

## 基于可见光-红外跨域迁移的红外弱小目标检测

薛如翔, 卫俊杰, 周华伟, 杨海, 王喆

引用本文

薛如翔, 卫俊杰, 周华伟, 杨海, 王喆. 基于可见光-红外跨域迁移的红外弱小目标检测[J]. 计算机科学, 2024, 51(10): 287-294.

XUE Ruxiang, WEI Junjie, ZHOU Huawei, YANG Hai, WANG Zhe. [Infrared Dim and Small Target Detection Based on Cross-domain Migration of Visible Light and Infrared](#) [J]. Computer Science, 2024, 51(10): 287-294.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

### [基于近端线性组合的信号识别神经网络黑盒对抗攻击方法](#)

Black-box Adversarial Attack Methods on Modulation Recognition Neural Networks Based on Signal Proximal Linear Combination

计算机科学, 2024, 51(10): 425-431. <https://doi.org/10.11896/jsjcx.230900054>

### [基于生成对抗网络的系统调用主机入侵检测技术](#)

System Call Host Intrusion Detection Technology Based on Generative Adversarial Network

计算机科学, 2024, 51(10): 408-415. <https://doi.org/10.11896/jsjcx.230700014>

### [基于深度学习的病理切片质量控制算法综述](#)

Review of Quality Control Algorithms for Pathological Slides Based on Deep Learning

计算机科学, 2024, 51(10): 276-286. <https://doi.org/10.11896/jsjcx.231000167>

### [基于深度学习的工业缺陷检测研究进展](#)

Research Progress in Industrial Defect Detection Based on Deep Learning

计算机科学, 2024, 51(10): 261-275. <https://doi.org/10.11896/jsjcx.230800158>

### [主观题自动评判算法研究综述](#)

Survey of Research on Automated Grading Algorithms for Subjective Questions

计算机科学, 2024, 51(10): 33-39. <https://doi.org/10.11896/jsjcx.240400008>

# 基于可见光-红外跨域迁移的红外弱小目标检测

薛如翔<sup>1</sup> 卫俊杰<sup>2</sup> 周华伟<sup>2</sup> 杨海<sup>1</sup> 王喆<sup>1</sup>

<sup>1</sup> 华东理工大学 信息科学与工程学院 上海 200237

<sup>2</sup> 上海航天控制技术研究院 上海 201109

(y30221070@mail.ecust.edu.cn)

**摘要** 红外弱小目标检测任务是红外探测领域的重点研究内容之一。然而由于其应用场景的特殊性,包含红外弱小目标的数据并不多见,且标注往往并不充分,这给由数据驱动的深度学习目标检测模型带来了挑战和困难。针对红外弱小目标数据集少、缺乏标记信息等问题,提出一种基于可见光-红外跨域迁移的红外弱小目标检测模型,将数据量更丰富的可见光域监督信息迁移到红外域中,实现红外域的无监督训练。首先,在YOLOv5的基础上设计通道增强的数据处理方法,利用低成本的通道分离技巧将可见光图像转换成类红外图像,缩小可见光域和红外域之间的模态差异。然后,构建多尺度域自适应模块,采用对抗训练的方式,对骨干网络提取得到的不同尺度特征在特征空间中进行域混淆以减小域偏移的影响,提高模型对弱小目标的检测性能。实验结果表明,所提方法改进后的模型相比各版本的YOLOv5模型检测精度均有所提升;与其他现有的无监督域自适应目标检测算法相比,所提方法在红外弱小目标的检测精度上明显占优。

**关键词:** 红外弱小目标;目标检测;深度学习;域自适应;无监督

**中图分类号** TP391

## Infrared Dim and Small Target Detection Based on Cross-domain Migration of Visible Light and Infrared

XUE Ruxiang<sup>1</sup>, WEI Junjie<sup>2</sup>, ZHOU Huawei<sup>2</sup>, YANG Hai<sup>1</sup> and WANG Zhe<sup>1</sup>

<sup>1</sup> School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

<sup>2</sup> Shanghai Academy of Spaceflight Technology, Shanghai 201109, China

**Abstract** The task of infrared dim and small target detection is one of the key research contents in the field of infrared detection. However, due to the particularity of its application scenarios, the data containing infrared dim and small targets is rare, and often not fully labeled, which poses challenges and difficulties for data-driven deep learning object detection models. In order to solve the problems of limited datasets and lack of label information, an infrared dim and small target detection model based on cross-domain migration of visible light and infrared is proposed to migrate the more abundant visible light domain supervision information to the infrared domain, so as to achieve unsupervised training in the infrared domain. First, a channel augmentation data processing method is designed on the basis of YOLOv5, utilizing low-cost channel separation techniques to convert visible light images into infrared like images, reducing the modal differences between the visible and infrared domains. Then, a multi-scale domain adaptive module is constructed, and the features of different scales extracted by the backbone network are used in the way of adversarial training. Domain confusion is performed in the feature space to reduce the impact of domain shift and improve the detection performance of dim and small target detection. Experimental results show that the improved model by the proposed method can improve the average detection precision compared to various versions of the YOLOv5 original model. Compared with other existing unsupervised domain adaptive target detection algorithms, the proposed method is obviously superior in the detection accuracy of small infrared targets.

**Keywords** Infrared dim and small targets, Object detection, Deep learning, Domain adaptive, Unsupervised

到稿日期:2023-08-03 返修日期:2023-12-28

基金项目:中国科技国防计划(2021-JCJQ-JJ-0041);中国航天科技集团有限公司第八研究院产学研合作基金(SAST2021-007)

This work was supported by the Chinese Defense Program of Science and Technology(2021-JCJQ-JJ-0041) and China Aerospace Science and Technology Corporation Industry-University-Research Cooperation Foundation of the Eighth Research Institute(SAST2021-007).

通信作者:王喆(wangzhe@ecust.edu.cn)

## 1 引言

红外探测可以穿透云层和烟雾等障碍,在白天或夜间场景下都能做到稳定探测。优良的隐蔽性和稳健的抗干扰能力使得红外成像技术在红外制导、预警防御等领域都得到广泛的应用<sup>[1]</sup>,成为当前重点研究和发展的方向之一<sup>[2]</sup>。但由于红外探测技术特殊的成像原理,远距离探测的目标在最终成像上呈现微小、无显著形态轮廓、缺乏颜色纹理等信息,容易淹没在背景中<sup>[3]</sup>,这些问题都给检测任务带来了极大的挑战和困难。在以往传统的红外弱小目标检测研究中,有许多通过基于滤波的方法来抑制背景、突出目标,实现对红外弱小目标的检测<sup>[4]</sup>;利用红外图像中目标及其局部区域灰度值突变的特点,不少研究者提出基于局部对比机制<sup>[5]</sup>和基于局部熵<sup>[6]</sup>的方法;从目标的稀疏性和背景低秩的角度出发,有研究者设计基于图像数据结构的检测算法<sup>[7]</sup>。然而,这些传统方法需要对小目标预先做出强有力的先验假设,并且当红外图像背景复杂或目标信噪比较低时无法进行有效的检测与识别。

近年来,随着数据驱动的深度学习技术的快速发展,诸如 Faster R-CNN<sup>[8]</sup>,YOLO 系列<sup>[9]</sup>的目标检测模型对常规的可见光物体的检测已经达到非常高的识别精度,基于深度学习的红外弱小目标检测研究也越来越多。为了更好地突出弱小目标,Dai 等<sup>[10]</sup>在 ACNNet 中提出非对称的上下文调制模块,除了自顶向下的全局上下文反馈外,还添加了基于点向通道注意的自底向上调制路径,以更好地交换高层语义和底层细节。为实现从信噪比低的背景中提取红外弱小目标,Zhang 等<sup>[11]</sup>设计了 ISNet 检测模型,将提取目标边缘信息的模块和注意力机制融入网络当中。Li 等<sup>[12]</sup>通过设计一种稠密嵌套注意网络 DNA-Net 来实现高层和低层特征间的交互,保持深层特征中的红外小目标信息。然而,这些方法都未考虑到红外数据稀缺且难以获取的事实。目前红外弱小目标检测技术主要应用于军事领域,可获得的相关红外图像数据集少,携带标签的数据则更少,而这会直接导致有监督的模型训练无法正常进行。

迁移学习是一种知识迁移再利用的技术<sup>[13]</sup>,利用相关领域的知识去提升目标领域任务的性能,可以减少对目标数据的依赖。通常将使用到的相关大规模数据集称为“源域”,而将目标任务的数据集称为“目标域”。域自适应作为迁移学习<sup>[14]</sup>的一种,旨在优化源域和目标域之间客观存在的分布差异,提升模型的泛化性能。Chen 等<sup>[15]</sup>首次将无监督域自适应方法应用于目标检测领域,提出的 DA Faster R-CNN 模型在图像级和实例级两个层面上利用对抗训练的方式最小化源域和目标域之间的距离。Saito 等<sup>[16]</sup>在进行域自适应时对不同层次的特征各有侧重,对于局部特征采用强对齐的方式,而对全局特征采用弱对齐,以更好地学习域不变特征。Liu 等<sup>[17]</sup>注意到带噪标签的数据集会导致有偏差的源域分布,因此提出 NLTE 模型来解决带噪标签带来的问题。相对于红外图像,可见光图像数据

丰富且易获取,可见光图像中的纹理、颜色等图像信息也有利于深度网络模型的学习。借助于跨域迁移学习,可以将可见光域中的弱小目标知识迁移到红外域中,以实现无需标注工作就可以在红外弱小目标的识别任务上达到较为理想的结果。

然而,上述域自适应迁移方法在可见光跨红外域上还缺乏应用研究。因此,本文提出一种基于可见光-红外跨域迁移的红外弱小目标检测算法,在 YOLOv5 检测模型的基础上,设计通道增强的数据处理方法和多尺度域自适应模块,优化可见光源域和红外目标域之间的分布差异,从而解决红外数据样本少和无标签的问题。

本文的主要贡献如下:

- 1)设计通道增强的数据处理方法,对可见光图像的 3 个通道进行随机选择分离,模拟“单通道”的红外图像风格。
- 2)设计多尺度域自适应模块,对骨干网络提取到的特征进行域判别,采用对抗训练混淆源域和目标域特征,拉近两域的特征分布。
- 3)利用公开的可见光和红外数据集验证了本文方法的有效性,并且与当下流行的域自适应方法相比,其更适用于可见光跨红外域的应用场景。

## 2 本文方法

本文提出一种基于可见光-红外跨域迁移的红外弱小目标检测模型,利用大规模有标签的可见光源域数据和有限的无标签红外域数据,在红外域标签稀缺的情况下实现对红外弱小目标的准确检测与识别。模型的整体框架如图 1 所示。其在 YOLOv5 检测模型的基础上进行了两大改进:1)设计通道增强策略,对输入的可见光源域数据进行通道分离和随机选择以生成“单通道”的类红外图像,缩小可见光模态和红外模态之间的差异;2)设计多尺度域自适应模块,对于主干网络提取到的不同尺度特征图,经多次卷积操作后进行像素级细粒度的域分类判别,实现在高维特征空间中拉近可见光源域和红外目标域之间的分布距离。

就具体流程而言,在模型训练阶段,有标签的可见光源域和无标签的红外目标域数据会一同输入网络中。而对于三通道的可见光图像,首先会进行通道增强的数据处理,生成类似红外图像的“单通道”图像;然后,主干网络对源域和目标域图像进行特征提取,得到不同尺度的特征图,多尺度域自适应模块则对这些特征进行域分类,判断其特征是来自于源域还是目标域,并计算域自适应损失;最后,只有带标签的可见光源域数据才会进入特征融合的颈部网络以及最后的检测头网络,进行检测损失的计算。而在模型测试阶段,输入的红外图像仅经过常规的 YOLOv5 检测网络,即不会进入多尺度域自适应模块。通道增强和多尺度域自适应模块仅在训练阶段使用,目的是提升骨干网络在红外目标域上的泛化性,更好地提取到域不变特征,从而提高对红外弱小目标的预测精度。

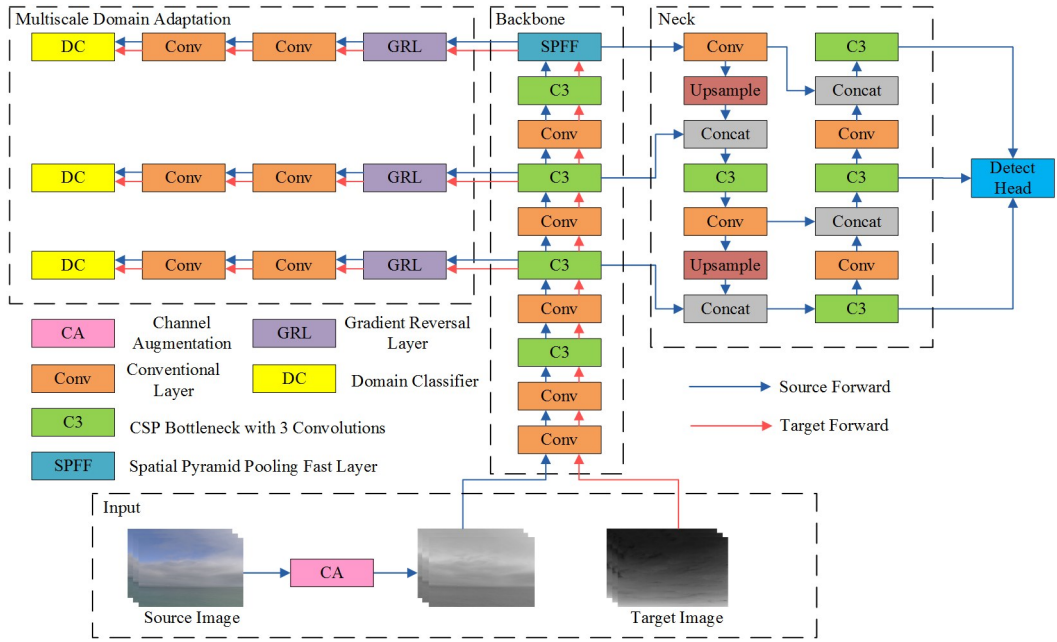


图1 可见光-红外跨域迁移的红外弱小目标检测模型

Fig. 1 Infrared dim and small target detection model based on cross-domain migration of visible light and infrared

## 2.1 YOLOv5 检测模型

YOLOv5 是一种通用的单阶段目标检测算法。其与先前的 YOLO 系列算法一样,都是对图像进行网格划分,然后进行目标的分类和位置回归预测。YOLOv5 的基本网络结构由骨干网络、颈部网络和检测头网络 3 部分组成。YOLOv5 的骨干网络集合了残差连接<sup>[18]</sup>、空间金字塔池化(Spatial Pyramid Pooling, SPP)<sup>[19]</sup>等结构的优点,可以高效地提取特征,增强网络的特征表示能力。在颈部网络中,除了使用常规的自顶向下的特征金字塔网络结构(Feature Pyramid Network, FPN)<sup>[20]</sup>,还结合自底向上的路径聚合网络结构(Path Aggregation Network, PANet)<sup>[21]</sup>,将深层和浅层的特征更好地融合在一起,获得信息更加丰富、完整的特征图。检测头网络结构设计简单,由 3 个  $1 \times 1$  卷积组成,对应于提取到的 3 个特征层。最终检测损失  $L_{det}$  包括分类损失  $L_{cls}$ 、定位损失  $L_{loc}$  和置信度损失  $L_{obj}$  3 部分组成,如式(1)所示:

$$L_{det} = L_{cls} + \lambda L_{loc} + \gamma L_{obj} \quad (1)$$

其中,  $\lambda$  和  $\gamma$  是两个超参数,用于平衡不同部分的损失。

YOLOv5 除了在网络结构上做了许多优化外,在输入数据端也做了很多的改进。例如 Mosaic 数据增强<sup>[22]</sup>和自适应锚框计算,前者将 4 张图片随机剪裁并拼接到一张图上,由于每张图都会有不同程度的缩小,因此会更有益于小目标的检测;而后者会自适应地迭代计算最适合目标任务的锚框值,也会有助于小目标的准确识别。这是本文选择 YOLOv5 作为基模型的原因。

## 2.2 通道增强策略

三通道的可见光图像相比单通道红外图像具有更丰富的颜色信息,因此两者间具有较大的模态差异。为了减少图像风格上的差距,有研究者利用生成对抗网络(Generative Adversarial Network, GAN)将可见光图像转变为类红外图像<sup>[23]</sup>。但这种生成方法需要额外的较大的计算成本且容易受噪声影响,生成的图像质量不稳定。Ye 等<sup>[24]</sup>发现,将可见

光图像的 RGB 三通道拆分成 3 个独立通道后,可以将其视为 3 个“单通道”图像,其与红外图像之间的差异比原本 RGB 图像更小,如图 2 所示。因此,本文提出通道增强的数据处理策略。如图 1 中输入端所示,在可见光源域图像输入网络前,先进行通道增强操作,转换生成类似红外图像的单通道灰度图像,再输入网络当中进行训练。

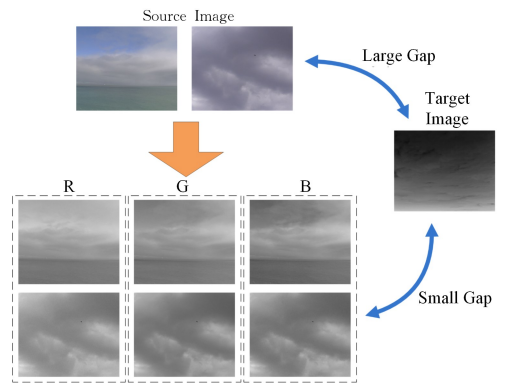


图2 通道增强

Fig. 2 Channel augmentation

通道增强策略的具体流程如图 3 所示。对于输入网络模型的可见光三通道图像,分离拆解其各个通道,并以相同的概率随机选择 RGB 中的某一通道值。由于使用的 YOLOv5 基模型要求输入的图像数据格式为三通道图像,因此对选中的通道值进行复制,代替原来的其余两个通道。最终生成的是 3 个通道信息一致的输出结果,其会代替原先的可见光图像输入网络模型中参与训练。式(2)为公式化的流程说明。

$$\begin{aligned} \tilde{x}_{s,i}^R &= (x_{s,i}^R, x_{s,i}^R, x_{s,i}^R) \\ \tilde{x}_{s,i}^G &= (x_{s,i}^G, x_{s,i}^G, x_{s,i}^G) \\ \tilde{x}_{s,i}^B &= (x_{s,i}^B, x_{s,i}^B, x_{s,i}^B) \end{aligned} \quad (2)$$

其中,  $\tilde{x}_{s,i}^R$ ,  $\tilde{x}_{s,i}^G$ ,  $\tilde{x}_{s,i}^B$  为通道增强后的图像,  $s$  表示图像来自于源域,  $i$  表示第  $i$  张输入图像,  $x_{s,i}$  表示第  $i$  张源域的输入图像。

图2给出了一些例图经通道增强后的可视化结果,观察发现,相较于原始的RGB图像,通道增强后的图像在视觉外观上更接近红外图像。

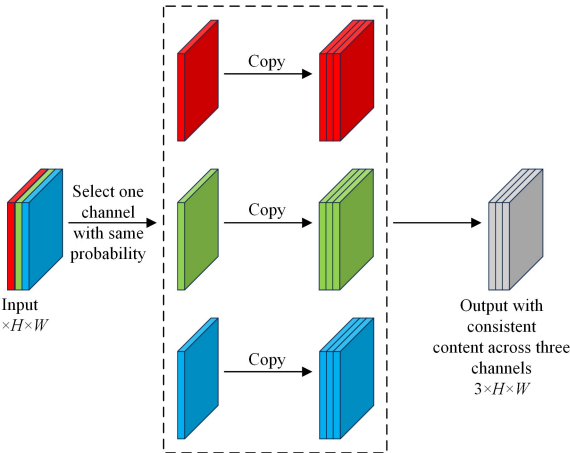


图3 通道增强流程图

Fig. 3 Channel augmentation flowchart

通道增强方法是一种简单且有效的数据处理策略,它可以将可见光图像转化成与红外图像更相似的单通道灰度图像,进一步缩小两个域之间的模态差异,令检测模型能更好地学习两域间的相似特征,最终提升模型的检测性能。同时这种方法保留了原图中的亮度分布和对比度等结构信息,图片质量稳定,受干扰少。第3.3节中的实验也可表明这种简单的通道增强策略能够有效地缩小可见光域和红外域之间的模态差异,提升模型对红外弱小目标的检测性能。值得一提的是,通道增强方法易于实现,且几乎没有额外的计算开销,省去了加权平均计算灰度图的成本,而且可以在数据加载器中与其他基础的数据增强操作无缝集成在一起,为目标检测任务提供一种简便而有效的预处理方法。

### 2.3 多尺度域自适应模块

由于可见光图像和红外图像的成像机理不同,因此两域数据样本分布之间必然存在一定的差异,这种分布不一致也被称作域偏移(Domain Shift)<sup>[25]</sup>。因此,若仅用有标签注释的可见光数据作为训练集进行训练,最终在红外测试集上的检测性能往往会有所下降,不及预期。

为了减小域偏移的影响,根据YOLOv5网络模型的特点,提出多尺度域自适应模块,充分利用可获得到的、有限的红外图像数据,辅助网络更好地提取域不变特征,在高维特征空间中将可见光域和红外域的特征进行对齐,拉近两域的分布距离,提升模型在红外域上的泛化性。值得注意的是,这里训练所使用的红外数据均是不携带标签的,因此对于红外域来说实际采取的是一种无监督的训练方式。

多尺度域自适应模块的设计如图1左侧网络所示,其输入为骨干网络提取得到3个不同尺度特征。对不同层级的特征进行特征对齐,可以更好地从局部和全局信息上拉近两域距离,更有效地提升模型的鲁棒性。选择这3个特征的原因是其会参与颈部网络特征融合操作,是YOLOv5骨干网络中最主要的3个特征,同时也能避免在全部尺度上进行域自适应计算的冗余开销。对输入的每一特征而言,其流程如图4

所示,输入的特征经过梯度反转层(Gradient Reversal Layer, GRL)<sup>[26]</sup>后,会再卷积两次。第一次的卷积操作是将各尺度的特征图通道数减半;第二次卷积操作是将各特征图的输出通道数变为1。先后两次卷积操作都不改变输入的特征图尺寸,只是减少通道数。域分类器分别对每张特征图上的每个特征点进行域分类,判别其是来自可见光源域还是红外目标域。对于最终进行域分类的特征来说,其上的每个特征点都对应于原输入图像上不同大小的感受野。因此,对每个特征点分别进行域分类判别等同于对原图像中的每个不同区域进行域标签预测,这种像素级别的细粒度特征对齐有助于消除图像风格带来的域偏移。

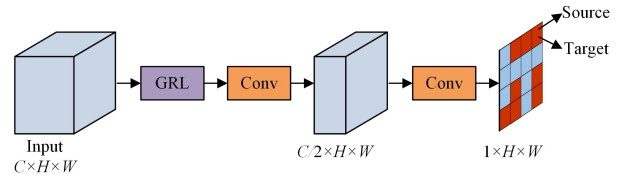


图4 域自适应模块结构

Fig. 4 Structure of domain adaptive module

域自适应损失采用交叉熵损失,如式(3)所示:

$$L_{da} = - \sum_{i,l,h,w} [d_i \log p_{i,l}^{(h,w)} + (1-d_i) \log(1-p_{i,l}^{(h,w)})] \quad (3)$$

其中, $d_i$ 表示第*i*个训练样本的域标签,若为可见光源域样本,则 $d_i=0$ ;若为红外目标域样本,则 $d_i=1$ 。对于第*i*个图像的第*l*层特征图在( $h,w$ )处的域分类器输出表示为 $p_{i,l}^{(h,w)}$ 。

整个检测模型的总损失函数如式(4)所示,包括检测损失和域自适应损失两部分。

$$L_{total} = L_{det} + \alpha L_{da} \quad (4)$$

其中, $\alpha$ 为超参数用于平衡损失。

在多尺度域自适应模块中,不同尺度的特征图进行两次卷积操作前会先经过GRL这个模块。GRL的功能十分简单,其所做的任务就是将传到本层的梯度乘以一个负数,这样会使GRL前后的网络训练目标相反,达到对抗训练的效果。可以说,GRL对前向传播不起作用,只有在梯度的反向传播中才起作用。按Ben-David等<sup>[27]</sup>提出的H-divergence计算可见光源域和红外域之间的分布距离,如式(5)所示:

$$dis_H(S, T) = 2(1 - \min(e_S(D(x)) + e_T(D(x)))) \quad (5)$$

其中, $x$ 为提取到的特征, $D$ 为域分类器, $e_S$ 和 $e_T$ 是域分类器分别在源域和目标域上的预测误差。域分类器的目标是尽可能正确地判断特征图的域标签,即最小化域分类损失。而梯度回传经过GRL模块时,其训练期望也反向了,此时骨干网络的目的是要最大化这个域分类损失,如式(6)所示:

$$\max_f \min(e_S(D(x)) + e_T(D(x))) \Leftrightarrow \min_f dis_H(S, T) \quad (6)$$

特征提取骨干网络和域分类器形成了一个对抗博弈的局面,最终会使两域之间的分布距离拉近,骨干网络朝提取域不变特征的方向训练,网络的泛化移植性也会进一步得到增强。

## 3 实验

### 3.1 数据集和实验设置

本文研究的是可见光-红外跨域迁移的红外弱小目标检测,其具体场景为空中红外弱小目标的检测,旨在实现红外

数据集有限以及标签注释缺失的情况下,利用大量的可见光图像及其监督信息作为补充以提高模型对红外弱小目标的检测性能。因此,实验使用公开的可见光弱小目标数据集和红外弱小目标数据集。模型训练时使用带标签的可见光数据和无标签的红外数据;在测试阶段则全部使用红外数据,以实现红外弱小目标检测性能的评估。

实验选择的可见光弱小目标数据集是 Drone-vs-Bird Detection Challenge<sup>[28]</sup>。该数据集中包含大量远距离的成像尺寸小的无人机目标。本实验的检测对象是红外弱小目标,因此在源域数据的选择上排除了包含大目标的视频帧。由于相邻视频帧之间相似性极高,因此进一步对视频帧序列进行不连续采样,最终筛选 5 000 张可见光图像作为源域的训练数据。图 5 展示了部分源域训练集的图片,可见其上的无人机目标尺寸均非常微小。

实验选择公开的复杂背景下红外弱小运动目标检测数据集 SAITD<sup>[29]</sup>,该数据集包括红外弱小目标仰视天空、俯视植被、俯视建筑物等多种场景,并且每种场景包含不同目标数目(1 个、2 个和 3 个)的划分。实验研究的是空中弱小目标的检测,所以挑选符合天空背景和微小目标尺寸的 10 个不同序列构建最终的评估测试集,其上的目标大小均为几像素到十几像素不等<sup>[12]</sup>,且对每种目标数目的场景序列均有所挑选,以保证实验的多样性和合理性。对挑选出来的 10 个序列用 Labelimg 进行目标标注,这些标注信息仅在测试阶段使用,用于衡量模型的准确性。此外,实验还需要部分红外图像数据作为训练集,所以在选出测试集的 10 个序列之外又分别从不同场景序列下挑选出 2 000 张图像作为训练集,这部分的训练集是无标注的。

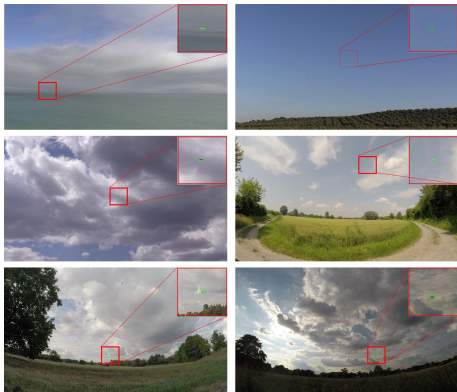


图 5 可见光数据集

Fig. 5 Visible light dataset

实验所使用的操作系统为 Linux 系统,基于 Pytorch-1.10.0 的框架实现,GPU 采用的是 NVIDIA GeForce RTX 3090。在消融实验和对比实验中,基模型采用 YOLOv5s 版本,实验设置的总训练轮次为 100,批处理大小为 16,一个批次中可见光和红外图像数目均为 8。训练采用 Adam 优化器,初始学习率为 0.01,动量为 0.937,权重衰减项为 0.0005,设置网络图像的输入大小为  $640 \times 640$ ,式(4)中用于平衡域自适应损失的超参数  $\alpha$  根据实验经验取 0.1。

### 3.2 性能指标

实验性能指标采用的是目标检测常用的评价指标:准确

率(Precision, P)、召回率(Recall, R)、F1 值、平均精度(Average Precision, AP)和平均精度均值(Mean Average Precision, mAP)。由于本实验仅针对无人机一类目标进行检测,因此 mAP 值等同于 AP 值。这些指标的具体计算式如下:

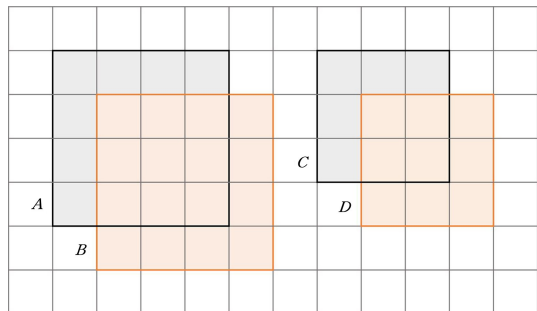
$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (9)$$

$$AP = \int_0^1 P(R) dR \quad (10)$$

其中,TP,FP, FN 分别为真正例、假正例、反例,准确率 P 为正确检出数占总检测数的比例,召回率 R 为正确检出数占有真实数的比例,F1 与 AP 可以客观地衡量模型的综合性能。P, R, F1, AP 数值越大代表算法性能越好。上述指标通常以预测框与真实框的交并比(IOU)为 0.5 时计算,但由于本实验数据的目标对象太过微小,如图 6 所示,预测框即使仅有一个像素的水平或垂直方向偏移,与真实框的 IOU 值都将有很大程度上下降<sup>[30]</sup>,因此对于只有几像素到几十像素的目标,较低的 IOU 也是可以接受的<sup>[31]</sup>。故在本实验中,将 IOU 的判定阈值下调为 0.25,放宽对正例的判断。由图 7 可直观看出,符合新设定的 IOU 值条件的预测框全为正确检测出目标的锚框。



A 与 B 的 IOU 计算:

$$IOU = \frac{|A \cap B|}{|A \cup B|} = 0.39$$

C 与 D 的 IOU 计算:

$$IOU = \frac{|C \cap D|}{|C \cup D|} = 0.28$$

图 6 小目标的 IOU 值对位置偏移敏感

Fig. 6 IOU of small targets is sensitive to position offset

### 3.3 消融实验

为了验证本文提出的通道增强策略以及多尺度域自适应模块的有效性,进行了消融实验,在 10 个不同场景的序列测试集上进行性能评估,其结果如表 1 所列。

表 1 中记录了 YOLOv5 基模型、单独添加通道增强策略的模型、单独添加多尺度域自适应模块的模型以及同时采用通道增强与多尺度域自适应模块的改进后的跨域迁移模型,在 10 个不同序列上检测精度的表现情况及其平均总体表现情况。对比第一行和第二行数据可以发现,添加通道增强策略的模型在 9 个序列上检测精度都高于 YOLOv5 基模型。其中在 Seq3 和 Seq10 上取得了最高的检测精度,平均结果相比基模型约有 4% 的精度提升,证明了通道增强策略的有效性,能够缩小两个模态之间的差异,提高检测性能。表 2 列出了定量描述通道增强策略能改善可见光域和红外域模态之间

的分布差异而计算的弗雷歇距离。弗雷歇距离是一种度量两个分布之间差异程度的指标,数值越小表示两个分布越接近。从表 2 可以看到通道增强后的可见光域与红外域两模态间的分布差异有所减小。对比第一行和第三行可以发现,添加多尺度域自适应模块的模型在 8 个序列上的检测精度均高于 YOLOv5 基模型,其中,在 Seq6 上取得了最高的检测精度,平均结果相比基模型约有 4% 的精度提升,证明了多尺度域

表 1 在 10 个序列上平均精度的消融实验结果

Table 1 Ablation experiment results of AP across 10 sequences

模型	通道增强	多尺度域自适应	AP										
			Seq1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	Seq8	Seq9	Seq10	Average
YOLOv5	×	×	0.550	0.752	0.666	0.556	0.445	0.829	0.448	0.887	0.492	0.906	0.653
	✓	×	0.616	0.703	<b>0.755</b>	0.700	0.520	0.840	0.451	0.895	0.552	<b>0.927</b>	0.696
	×	✓	0.551	0.760	0.707	0.694	0.571	<b>0.856</b>	0.496	0.861	0.562	0.867	0.693
	✓	✓	<b>0.729</b>	<b>0.861</b>	0.724	<b>0.724</b>	<b>0.619</b>	0.807	<b>0.497</b>	<b>0.913</b>	<b>0.602</b>	0.884	<b>0.736</b>

表 2 两模态间的弗雷歇距离

Table 2 Fréchet distance between two modes

模态 1	模态 2	弗雷歇距离
可见光域	红外域	197.92
通道增强后的可见光域	红外域	182.98

表 3 列出了 YOLOv5 基模型、单独添加通道增强策略的模型、单独添加多尺度域自适应模块的模型以及所提的跨域迁移方法在 10 个序列上  $P$ ,  $R$ ,  $F_1$  和  $AP$  4 个性能指标的平均表现结果。通过对比实验结果可以发现,通道增强策略和多尺度域自适应模块均可以有效提高召回率值,  $F_1$  值较 YOLOv5 基模型均有所提高,说明提出的两个子模块能更好地平衡准确率和召回率之间的关系。而融合通道增强策略和多尺度域自适应模块的跨域迁移方法在准确率  $P$  和召回率  $R$  上相比基模型都有一定程度提升,  $F_1$  和  $AP$  均最高,表明所提模型能够提升对红外弱小目标的检测性能。

表 4 不同模型在 10 个序列上平均精度的对比实验结果

Table 4 Comparative experimental results of average accuracy of different models on 10 sequences

模型	AP										
	Seq1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	Seq8	Seq9	Seq10	Average
YOLOv5	0.550	0.752	0.666	0.556	0.445	<b>0.829</b>	0.448	0.887	0.492	<b>0.906</b>	0.653
DAF	0.164	0.253	0.697	0.427	0.358	0.657	0.461	0.754	0.562	0.821	0.515
NLTE	0.189	0.304	0.659	0.450	0.357	0.702	0.466	0.805	0.593	0.856	0.538
Ours	<b>0.729</b>	<b>0.861</b>	<b>0.724</b>	<b>0.724</b>	<b>0.619</b>	0.807	<b>0.497</b>	<b>0.913</b>	<b>0.602</b>	0.884	<b>0.736</b>

YOLOv5 是本文的基线方法,其训练仅采用了源域可见光数据进行有监督训练,体现了 YOLOv5 模型本身的检测性能和泛化性能。DAF 模型是目标检测领域无监督域自适应模型的开山之作,是基于 Faster R-CNN 目标检测模型设计的跨域模型,也是许多跨域目标检测模型的对比基线。NLTE 模型是 DAF 模型的一种改进,也是最新的跨域目标检测模型。对于 DAF、NLTE 以及本文提出的跨域迁移模型,其训练集均采取了相同的设置,详见 3.1 节中的实验设置。对比表 4 中的各行数据可以发现,本文方法明显优于 YOLOv5 基模型,且在 10 个序列上的平均精度均优于并大幅度优于 DAF 和 NLTE 两个无监督域自适应模型,

自适应模块的有效性。表 1 第 4 行是集成上述两种模块提出的可见光-红外跨域迁移模型,可以发现提出的模型在 8 个序列上均优于 YOLOv5 基模型,仅在 2 个序列上稍逊色于基模型,同时在 7 个序列上取得了最高的精度,10 个序列的平均结果相比基模型有 8.3% 的大幅度提升,证明所提模型能够在跨域条件下提高对红外弱小目标的检测性能。

表 3 不同性能指标的消融实验结果

Table 3 Ablation experiment results of different performance indicators

模型	通道增强	多尺度域自适应	$P$	$R$	$F_1$	$AP$
YOLOv5	×	×	0.784	0.592	0.675	0.653
	✓	×	0.752	0.665	0.706	0.696
	×	✓	0.720	<b>0.673</b>	0.696	0.692
	✓	✓	<b>0.837</b>	0.658	<b>0.737</b>	<b>0.736</b>

### 3.4 对比实验

本文还将提出的可见光-红外跨域迁移模型与目前现有的无监督域自适应模型进行对比实验,以验证本文提出的方法是更适用于红外弱小目标检测的模型,能充分拉近可见光域和红外域之间的差异。

表 4 列出了本文提出的方法和 YOLOv5、DA Faster R-CNN(DAF)<sup>[15]</sup>、NLTE 模型<sup>[17]</sup>在 10 个红外测试序列上平均精度的表现结果。

表明了所提方法的优越性。

图 7 展示了不同模型在 SAITD 红外数据集上的检测结果图,图中红色锚框为各模型的预测框,绿色锚框为真实标注框。

由于所检测的目标非常弱小,因此在图中将存在真实框的区域局部放大。图 7 中每一行均为相同的检测图,每列图为不同的模型所检测的结果图。图 7 第一列为 YOLOv5 基模型所检测的结果图,观察图中的红色预测框和绿色真实框可以发现,YOLOv5 可以部分准确识别出弱小目标的位置,但还存在明显的漏检情况;第二列和第三列为现有的跨域目标检测模型 DAF 和 NLTE 的检测结果图。对比预测框与真实框可以发现,模型对于第一行和第二行的图片均不能准确

识别出其上的弱小目标,而且存在错检情况;对于第三行和第四行的图片能正确识别出部分弱小目标,但依然存在错检的情况。第四列为本文提出的可见光-红外跨域迁移模型的检测结果,观察图中结果可以发现,所提的方法基本能准确识别出图中的所有弱小目标,这也说明本文方法在未获取红外图像标签的情况下能够有效提高对红外弱小目标的检测能力。

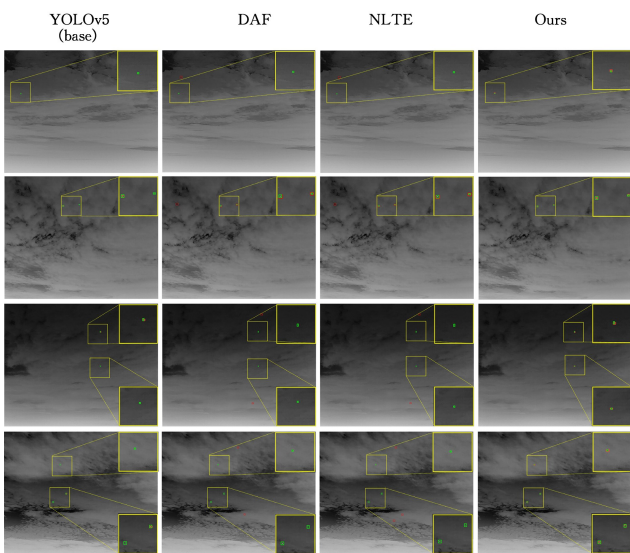


图7 不同模型的检测结果(电子版为彩图)

Fig.7 Detection result graphs of different models

### 3.5 不同 YOLOv5 版本上的性能提升表现

YOLOv5 模型有多个版本,这些版本主要在模型规模和复杂度上有所区别,在检测速度和准确性上进行了不同的权衡,以适应不同的计算能力和实时性需求。从表 5 和表 6 可以看出,YOLOv5s, YOLOv5m, YOLOv5l 这 3 个版本的计算量和参数量依次变大,推理速度依次变慢。为进一步验证本文提出的通道增强策略和多尺度域自适应模块的有效性和通用性,对改进前后的 YOLOv5m 和 YOLOv5l 分别进行实验。

表 5 YOLOv5 各版本的计算量和参数量

Table 5 Calculation and parameter quantity of YOLOv5 versions

模型版本	计算量	参数量
YOLOv5s	$16.5 \times 10^9$	$7.2 \times 10^6$
YOLOv5m	$49.0 \times 10^9$	$21.2 \times 10^6$
YOLOv5l	$109.1 \times 10^9$	$46.5 \times 10^6$

表 6 红外测试数据在 YOLOv5 各版本上的检测性能结果

Table 6 Detection performance results of infrared test data on various versions of YOLOv5

模型版本	P	R	AP	Inference speed/ms
YOLOv5s	0.784	0.592	0.653	1.2
改进 YOLOv5s	0.837	0.658	0.736	1.2
YOLOv5m	0.718	0.613	0.661	3.4
改进 YOLOv5m	0.866	0.733	0.789	3.4
YOLOv5l	0.701	0.621	0.631	5.4
改进 YOLOv5l	0.821	0.670	0.717	5.4

改进前的各版本基线均是利用单源域的有标签的可见光数据进行训练;改进后的模型为跨域迁移模型,同时利用有标签的可见光源域和无标签的红外目标域数据进行训练。

实验结果如表 6 所列,可以发现改进后模型的检测性能均优于相应的基模型,YOLOv5s, YOLOv5m, YOLOv5l 各版本模型的 AP 值分别有 8.3%, 12.8%, 8.6% 的提升,证明了本文方法的有效性和通用性。对比各个基模型可以发现,YOLOv5m 模型的 AP 达到最高,YOLOv5l 模型的 AP 并非最高,说明对于小目标的检测而言,并不是模型越大、网络越深越好。改进后的 YOLOv5s 模型显著优于原 YOLOv5m 模型的检测性能,AP 超出 7.5%,再次说明所提方法模块的优越性。此外从推理速度上也可以发现,改进前后的推理速度是相同的,因为所提方法仅作用于模型训练阶段,不会在测试阶段参与推理过程,因而不会产生额外的时间消耗。

**结束语** 本文提出一种基于可见光-红外跨域迁移的红外弱小目标检测模型,充分利用大规模可获得的可见光图像监督信息,以在红外数据集少、标签稀缺的情况下依旧能够准确识别出红外弱小目标。针对可见光与红外图像之间模态差距大的特点,设计通道增强的处理方法,通过随机选择可见光图像三通道中的某一通道信息来模拟“单通道”类红外图像。同时,为进一步拉近可见光域与红外域之间的分布距离,设计多尺度域自适应模块,在多个层级的高维特征空间中,通过对抗训练的方式缩小域间差异以提升模型在红外域上的泛化性。在公开数据集上的实验证明,所提方法能够有效提升模型对红外弱小目标的检测识别能力。同时可以发现检测精度依然存在可提升的空间,因此下一步将会继续研究如何增强对弱小目标的特征提取能力,以进一步提高对红外弱小目标的检测精度。本文实验测试评估中下调 IOU 阈值的措施,在背景更加复杂、目标更加密集的场景下,会引起虚警率的增加,因此如何设置更合理的评价指标也是未来研究小目标检测所需要进一步考虑的问题。

### 参考文献

- [1] ZHU Y, HAO Y G, WANG H Y. Deep learning based salient object detection in infrared video [J]. Computer Science, 2023, 50(9): 227-234.
- [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] QIN Y, LI B. Effective infrared small target detection utilizing a novel local contrast method [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1890-1894.
- [4] JIAO J, XIE Y J, ZHANG H L, et al. Infrared dim and small target detection based on background prediction by wavelet filter [J]. Computer Science, 2016, 43(2): 60-63.
- [5] CHEN C P, LI H, WEI Y T, et al. A local contrast method for small infrared target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(1): 574-581.
- [6] DENG H, SUN X, LIU M, et al. Infrared small-target detection using multiscale gray difference weighted image entropy [J]. IEEE Transactions on Aerospace and Electronic Systems, 2016, 52(1): 60-72.
- [7] GAO C, MENG D, YANG Y, et al. Infrared patch-image model for small target detection in a single image [J]. IEEE Transac-

- tions on Image Processing, 2013, 22(12):4996-5009.
- [8] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.
- [9] 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. Las Vegas, NV, USA; IEEE, 2016:779-788.
- [10] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 2021:950-959.
- [11] ZHANG M, ZHANG R, YANG Y, et al. ISNet: Shape matters for infrared small target detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA; IEEE, 2022:877-886.
- [12] LI B, XIAO C, WANG Y, et al. Dense nested attention network for infrared small target detection [J]. IEEE Transactions on Image Processing, 2022, 32:1745-1758.
- [13] ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning [J]. Proceedings of the IEEE, 2020, 109(1):43-76.
- [14] FARAHANI A, VOGHOEI S, RASHEED K, et al. A brief review of domain adaptation [C] // Advances in Data Science and Information Engineering: Proceedings from ICDATA 2020 and IKE 2020. 2021:877-894.
- [15] CHEN Y, LI W, SAKARIDIS C, et al. Domain adaptive faster r-cnn for object detection in the wild [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA; IEEE, 2018:3339-3348.
- [16] SAITO K, USHIKU Y, HARADA T, et al. Strong-weak distribution alignment for adaptive object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA; IEEE, 2019:6956-6965.
- [17] LIU X Y, LI W Y, YANG Q S, et al. Towards robust adaptive object detection under noisy annotations [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA; IEEE, 2022:14207-14216.
- [18] 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. Las Vegas, NV, USA; IEEE, 2016:770-778.
- [19] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1904-1916.
- [20] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA; IEEE, 2017:2117-2125.
- [21] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA; IEEE, 2018:8759-8768.
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M, et al. YOLOv4: Optimal Speed and Accuracy of Object Detection [J]. arXiv; 2004. 10934, 2020.
- [23] XIA D X, LIU H J, XU L L, et al. Visible-infrared person re-identification with data augmentation via cycle-consistent adversarial network [J]. Neurocomputing, 2021, 443:35-46.
- [24] YE M, RUAN W J, DU B, et al. Channel augmented joint learning for visible-infrared recognition [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021:13567-13576.
- [25] LUO Y W, ZHENG L, GUAN T, et al. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA; IEEE, 2019:2507-2516.
- [26] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks [J]. The Journal of Machine Learning Research, 2016, 17(1):2096-2030.
- [27] BEN-DAVID S, BLITZER J, GRAMMAR K, et al. A theory of learning from different domains [J]. Machine Learning, 2010, 79:151-175.
- [28] COLUCCIA A, FASCISTA A, SCHUMANN A, et al. Drone-vs-bird detection challenge at IEEE AVSS2021 [C] // 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance(AVSS). IEEE, 2021:1-8.
- [29] SUN X L, GUO L C, ZHANG W L, et al. A dataset for small infrared moving target detection under clutter background [EB/OL]. <https://doi.org/10.11922/sciencedb.j00001.00>.
- [30] XU C, WANG J W, YANG W, et al. Dot distance for tiny object detection in aerial images [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021:1192-1201.
- [31] LIU T, ZHAO Y, WEI Y C, et al. Concealed object detection for activate millimeter wave image [J]. IEEE Transactions on Industrial Electronics, 2019, 66(12):9909-9917.



**XUE Ruxiang**, born in 1999, postgraduate. Her main research interests include image processing and object detection.



**WANG Zhe**, born in 1981, Ph.D, professor, is a member of CCF(No. 16666M). His main research interests include pattern recognition and image processing.