引用格式：LI Zongzhe, WANG Qicheng, YANG Weihua, et al. Quantifying the Effects of Sensor Noise and Spatial Resolution on Small Target Detection in Multispectral Remote Sensing Imagery[J]. Acta Photonica Sinica, 2026, 55(1):0110002
李宗哲,王琪程,杨伟华,等. 光学遥感图像噪声和空间分辨率对小目标检测精度影响机制研究[J].光子学报,2026,55(1):0110002

光学遥感图像噪声和空间分辨率对小目标检测精度影响机制研究

李宗哲[1],王琪程[1],杨伟华[2],韩晓腾[2]
(1 空军工程大学 装备管理与无人机工程学院,西安 710051)
(2 中国科学院西安光学精密机械研究所 光谱成像技术研究室,西安 710119)

摘　要：多光谱图像因载荷平台视角差异和背景复杂多变导致图像质量和空间尺度具有较大差异性，在各向异质的场景小目标检测性能受到严重挑战。针对该问题，重点分析了SuperYOLO模型在不同场景数据集的检测精度以及多光谱图像分辨率和质量对SuperYOLO小目标检测的影响。研究结果表明：1)SuperYOLO检测精度在独立数据集(VEDAI和HRRSD数据集)优于混合数据集；2)仪器噪声对SuperYOLO检测精度影响剧烈，当增加了5%和10%仪器噪声时，mAP@0.5由0.813降低至0.420和0.058，而在数据处理模块加入去噪的卷积神经网络可有效提升SuperYOLO模型检测精度；3)多光谱图像分辨率对SuperYOLO影响较小，降采样后mAP@0.5由0.813降至0.741和0.464。因而，仪器噪声显著影响SuperYOLO中多分支网络结构中光谱特征的提取和融合，影响SuperYOLO对于多光谱遥感图像小目标的检测精度和实用性，而提升光学图像质量，在SuperYOLO模型耦合降噪算法模块可有效推进多光谱遥感图像小目标探测与识别算法的检测能力。

关键词：多光谱遥感图像；目标检测；SuperYOLO；深度学习
中图分类号：TP120.10　　**文献标识码**：A　　**doi**：10.3788/gzxb20265501.0110002

0　引言

随着遥感技术的快速进步，遥感图像在环境监测和资源调查等各领域已成为不可或缺的信息资源[1]。但光学遥感数据的海量性、场景复杂性等会造成信息过载、检测算法计算复杂度增加、目标特征提取难度增大等问题，给地面小目标的检测造成了严重的干扰和巨大的挑战[2]。此外，光学遥感图像中的目标在复杂和广阔背景中仅占几十个像素，空间异质性特征显著，且由于小目标的纹理差异大、对比度低、方向任意、密集分布等特性，实现对遥感图像中地面小目标的高精度检测仍具备挑战性[3-4]。

传统的目标检测采用基于统计的方法，但是其无法应对海量的数据。而深度神经网络通过多个卷积和池化层提取图像的高级特征解决了这一问题[5]。REDMON J等于2016年提出了基于深度学习的单阶段目标检测方法YOLO[6]，随后涌现出大量用于特定情况下目标检测的改进版YOLO (You Only Look Once)算法。在遥感图像目标检测领域，ZHANG Jiaqing等提出了一种准确且快速的针对于多模态遥感图像的目标检测方法SuperYOLO[7]。该方法融合了多模态数据，并通过辅助超分辨率(Super Resolution Assisted, SRA)学习对多尺度目标进行高分辨率目标检测，兼顾了检测准确性和计算成本。

但在实际应用中，不同应用场景下应用多光谱图像进行小目标检测的效果波动较大[8]，其小目标的辐射强度相对于背景噪声和干扰物(如纹理复杂的农田、森林、建筑群)往往非常微弱[9]，容易被淹没。由于模型

基金项目：国家自然科学基金(42576184, 42176182, 42306202)
第一作者：李宗哲, lzz144@163.com
通讯作者：王琪程, m19824443084@163com
收稿日期：2025-07-02；**录用日期**：2025-09-15

是恒定的，那么效果波动的原因应来自输入图像的差异[10]。不同应用场景下多光谱图像波段、分辨率、图像质量、背景复杂差异显著[11]。特别是在获取多光谱遥感图像时，会使用到多种光学传感器和载荷平台，其差异性以及拍摄过程中的诸多困难会使多光谱遥感图像中存在着不同仪器带来的噪声[12]。此外，获取的可见光和近红外波段图像的空间分辨率也通常不同（例如，全色波段分辨率高，热红外波段分辨率低）[13]。因而，如何对齐和处理多光谱图像分辨率差异直接影响小目标的清晰度和定位精度。此外，不同场景下多光谱传感器可能引入不同类型的噪声，会淹没小目标的微弱特征，一方面，不同时间、天气条件下获取的图像，光照差异巨大，影响目标的表观特征；另一方面，大气散射、吸收等效应会影响不同波段的辐射值，改变目标的光谱特征。因此，本文将从图像质量和图像空间分辨率两个角度出发，探究影响SuperYOLO在多光谱遥感图像小目标检测时性能变化的主要因素。

1 模型与方法

1.1 SuperYOLO模型简介

SuperYOLO是在YOLO5s模型基础上的改进，可分为输入（input）、骨干（backbone）、颈部（neck）、头部（head）四大部分。图1为SuperYOLO的结构图。SuperYOLO的骨干网络使用CSPNet提取特征信息[14]，其中包括众多的跨阶段局部网络（Cross Stage Partial，CSP）结构，增强了深度卷积神经网络的学习能力，既保持了较低的参数量又提高了精度。同时，SuperYOLO将骨干网络中的Focus模块替换成了像素级融合模块（Multi-Scale Fusion，MF），其将可见光和近红外不同波段的特征融合，提炼出不同且互补的特征信息。此外，在训练阶段，SuperYOLO添加了辅助超分辨率（SR）模块，该模块重建高分辨率图像以指导骨干网络在空间维度的学习。在推理阶段，SR分支被移除，以减少计算量。在骨干网络的最后，引入了空间金字塔池化（Spatial Pyramid Pooling，SPP）结构。SPP模块是一个多尺度特征融合阶段，它使用四种不同大小的最大池化来捕捉多尺度信息[15]。SuperYOLO使用PANet作为颈部[16]，在FPN自顶而下结构的基础上新增从底层到顶层的侧连接，构建自底向上的路径，增强整体特征金字塔的定位表达能力。

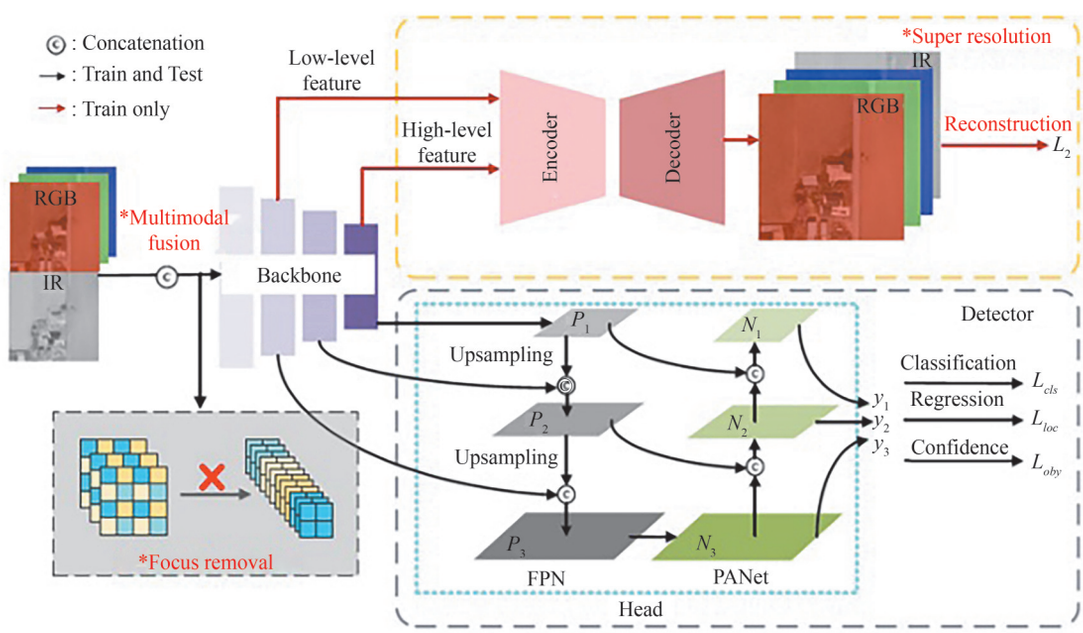


图1 SuperYOLO结构[7]

Fig. 1 The framework of SuperYOLO[7]

1.2 多模态融合

SuperYOLO采用了像素级多模态融合（MF）方法（如图2），以对称且紧凑的方式双向结合多模态信息。首先，将输入的RGB图像和IR图像规范化到[0,1]区间。然后，将输入的模态X_{RGB}，$X_{IR}\in R^{C\times H\times W}$下采样得

到I_{RGB}，I_{IR}长宽缩小为原来的$1/n$。下采样后的图像被输入到SE模块中提取内在特征信息，生成F_{RGB}，F_{IR}，并对其进行1×1卷积得到m_{RGB}，m_{IR}。不同模态之间的内在空间信息可表示为

$$\begin{cases} F_{in1} = m_{RGB} \otimes F_{RGB} \\ F_{in2} = m_{IR} \otimes F_{IR} \end{cases} \tag{1}$$

为了结合内部视图信息和空间纹理信息，将这些特征与原始输入模态相加，再进行1×1卷积。

$$\begin{cases} F_{full1} = f(F_{in1} + I_{RGB}) \\ F_{full2} = f(F_{in2} + I_{IR}) \end{cases} \tag{2}$$

最后，进行特征融合，即

$$F_{o} = \mathrm{SE}\left(\mathrm{Concat}(F_{full1}, F_{full2})\right) \tag{3}$$

结果F_{o}随后被输入到骨干网络中以提取多层次的特征。

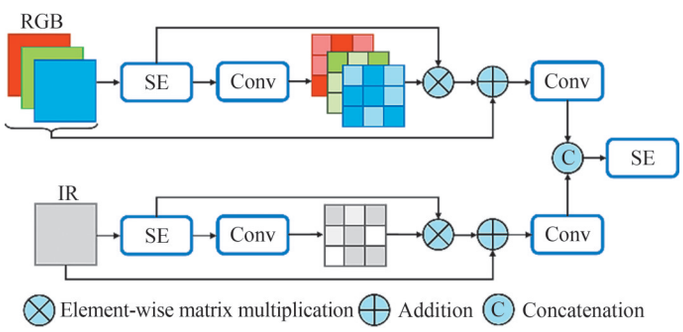


图2　SuperYOLO多模态融合流程

Fig. 2　SuperYOLO multi-modal fusion flowchart

1.3　超分辨率重建

用于多尺度检测的特征图尺寸远小于原始输入图像的尺寸，现有的大多数方法是通过上采样操作来恢复特征尺寸的。但是，该方法会导致信息丢失，效果有限，特别是在小目标检测中，更需要保留高分辨率的特征。

为了解决这个问题，SuperYOLO引入了一个辅助的超分辨率（SR）分支，其可以看作是一个简单的编码器-解码器模型。分别选择第四和第九模块的结果作为低级和高级特征，融合局部纹理和模式以及语义信息。在编码器中，对低级特征进行CR（卷积和ReLU激活函数）操作；对于高级特征，进行上采样操作来匹配低级特征的空间大小。接着，将低级特征和高级特征通过两个CR模块并合并，实现特征整合。解码器使用三个卷积层来实现。SR指导空间维度的相关学习，在损失函数中，也相应地添加了SR重建损失函数。

1.4　DnCNN去噪模型

由于不同光学载荷的杂散光、暗电流、光子噪声及探测器响应异常噪声等导致图像退化，严重影响图像质量和目标检测精度。本研究在分析环境高光谱卫星（HJ-2A）定标采集的数据噪声特性的基础上，在SuperYOLO处理之前增加了去噪模块，即去噪的卷积神经网络（Denoising Convolutional Neural Network，DCNN），用以分析SuperYOLO在不同数据集的场景适用性。设理想的无噪声图像为X，每个像素都有一个噪声偏移量V，得到的噪声图Y可表示为

$$Y = X + V \tag{4}$$

DnCNN在VGG的基础上进行修改，网络结构是（卷积、BN、ReLU）级联的结构，并在网络的输出使用残差学习。依据噪声图像Y估算干净图像X，而模型的直接输出的噪声图像V，最终干净图的获取过程为$X = Y - R(Y)$，这里$R(y)$就相当于模型。模型训练时，输出的误差图和真实误差图之间采用SME损失函数进行约束。

2　实验设置

2.1　数据集

本文在遥感目标检测常用数据集航空图像的车辆检测（Vehicle Detection in Aerial Imagery，VEDAI）

上[17]，开展了相关实验。该数据集包含了从犹他州自动地理参考中心(Utah Automated Geographic Reference Center, AGRC)数据集中裁剪的图像，每张从AGRC中收集的原始图像尺寸为16 000×16 000像素，像元分辨率12.5 cm × 12.5 cm，包含RGB和IR两种模态的数据。VEDAI数据集包括1 246张较小的图像，背景丰富，包括草地、高速公路、山脉和城市区域。所有图像的大小为1 024×1 024或512×512像素，检测目标为11类不同的车辆，均为小目标。在本文中，训练集与验证集的比例设置为9∶1。

为了提升SuperYOLO在不同场景的适用性，采用高分辨率遥感检测(High-resolution Remote Sensing Detection, HRRSD)数据集进行了验证，HRRSD数据集涵盖了从0.15～1.2 m的多种空间分辨率。该数据集包含21 761张图像，标注了55 740个对象实例，涵盖13类遥感目标。

2.2 实验过程

利用PyTorch深度学习框架进行模型的训练与推理，硬件为NVIDIA 4090 GPU。为了实现SR辅助分支，在训练过程中将1 024×1 024大小的输入图像统一下采样到512×512。数据增强方面，使用了HSV、尺度变化、平移、左右翻转和马赛克数据增强等方法。超参数方面，使用标准的随机梯度下降(Stochastic Gradient Descent, SGD)来训练网络，动量为0.937，权重衰减为0.000 5；epoch为300，batchsize为2，lr初始设置为0.01。

为了研究影响多模态遥感图像小目标检测的主要因素，开展了两组不同的实验：第一组参照以环境卫星HJ-2A实验室辐射定标采集的噪声数据为基础，在原始图像中添加不同强度的仪器噪声，用于探究仪器噪声与检测性能之间的关系。第二组将改变原始图像的空间分辨率，用于探究空间分辨率与对模型中SR模块的影响，进而探究出与检测模型性能之间的关系。

2.3 评价指标

利用准确性评估测量了检测结果与真实之间的一致性。用于评估检测性能的准确性指标包括召回率(recall)、准确率(precision)、$F1$分数、AP(平均精度)和mAP(均值平均精度)。准确率和召回率可分别表示为

$$P_{\text{recision}}=\frac{T_{\text{P}}}{T_{\text{P}}+F_{\text{P}}} \tag{5}$$

$$R_{\text{ecall}}=\frac{T_{\text{P}}}{T_{\text{P}}+F_{\text{N}}} \tag{6}$$

式中，真阳性(True Positive, T_{P})为模型正确预测为正类的正样本数量；真阴性(True Negative, T_{N})为模型正确预测为负类的负样本数量；假阳性(False Positive, F_{P})为模型错误预测为正类的负样本数量；假阴性(False Negative, F_{N})为模型错误预测为负类的正样本数量。准确率与误检(commission)错误相关，召回率与漏检(omission)错误相关。F_1分数是同时考虑准确率与召回率的一个指标，具体可表示为

$$F_1=\frac{P_{\text{recision}}R_{\text{ecall}}}{0.5\left(P_{\text{recision}}+R_{\text{ecall}}\right)} \tag{7}$$

AP可通过积分方法计算Precision-Recall曲线下的面积得到。mAP是一个综合指标，通过对所有检测类别的AP进行平均得到，所有类别的mAP可以通过式(8)计算。

$$\text{mAP}=\frac{\text{AP}}{N}=\frac{\int_0^1 p(r)\text{d}_r}{N} \tag{8}$$

式中，N代表类别的数量，$p(r)$为各类别的检测精度，$\int_0^1 p(r)\text{d}_r$为计算Precision-Recall曲线下的面积。

3 结果与讨论

3.1 SuperYOLO模型检测性能

为了研究SuperYOLO在不同场景的适用性，分析了该模型在HRRSD数据集及在HRRSD数据集与VEDAI混合数据集的检测效果。研究结果表明，SuperYOLO在HRRSD数据集与VEDAI混合数据集的检测结果急剧下降，应用场景的不同给SuperYOLO在小目标检测带来较大的挑战(如图3(d)～(f))。但SuperYOLO在HRRSD数据集或VEDAI数据集进行多模态遥感图像小目标检测结果较高。例如，SuperYOLO单独在VEDAI数据集或HRRSD数据集中汽车目标检测精度为0.91和0.94，而在混合数据集中

汽车目标整体上AP为0.63。可以看出，虽然有部分相似度过高的目标被误检，但总体上来说SuperYOLO能够较为准确地检测地面多类小目标。图4为不同置信度下SuperYOLO模型在不同数据集下各类别F_1和Precision分数随置信度的变化曲线。以VEDAI数据集为例，当置信度阈值为0.87时，所有类别的平均精确度

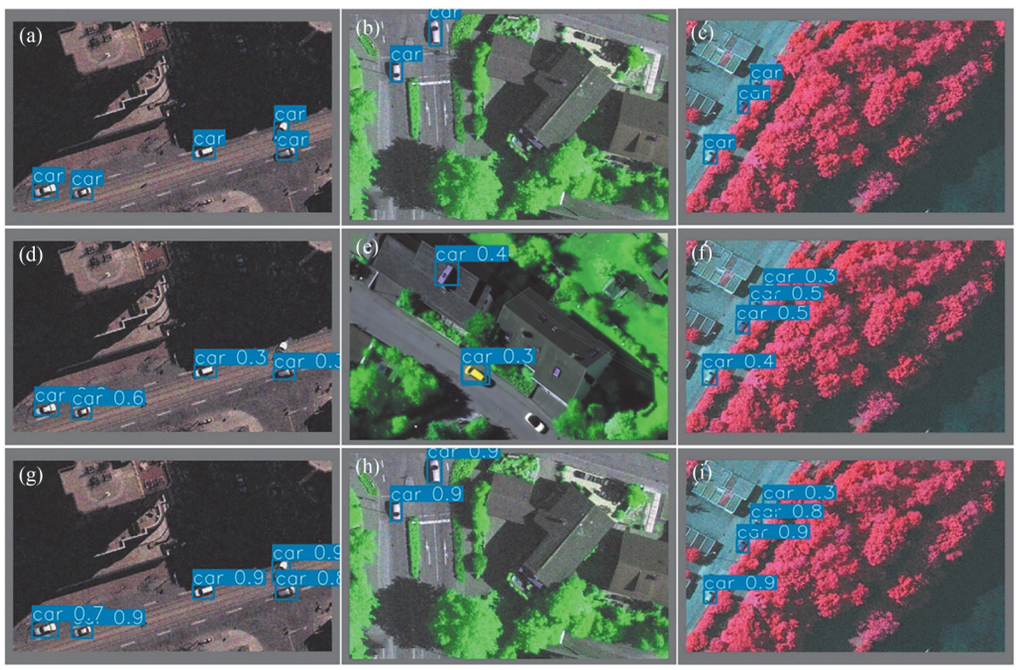


图3　SuperYOLO小目标检测结果。(a)～(c)原始图像；(d)～(f)混合数据集检测结果；(g)～(i)HRRSD数据集检测结果

Fig. 3　Small target detection results of SuperYOLO. (a)～(c) The original images; (d)～(f) The detection results on the mixed dataset; (g)～(i) The detection results on the HRRSD dataset

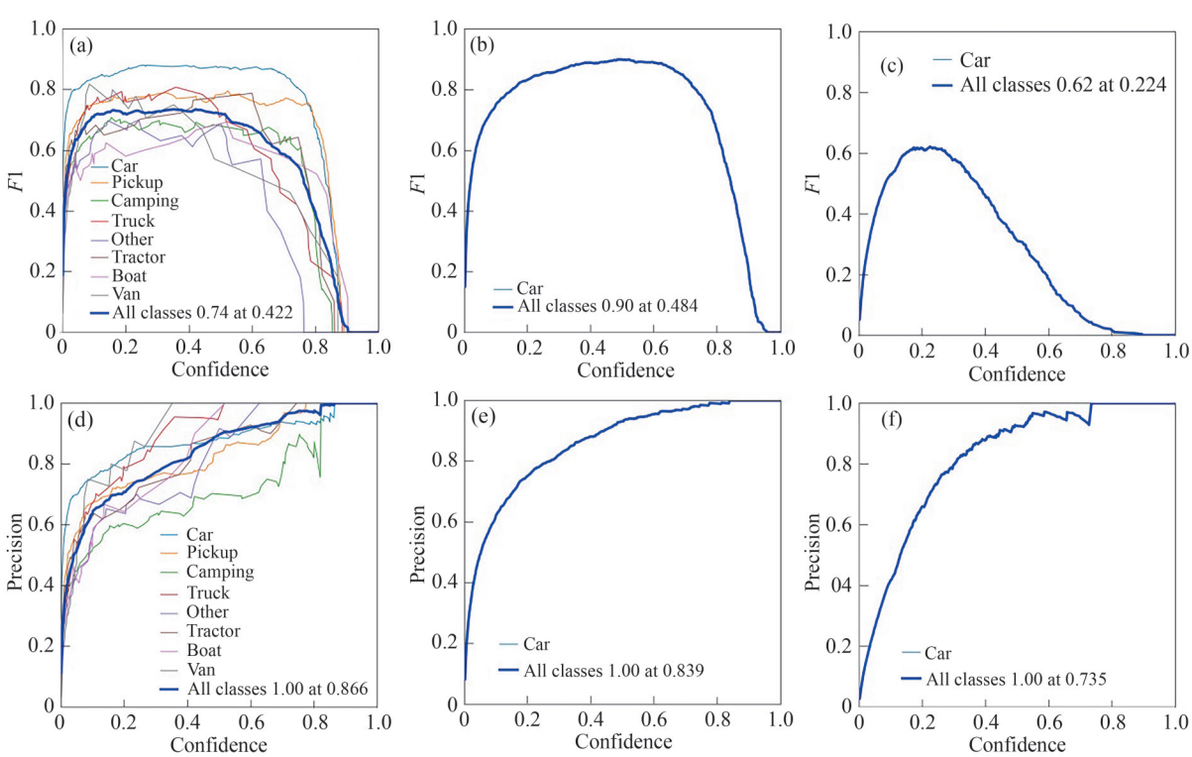


图4　不同置信度下各类别$F1$和Precision分数的变化规律。(a)～(c)原始图像；(d)～(f)混合数据集检测结果；(g)～(i) HRRSD数据集检测结果

Fig. 4　Patterns of change in F1 and Precision scores for each category at different confidence level. (a)～(c) The original images; (d)～(f) The detection results on the mixed dataset; (g)～(i)The detection results on the HRRSD dataset

达到了近100%，这表明检测结果较为准确。表1中各类别的AP值都保持了较高的水平，说明检测的召回率和准确率都较高。而在HRRSD数据集中，当置信度阈值为0.84时，所有类别的平均精确度达到了近100%。但是在混合数据集，平均精确度达到100%时，置信度阈值则需降低至0.74。以上结果证实了SuperYOLO是基于多模态信息互补的有效目标检测策略，在多光谱遥感图像小目标检测中具有较高检测性能，但是该模型在不同应用场景下也存在适用性不足的缺点。一方面，模型在单一数据集学习到特定捷径特征，如VEDAI依赖红外热特征，而HRRSD依赖纹理细节；另一方面，混合训练时梯度冲突导致特征表示退化。

表 1 各类别的AP与总类别的mAP

Table 1 AP of each category and mAP of all categories

Category	Car	Pickup	Camping	Truck	Other	Tractor	Boat	Van	mAP@0.5
AP@0.5	0.913	0.872	0.739	0.874	0.731	0.790	0.727	0.858	0.813

3.2 噪声对SuperYOLO检测性能的影响

3.2.1 图像噪声在不同数据集的影响差异性

多光谱遥感图像由于受到仪器设备的物理限制和各因素的干扰（如电子学噪声、光学噪声等），图像在获取、传输、存储过程中不可避免地发生退化，最终导致多光谱图像质量下降和光谱失真，表现为图像上的条带噪声、高斯噪声、坏线、光谱分辨率下降、谱线偏移等[18-19]。多光谱图像质量的下降导致图像的视觉效果变差，严重影响后续遥感目标探测和定量遥感应用潜力[20]。

本文以环境卫星HJ-2A实验室辐射定标采集的噪声数据为基础，通过在HRRSD数据集数据集中添加不同程度的高斯噪声的方式模拟在实际应用中多光谱遥感图像受到的噪声，并以5%、10%噪声为节点，具体分析了图像噪声对于使用SuperYOLO在不同数据集进行多光谱遥感图像小目标检测的效果影响。图5



图5 SuperYOLO添加5%和10%噪声后小目标检测结果，红色框代表目标漏检，白色框代表错检，第一列为原始图像，中间为5%噪声，第三列为10%噪声。(a)～(c)VEDAI数据集；(d)～(f)HRRSD数据集；(g)～(i)混合数据集检测结果

Fig.5 Small-object detection results of SuperYOLO after adding 5 % and 10 % noise. Red boxes=missed targets, white boxes=false alarms; Column 1 is original images, Column 2 is 5% noise; Column 3 is 10% noise. (a)～(c) fVEDAI dataset; (d)～(f) HRRSD dataset; (g)～(i) Mixed-dataset results

为添加5%、10%噪声后的检测结果。从图中可直观地看出，当添加噪声后，误检和漏检的目标数量均有所增加，甚至当噪声增至10%时，几乎检测不到目标。图5(d)～(i)分别展示了图像噪声在HRRSD数据集和混合数据集检测结果。结果表明，在HRRSD数据集在添加不同等级噪声后汽车检测精度仅为0.91和0.32，而对于混合数据集汽车检测精度为0.56和0.36。图6为添加5%、10%噪声后不同置信度下各类别F_1和P分数随置信度的变化图，发现相较于图4，F_1和P分数均有明显下降。从表2可以定量地得出，AP和mAP的大幅降低代表着检测性能明显变差。

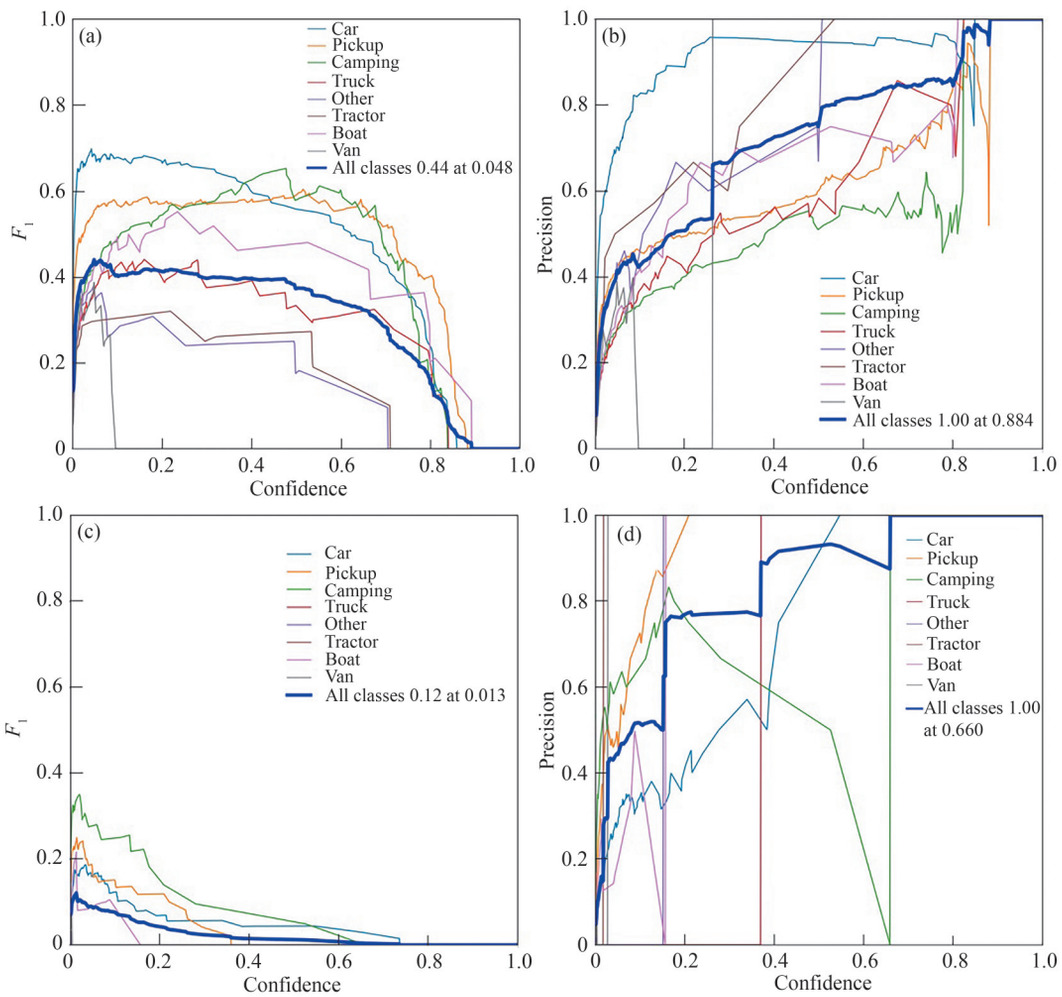


图6　不同置信度下各类别$F1$和P分数的变化规律。(a)～(b) 5%噪声；(c)～(d) 10%噪声

Fig.6　Variation patterns of F1 and P scores for different categories under various confidence levels. (a)～(b) Correspond to the case with 5% noise; (c)～(d) Correspond to the case with 10% noise

表2　VEDAI数据集添加噪声后各类别的AP与总类别的mAP

Table 2　AP of each category and mAP of all categories after adding noise for VEDAI data

Object categories (resolution)	Car	Pickup	Camping	Truck	Other	Tractor	Boat	Van	mAP@0.5
AP@0.5(0%)	0.913	0.872	0.739	0.874	0.731	0.790	0.727	0.858	0.813
AP@0.5(5%)	0.741	0.553	0.578	0.282	0.261	0.125	0.499	0.319	0.420
AP@0.5(10%)	0.096	0.065	0.264	0.000	0.000	0.000	0.030	0.010	0.058

图7为mAP随添加噪声强度的变化曲线，可以发现，mAP@0.5随噪声的增加而剧烈下降。模型的检测性能与图像噪声存在着极强的负相关关系。SuperYOLO模型通过多分支网络结构、特征融合机制和注意力模块捕获小目标的几何特征和光谱特征，如目标的形状、边缘、纹理信息及谱段差异性，但SuperYOLO模型对数据质量更加敏感，该模型在小目标检测时更多的是依赖小目标的光谱特征信息，因为噪声干扰了光

谱模型的学习，导致模型特征融合模块难以进行光谱和几何特征的融合，导致多光谱遥感图像因噪声干扰无法有效对小目标进行检测。此外，机器学习通常需要高质量数据进行训练以获得良好的泛化能力。如果训练数据包含大量低质量图像，模型可能无法学习到准确的特征表示，导致其在测试数据上的性能下降。不仅如此，低质量图像对小目标检测尤其有害，小目标本身包含的信息量有限，进一步的模糊或噪声会导致目标特征消失或与背景特征融合，模型SuperYOLO难以识别或定位这些目标。

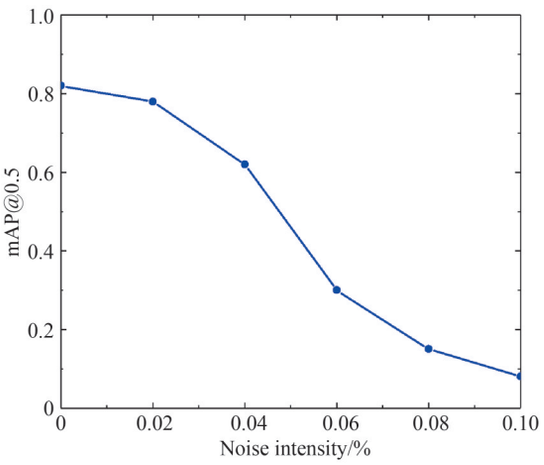


图7 mAP随添加噪声强度的变化曲线

Fig. 7 Variation curve of mAP with added noise intensity

由上述结果可得，图像噪声对使用SuperYOLO进行多光谱遥感图像小目标检测的性能影响严重，无论是检测的精确率还是召回率都有明显的下降。甚至当噪声增加至一定程度时，多个类目标无法被检测。因此，图像噪声的增强会掩盖目标的特征，使得SuperYOLO从图像中提取有用的特征变得更加困难，严重影响小目标的检测。

3.2.2 图像噪声在不同数据集的影响差异性

本研究利用DCNN对噪声数据进行了数据预处理，研究结果表明，DCNN在一定程度上去除了图像噪声的干扰（如图8）。整体上当添加5%噪声之后，汽车检测精度有一定程度较低，而当噪声水平增加至10%后，出现较多的目标无法检测。总体上，HRRSD数据集检测结果优于混合数据集。而对于HRRSD数据集，对比图8(b)和(d)可以发现经过DCNN噪声去除后整体上目标检测置信度大幅提升。同样，对比图8(c)和(e)添加10%噪声结果，可以发现经过噪声预处理后部分汽车目标能够成功检测。同样的对比结果在混合数据集也较为显著，例如图8(g)和(i)DCNN噪声处理也在一定程度上提升了目标检测的置信度，并且针对漏检目标也可成功检测。DCNN在去除图像噪声方面表现出色，能够在一定程度上提高SuperYOLO目

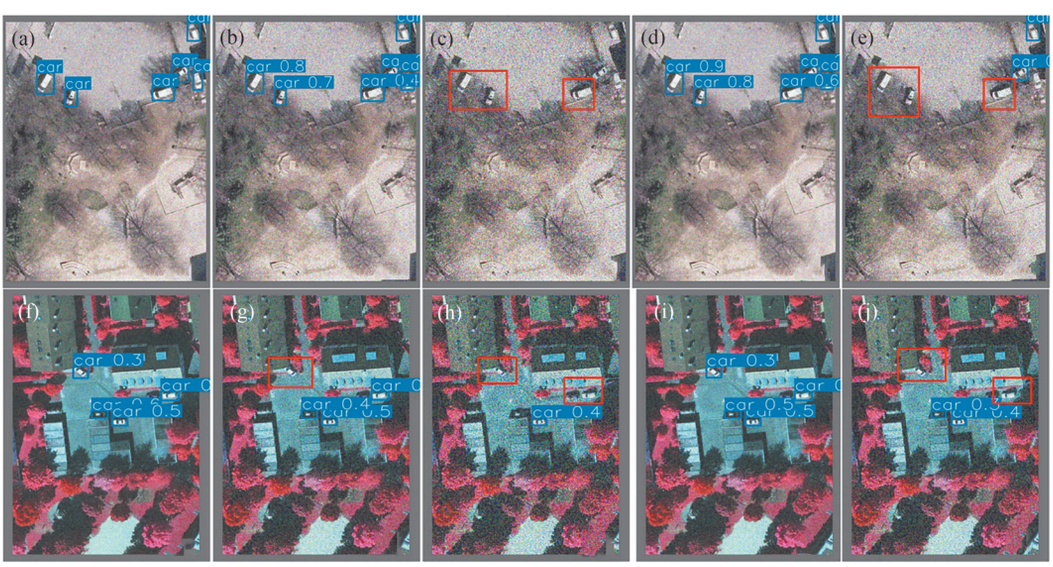


图8 SuperYOLO在数据集添加5%和10%噪声后小目标检测结果，第一行为HRRSD数据集，第二行为混合数据集检测结果。(a)和(f)为原始图像；(b)和(g)添加5%噪声；(e)和(h)为添加10%噪声的结果；(d)～(j)分别为去噪之后目标检测结果

Fig. 8 Small object detection results of SuperYOLO after adding 5% and 10% noise to the dataset, the first row shows results from the HRRSD dataset, and the second row shows results from the mixed dataset. (a)～(f) are the original images; (b) and (g) are the results with 5% noise added; (c) and (h) are the results with 10% noise added; (d)～(j) are the detection results after denoising

标检测的精度和置信度，尤其是在高噪声水平下，其优势更为明显。HRRSD数据集的检测结果优于混合数据集，这表明数据集的特性对DCNN的性能有重要影响。不同数据集可能具有不同的噪声特性、目标特性和背景复杂度，这些因素都会影响SuperYOLO检测效果。

3.3 空间分辨率与SuperYOLO检测性能的关系

多光谱遥感图像因成像载荷的硬件指标和载荷平台飞行高度的差异性，其地面分辨率差异显著。因而，地面小目标在不同分辨率的光学图像中所占的像素个数也不尽相同[21-22]。为了探究这种空间分辨率差异性是否会影响SuperYOLO的检测性能，通过下采样的方式模拟实际应用中图像空间分辨率的变化，并将原始图像空间分辨率的75%和50%，具体分析了图像空间分辨率对于使用SuperYOLO进行多光谱遥感图像小目标检测的效果影响。

图9和图10分别为VEDAI数据集图像空间分辨率为原图75%时的检测结果和各类别F_1和P分数随置信度的变化图。研究结果表明，不同类别的地面目标AP值分别为0.85、0.79、0.75、0.79、0.41、0.76、0.85、0.74，各类别的检测精度分别为0.66、0.74、0.65、0.83、0.74、0.86、0.71、0.90(如表3)，在0.5的交并比(LoU)阈值下的总体的mAP为0.76。而当空间分辨率降低至原始分辨率的50%时不同类别的地面目标AP值分别为0.65、0.53、0.63、0.48、0.30、0.62、0.24、0.27，而各类别的检测精度分别为0.70、0.51、0.54、0.48、0.69、1.00、0.76、0.66。与图3中原始图像的检测结果对比可得，仅有个别原来被正确检测的目标被误检或漏检；F_1和P分数也仅有小幅度的下降。从表3中可以定量地看出，小幅度的空间分辨降低仅造成mAP有少量降低，说明检测性能仅有轻微下降。图11为mAP随图像空间分辨率的变化曲线，可以看出空间分辨率的降低会造成检测性能的减弱，但是在较大的范围内，这种减弱程度都是轻微的，所以模型的检测性能与图像的空间分辨率存在着较弱的负相关关系。

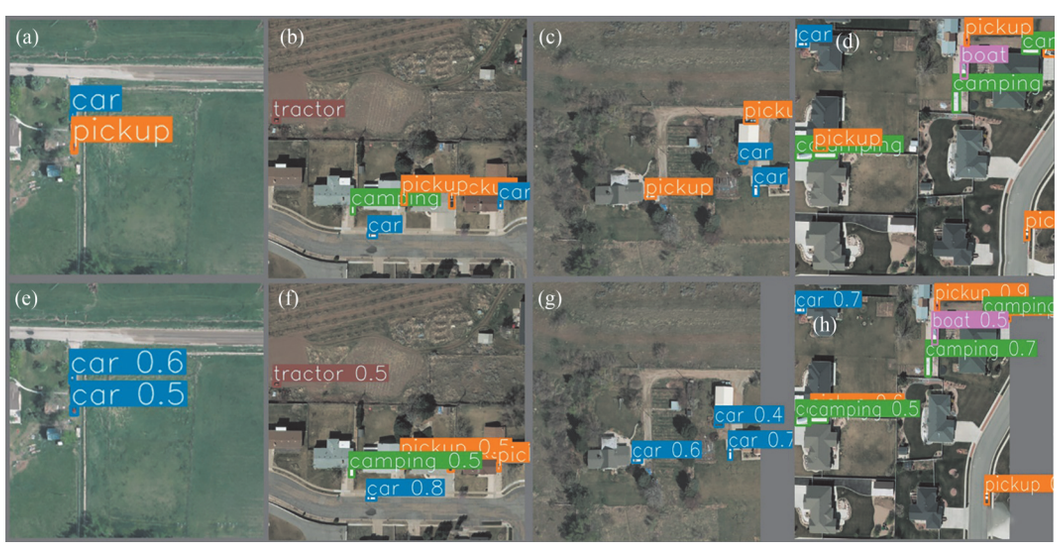


图9 降采样SuperYOLO模型视觉目标检测结果。(a)～(d)为真实标注结果；(e)～(h)为图像空间分辨率降采样至75%后的检测结果

Fig.9 Down-sampled SuperYOLO visual detection results. (a)～(d) ground-truth annotations; (e)～(h) detection results after reducing the image spatial resolution to 75%

在HRRSD及混合数据集也得到类似的结果。从以上结果可以推断出，SuperYOLO的SR分支结构在处理降分辨率图像时在一定程度上弥补了空间分辨率降低对整体模型对于小目标检测的影响，模型性能受空间分辨率的影响不大。SuperYOLO采用多尺度特征提取模块，能够捕捉小目标在低分辨率图像中的显著特征[23]。因而，即使图像空间分辨率降低，小目标的核心特征仍可以被提取，模型的检测能力明显不显著。一方面，SuperYOLO的目标检测机制能自动聚焦于关键区域，尤其是对小目标具有强烈对比的背景场景。另一方面，虽然数据集空间分辨率降低可能使边界模糊，但SuperYOLO模型的损失函数(如交并比IoU)设计使其对边界的轻微模糊较为宽容，因此，相比于图像质量的影响，图像分辨率对SuperYOLO模型检测精度的影响较低。

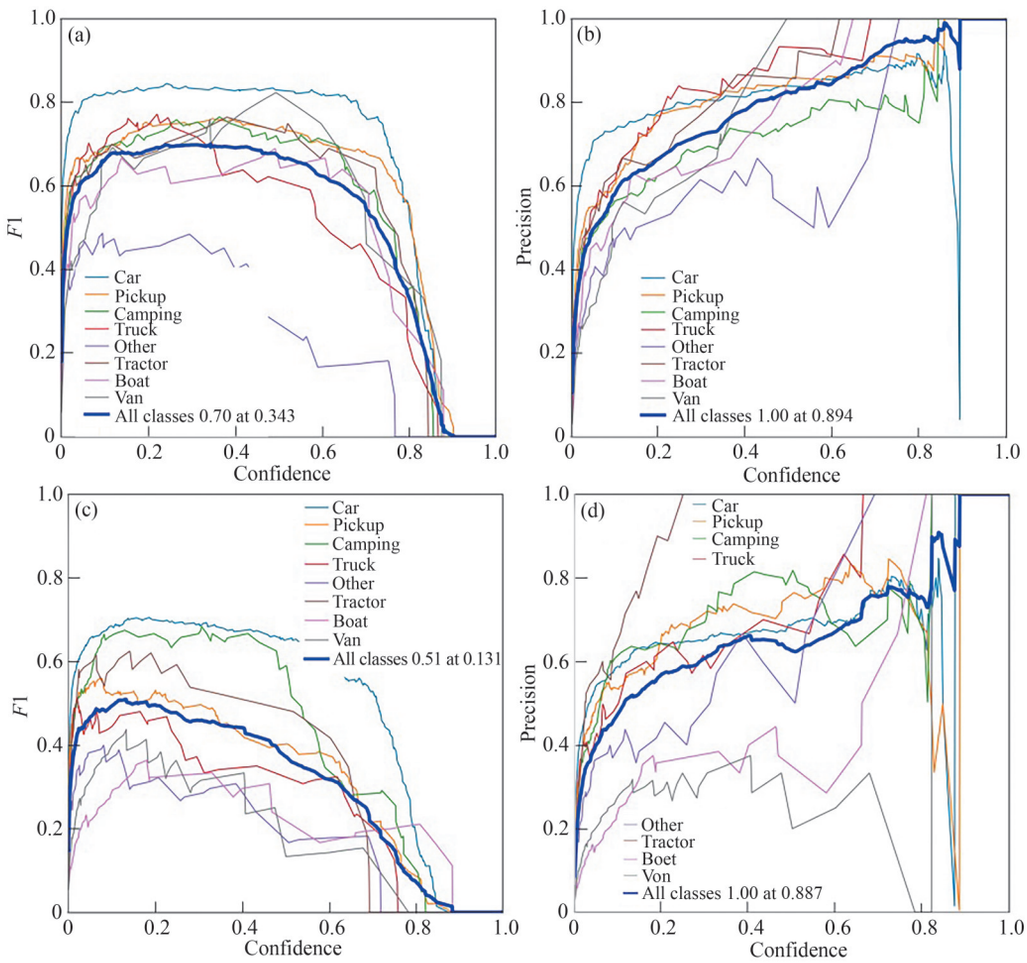


图 10 不同置信度下各类别 $F1$ 和 P 分数的变化规律。(a)和(b)为降采样至 75%；(c)和(d)为降采样至 50%

Fig. 10 Variation patterns of F1 and P scores for different categories under varying confidence levels. (a) and (b) downsampling to 75%; (c) and (d) downsampling to 50%

表 3 改变空间分辨率后各类别的 AP 与总类别的 mAP

Table 3 AP of each category and mAP of all categories after changing spatial resolution

Object categories(resolution)	Car	Pickup	Camping	Truck	Other	Tractor	Boat	Van	mAP@0.5
AP@0.5(0%)	0.913	0.872	0.739	0.874	0.731	0.790	0.727	0.858	0.813
AP@0.5(75%)	0.849	0.787	0.745	0.794	0.406	0.761	0.730	0.851	0.741
AP@0.5(50%)	0.648	0.531	0.625	0.480	0.299	0.616	0.237	0.274	0.464

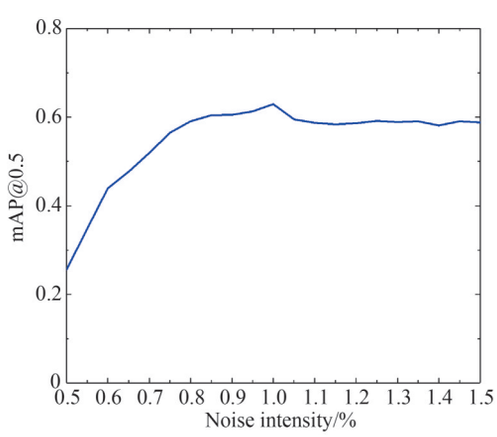


图 11 mAP 随空间分辨率的变化曲线

Fig. 11 Variation curve of mAP with spatial resolution

4 结论

本文旨在研究多光谱遥感图像不同应用场景的检测效果，探究了在实际使用中SuperYOLO对多光谱遥感图像中小目标检测效果的主要影响因素（图像噪声和图像空间分辨率）。通过在图片中添加不同程度的仪器噪声、进行不同尺度下采样的方式模拟了多光谱图像在采集、传输等过程中受到的噪声和空间分辨率的变化，定性地分析了在不同数据集（VEDAI、HRRSD及混合数据集）中SuperYOLO的性能变化规律，定量地分析了图像噪声和地面分辨率两者影响下SuperYOLO的性能变化程度。研究结果表明：1）SuperYOLO检测精度在独立数据集（VEDAI和HRRSD数据集）优于混合数据集；2）相较于空间分辨率，图像噪声会使目标特征不清晰，严重影响模型对目标特征的提取，对SuperYOLO检测多光谱遥感图像小目标有着不可忽略的影响，而在SuperYOLO增加去噪模块可以显著提升小目标检测精度；3）多光谱图像分辨率对SuperYOLO影响较小，降采样后mAP@0.5由0.813降至0.741和0.464。因此，利用SuperYOLO等深度学习算法检测多光谱遥感图像小目标时需重点关注各种环节给图像带来的噪声，要从抑制噪声和去除噪声两个方向降低多光谱遥感图像中含有的噪声，提高图像质量，保持网络模型具有良好的检测效果。

参考文献

[1] SUN Chao, ZHOU Yongkang, CHEN Zhengchao, et al, Improved small object detection in remote sensing images based on YOLOv7[J]. Remote Sensing Information, 2024, 39(4): 53-60.
孙超，周永康，陈正超，等. 改进YOLOv7的遥感图像小目标检测[J]. 遥感信息，2024，39(4)：53-60.

[2] ZHONG Yanfei, WANG Xinyu, XU Yao, et al. Mini-UAV-borne hyperspectral remote sensing: from observation and processing to applications[J]. IEEE Geoscience and Remote Sensing Magazine, 2018, 6(4): 46-62.

[3] CUI Chunhui, WANG Rugang, WANG Yuanyuan, et al. Research on optical remote sensing image target detection technique based on DCH-YOLOv7 algorithm[J]. IEEE Access, 2024, 12: 34741-34751.

[4] TU Bing, HE Wangquan, LI Qianming, et al. A new context-aware framework for defending against adversarial attacks in hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.

[5] TANG Guangyi, NI Jianjun, ZHAO Yonghao, et al. A survey of object detection for UAVs based on deep learning[J]. Remote Sensing, 2024, 16(1): 149.

[6] 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.

[7] ZHANG Jiaqing, LEI Jie, XIE Weiying, et al. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.

[8] JANG J, LEE J, KIM S. Two-phase approach for monocular object detection and 6-DoF pose estimation[J]. Journal of Electrical Engineering & Technology, 2024, 19(3): 1817-1825.

[9] FENG Zhichen, WEN Huiqing, HAN Xu, et al. A deep reinforcement learning framework for 3L-NPC-CAB converters with multiple-degree-of-freedom phase-shift control[J]. IEEE Transactions on Transportation Electrification, 2025, 11(4): 9346-9358.

[10] SENGÜL F, ADEM K. Detection of military aircraft using YOLO and transformer-based object detection models in complex environments[J]. Bilişim Teknolojileri Dergisi, 2025, 18(1): 85-97.

[11] GUO Dudu, ZHAO Chao, SHUAI Hongbo, et al. Enhancing sustainable traffic monitoring: leveraging nanosight-YOLO for precision detection of micro-vehicle targets in satellite imagery[J]. Sustainability, 2024, 16(17): 7539.

[12] PENG Hongxing, LI Zhenxing, ZOU Xiangjun, et al. Research on litchi image detection in orchard using UAV based on improved YOLOv5[J]. Expert Systems with Applications, 2025, 263: 125828.

[13] VIVONE G. Multispectral and hyperspectral image fusion in remote sensing: a survey[J]. Information Fusion, 2022, 89: 405-417.

[14] WANG Jianyao, LIAO Hongyuan, WU Yuehua, et al. CSPnet: a new backbone that can enhance learning capability of CNN[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 1571-1580.

[15] WANG Lei, HAO Yongding, PAN Mingran, et al, Improved defect detection algorithm in power inspection based on YOLOv5s[J], Computer Engineering and Applications, 2024, 60(10): 256-265.
王磊，郝涌汀，潘明然，等. 电力巡检中改进YOLOv5s的缺陷检测算法研究[J]. 计算机工程与应用，2024，60(10)：256-265.

[16] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[17] ZHANG Kai, ZUO Wangmeng, CHEN Yunjin, et al. Beyond a gaussian denoiser: residual learning of deep CNN for

image denoising[J].IEEE Transactions on Image Processing,2016,26(7):3142-3155.
[18] RAZAKARIVONY S,JURIE F.Vehicle detection in aerial imagery:a small target detection benchmark[J].Journal of Visual Communication and Image Representation,2016,34:187-203.
[19] GEWALI U B,MONTEIRO S T,SABER E.Spectral super-resolution with optimized bands[J].Remote Sensing,2019,11(14):1648.
[20] MA Tao, MA Jie, YU Kun. A local feature descriptor based on oriented structure maps with guided filtering for multispectral remote sensing image matching[J].Remote Sensing,2019,11(8):951.
[21] SUN Xuejian,ZHANG Lifu,YANG Huang, et al.Enhancement of spectral resolution for remotely sensed multispectral image[J].IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2015,8(5):2198-2211.
[22] ZHANG Weihang, SUO Jinli, DONG Kaiming, et al. Handheld snapshot multi-spectral camera at tens-of-megapixel resolution[J].Nature Communications,2023,14(1):5043.
[23] KEUM J, LEE H. Speech/music discrimination based on spectral peak analysis and multi-layer perceptron[J]. IEEE Computer Society,2006,2:56-61.

Quantifying the Effects of Sensor Noise and Spatial Resolution on Small Target Detection in Multispectral Remote Sensing Imagery

LI Zongzhe[1], WANG Qicheng[1], YANG Weihua[2], HAN Xiaoteng[2]
(1 *Equipment Management and UAV Engineering School, Air Force Engineering University, Xi'an* 710051, *China*)
(2 *Spectroscopic Imaging Technology Laboratory, Xi'an Institute of Optical Precision Machinery, Chinese Academy of Sciences, Xi'an* 710119, *China*)

Abstract: In the field of high-resolution optical remote sensing image analysis, target detection and recognition technologies serve as foundational pillars for critical applications such as environmental monitoring, urban planning, agricultural yield estimation, and military reconnaissance. In recent years, deep learning algorithms have revolutionized this domain by leveraging their exceptional ability to extract intricate features: they efficiently capture both the geometric characteristics of ground objects (e.g., shape contours, spatial distribution patterns, and structural details) and spectral attributes (e.g., band reflectance variations, spectral curve uniqueness across land covers). This capability has driven transformative breakthroughs in target detection for high-resolution optical remote sensing images, enabling more accurate and automated identification of diverse ground targets compared to traditional handcrafted-feature-based methods.

However, the practical application of deep learning in multispectral remote sensing image analysis is hindered by inherent challenges stemming from data acquisition and scene complexity. First, multispectral images exhibit significant variability in both image quality and spatial resolution due to differences in sensor platforms—ranging from satellite-borne sensors (with meter to sub-meter resolution) to Unmanned Aerial Vehicle (UAV) cameras and airborne imaging systems. These hardware discrepancies lead to inconsistencies in Signal-to-Noise Ratio (SNR), radiometric accuracy, and spatial granularity, directly impacting the reliability of feature extraction. Second, remote sensing scenes are inherently complex and variable: factors such as topographic relief, cloud cover, atmospheric scattering, and heterogeneous land cover (e.g., intermingled vegetation, buildings, and water bodies) create a cluttered background that obscures small targets (e.g., small vehicles, individual buildings, or vegetation patches). These challenges are particularly acute for small target detection in heterogeneous scenes, where traditional deep learning models often suffer from high false detection rates, missed detections, or degraded localization accuracy due to insufficient discriminative features or severe background interference.

To address these issues, this study focuses on the SuperYOLO model—a state-of-the-art object detection architecture optimized for remote sensing scenarios—and conducts a systematic investigation to evaluate its performance in small target detection across diverse multispectral datasets. The research specifically analyzes two core factors influencing detection capability: the characteristics of scene datasets and the quality/resolution of multispectral images. For dataset analysis, the study employs two types of experimental data: independent datasets (VEDAI and HRRSD) and a mixed dataset. The VEDAI dataset

specializes in small ground targets (primarily vehicles) with consistent data distribution and high annotation precision, while the HRRSD dataset covers multiple small target categories (e.g., buildings, roads, and small vegetation clusters) in complex urban and natural scenes. In contrast, the mixed dataset is constructed by fusing multispectral data from different sensors, geographic regions, and acquisition times, simulating the high data heterogeneity encountered in real-world remote sensing applications.

The research findings yield three key insights. First, dataset homogeneity significantly impacts SuperYOLO′s detection accuracy: the model achieves notably higher performance on independent datasets (VEDAI and HRRSD) compared to the mixed dataset. This is attributed to the consistent data distribution and reduced background variability in independent datasets, which minimize the model′s sensitivity to irrelevant background features; in contrast, the mixed dataset′s diverse land cover types and spectral variations increase the difficulty of distinguishing small targets from cluttered backgrounds, leading to reduced detection stability. Second, instrument noise emerges as a critical limiting factor for SuperYOLO′s precision. When 5% and 10% instrument noise (simulating real-world sensor electronic noise, radiometric distortion, and atmospheric interference) are introduced into multispectral images, the model′s mean Average Precision (AP) at an Intersection-over-Union (IoU) threshold of 0.5 (mAP@0.5) plummets sharply from the original 0.813 to 0.420 and 0.058, respectively. This degradation occurs because noise disrupts the integrity of spectral features in multispectral bands, interfering with the multi-branch network architecture of SuperYOLO—specifically impairing its ability to extract and fuse discriminative spectral and geometric features for small targets. Notably, integrating a Denoising Convolutional Neural Network (DCNN) into the data preprocessing module effectively mitigates this issue: the DCNN′s adaptive noise-filtering capability restores spectral feature integrity, leading to a significant recovery of SuperYOLO′s detection accuracy. Third, the spatial resolution of multispectral images exerts a relatively minor impact on SuperYOLO′s performance. After downsampling the original images to simulate lower-resolution scenarios, the model′s mAP@0.5 decreases from 0.813 to 0.741 (for mild downsampling) and 0.464 (for severe downsampling). This resilience is attributed to SuperYOLO′s feature pyramid structure, which enables multi-scale feature fusion to compensate for partial resolution loss—though severe downsampling still causes accuracy degradation due to the loss of fine-grained details in small targets.

In conclusion, instrument noise significantly impairs the extraction and fusion of spectral features in SuperYOLO′s multi-branch architecture, ultimately limiting its detection accuracy and practical utility for small targets in multispectral remote sensing images. Enhancing optical image quality (e.g., optimizing sensor hardware to reduce inherent noise) and incorporating denoising algorithms (such as DCNN) into the SuperYOLO framework can substantially improve the detection capability, accuracy, and applicability of small target detection algorithms in multispectral remote sensing imagery. This study emphasizes that maximizing the potential of deep learning models in remote sensing applications requires prioritizing solutions to sensor-induced noise and optimizing image preprocessing strategies. The proposed integration of denoising techniques with advanced detection architectures like SuperYOLO not only provides a more practical technical solution for small target detection in multispectral remote sensing but also points to a promising direction for future research (e.g., multi-modal data fusion detection, real-time remote sensing image processing) and engineering applications (e.g., UAV inspection systems, satellite remote sensing monitoring platforms) in this field.

Key words: Multispectral remote sensing images; Target detection; SuperYOLO; Deep learning
OCIS Codes: 100.2960; 100.4996; 110.0297; 120.4820; 120.0280
CSTR: 32255.14.gzxb20265501.0110002

Foundation item: National Natural Science Foundation of China (42576184, 42176182, 42306202)