

文章编号: 1672-3392(2024)01-0010-08

基于重参数化的红外小目标检测网络

朱翊翔¹ 闵志方¹ 朱雪琼² 王翔¹

(1 华中光电技术研究所—武汉光电国家研究中心, 湖北 武汉 430223;

2 武汉设计工程学院信息工程学院, 湖北 武汉 430225)

摘要 红外小目标检测通常受制于较远的成像距离,使得提取目标特征成为了一种困难,如何增强目标的特征表达是近些年的主要方向之一。而过于复杂的特征表达会损失推理速度,这对于有实时性要求的红外小目标检测任务是不利的。通过使用重参数化技术结合领域中常用的残差网络作为特征提取网络,再使用额外注意力与通道注意力作为特征增强模块与特征融合模块,在数据集上取得了较好的结果。提出的模型在SIRST与IRSTD-1K数据集上分别取得了0.734与0.638的mIoU值,同时参数量和计算复杂度只有0.306 M与1.114 G FLOPs。该模型能够在推理阶段保持较少参数的同时拥有和其他领先的方法相近甚至领先的性能,在串行运行的环境上有着明显的优势。

关键词 红外小目标检测;深度学习;卷积神经网络;模型压缩;注意力机制

中图分类号 TN219 **文献标识码** A

1 引言

红外光是依靠热辐射成像的,可以穿透云雾,通常被应用于长距离捕捉目标,而较长的成像距离导致捕获的目标在图像中的占比过小。根据定义,目标在红外图像中占比小于0.15%且对比度差异小于15%的检测任务称为红外小目标检测任务。

在这种限制下,如何准确地将目标从背景中分离出来成为了主要研究问题。过去的几十年里对于红外小目标检测的研究可大致分为两类:模型驱动(Model-driven)与数据驱动(Data-driven)^[1,2]。

模型驱动方法可分为三类:基于图像滤波的方法、基于局部信息的方法、基于数据结构的方法。基于图像滤波的方法^[3,4]设计思想来源于信号处理,利用各类图像滤波器去除噪声或孤立

点,并依据设定的阈值来选取特殊区域如TopHat;基于局部信息的方法^[5,6]通过整合局部信息度量值来衡量目标与背景的差异,从而分离出目标点如多尺度块对比度(Multi-scale Patch-based Contrast Measure, MPCM);基于数据结构的方法^[7,8]对图像进行建模并进行分解,可得到干净的背景分量与目标分量如红外图像模型(Infrared Patch Image model, IPI)。

以上的研究主要集中于目标的特征表达。图像滤波类方法将灰度视作目标特征;局部信息法将局部区域的对比度量值作为特征;数据结构法利用图像的数学结构作为特征。模型驱动方法检测结果取决于工程师对应用环境的假设,泛化性较弱。相比传统的模型驱动方法,数据驱动方法学习了训练数据中的先验知识,具有更强的鲁棒性与准确率。

数据驱动的方法根据模型输出的类型可分为两个类别:基于检测的方法与基于分割的方

收稿日期 2023-09-13 收修改稿日期 2023-11-20

作者简介 朱翊翔(1998—),男,硕士研究生,主要研究方向为图像处理与计算机视觉。

通讯作者 朱雪琼(1983—),女,硕士,工程师,主要研究方向为图像处理及软件测试。

法。基于检测的方法输出结果为包含目标的预测框的坐标和尺寸,在数据集的标注与实时地运行上都有着明显的优势;基于分割类的方法输出结果是与输入图相同尺寸的置信度图。这种处理方法便于在原有模型驱动方法的基础上进行替换,直接提升红外系统的性能。

基于分割的方法中,以 U-net 和特征金字塔网络(FPN)为首的一系列 U 型网络能够在提取抽象语义信息的同时保留浅层的图像信息来进行特征融合,在各种侧重小目标检测的任务中取得了显著的成绩。ACM-FPN^[9]基于 FPN 提出了一种非对称特征融合模块,更好地结合浅层与深层特征。ALCNet^[10]在 MPCM 的基础上提出 DLC 特征,将局部信息法与深度网络相结合。AGPCNet^[11]使用多尺度注意力模块 CPM 与双向非对称特征融合模块 AFM 获得了明显的提升。ISNet^[12]使用可变形卷积进行特征融合并使用门控卷积补充边缘信息。MTUNet^[13]结合卷积网络与自注意力块来提取鲁棒的目标特征。这些方法在设计中针对特征给出了各种各样的解决方案,但是红外小目标检测是一项需要高实

时性的任务,这些解决方案在增添模块的同时带来了大量的额外计算量,因此需要找出一个能够保持较低计算量的同时兼顾性能的结构。

本文提出了一种轻量化的基于分割的红外小目标解决方案:REP-IRSTD (Reparameterization Infrared Small Target Detection)。我们的方案有三个显著优势:

- 1) 使用重参数化(Reparameterization)技术使得网络训练效果更好,并在推理时进行无损压缩,加快推理速度;
- 2) 使用外部注意力模块 EAB 来增强特征,使网络的表达能力更强;
- 3) 提出 GFM 模块,整合浅层特征的空间信息与深层特征的通道信息,加强目标的特征表达。

2 基本原理

2.1 基于分割的小目标检测网络 REP-IRSTD

REP-IRSTD 的网络结构如图 1 所示。

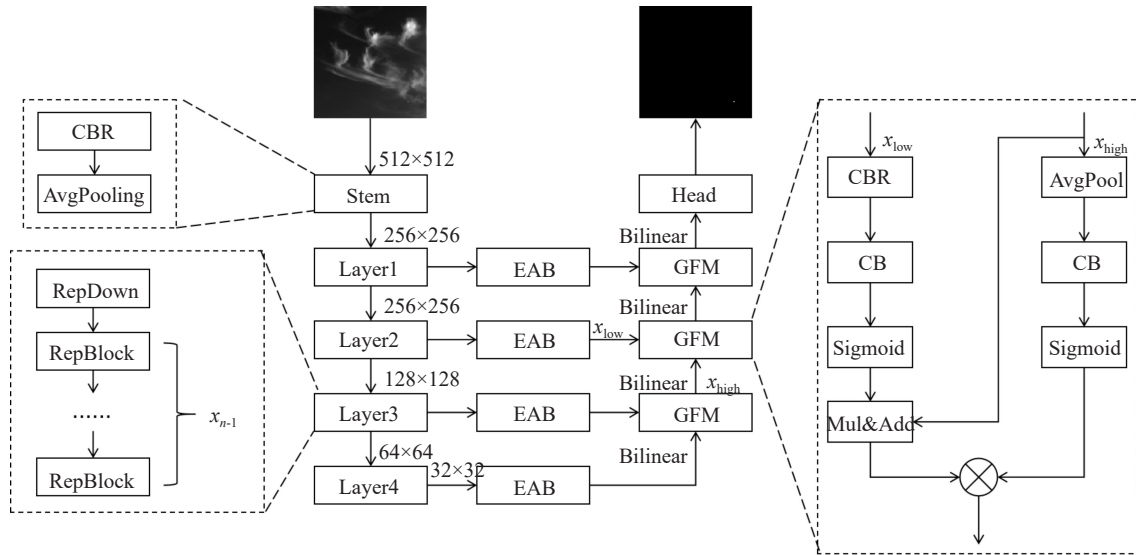


图 1 REP-IRSTD 网络结构,其中 CBR 分别代指卷积块、归一化块、线性整流单元(ReLu), x_{low} 、 x_{high} 分别表示浅层特征与深层特征

Fig. 1 Structure of Rep-IRSTD, CBR stands for convolution block, normalized block and linear rectifier unit (Relu) respectively, and x_{low} and x_{high} stand for shallow feature and deep feature respectively

网络经过由重参数化块(RepDown, RepBlock)组成的特征提取网络,将提取到的各层特征经过额外注意力模块 EAB 增强特征后传

输至全局特征融合模块(Global Feature module, GFM),全局特征模块会融合深层特征与浅层特征信息,使得网络在提取抽象信息的同时能够恢

复细节信息。

GFM 拥有两路输入图像,与基本的 FPN 加算或拼接操作不同,根据信息论的理论,在信息传递过程中信息会有所损耗,而检测任务中的空间信息更为重要,故我们选择提取低层次的空间注意力与高层次的通道注意力来对 FPN 的上采样模块进行改进,使得模型能够更加充分的融合

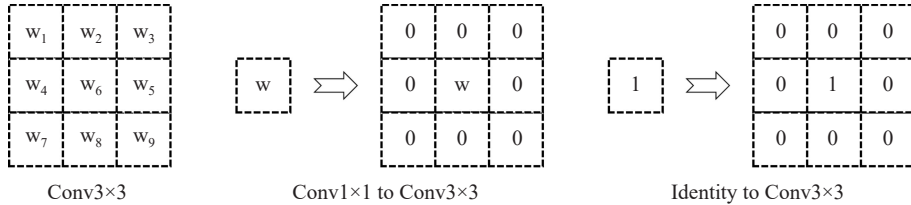


图 2 从左到右分别为尺寸为 3 的卷积核,尺寸为 1 的卷积核所等价的尺寸为 3 的卷积核, Identity 层等同于通道数相对应时的的卷积核

Fig. 2 From left to right, there are convolution kernels of size 3, convolution kernels of size 1 are equivalent to convolution kernels of size 3, and the Identity is equivalent to convolution kernels of size 3 when the number of channels is relative

而参数固定时一个核大小为 3 的卷积 (Conv)层与批量归一化层(BN)又能够通过数学推导重新转化为一个单独的尺寸为 3 的卷积层,

现有信息。

2.2 重参数化块(RepBlock)

丁霄汉等人在 RepVGG^[14]中应用了重参数化技术(Reparameterization),他们提出一个核大小为 1 的卷积层和直连层可以视为等同于一个核大小为 3 且参数固定的卷积核如图 2 所示。

从而使得原有的并行多通路卷积核结构转换为类似于 VGG 的直通路结构,流程如图 3 所示。

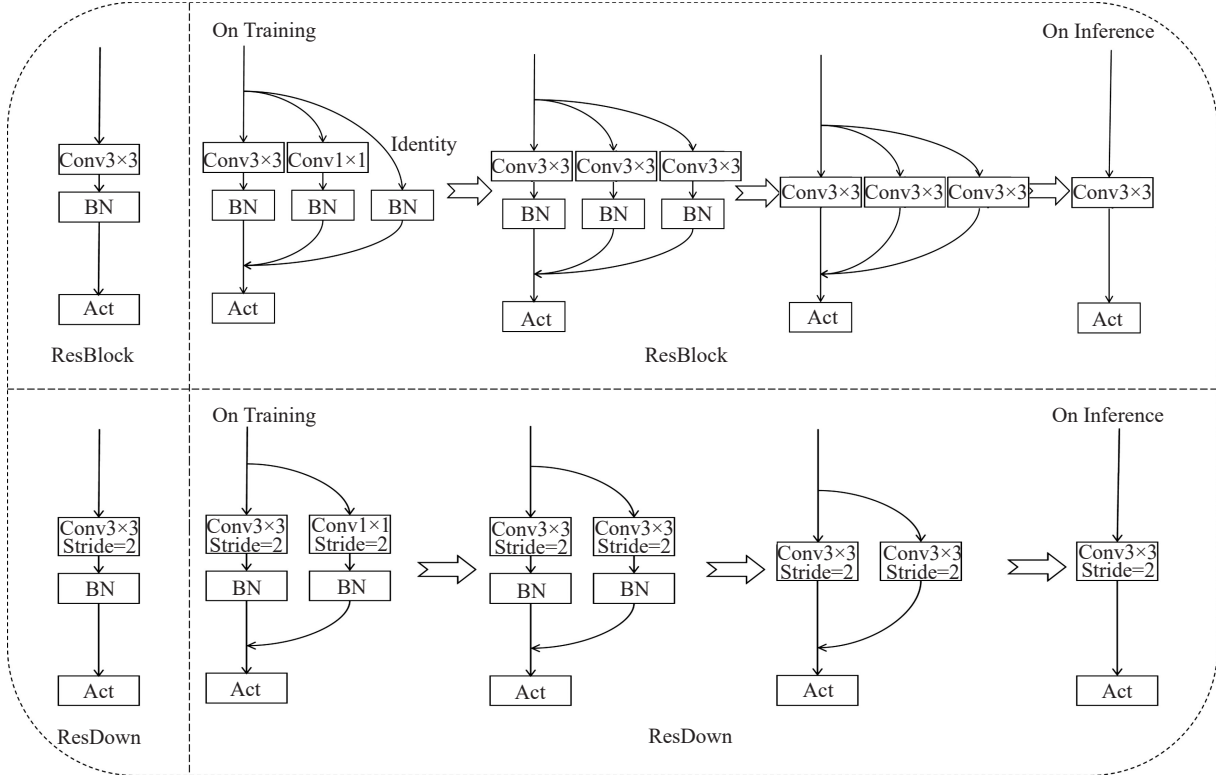


图 3 并行多通路卷积层结构转换为核大小为 3 的卷积核串联激活函数的结构

Fig. 3 Parallel multi-path convolution layer structure converted to the convolution kernel concatenated activation function structure with kernel size 3

无偏置的卷积核与批量归一化层在网络推理过程可以表示为式(1)、(2),其中 W 表示卷积层学习到的线性变换, $\gamma, \beta, \varepsilon, \sigma, \mu$ 为批量归一化层的可学习参数:

$$\text{Conv}(x) = W(x) \quad (1)$$

$$\text{BN}(x) = \gamma * \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (2)$$

两式合并可得串联的 ConvBN 层表达式(3):

$$\text{ConvBN}(x) = \gamma * \frac{W(x) - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta = \gamma * \frac{W(x)}{\sqrt{\sigma^2 + \varepsilon}} - \gamma * \frac{\mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta = W^*(x) + b^* \quad (3)$$

通过式(3)可推导出串联的卷积与归一化层可合并为一个带偏置的卷积核。综上一个并行多分支卷积归一化层可以合并为一个卷积层,将这种思想应用在残差网络的设计中,设计出如图 3 所示的两种重参数化块。

这种设计方法能够在训练时增加网络的复杂度与特征的多样性,使得网络的泛化性与性能能够获得明显的提升,而在实际应用时,网络会转化成为没有归一化层的残差网络,在推理时没有带来任何额外的计算量,这种无损模型压缩方法很适合红外小目标检测这种对实时性有要求

的视觉任务。在提升网络性能的同时不增加推理速度。

2.3 外部注意力块 (External Attention Block, EAB)

U 型结构能够融合低层次的特征与高层次的特征的信息,但是直接将提取的特征送入复杂的视觉任务中的效果不好,使用特征改进模块来增强信息的表达能力是红外小目标检测领域重要的研究方向之一,扩张卷积、DLC 特征、门控卷积等,在这里我们选取外部注意力机制。

注意力(Attention)机制有着极强的信息关联的能力,大量应用于建立上下文联系与增强特征。但是其中高复杂度的计算使得注意力机制的使用成为了难点。近些年来有着相当数量的实验针对注意力的复杂度进行优化,外部注意力机制(External Attention)就是其中低复杂度的计算方法之一。

我们基于额外注意力机制设计 EAB 模块如图 4 所示,首先对特征图进行信息整合,然后转换为索引向量形式,避免信息丢失构建逆瓶颈结构,通过标准层(Norm)后变换输入时的结构并作为残差结构的一部分与原特征相加来增强特征。

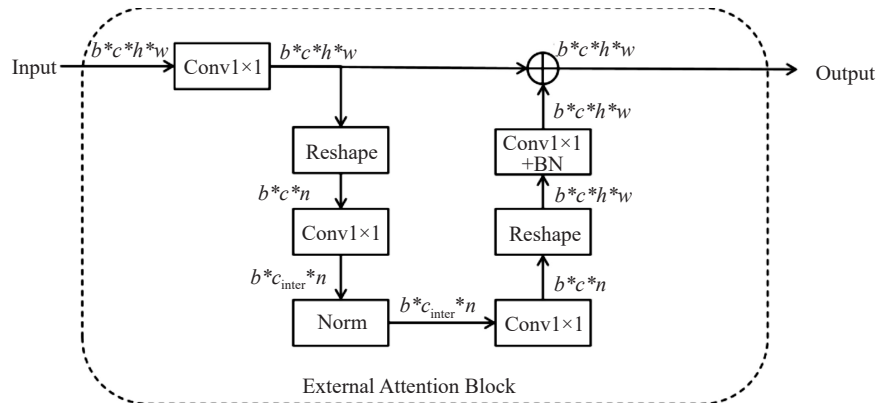


图 4 外部注意力块 EAB,图中 b, c, h, w, n 分别为批量大小、通道数、特征图的高、特征图的宽和索引向量数

Fig. 4 External attention block EAB, where b, c, h, w, n are batch size, number of channels, height of feature maps, width of feature maps and number of index vectors respectively

2.4 CBSLoss 激活函数

基于分割的红外小目标检测方法通常使用 SoftIoU 与 BCELoss 两种损失函数,而这两种损失函数都各自有自身的局限性,SoftIoU 的计算

集中于正样本的表现如式(4)所示,从而导致训练后期的驱动力不足, BCELoss 如式(5)所示,同时关注全局正负样本但是缺少对局部区域的关注,使得网络收敛结果不佳。式中 P_k, G_k 分别为

第 k 个输入的预测值和真实值。

$$L_{\text{softIoU}} = 1 - \frac{1}{n} \sum_{k=1}^n \frac{P_k \cap G_k}{P_k \cup G_k} \quad (4)$$

$$L_{\text{BCE}} = -\frac{1}{n} \sum_{k=1}^n (G_k \log P_k + (1 - G_k) \log (1 - P_k)) \quad (5)$$

针对以上问题,我们采用 CBSLoss^[15],结合了 SoftIoU 与 BCE 两种损失函数的特性,表达式如式(6)所示,CBSLoss 的导数如式(7)所示:

$$L_{\text{CBS}} = L_{\text{softIoU}} + \ln(1 + 100L_{\text{BCE}}) \quad (6)$$

$$\frac{\partial L_{\text{CBS}}}{\partial x} = \frac{\partial L_{\text{softIoU}}}{\partial x} + \frac{100}{1 + 100L_{\text{BCE}}} * \frac{\partial L_{\text{BCE}}}{\partial x} \quad (7)$$

通过 CBSLoss 的导数可以看出,在训练初期梯度较大的时候两边的系数都接近于 1,在训练后期梯度较小的时候,SoftIoU 饱和失去驱动力,而 BCELoss 的系数会变大成为梯度主要成分,在梯度较小时依旧能够为网络的训练提供驱动力。

3 实验及结果分析

3.1 实验设置

环境设置方面,所有的实验结果都基于 Python3.8、Cuda11.7 和 Matlab2022a。硬件配置

为 NVIDIA Geforce RTX3080 laptop GPU 及 AMD R9 5900HX CPU。

在网络训练方面,我们选择 Adagrad 优化器且衰减系数为 0.000 4,网络的学习率为 0.1,在每一个迭代次数后乘一个 $(1 - \frac{\text{iter}}{\text{total_iter}})^{0.9}$ 来使得学习率衰减至 0。训练时输入图像尺寸为 512,批量大小为 8 且训练周期为 500。

实验设置方面,我们选择 NUAA-SIRST 与 IRSTD-1K 作为实验数据集,选择 mIoU 和 F1-score 作为性能指标,选择 CPU 上的推理速度和模型的参数量(Parameters)与复杂度(FLOPs)作为速度指标。

3.2 对比实验结果

我们选择图像滤波方法 TopHat、局部对比度方法 MPCM、数据结构方法 IPI 和深度学习方法 FPN、AlcNet、AgpcNet 进行性能对比,其中深度学习按照原文设置^[10,11]都为三层特征提取,通道数分别为 [16, 32, 64]、[16, 32, 64] 和 [64, 128, 256]。我们的方法 Rep-IRSTD 的四个特征提取层(layer1-4)的通道数量为 [8, 16, 32, 64],且每一层都包含三个重参数化块。各种方法在数据集上的表现如表 1 所示。

表 1 各种红外小目标检测方法在数据集上的表现

Tabel 1 Performance of IRSTD methods on the datasets

	NUAA-SIRST		IRSTD-1K		Speed Metrics		
	mIoU	F1-score	mIoU	F1-score	Cost	Params	FLOPs
TopHat	0.086	0.163	0.060	0.112	0.005 s	—	—
MPCM	0.123	0.219	0.103	0.182	0.111 s	—	—
IPI	0.339	0.506	0.015	0.031	52.71 s	—	—
FPN	0.659	0.795	0.586	0.793	0.103 s	0.374M	0.948G
ALCNet	0.690	0.816	0.589	0.741	0.626 s	0.373M	3.618G
AGPCNet	0.632	0.777	0.595	0.746	1.331 s	12.45M	43.26G
OURS	0.734	0.846	0.638	0.778	0.103 s	0.306M	1.144G

通过对比可以看出,本文提出的方法提升了性能,同时保持着较少的推理延时和计算量。

3.3 实验可视化与分析

为了得到更直观的对比结果,我们选取

SIRST 数据集的验证集里的一些未参与训练的场景(a)~(e)作为实验的可视化结果,实验结果如图 5 所示。

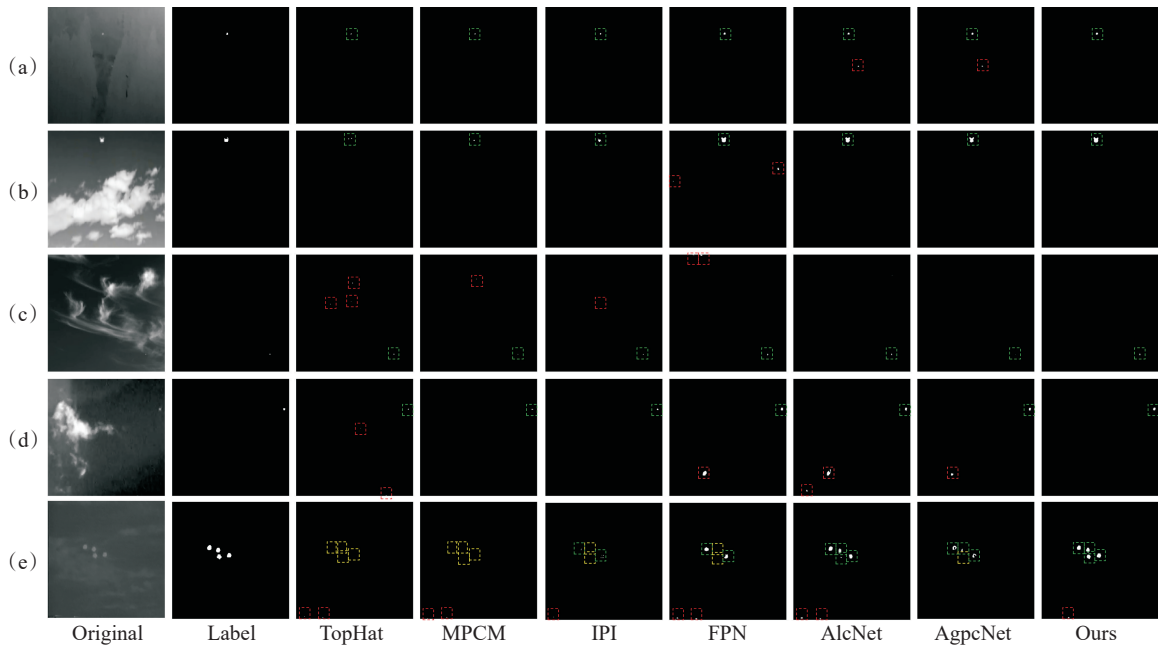


图 5 可视化结果,其中被绿框包围的是真实目标,被红框包围的是虚警目标,被黄框包围的是漏检目标

Fig. 5 Results of experiment, surrounded by the green box is the real target, surrounded by the red box is the false alarm target, surrounded by the yellow box is the missed target

图 5(a)、5(b)、5(d)三个场景中的目标处于一个孤立的位置,且目标与背景有明显的差异。针对这种应用场景,模型驱动方法都有着较好的处理效果,但是实际的处理结果只能得到目标的大致方位,在滤波或分解的过程中丢失了大量的图像信息,导致最终检测的目标只有所剩无几的几个像素点。当面临图 5(c)这种背景混乱的场景时,三种模型驱动方法都出现了虚警目标,不平滑的背景云层中几处尖锐的亮度变化误导了模型将其识别为目标。面对图 5(e)场景时,距离较近的多目标成为了难点,这是因为模型驱动方法对于目标的识别直接受制于步长等超参数的设定,在面对不同场景时固定的超参数会导致不理想的结果。

不同于传统模型驱动方法提取的少量目标点,深度学习方法在目标的提取上有着明显的优势,通常能够将目标分离出来。在图中的五个场景中,深度学习方法提取的目标明显更接近于真实标签。

FPN 作为实验所采用的骨干网络,在拟合能力上有着明显的不足,在五个场景上都出现了虚警的问题,不充足的特征导致 FPN 倾向于提取亮度与背景有明显差异的像素团。面对这个问

题 AlcNet 使用了 DLC 特征使得特征图的计算更集中于亮度差值,AgpcNet 使用注意力机制联系上下文,本文的方法使用外部注意力来增强特征,而重参数化技术带来的更加平滑的梯度与 CBSLoss 带来的低梯度条件下的驱动力合理使得我们的模型更好地拟合数据集,检测的目标更加接近真实目标。

3.4 消融实验

这部分将本文所提的模块进行比较,实验基于 NUAA-SIRST 数据集。首先是针对骨干网络的比较,与相同参数的残差块进行比较如表 2 所示,表中 ResBlock 是相同参数的残差块,Rep (Original) 是重参数化前的重参数化块,Rep (Reparam) 是重参数化后的重参数化块。

表 2 文中所使用重参数化块与残差块的实验对比
Tabel 2 Experimental comparison between the reparameterization block and the residual block used in this paper

	mIoU	F1-score	Speed	Params	FLOPs
ResBlock	0.702	0.826	0.112 s	0.307M	1.168G
Rep(Original)	0.734	0.846	0.147 s	0.339M	1.308G
Rep(Reparam)	0.734	0.846	0.103 s	0.306M	1.144G

通过表格可以看出,在训练阶段,多分支结构提升了模型的性能,但是带来了更多的参数与计算复杂度;在推理阶段,卷积层的参数与归一化层的参数得到了固定,可以利用无损压缩的重参数化技术来减少参数量,重参数化后的模型保持了训练时的性能,相比使用残差块堆叠的网络不仅获得了性能提升还带来了更快的推理速度。

为了体现本文使用的特征改进外部注意力模块 EAB 的作用,我们选取几种 SOTA 方法使用的特征改进模块来进行实验与比较,分别为无处理、扩张局部注意力 Dilated local contrast (DLC)、扩张卷积 Dilated Convolution(DC)。实验除了特征改进模块外,其他模块完全相同。结果如表 3 所示。

表 3 文中所使用特征改进块与其他特征改进方法的比较

Tabel 3 Comparison of feature improvement block used in this paper with other feature improvement methods

	mIoU	F1-score	Speed	Params	FLOPs
No Conv	0.712	0.831	0.071 s	0.280M	0.985G
DC	0.703	0.826	0.082 s	0.332M	1.158G
DLC	0.707	0.828	0.131 s	0.280M	0.981G
EAB	0.734	0.846	0.103 s	0.306M	1.144G

根据实验结果可以看出我们使用的外部注意力模块 EAB 相比别的模块有明显的性能提升,随着网络深度与下采样倍数提升,类似于 DC 的扩张卷积计算方式在扩张算子的空隙中丢失了信息,导致特征增强效果不如建立像素间关联性的 EAB。

4 结 论

本文提出了一种基于重参数化技术的 U 型网络,重参数化技术为网络的训练带来了更好的效果且能够通过合并卷积与归一化层来降低推理耗时。同时本文结合了外部注意力来增强特征,并使用通道信息来加强特征融合,取得了较为明显的性能提升。在 NUAA-SIRST 数据集上取得了 0.734 的 mIoU 与 0.846 的 F1-Score,同时模型参数量与浮点数计算量只有 0.306M 与 1.144G。实验结果表明,重参数化技术切实有效的在不增加推理延迟的同时提升了网络的性能,在 CPU 等串行计算的环境中有着明显的优势。相信在以后的研究中,重参数化技术会成为保证红外小目标检测任务实时性的重要手段之一。

参 考 文 献

- [1] 刘征, 杨德振, 李江勇, 等. 红外单帧弱小目标检测算法研究综述[J]. 激光与红外, 2022, 52(2): 154-162.
LIU Z, YANG D, LI J, *et al.* A review of infrared single-frame dim small target detection algorithms[J]. Laser & Infrared, 2022, 52(2): 154-162.
- [2] ZHAO M, LI W, LI L, *et al.* Single-frame infrared small-target detection: A survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(2): 87-119.
- [3] ZENG M, LI J, PENG Z. The design of Top-Hat morphological filter and application to infrared target detection[J]. Infrared Physics & Technology, 2006, 48(1): 67-76.
- [4] BAE T W, LEE S H, SOHNG K I. Small target detection using the bilateral filter based on target similarity index [J]. IEICE Electronics Express, 2010, 7(9): 589-595.
- [5] CHEN C L P, LI H, WEI Y, *et al.* A local contrast method for small infrared target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(1): 574-581.
- [6] WEI Y, YOU X, LI H. Multiscale patch-based contrast measure for small infrared target detection [J]. Pattern Recognition, 2016, 58: 216-226.
- [7] GAO C, MENG D, YANG Y, *et al.* Infrared patch-image model for small target detection in a single image[J]. IEEE Transactions on Image Processing, 2013, 22(12): 4996-5009.
- [8] ZHANG T, PENG Z, WU H, *et al.* Infrared small target detection via self-regularized weighted sparse model[J].

- Neurocomputing, 2021, 420: 124-148.
- [9] DAI Y, WU Y, ZHOU F, *et al.* Asymmetric contextual modulation for infrared small target detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2021: 949-958.
- [10] DAI Y, WU Y, ZHOU F, *et al.* Attentional local contrast networks for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11): 9813-9824.
- [11] ZHANG T, LI L, CAO S, *et al.* Attention-guided pyramid context networks for detecting infrared small target under complex background[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023: 1-13.
- [12] ZHANG M, ZHANG R, YANG Y, *et al.* ISNet: Shape matters for infrared small target detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 867-876.
- [13] WU T, LI B, LUO Y, *et al.* MTU-Net: Multilevel TransUNet for space-based infrared tiny ship detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.
- [14] DING X, ZHANG X, MA N, *et al.* RepVGG: Making VGG-style ConvNets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, 2021: 13728-13737.
- [15] CHEN G, WANG W, TAN S. IRSTFormer: A Hierarchical Vision Transformer for Infrared Small Target Detection [J]. Remote Sensing, 2022, 14(14): 3258.

Infrared Small Target Detection Network Based on Reparameterization

ZHU Yi-xiang¹ MIN Zhi-fang¹ ZHU Xue-qiong² WANG Xiang¹

(¹ Huazhong Institute of Electro-Optics—Wuhan National Laboratory for Optoelectronics, Wuhan 430223, China;

² Wuhan Institute of Design and Science, School of Information Engineering, Wuhan 430205, China)

Abstract Infrared small target detection is usually limited by a long imaging distance, which makes it difficult to extract target features. How to enhance target feature expression is one of the main research directions in recent years. However, too complex feature representation will lose the speed of inference. In this paper, we use reparameterization technology and residual network as feature enhancement module and feature fusion module, and achieve good results on the datasets. On SIRST and IRSTD-1K datasets, the proposed method achieves 0.734 and 0.638 mIoU, while having only 0.306M and 1.114G FLOPs in parameter number and computational complexity. Our model can maintain fewer parameters in the inference stage while having performance similar to or even leading other leading methods, which has obvious advantages in a serial environment.

Key words infrared small targets detection; deep learning; convolutional neural networks; model compression; attention mechanism