANG Huo. Small target detection in infrared videos based on spatio-temporal tensor model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(12): 8689-8700.

[3] LUO Kaiping. Space-based infrared sensor scheduling with high uncertainty: issues and challenges[J]. Systems Engineering, 2015, 18(1): 102-113.

[4] ZHU Yu, FANG Guanshou, ZHENG Bingbing, et al. Research on detection method of refined rotated boxes in remote sensing[J/OL]. Acta Automatica Sinica, 2020, doi: 10.16383/j.aas.c200261.
朱煜, 方观寿, 郑兵兵, 等. 基于旋转框精细定位的遥感目标检测方法研究[J/OL]. 自动化学报, 2020, doi: 10.16383/j.aas.c200261.

[5] REDMON J, DIVVALA S, GIRCHICK 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] REN Shaoqing, HE Kaiming, GIRSHICK R B, et al. towards real-time object detection with region proposal networks[J]. CoRR (2015), arXiv preprint arXiv:1506.01497.

[7] LI Zuoxin, ZHOU Fuxiang. FSSD: feature fusion single shot multibox detector[J]. arXiv preprint arXiv: 1712.00960, 2017.

[8] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[9] LIANG Xi, ZHANG Jing, ZHOU Li, et al. Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(6): 1758-1770.

[10] CHEN C L P, LI Hong, WEI Yantao, et al. A local contrast method for small infrared target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52: 574-581.

[11] HAN Jinghui, LIANG Kai, ZHOU Bo, et al. Infrared small target detection utilizing the multiscale relative local contrast measure, [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 15: 612-616.

[12] WEI Yaotao, YOU Xinge, LI Hong. Multiscale patch-based contrast measure for small infrared target detection[J]. Pattern Recognition, 2016, 58: 216-226.

[13] WU Lang, MA Yong, FAN Fan, et al. A double-neighborhood gradient method for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020,18(8): 1-5.

[14] DENG Lizhen, ZHU Hu, TAO Chao, et al. Infrared moving point target detection based on spatial-temporal local contrast filter[J]. Infrared Physics & Technology, 2016, 76: 168-173.

[15] WANG Huaichao, LI Haifeng, ZHOU Hai, et al. Low-altitude infrared small target detection based on fully convolutional regression network and graph matching[J]. Infrared Physics & Technology, 2021, 115: 103738.

[16] JIANG Bitao, MA Xiaofeng, LU Yao, et al. Ship detection in spaceborne infrared images based on Convolutional Neural Networks and synthetic targets[J]. Infrared Physics & Technology, 2019, 97: 229-234.

[17] SHI Manshu, WANG Huan. Infrared dim and small target detection based on denoising autoencoder network[J]. Mobile Networks and Applications, 2020, 25(4): 1476-1480.

[18] JU Moran, LUO Jiangning, LIU Guangqi, et al. ISTDet: An efficient end-to-end neural network for infrared small target detection[J]. Infrared Physics & Technology, 2021, 114: 103659.

[19] LIN Jingdong, WU Xinyi, CHAI Yi, et.al. Structure optimization of convolutional neural networks: a survey[J]. Acta Automatica Sinica, 2020, 46(1): 24-37.
林景栋, 吴欣怡, 柴毅, 等. 卷积神经网络结构优化综述[J]. 自动化学报, 2020, 46(1): 24-37.

[20] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[21] HE Kaiming, ZHANG Xianyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[22] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference

on Computer Vision and Pattern Recognition, 2015: 1-9.
[23] HUANG Gao, LIU Zhuang, MAATEN L V D, et al. Densely connected convolutional networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
[24] LIU Yinping, LI Junmin. Comparing the effectiveness of two convolutional neural networks methods on fault diagnosis[C]. 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), 2020:350-354.
[25] JI L, FU C, SUN W. Soft fault diagnosis of analog circuits based on a resnet with circuit spectrum map[J]. IEEE Transactions on Circuits and Systems, 2021, 68(7): 2841-2849.
[26] YANG Yu, GAO Xiaoguang, GUO Zhigao. Learning BN parameters with small data sets based by data reutilization[J]. Acta Automatica Sinica, 2015, 41(12): 2058-2071.
杨宇, 高晓光, 郭志高. 小数据集条件下基于数据再利用的BN参数学习[J]. 自动化学报, 2015, 41(12): 2058-2071.
[27] ZHANG Ting, Li Yujian, HU Haihe, et al. A gender classification model based on cross-connected convolutional neural networks[J]. Acta Automatica Sinica, 2016, 42(6): 858-865.
张婷,李玉鑑,胡海鹤,等.基于跨连卷积神经网络的性别分类模型[J]. 自动化学报, 2016, 42(6): 858-865.
[28] WANG Weining, WANG Li, Zhao Mingquan, et al. Image aesthetic classification using parallel deep convolutional neural networks[J]. Acta Automatica Sinica, 2016, 42(6): 904-914.
王伟凝, 王励, 赵明权, 等. 基于并行深度卷积神经网络的图像美感分类[J]. 自动化学报, 2016, 42(6): 904-914.
[29] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
[30] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1520-1528.
[31] LEVIN A, WEISS Y, DURAND F, et al. Understanding and evaluating blind deconvolution algorithms[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009: 1964-1971.
[32] REN Dongwei, ZHANG Kai, WANG Qilong, et al. Neural blind deconvolution using deep priors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 3341-3350.
[33] ODENA A, DUMONLIN V, OLAH C. Deconvolution and checkerboard artifacts[J]. Distill, 2016, 1(10): e3.
[34] Learning Multiple Layers of Features from Tiny Images[M]. Alex Krizhevsky, 2009.
[35] 第三届"空天杯"全国创新创意大赛初赛, 命题二, 赛道二, 光学目标识别, 初赛数据集[EB/OL]. https://www.atrdata.cn/#/customer/match/2cdfe76d-de6c-48f1-abf9-6e8b7ace1ab8/bd3aac0b-4742-438d-abca-b9a84ca76cb3?questionType=model.
[36] LV Xiaojun, XIANG Wei, LIU Yunpeng. Small object detection algorithm on UAV aerial images based on enhanced lower feature[J]. Application Research of Computers, 2021, 38(5): 1567-1571.
吕晓君, 向伟, 刘云鹏. 基于强化底层特征的无人机航拍图像小目标检测算法[J]. 计算机应用研究, 2021, 38(5): 1567-1571.

Infrared Small Vehicle Detection Based on Parallel Fusion Network

ZHU Zijian[1], LIU Qi[2], CHEN Hongfen[1], ZHANG Guiyang[3], WANG Fukuan[4], HUO Ju[2]
(1 *School of Astronautics*, *Harbin Institute of Technology*, *Harbin* 150001, *China*)
(2 *School of Electrical Engineering and Automation*, *Harbin Institute of Technology*, *Harbin* 150001, *China*)
(3 *School of Mechanical Engineering*, *Changshu Institute of Technology*, *Suzhou*, *Jiangsu* 215500, *China*)
(4 *School of Mechanical Engineering*, *Guangxi University*, *Nanning* 530004, *China*)

Abstract: The high-precision infrared small vehicle detection under disturbed background has a high application value. The existing small IR vehicle detection methods usually fail or cause a high probability of false alarm in the complex backgrounds, fuzzy targets and overlapping targets. A detection method base on parallel fusion network is proposed to solve the above problems. First, a network based on the parallel residual block is designed as the backbone to complete high-precision and robust detection of the targets. Then, an improving YOLOv3 algorithm based on cross-layer connection is established which can make full use of the underlying information to complete high-precision detection and positioning of small infrared vehicle targets. Finally, the soft-NMS is employed instead of the NMS to alleviate the problem of overlapping targets. Experimental results show that the method in this paper can accurately detect infrared small vehicles in complex moving backgrounds and achieve a high detection rate in the case of a low false

alarm rate. The false alarm rate is only 0.01% and the missed detection rate is only 1.36%. The PaRNet backbone network proposed in this paper has high accuracy and convergence rate, and more importantly, it has a certain degree of robustness, and also has a certain ability to recognize fuzzy targets in complex environments. On the other hand, our improved algorithm adds cross-layer connections and deeper feature fusion, so that the network has a fuller understanding of the low-level small goals. The finally introduced soft-NMS algorithm also solves the overlap problem caused by the overlapping cars to a certain extent.

Key words: Pattern recognition; Infrared image; Vehicle detection; Small object detection; Parallel fusion

OCIS Codes: 100.4996; 280.3640; 100.3008

Foundation item: National Natural Science Foundation of China (No. 61473100), National Natural Science Foundation of China (No. 61473100)