

## J-MSF: 一种新的多通道多尺度红外弱小目标检测算法

王国刚<sup>1</sup>, 孙召进<sup>1</sup>, 刘云鹏<sup>2</sup>

(1. 沈阳化工大学 信息工程学院, 辽宁 沈阳 110142;  
2. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016)

**摘要:** 针对经典的基于深度学习的红外弱小目标检测算法存在目标信息在高层感受野消失导致无法检出的问题, 提出一种新的基于多通道多尺度特征融合的红外弱小目标检测算法 (J-MSF)。首先, 该算法提出了一种新的多通道 JAnet 结构, 基于此结构搭建了主干特征提取网络; 其次, 设计了下降门限式特征金字塔池化结构 (DSPP), 并提出了多尺度融合检测策略; 最后, 设计了高斯损失优化函数。实验结果表明, 所提出的算法在“地/空背景下红外图像弱小飞机目标检测跟踪数据集”上的检测效果与 YOLOv3、YOLOv4 算法对比, 检出率、整体 AP 值分别提升 9.07%、9.89% 和 1.67%、3.16%, 提出算法优于目前主流的检测算法, 体现出了良好的鲁棒性和适应性, 可以有效的应用于红外弱小目标的检测。

**关键词:** 红外弱小目标; JAnet; 多通道; 多尺度; YOLO; 高斯; 池化

**中图分类号:** TP219; TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20210459

## J-MSF: A new infrared dim and small target detection algorithm based on multi-channel and multiscale

Wang Guogang<sup>1</sup>, Sun Zhaojin<sup>1</sup>, Liu Yunpeng<sup>2</sup>

(1. College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;  
2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China)

**Abstract:** An novel infrared dim and small target detection algorithm, called J-MSF, based on multi-channel and multi-scale feature fusion was proposed, which solved the problem that the classical infrared dim and small target detection algorithm based on deep learning cannot detect because the target information disappeared in the upper receptive field. Firstly, a new multi-channel Janet structure was proposed to design the J-MSF backbone extraction framework. Secondly, a descending threshold feature pyramid pooling structure (DSPP) was exploited, and a multi-scale fusion detection strategy was conducted. Finally, the Gauss loss optimization function was designed. The experimental results show that the recall rate and the AP value of the proposed algorithm are improved by 9.07%, 9.89% and 1.67%, 3.16%, respectively, compared with those of YOLOv3 and YOLOv4 algorithms in "a dataset for infrared detection and tracking of dim and small aircraft targets underground/air background". The proposed algorithm can be effectively applied to infrared dim and small target detection, shows good robustness and adaptability, and is better than the state of the art algorithms.

**Key words:** dim and small; JAnet; multi-channel; multi-scale; YOLO; Gaussian; pooling

收稿日期: 2021-07-06; 修订日期: 2021-08-05

基金项目: 国家重点研发计划 (2018FYB1700200)

作者简介: 王国刚, 男, 教授, 硕士生导师, 博士, 主要从事图像信息处理、工业自动化优化与控制, 目标检测、视觉跟踪和模式识别方面的研究。

## 0 引言

红外弱小目标及小目标检测是计算机视觉中的一个关键问题,因为目标上几乎没有可利用的信号,一般小目标于文中研究的红外小目标相比,其内在特征的稀缺性和背景复杂性有所不同,红外弱小目标的局部信噪比低。一般视觉任务中的小物体通常占图像区域的 1% 左右,而红外弱小目标的像素可能只占图像区域的 0.1% 或者更少。

红外弱小目标检测是指通过红外系统对打击目标进行识别检测和跟踪,主要应用在预警技术、导弹精确制导、海上巡逻预警监测等领域。如何在信噪比低、红外目标形状不规则及目标在像平面上像素占比较少等限制条件下能够精准、快速、稳定地检测出红外弱小目标是对武器的系统信息化和智能化的重要体现。

目前,红外弱小目标检测的传统算法中基于滤波的算法有:基于最恰对比度显著性、背景抑制中的二维空域廓线法等<sup>[1-2]</sup>,通过滤波的方式增强目标特征和抑制背景信息,总体计算量小,但检出率较低。基于人类视觉系统的方法有:基于局部对比度的算法、基于低秩稀疏恢复等<sup>[3-4]</sup>,通过人类感知特性对图像局部差分形成显著化目标特征,准确率高,但存在建模过程复杂,泛化能力较差等问题。红外图像数据集背景和特征多样化,使得传统检测算法必须与数据集高度匹配,且传统数学建模算法设计过程复杂,泛化性低,使得检测红外弱小目标成本较高。通过深度卷积神经网络识别检测红外弱小目标具有泛化能力强、建模过程更智能、鲁棒性更强等优点,运用深度学习建立通用性较强的红外检测系统成为越来越多学者的研究方向。

深度学习在红外弱小目标检测的研究主要有:赵琰等人使用 YOLOv3 的方法在自建数据集上进行检测<sup>[5]</sup>,但是在公开数据集上准确率较低;冯小雨等人使用 Faster RCNN 进行检测<sup>[6]</sup>,实时性低,且准确率有待提高。总体而言,深度学习在检测红外弱小目标方面需要解决以下问题:(1)卷积神经网络结构在对局部信噪比较低的图像进行特征提取时具有敏感度低、特征信息消失等不利因素;(2)深度学习对复杂背景下的弱小目标检测能力较差,误检率较高;(3)红外数据

集作为深度学习的训练数据集时,不同背景信息的图像数量需符合正比例分布。

论文在深入研究 YOLO 系列算法<sup>[7-10]</sup>的基础上,提出一种基于红外弱小目标检测的多尺度特征融合检测算法 J-MSF,以采用单帧目标检测的实时算法 YOLO 框架为基础,以地/空背景下红外图像弱小飞机目标检测跟踪数据集<sup>[11]</sup>为研究对象,通过实验结果证明了所提出的红外弱小目标检测算法 J-MSF 的优越性,较 YOLOv3 系列算法和 YOLOv4 算法都有一定提升。

## 1 红外弱小目标检测 YOLO 系列算法

YOLO 检测目标流程可分为三大部分,待检测的图像作为输入,第一部分首先通过主干提取网络采集多层特征图信息;第二部分对提取到的不同特征图进行上采样并与对应特征融合叠加,获得精细化的语义特征表示;第三部分为分类与回归输出模块,主要对目标进行分类及边框预测。如图 1 所示。



图 1 YOLOv3 检测流程图

Fig.1 YOLOv3 flow chart of detection

YOLOv3 主干特征提取网络采用的是 Darknet-53, YOLOv4 的主干提取框架是 CSPDarknet-53,后者在前者基础之上对每个残差块增加自上而下的并行连接,增加特征的初始特性,并与 PAN 自下向上的路径相互作用,使得梯度流分裂,梯度流通过不同的网络路径进行传播,增加特征提取的丰富度。YOLO 颈部的作用为不同融合尺度特征图对不同大小的目标进行多尺度检测。YOLOv4 在特征融合部分在 YOLOv3 的基础上增加 SPP、SAM、PAN 等模块。YOLO 检测的头部是对每个预测目标输出目标边框、置信度、类别。YOLO 采用的对 bbox 的损失计算是对预测框和实际真实框之间的平方误差损失之和, YOLO 的预测边界框总损失为公式 (1):

$$L = \text{loss}_1 + \text{loss}_2 + \text{loss}_3 \quad (1)$$

式中:  $\text{loss}_1$  为预测框位置损失,即中心点坐标  $x$  和  $y$  以及预测框高度  $h$  和宽度  $w$  的损失;  $\text{loss}_2$  为置信度损失和  $\text{loss}_3$  分类损失。

YOLOv4 在 YOLOv3 的基础上进行改进补充, 将 Darknet-53 改进升级为 CSPDarknet-53, 以及增加 SPP 模块, 增加注意力机制模块, 优化训练框架等, 检测模块边界框回归依然采用原法。检测模块和主干提取网络部分采用的 PAN 通过自下向上的路径增强浅层中准确的定位信息流, 建立浅层特征和深层特征的交流通道。

## 2 J-MSF 的红外弱小目标检测

红外弱小目标相比于可见光图像小目标更难以检测, 主要原因是红外弱小目标的像素占比极少, 相互作用的卷积核之间的局部性质会强加有限感受野, 8 倍降采样特征图像对弱小目标的信息就很难提取到, 进而造成目标信息丢失。提出算法整体检测流程如图 2 所示, 首先, 主干提取框架进行多级特征提取; 其次, 特定融合策略作为检测网络颈部, 在流程图中用 F 来表示; 最后, 检测网络头部回归输出目标。为解决高层感受野目标易消失问题, 提高弱小目标检出率, 在主干提取网络部分提出 JAnet 结构来搭建多通道提取网络, 使用 DSPP 等量整合多级融合特征的策略, 对算法损失函数采用特定高斯优化。

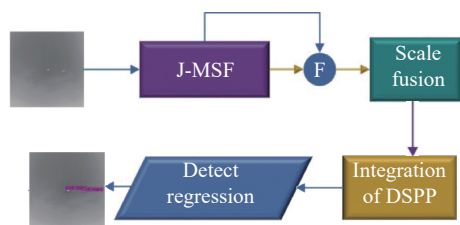


图 2 J-MSF 检测流程图

Fig.2 Detection flow chart of J-MSF

### 2.1 J-MSF 主干网络结构

可见光图像具有对比度高, 层次感强, 纹理信息明显, 相对于可见光图像, 红外图像可较好地体现远距离的物体的边缘信息, 但红外图像灰度值密集地分布在某些相邻灰度级上。经典算法 CNN 结构相互作用的卷积核之间的局部性质会强加有限感受野, 且单通道结构使得灰度级较少的红外图像中弱小目标信息极易丢失。为了解决在深层的网络结构中避免弱小红外目标像素信息丢失, 提高对弱小目标的检出率, 提出了一种分流聚合双通道结构 JAnet(Junction Aggregation Net)。

分流映射通过把图像  $x \in \mathbb{R}^{H \times W \times C}$  输入并进行 JA 操作来保留原始信息流, 如图 3 所示。其中,  $\mu_{ji}$  是所有预测向量的加权和,  $\eta_{ji}$  是前一层网络和权重矩阵的乘积,  $c_{ij}$  是卷积核系数。计算公式如下:

$$M_j = \sum_i c_{ij} (\mu_{ji} - \eta_{ji}) \cdot \mu_{ji} \quad (2)$$

JAnet 采用 mish 激活函数进行非线性操作之后进行输出, mish 激活函数<sup>[12]</sup> 在负值时允许较小的负梯度流入, 保证信息不会中断。

$$M(x) = x \cdot \tanh(\ln(1 + e^x)) \quad (3)$$

JAnet 网络期望通过相同底层卷积对图像特征进行传递并分流映射实现映射输出, 期望映射关系为  $G(x) = M(M(x) + x)$ , 最终得到的期望映射输出为  $M(x)$  和  $G(x) + M(x)$ , 保留更丰富的梯度流, 图 3 描述了 JAnet 结构单元信息。

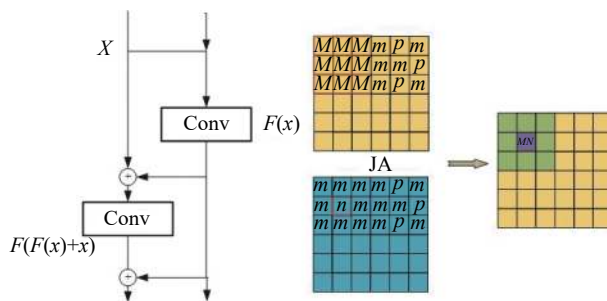


图 3 JAnet 结构单元

Fig.3 Structural unit of JAnet

JAnet 结构优势:

- (1) 增加感受野中目标信息的完整性, 并保持局部与整体空间关系的不变性。
- (2) 分流聚合的结构更有利于信息的传递和汇总, 可以有更深的网络层数, 并可更好地描述复杂的红外背景。
- (3) 网络参数的层更少, 训练时长和检测速度更优。

根据 J-MSF 主干网络结构, 文中设计更适用于提取红外弱小目标 J-MSF 的主干提取网络, 深的网络结构可以对复杂的背景信息进行描述, 某些红外图像背景信息较为复杂, 主干提取框架选取 5 个卷积提取块。在实时检测中, 目标是移动的, 显现在红外图像中目标所占像素有大有小, 多级特征可以兼顾大小目标进行检测, 主干特征图提取框架结构及 J-MSF 整体结构如图 4 所示。

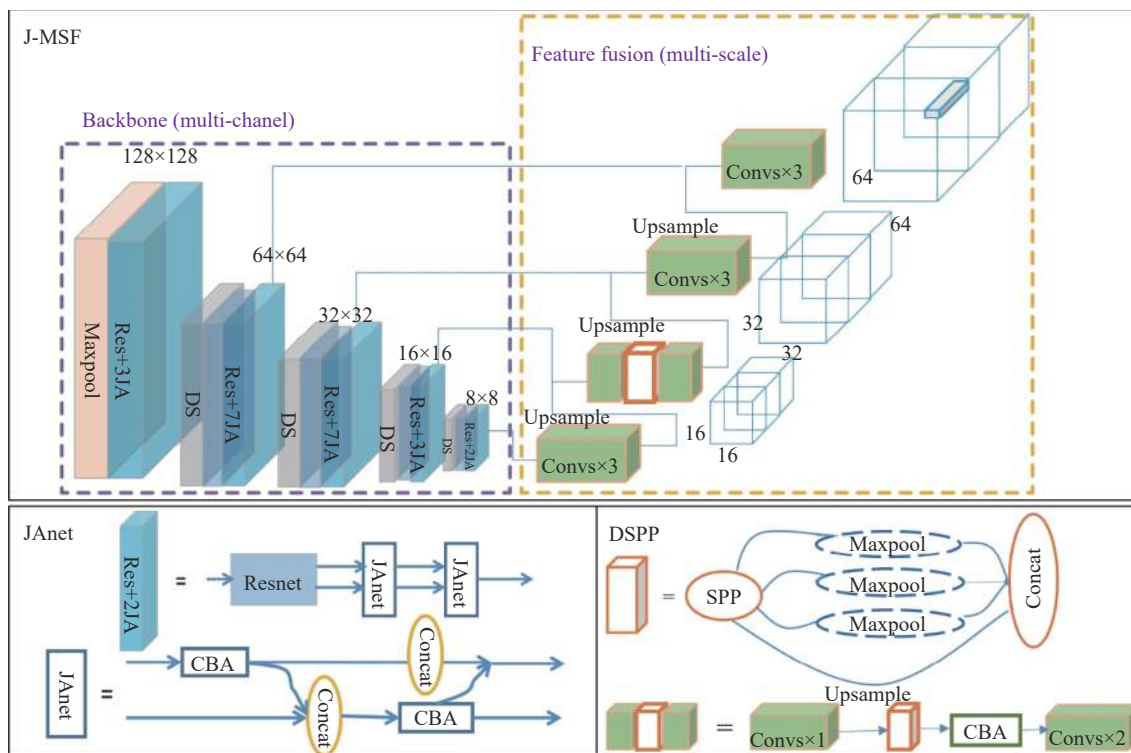


图 4 J-MSF 网络结构图

Fig.4 Network structure of J-MSF

2.2 弱小目标检测的多尺度融合策略

跨层特征融合已被证明是缓解计算机视觉中尺度变化问题的一种有效方法,它通常通过线性组合来实现,如加法或级联<sup>[13-14]</sup>、或非线性自顶向下调制<sup>[15]</sup>。然而,这些方法不是为红外弱小目标检测任务设计的,与一般的计算机视觉任务不同,红外弱小目标检测的瓶颈是如何在深层保持和突出弱小目标的特征,而不是在浅层缺乏高层语义。为此,文中采用了新的多级特征融合策略,将主干网络的降 32 倍特征图上采样与降 16 倍特征图融合作为基础融合特征图。基础融合

特征图再经上采样与降 8 倍特征图融合作为动脉融合特征图,在此处的动脉特征是检测的信息流的隘口,因此,增加 DSPP 模块用于整合融合特征,抑制弱小目标信息的丢失及掩盖。表 1 为基础特征图、动脉特征图、DSPP 及各输出特征图网络参数,最终输出特征图较原图像尺度为 16×16、32×32、64×64,融合更底层的特征有助于弱小目标检测,但仅增加浅层语义同样会带来更高的误检率,因此,弱小目标的多尺度策略如图 4 所示。并通过 K-means 聚类<sup>[9]</sup>得到 3 个不同尺度的预选框,即 (1,1), (2,2), (4,3), (3,4), (4,4),

表 1 网络参数维度

Tab.1 Dimensions of network parameters

Fusion map/layer	Kernel size	Output size	Stride	Channel
Basic-feature map	-	8×8	-	1024
Artery-feature map	-	16×16	-	768
Detection map 1	-	32×32	-	30
Detection map 2	-	64×64	-	30
Detection map 3	-	128×128	-	30
Maxpooling 1	3	32×32	1	128
Maxpooling 2	5	32×32	1	128
Maxpooling 3	7	32×32	1	128

(5,5), (7,6), (13,6), (25,25), 算法初值的设置主要用于预测部分遮挡目标的检测框, 该预选框初值同样也用于基准对比算法。

### 2.3 高斯损失优化函数的设计

J-MSF 损失函数中的 bbox 回归是提取 bbox 中心点的信息  $t_x$  和  $t_y$  以及 bbox 的高宽信息  $t_w$  和  $t_h$ 。在提出的算法中对最终预测框的坐标信息建立高斯模型, 给定预测输入  $x$  和输出  $y$  都需包含高斯参数, 如公式 (4) 所示:

$$P(y|x) = N(y; m(x), \sum(x)) \quad (4)$$

YOLO 损失函数是使用 bbox 信息进行回归建模, 值只包含了边界框的参数, 并没有对预测的边界框参数的置信度进行回归, 所以 J-MSF 的 bbox 边界框需要对中心点和高宽进行均值操作, 如公式 (5)~(7) 所示:

$$\mu_{t_x} = \sigma(\widehat{\mu}_{t_x}), \quad \mu_{t_y} = \sigma(\widehat{\mu}_{t_y}) \quad (5)$$

$$\mu_{t_w} = \sigma(\widehat{\mu}_{t_w}), \quad \mu_{t_h} = \sigma(\widehat{\mu}_{t_h}) \quad (6)$$

$$\sigma(x) = \frac{1}{(1 + \exp(-x))} \quad (7)$$

但是由于均值必须表示为网格内的中心点坐标, 实际值为在整张图片中的比例值, 所以使用 Sigmoid 激活函数为输出方差即为增加的预测边框置信度值, 每个方差的大小代表了不同坐标的不确定性, 方差函数同样需要经过 Sigmoid 函数进行激活, 如公式 (8)、(9) 所示:

$$\sum t_x = \sigma(\sum \widehat{t}_x), \quad \sum t_y = \sigma(\sum \widehat{t}_y) \quad (8)$$

$$\sum t_w = \sigma(\sum \widehat{t}_w), \quad \sum t_h = \sigma(\sum \widehat{t}_h) \quad (9)$$

在原损失函数中对于 bbox 的损失计算结果是对预测框和实际真实框之间的平方误差损失之和, 但是引入高斯模型之后, 要对原 bbox 目标框位置损失函数进行重新设计, 通过负对数似然损失进行计算。而置信度损失依然采用二元交叉熵损失, 公式 (10) 为 J-MSF 高斯损失函数表达式:

$$L(O, C, l, g) = \lambda_1 L_{conf}(o, c) + \lambda_2 L_{conf}(l, g) \quad (10)$$

$$L_{conf}(O, C) = \frac{\sum_i (o_i \ln(\hat{c}_i) + (1 - \hat{c}_i) \ln(1 - \hat{c}_i))}{N} \quad (11)$$

$$L_{conf}(l, g) = \frac{-\sum_{i \in pos} \sum_{m \in \{x, y, w, h\}} \log(N(g_i^m, \mu(l_i), \sum l_i) + \varepsilon)}{N_{pos}} \quad (12)$$

式中:  $L_{conf}(o, c)$  为置信度损失;  $L_{conf}(l, g)$  为高斯定位损失。公式 (11) 进行表示置信度损失, 其中,  $o_i \in [0, 1]$  表示预测目标边界框与真实目标边界框的 IOU,  $c$  为预测值。公式 (12) 中  $l$  为预测位置坐标信息,  $g$  为目标真实值信息。

## 3 实验与结果

首先, 通过目标检测评价标准定量分析 JAnet 对检测弱小目标的性能影响, 利用多尺度融合模块及特定损失函数检测弱小目标的灵敏性, 进行消融验证分析, 提出检测红外弱小目标的算法被记为 J-MSF 算法。其次, 用弱小目标数据集验证算法的性能, 与 YOLO 系列目前衍生算法以及主流算法进行对比定性分析。

### 3.1 红外弱小目标数据集

弱小目标图像的局部信噪比可以评价弱小目标检测的难易程度, 目标的大小可以通过所占像素的大小进行判断, 信噪比的数值可以判断目标信息的强弱。

描述图像中的弱小目标的信噪比称为局部信噪比<sup>[16]</sup>(Local Signal to Noise Ratio), 描述为公式 (13):

$$LSNR = 10 \times \log_{10}^{(F_r - F_B)/\delta_B} \quad (13)$$

式中:  $F_r$  为目标区域的均值;  $F_B$  为背景区域的均值;  $\delta_B$  为背景区域的标准差。一般背景区域大小为目标区域大小的 3 倍。

以下所有实验在地/空背景下红外图像弱小飞机目标检测跟踪数据集<sup>[11]</sup>下进行研究, 进而对深度学习模型性能进行标准判断。此数据集一共由 22 段数据组成, 30 条航迹、16 177 帧图像、16 944 个目标, 图中目标均为空中固定翼无人机 (燃油动力), 飞机长度为 2.0 m, 翼展长度为 2.6 m, 飞机速度为空速 30 km/h。所拍摄的图像飞行的航路近端到远端的范围为 500~5 000 m, 高度的范围为 50~500 m。数据集中第 2 段、第 4 段为双目标, 其余为远近不同的单目标和不同背景下的数据段。图像像素大小均为 256×256。图 5 为检测测试集目标占像素大小分布图, 表 2 为测试集的信噪比分布表。

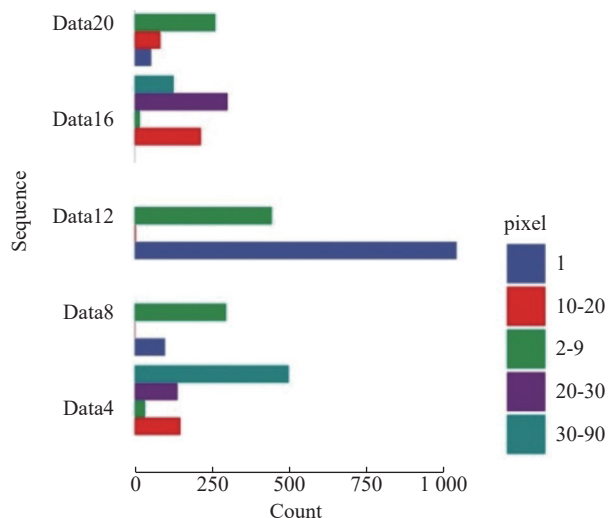


图 5 测试集序列中目标所占像素值  
Fig.5 Pixel value of target in test set sequence

表 2 测试集信噪比数据分布表

Tab.2 SNR data distribution table of test set

SNR region	3.26-3	3-2	2-1	1-0	0-(-1.97)	-3-(-20)
Data4	0	5	209	379	204	2
Data8	2	39	108	94	101	55
Data12	5	84	407	424	341	238
Data16	5	247	214	15	1	12
Data20	0	12	155	197	29	8
Total	12	387	1093	1109	676	315

### 3.2 评价标准

在检测过程中, 红外弱小目标的检测算法召回率(检出率)是一个重要指标,  $P-R$  曲线是根据不同置信度的召回率值对应的准确率来描述模型性能的指标。目标召回率(检出率) $R$  和检测的准确率  $P$  可分别表示为:

$$R = \frac{X_{TP}}{X_{TP} + X_{FN}} \quad (14)$$

$$P = \frac{X_{TP}}{X_{TP} + X_{FP}} \quad (15)$$

式中:  $X_{TP}$  表示正确检测出来的目标数;  $X_{FN}$  表示没有被检测出来的目标数;  $X_{FP}$  表示被错误检出的目标数。

### 3.3 训练参数的设置

所有算法模型均采用 4 : 1 的比例进行训练集和测试集的划分, 即 Data4、Data8、Data12、Data16、Data20

作为测试集数据, 其余作为训练集数据。

所以训练均使用的超参数如下: 训练学习率以 Step 策略衰减的方式进行设置, 即开始系数为 0.0001, 当训练迭代次数为 40000 次和 45000 次是分别将学习率减少为 0.00001 和 0.000001。训练损失曲线如图 6 所示。实验条件为: 在 Ubuntu16.04 系统, 采用的深度学习框架是 Darknet 框架; GPU 采用的是 NVIDIA GeForce GTX 1080 Ti, 内存为 11 GB。

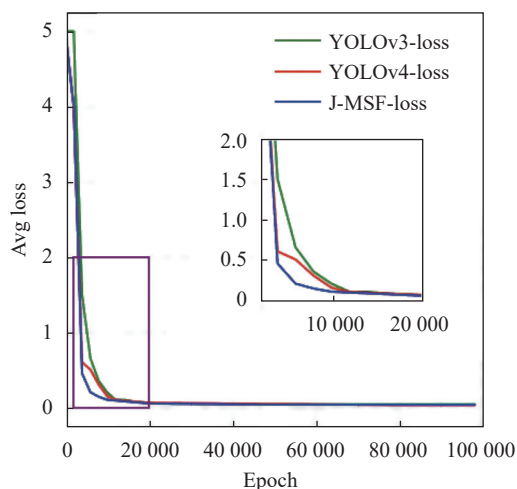


图 6 训练损失曲线

Fig.6 Training loss curve

### 3.4 数据分析

针对第 1 章所提的地/空背景下红外图像弱小飞机目标检测跟踪数据集<sup>[1]</sup>中 Data4、Data8、Data12、Data16、Data20 作为测试集, 测试集共 3592 个目标, 和相同的训练参数进行测试, 在定性对比实验中, 与当前 YOLO 系列算法及衍生算法检测输出结果进行对比。在定量对比实验过程中, 通过 J-MSF 和 YOLO 系列算法进行对比, 对召回率(检出率)、AP (AP50) 值、 $P-R$  曲线综合评估模型的检测性能。最后, 与目前主流算法进行对比实验。

#### 3.4.1 目标检测定性结果分析

为验证所提算法在实际任务中的有效性, 抽取测试集每个序列目标检测较为困难的实际背景图像进行测试, 评估算法检测结果并进行可视化分析。可视化检测结果展示如图 7 所示。粉色标记为未检出的真实目标, 红色标记为在检测错误的目标, 绿色标记为检出目标, 所提算法 J-MSF 都可以进行正确检出。

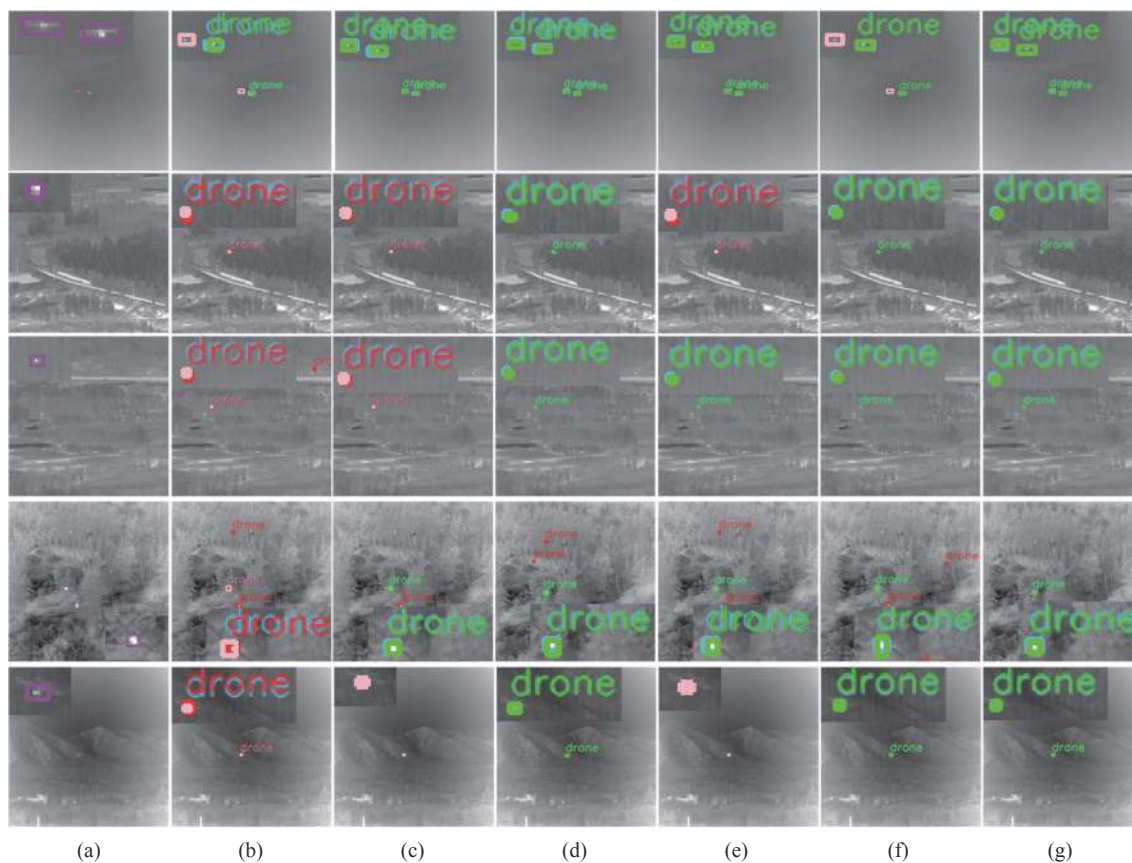


图 7 (a) 标记对照框; (b) YOLO-Tiny 检测结果; (c) YOLOv3 检测结果; (d) YOLOv3+SPP 模块检测结果; (e) Gaussian YOLOv3+SPP 检测结果; (f) YOLOv4 检测结果; (g) J-MSF 检测结果

Fig.7 (a) Mark contrast box; (b) YOLO-Tiny detection result; (c) YOLOv3 detection result; (d) YOLOv3+SPP detection result; (e) Gaussian YOLOv3+SPP detection result; (f) YOLOv4 detection result; (g) J-MSF detection result

### 3.4.2 算法定量验证分析

为检验所提出的 JAnet 网络结构对弱小目标的提取能力, 在采用相同网络框架 YOLOv3 上进行验证, 即只更换主干提取网络进行实验。Darknet-53-JA 将网络 ResNet 更换为 JAnet 结构, J-MSF 为最终设计提取弱小目标的主干网络。如表 3 所示, 多通道的 JAnet 对检出率有显著提升, 验证多通道的网络结构可以缓解弱小目标信息的丢失。

通过消融实验来验证算法模型的性能以及各个模块对所提数据集的检测能力, 即对检测性能和检测

速度的影响。表 3 中的代表的含义如下: (1) Darknet-53: YOLOv3 使用的主干提取框架。(2) J-MSF: 提出的红外弱小目标的主干提取网络。(3) Loss: YOLO 系列算法采用的损失函数称为 D, J-MSF 算法改进的高斯损失函数称为 M。(4) Fusion: 针对检测红外弱小目标采用的新的特征融合策略。消融实验是通过加减关键模块来分析各模块对算法性能的影响, 文中针对检测弱小目标设计的主干提取框架、融合策略、特定损失等组件进行消除实验。J-MSF 的 backbone 和 Darknet-53 作对比, 随着组件的添加, 算法的性能逐步提升。如表 4 所示, 采用多通道 JAnet 网络结构的检出率和整体 AP 值均有显著提升。同时, 所提的多尺度融合策略对弱小目标检测性能也都有所提升。提出的优化高斯损失函数可以有效提高检出率, 但由于噪声的影响, 随着检出目标的增多, 准确率出现一定程度的下降。最后, 与 YOLOv3 相比, 检出率提升 9.07 个百分点, AP 值提升 9.91。且同时加入 JAnet 结构、

表 3 JAnet 网络对比实验

Tab.3 Contrast experiment of JAnet network

Model	$X_{FN}$	$R$	AP
Darknet-53	458	87.2%	86.38%
Darknet-53-JA	343	90.0%	88.43%
J-MSF	217	94.0%	93.13%

融合策略和特定损失模块带来的提升, 要比各个模块带来的涨点都高, 这说明各组件间存在促进互补关系。

还与基于红外弱小目标检测 YOLO 系列衍生算法进行了对比实验, 它们分别是 YOLO-tiny、YOLOv3+

SPP<sup>[17]</sup>、Gaussian YOLOv3+SPP, 还有 YOLOv3 和 YOLOv4。图 8 为它们的 *P-R* 曲线, 可以看出所提算法 J-MSF 有显著的提升。从表 5 可以看出, J-MSF 漏检总数优于其他算法, 验证了算法的优越性。

表 4 消融实验

Tab.4 Ablation study

Darknet53	J-MSF	Loss	Fusion	Precision	R	AP	FPS
√	-	D	-	86%	87.20%	86.38%	66.3
√	-	D	√	92%	92.20%	92.74%	57.5
√	-	M	-	89%	94.04%	93.88%	71.9
√	-	M	√	82%	95.00%	93.47%	71.6
-	√	D	-	90%	94.00%	93.13%	59.0
-	√	D	√	90%	94.10%	93.46%	73.4
-	√	M	-	86%	95.85%	94.80%	66.8
-	√	M	√	88%	96.27%	96.29%	67.6

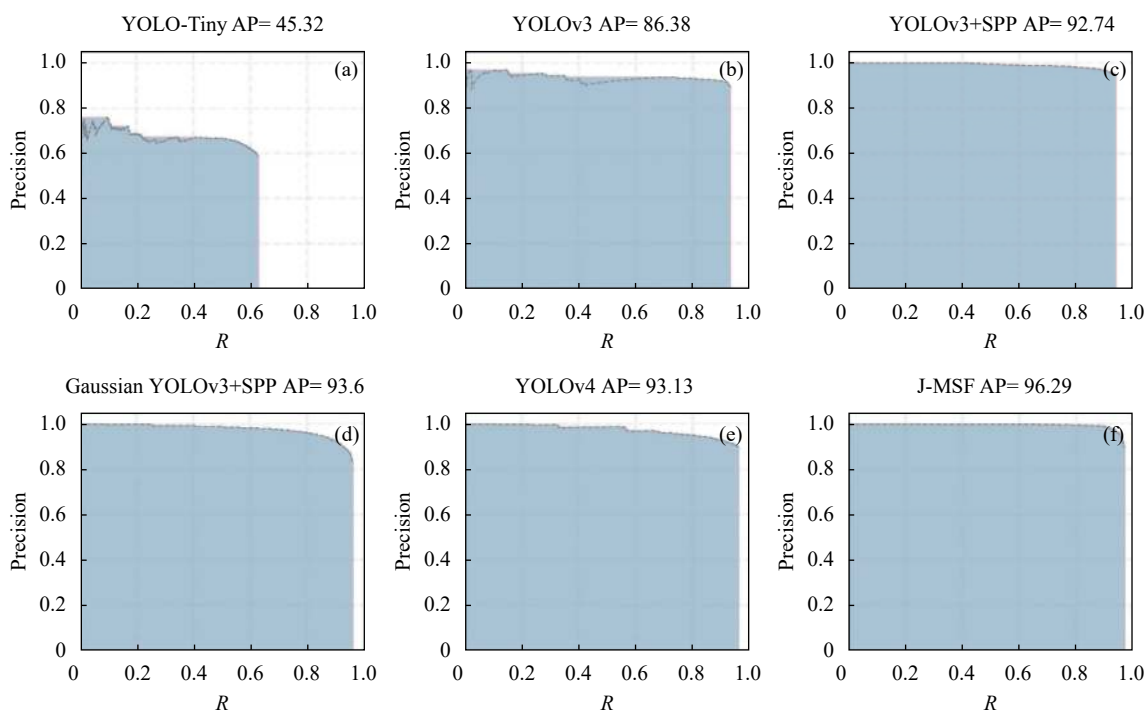


图 8 (a) YOLO-Tiny Precision-R 曲线; (b) YOLOv3 Precision-R 曲线; (c) YOLOv3+SPP 模块 Precision-R 曲线; (d) Gaussian YOLOv3+SPP Precision-R 曲线; (e) YOLOv4 Precision-R 曲线; (f) J-MSF Precision-R 曲线

Fig.8 (a) YOLO-Tiny Precision-R curve; (b) YOLOv3 Precision-R curve; (c) YOLOv3+SPP Precision-R curve; (d) Gaussian YOLOv3+SPP Precision-R curve; (e) YOLOv4 Precision-R curve; (f) J-MSF Precision-R curve

表 5 YOLO 系列模型对红外目标检测结果

Tab.5 Results of infrared target detection by YOLO serial model

Detection algorithm	$X_{TP}$	$X_{FP}$	$X_{FN}$	Precision	R	AP
YOLO-Tiny	2389	1355	1203	59%	64%	45.32%
YOLOv3	3309	435	283	88%	92%	86.38%
YOLOv3+SPP <sup>[17]</sup>	3318	283	274	92%	92%	92.74%
Gaussian YOLOv3 <sup>[18]</sup> +SPP	3407	758	185	78%	95%	93.60%
YOLOv4	3397	446	195	88%	95%	93.13%
J-MSF	3443	451	149	88%	96%	96.29%



### 3.4.3 主流检测算法对比实验

对不同主流的目标检测网络通过相同的数据训练得到的模型进行评价<sup>[19]</sup>,主要比较不同目标检测算法对红外无人机目标的检测识别能力,并验证文中算法的检测性能。对比算法包含在实验中,分别采用 AP50、帧率 (FPS) 评价指标进行检测准确度和检测速度的定量分析。

如表 6 所示,所提算法与主流目标检测算法对比,从目标检测的精度指标来看,所提算法的 AP 值有了较大的提高。图 9 为主流算法测试结果对比图,在检测效果和检测实时性方面,文中算法检测速度与目标检测算法 SSD300 相比,虽未达到高度实时性,但与 YOLO 系列算法相比,依然占优势,其主要原因是 JAnet 结构相比于 ResNet 降低了网络参数量,提升算法的运行效率。

表 6 主流算法测试结果对比

Tab.6 Comparison of mainstream algorithms

Detection algorithm	AP	FPS
Faster R-CNN <sup>[20]</sup>	43.7%	35.2
SSD300 <sup>[21]</sup>	52.3%	154.7
RefineDet <sup>[22]</sup>	63.9%	70.1
RetinaNet <sup>[23]</sup>	65.4%	80.3
YOLOv3	86.4%	66.3
YOLOv4	93.1%	66.8
J-MSF	96.3%	67.6

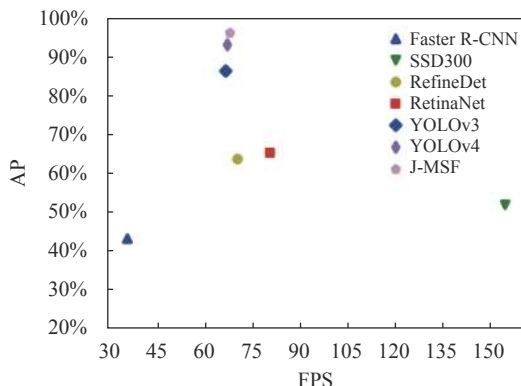


图 9 主流算法 FPS-AP 曲线

Fig.9 Mainstream algorithm FPS-AP curve

## 4 结 论

文中对红外弱小目标检测问题进行研究,提出一

种多通道多尺度特征融合算法 J-MSF,对红外弱小目标检测。通过提出的 JAnet 网络结构,搭建主干提取网络,并将底层特征与输出特征进行融合,同时提出算法采用改进高斯损失函数,利用 DSPP 策略整合融合特征,进而提高弱小目标的检出率。提出算法主要有如下研究成果:

(1) 针对红外弱小目标的图像特点,提出 JAnet 结构来设计更适合红外弱小目标特征提取的主干网络提取框架。

(2) 设计适用于检测弱小目标的多尺度特征融合策略,采用 DSPP (Down Spatial Pyramid Pooling) 对融合特征等量进行整合。

(3) 设计损失优化函数,对检出目标位置损失采用高斯进行优化,进一步提高对弱小目标的检出率,并对损失函数进行剪枝。在检测感受野中,增强弱小目标特征的曝光尺度。

通过公开数据集进行算法有效性的验证,即对弱小目标的检测能力有着明显的提升。但是红外图像成像等因素的影响对弱小目标的识别难度依然很大,通过网络改进难以实现,下一步工作是针对干扰去除进而实现更精准识别进行研究。

### 参考文献:

- [1] Jiang Guoqing, Wan Lanjun. Detection of dim and small infrared targets based on the most appropriate contrast saliency analysis [J]. *Infrared and Laser Engineering*, 2021, 50(4): 20200377. (in Chinese)
- [2] Liu Gaoru, Sun Shengli, Lin Changqing. Two-dimensional spatial profile method for infrared dim point target background suppression [J]. *Infrared Technology*, 2019, 41(4): 329-334. (in Chinese)
- [3] Zhang Congcong. Infrared dim small target detection method based on low rank background and sparse target characteristics [D]. Nanjing: Nanjing University of Science and Technology, 2018. (in Chinese)
- [4] Huang Yuanyuan. Research on infrared dim small target detection algorithm based on local contrast [D]. Chongqing: Chongqing University of Posts and Telecommunications, 2020. (in Chinese)
- [5] Zhao Yan, Liu Di, Zhao Lingjun. Infrared dim and small target detection based on YOLOv3 in complex environment [J]. *Aero Weaponry*, 2019, 26(6): 29-34. (in Chinese)
- [6] Feng Xiaoyu, Mei Wei, Hu Dashuai. Air target detection based

- on improved fast R-CNN [J]. *Acta Optica Sinica*, 2018, 38(6): 0615004. (in Chinese)
- [7] 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 (CVPR), 2016: 779-788.
- [8] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer vision and Pattern Recognition (CVPR), 2017: 7263-7271.
- [9] Redmon J, Farhadi A. Yolov3: An incremental improvement [J]. *arXiv*, 2018: 1804.02767.
- [10] Bochkovskiy A, Wang C Y, Liao H Y. YOLOv4: Optimal speed and accuracy of object detection [J]. *arXiv*, 2020: 2004.10934.
- [11] Hui B, Song Z, Fan H. A dataset for infrared detection and tracking of dim-small aircraft targets underground/air background [J]. *China Scientific Data*, 2020, 5(3): 291-302.
- [12] Misra D. Mish: A self-regularized non-monotonic neural activation function [J]. *arXiv*, 2019: 1908.08681.
- [13] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer Assisted Intervention(MICCAI), 2015: 234-241.
- [14] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer vision and Pattern Recognition (CVPR), 2017: 2117-2125.
- [15] Yuan W, Wang S, Li X, et al. A skip attention mechanism for monaural singing voice separation. [J]. *IEEE Signal Processing Letters*, 2019, 26(10): 1481-1485.
- [16] Fan Xiangsuo. Research on small target detection and tracking algorithm in image sequences[D]. Chengdu: University of Electronic Science and Technology, 2019. (in Chinese)
- [17] 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.
- [18] Choi J, Chun D, Kim H, et al. Gaussian YOLOv3: An accurate and fast object detector using localization uncertainty for autonomous driving[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 502-511.
- [19] Chen L, Shi W, Deng D. Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images [J]. *Remote Sensing*, 2021, 13(4): 660.
- [20] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *arXiv*, 2015: 1506.01497.
- [21] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multi-box detector[C]//European Conference on Computer Vision (ECCV), 2016: 21-37.
- [22] Zhang S, Wen L, Bian X, et al. Single-shot refinement neural network for object detection[C]//Proceedings of the IEEE Conference on Computer vision and Pattern Recognition (CVPR), 2018: 4203-4212.
- [23] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.