

引用格式: CAO Zhaorui, Li Xinye, PAN Qian, et al. Research on Multi-source Image Transfer Fusion and Target Recognition Algorithm Based on Generative Adversarial Network[J]. Acta Photonica Sinica, 2026, 55(8):0810003

曹昭睿,李昕烨,潘骞,等. 基于生成对抗网络的多源图像迁移融合与目标识别算法研究[J]. 光子学报, 2026, 55(8):0810003

基于生成对抗网络的多源图像迁移融合与目标识别算法研究

曹昭睿¹, 李昕烨², 潘骞¹, 叶明亮²

(1 沈阳理工大学 装备工程学院, 辽宁 沈阳 110159)

(2 沈阳理工大学 机械工程学院, 辽宁 沈阳 110159)

摘 要:针对现有红外迁移算法存在的泛化性弱、细节表现率不足和结构冗余等问题,面向无人武器低成本可见光系统探测光谱域拓展和低照度下可见光成像系统拓展光谱探测域的需求,提出一种基于可见光图像对红外图像迁移方法。通过构建一种强泛化红外迁移生成器,完成可见光特征向红外特征的转化与迁移;设计轻量特征敏感对抗鉴别器并利用二元交叉熵对数损失函数进行训练矫正,提升特征提取能力与捕捉图像细节的能力;通过梯度残差语义感知融合模块对伪红外图像与可见光图像进行融合,实现复合频段下多尺度目标的精确识别。结果表明,所设计的网络能够有效地生成目标红外物理特征,减小了伪红外图像与真实红外图像的差异,降低了低光环境下目标识别成本。在全天候复杂环境条件下,针对多尺度热源目标识别任务具有良好的计算精度与鲁棒性。

关键词:图像融合;生成对抗网络;多光谱成像;目标识别

中图分类号: TP391.41

文献标识码: A

doi: 10.3788/gzxb20265508.0810003

0 引言

随着现代战争向去中心化、多场景突发形态演进,诸如对“弹簧刀”、“柳叶刀”等低成本无人武器的需求日益凸显。俄乌战场昼夜持续消耗、美军舰载无人机全天候侦查等实战案例均表明,全天候作战与探测能力是无人武器掌握战场主动权的关键^[1]。高精度制冷型中远距离红外光电载荷虽探测性能优越,但仍存在成本高、体积重量大、装机门槛高的问题;而普通低成本可见光载荷成本低、体积小、适装性强,却缺少红外全天候探测能力,与低成本无人武器的载荷约束、性价比定位存在突出矛盾,成为制约其全天候作战效能释放的核心瓶颈。因此,研发低成本、高效率的方法以拓展无人武器光电系统全天候作战时域,是突破这一技术瓶颈、提升实战价值的关键方向^[2]。

无人武器察打任务可分为前序探测与后序指控两大关键环节:前序探测由武器端实施包含图像采集、特征提取、目标识别等步骤,完成战场信息获取与目标初步分析;后序指控为探测结果回传至后方,由人员依据结果下达指令。当前多数无人装备聚焦前序探测,后方人员仅依据最终识别结果决策,少参与前序过程。即便具备自主决策能力的装备,后方人员也仅评估任务有效性,不直接操控作战回路。此模式下,前序探测最终识别结果一致则后续指令趋同^[3]。这一现象揭示两者相对解耦特性,即探测结果相近时作战效果相似。因此,只要使无人武器端载低成本探测器的前序探测识别结果趋近高成本红外成像结果,即可实现作战效能等效。基于这一契机,利用可见光、微光图像向红外图像进行风格迁移融合与识别的策略,成为了当前国内外研究的焦点。

常规机器视觉低照度迁移主要基于深度卷积神经网络实现内容与风格解耦融合,构建内容风格联合损失函数监督,生成兼具目标内容与光谱特征的图像。这种方法存在迭代耗时、实时性差、泛化性弱、细节表

现力不足、高分辨率适配有限及内容风格融合失衡等问题。直到2014年, Ian Goodfellow等人开创性地提出生成式对抗网络(Generative Adversarial Networks, GAN)^[4], 核心机制是借助对抗训练策略, 促使生成器与判别器在相互博弈的动态过程中持续优化自身性能, 进而生成具备高度逼真性和高质量的样本数据, 为后续相关研究和应用奠定了坚实的理论基础。同年, 有条件的对抗生成网络cGAN在运行时不仅要生成图像, 还需要满足所设定的条件, 从而缓解了GAN网络生成结果的不确定性^[5]。2015年, Gatys等人提出的基于深度学习的图像风格迁移方法^[6], 为该领域带来了具有里程碑意义的重大突破。2016年, 快速风格迁移方法通过设计并训练特定的生成模型, 成功实现了实时风格迁移的效果, 极大地提升了图像风格迁移技术在实际应用场景中的可行性和实用性^[7]。2017年, 以CycleGAN为代表的算法能够在无监督的条件下使用非成对数据完成风格转换任务, 有效突破了传统方法对大规模标注数据的依赖, 显著拓宽了图像风格迁移技术的适用范围和应用场景^[8]。2022年CAST^[9]、Jojogan^[10]和Dct-net^[11]仍在GAN网络上改进不断取得成果。然而, 上述方法在训练效率、生成图像质量、复杂环境下高维图像的目标检测与多频段数据融合方面仍然存在一定缺陷^[12], 导致难以应用在无人武器装备低照度图像增强实战环境中应用。

因此, 本文面向无人武器低成本可见光系统探测光谱域拓展的需求, 针对现有的算法在泛化性弱、细节表现率不足和结构冗余等问题, 对目标位于红外和可见光频段下的图像特征进行深入研究, 设计了一种使用残差结构并引入注意力的多源图像迁移识别网络(Multi-source Image Migration Recognition Network, MIMRN)。该方法的主要创新点为: 1) 生成器部分使用多级连接残差代替原来的U型结构; 调整了鉴别器部分PatchGAN中卷积层的层数; 避免其结构带来的冗余, 提高了训练速率, 更易达到纳什均衡点, 增强了结构相似性; 2) 引入可变形卷积和瓶颈注意力机制, 通过更加灵活的采样, 提取到更具判别力的特征, 增强模型对各种图像变化的鲁棒性; 3) 设计轻量化双阶段卷积并将原来的L1范数损失替换为二元交叉熵对数损失, 对高频边缘信息与低频渐变信息有更好的捕捉效果。基于上述创新点, 所提出的MIMRN有效提升了伪红外图像的生成质量。同时, 通过采用伪红外图像替代真实红外图像开展模型训练, 提高了伪红外图像下的目标识别精度, 进一步将可见光图像与伪红外图像进行融合处理后, 目标识别性能得到额外增强, 最终实现以可见光相机弥补红外相机的技术目标, 强化了低成本武器平台的全天候目标侦察与探测能力。

1 算法结构

本文所设计的多源图像迁移识别网络由对可见光图像特征提取并生成的生成器、更易达到纳什均衡点的鉴别器、多图像融合模块和融合目标识别模块四个核心部分构成, MIMRN网络结构如图1所示。具体工作流程如下: 首先, 可见光图像(Input Visible Image, V_i)将输入强泛化红外迁移生成器(Strongly Generalized Infrared Transfer Generator, SGITG)中, 经其内部的目标偏移可变形卷积(Target Offset Deformable Convolution Network, TODCN)、轻量化双维度瓶颈注意力机制(Lightweight Dual-Dimension Bottleneck Attention Mechanism, LLDBAM)与多级连接残差模块(Multi-level Connected Residual Module, MCRM)提取目标特征, 引导模型聚焦目标物体所在区域并忽略背景干扰, 准确提取目标的形状、纹理等特征。将 V_i 与随机噪声向量进行映射转换, 生成伪红外图像(Fake Infrared Image, I_f)并使其尽可能地接近真实红外图像(Real Infrared Image, I_r)。随后MIMRN将 I_r 和 I_f 一同输入含有轻量化双阶段卷积(Lightweight Two-stage Convolution, LT Conv)的轻量特征敏感对抗鉴别器(Lightweight Feature-Sensitive Adversarial Discriminator, LFAD)中。

此时, SGITG与LFAD展开二元博弈对抗, 通过判别图像的真伪, 反向推动SGITG生成更为逼真的 I_f 。随着对抗的持续, 当SGITG与LFAD都无法进一步提升性能时, 二者达到纳什均衡状态, LFAD难以区分图像真伪, 对 I_r 和 I_f 的判断概率趋近于0.5。之后, 含有梯度残差语义融合模块(Gradient Residual Semantic Fusion Module, GRSFM)的梯度残差语义融合网络(Semantic Perception Fusion Network, SPFN)会对 V_i 与SGITG输出的 I_f 进行融合, 生成能同时包含最佳的高频纹理信息和低频热源信息的输出融合特征图(Output Fused Feature, F_o)。最后, 利用YOLO V8目标识别网络对 F_o 下的多尺度融合特征图进行识别, 从而精准确定融合伪红外图像中目标的位置与类别, 得到最终的识别图像(Fused detection, F_d)输出。

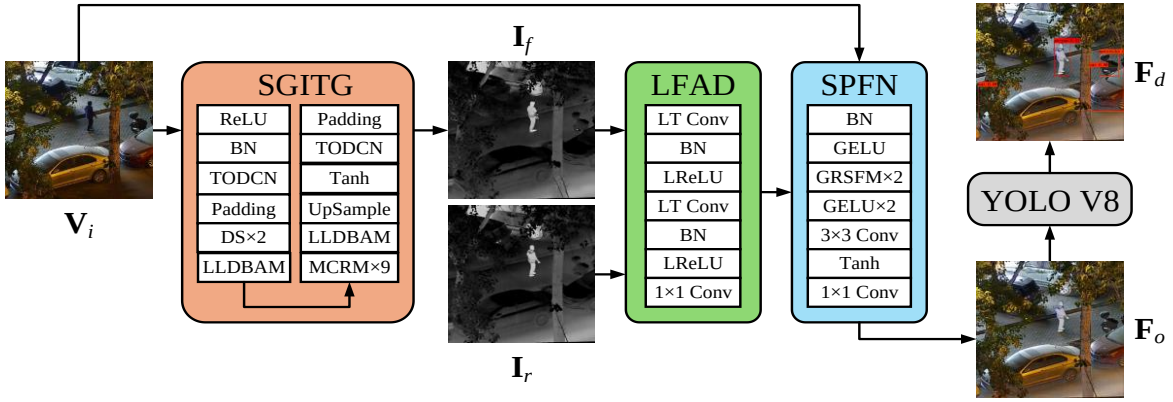


图1 MIMRN网络结构示意图
Fig.1 Network structure of MIMRN

2 算法实现

2.1 强泛化红外迁移生成器

SGITG作为首要处理模块,承担着MIMRN网络中多重关键任务,而在这一过程中特征提取尤为关键。只有对输入图像的特征进行极为细致入微的提取,充分捕捉图像中的每一处细节信息,生成器才有可能生成质量上乘、效果最佳的伪红外图像。这一过程不仅要求生成器具备强大的特征感知能力,更需要其在复杂的网络结构中,将提取到的特征准确转化为伪红外图像所需的特征表达,从而确保生成的伪红外图像能够在视觉效果和实际应用中达到最优状态^[12]。

为实现这一目标,SGITG首先通过TODCN进行初始特征提取,利用其自适应采样特性捕捉目标动态特征;随后经2层下采样模块扩大感受野,获取全局特征信息;紧接着通过LLDBAM进行通道与空间双重注意力加权,聚焦目标区域并抑制背景干扰;之后经9次MCRM循环计算,深化特征提取并缓解梯度消失问题;再通过1层LLDBAM强化关键特征筛选;最后经上采样模块恢复分辨率,由输出层TODCN精准捕捉目标边界细节,完成伪红外图像生成。基于上述结构与计算流程,SGITG即可实现可见光到红外特征的高效、精准模仿。

2.1.1 目标偏移可变形卷积

在无人武器对地探测任务中,低照度环境会导致图像信噪比降低、目标与背景对比度弱化,背景像素干扰权重放大,使得地面小型目标、低空飞行物等关键目标的边界与轮廓信息难以捕捉。常规卷积核形状固定,采样位置僵化,无法适配目标形态变化与低光环境下的特征分布,导致生成的伪红外图像存在小目标细节丢失、边缘模糊等问题,进而影响后续目标检测精度^[13]。

为了解决这一问题,本文在SGITG中引入原理如图2所示的TODCN。TODCN通过一个额外的卷积层计算出每个采样点的偏移量,然后将这些偏移量应用到后续的卷积操作中。因TODCN中卷积采样位置是可变的,所以可以更好地对图像中的重要目标进行特征提取、更精确地分割出不规则物体的边界、生成更符合目标形状和细节的图像^[14]。

对于一个 3×3 的卷积核,坐标点 R 如式(1)所示:

$$R = \{(-1, -1), (-1, 0), (-1, 1), \dots, (1, 0), (1, 1)\} \quad (1)$$

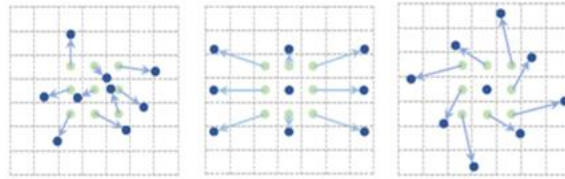


图2 目标偏移可变形卷积原理图
Fig.2 Principle of TODCN

则 TODCN 需要在该标准卷积基础上,赋予每个权重参数一个偏移量 Δp_n 。设 $\omega(p_n)$ 卷积核对应权重、 $x(p_0 + p_n)$ 为输入特征图上 $p_0 + p_n$ 位置处元素值,则 TODCN 每处元素值 $y(p_0)$ 可表示为:

$$y(p_0) = \sum_{p_n \in R} \omega(p_n) * x(p_0 + p_n + \Delta p_n) \quad (2)$$

Δp_n 由输入特征图通过一个卷积单元计算生成且通常为小数,而小数无法对应特征图上的实际像素点,因此需要通过双线性插值转换出对应点的最终像素值。TODCN 能够有效提升卷积核中小目标的占比,进而显著增强网络对目标边界以及小目标的特征提取能力。然而,若将网络中所有的标准卷积核均替换为可变形卷积核,将会导致网络运算量呈指数级增长,严重影响运算效率。因此,在综合考虑无人武器端载设备算力与 SGITG 图像生成需求的前提下,本文将按照如图 1 所示的结构,将 SGITG 中下采样层之前的第一层卷积以及上采样层之后的输出层中的标准卷积替换为 TODCN。与此同时,其余网络层依然使用标准卷积。通过这种方式,SGITG 在保障运算效率的同时,最大化发挥了 TODCN 对目标初始特征与最终细节的强化作用,实现了特征提取效果与算力开销的平衡优化。

2.1.2 多级连接残差模块

在低照度无人武器对地视觉探测任务中,特征提取网络过浅易导致网络整体在训练与性能优化中出现梯度消失、目标特征利用效率低下的问题,导致空对地视角下小目标特征淡化甚至丢失。而为追求小目标特征提取能力去深化网络层又将带来参数量、计算成本、过拟合几率提升的风险。因此,为增强 SGITG 在实际场景中应对特征提取需求复杂变化的能力,本文设计了结构如图 3 所示的 MCRM。

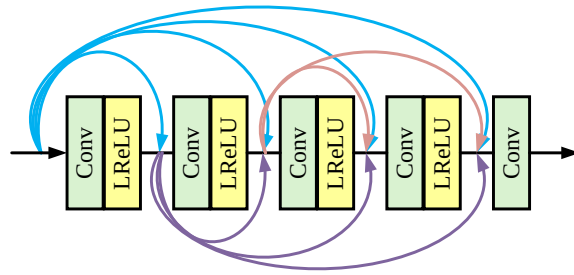


图 3 多级连接残差模块结构

Fig.3 Structure of MCRM

MCRM 的基础结构为含非线性变换的串联卷积层,每层由常规卷积与 Leaky ReLU (LReLU) 构成。MCRM 通过特征复用与梯度传递优化策略,使网络每层可直接调用浅层细粒度特征与深层语义特征,减少冗余的同时挖掘不同层级特征的互补价值。多路径梯度传递设计则让每层梯度可从输出层直接获取,缩短梯度传递距离,有效缓解深层网络梯度消失问题,保障参数同步高效更新。此外,MCRM 通道级联还能促使不同层级特征自动完成隐式融合,模型可通过训练自适应分配特征权重,无需额外设计显式融合策略,进而提升融合效率与灵活性。

基于上述原理,MCRM 在参数量更少的前提下实现更强特征提取能力,降低存储与计算成本。而在超分辨率等图像恢复任务中,MCRM 能够精准捕捉多尺度高频细节,为目标纹理、边缘恢复提供支撑,与残差学习协同强化特征传递稳定性。且该结构适配无 BN 层设计,避免低光场景下 BN 可能引入的图像伪影,保障伪红外图像生成质量,有效平衡了特征提取精度与算力开销,降低低光无人武器端载设备的算力资源开销。

2.1.3 轻量化双维度瓶颈注意力机制

低照度环境图像中目标与背景对比度低,小目标特征易被背景噪声掩盖,导致生成器难以精准聚焦关键目标区域。为解决这一问题,本文基于 pix2pix 的残差结构构建 LLDBAM,从通道与空间双维度强化网络对小目标信息的聚焦能力,同时保持轻量化特性以适配端载设备算力需求^[15]

LLDBAM 以空间尺寸为 $H \times W$ 、通道数为 C 的特征图 $U \in R^{C \times H \times W}$ 为输入,将依次通过通道注意力模块(Channel Attention Module, CAM)与空间注意力模块(Spatial Attention Module, SAM)^[16-17]。在 CAM 中将首先对 U 进行全局平均池化,得到编码了每个通道全局信息的通道向量,之后通过多层感知机对通道向量

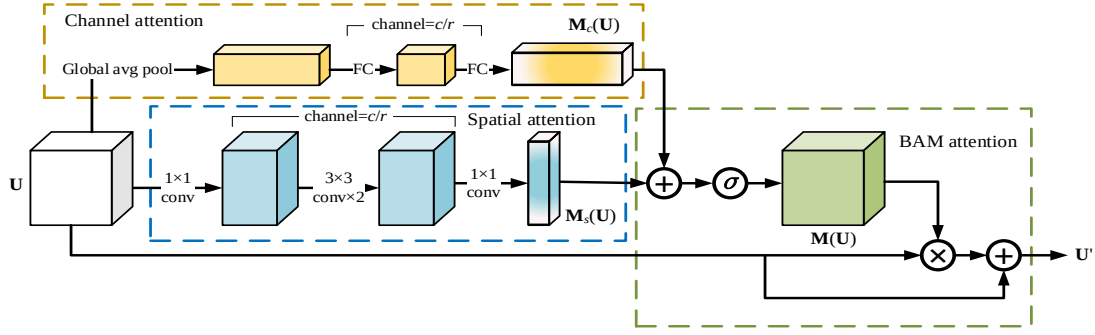


图4 轻量化双维度瓶颈注意力机制结构
Fig.4 Structure of LLDBAM

进行处理,其感知机内部隐层激活大小设为 $R^{c/r \times 1 \times 1}$, r 为以减少参数为目标的缩减比。最后归一化层调整尺度得到如式(3)所示的通道注意力向量 $M_c(U)$,其中 $W_0 \in R^{c/r \times C}$ 、 $W_1 \in R^{C \times C/r}$ 为多层感知机权重矩阵、 $b_0 \in R^{C/r}$ 、 $b_1 \in R^C$ 为偏置项。

$$M_c(U) = BN(W_1(W_0(AvgPool(U) + b_0) + b_1)) \quad (3)$$

在另一分支SAM中网络将聚焦目标在不同空间下的位置特征。SAM先使用 1×1 卷积将特征图投影到低维以整合和压缩通道维度;接着使用两个 3×3 的扩张卷积扩大感受野,提升上下文语义信息利用率;最后再用 1×1 卷积将特征降维,并在末尾添加批量归一化调整尺度,从而得到如公式(4)所示的空间注意力特征 $M_s(U)$,其中 $f_0^{1 \times 1}$ 、 $f_1^{3 \times 3}$ 、 $f_2^{3 \times 3}$ 、 $f_3^{1 \times 1}$ 分别为上述过程中产生的特征图。

$$M_s(U) = BN(f_3^{1 \times 1}(f_2^{3 \times 3}(f_1^{3 \times 3}(f_0^{1 \times 1}(U)))))) \quad (4)$$

最后LLDBAM将按照如公式(5)所示过程整合 $M_c(U)$ 与 $M_s(U)$,生成注意力特征 $M(U)$,并与 U 进行叠加与累加运算,得到如公式(6)所示的输出特征图 U' 。

$$M(U) = Sigmoid(M_c(U) + M_s(U)) \quad (5)$$

$$U' = U \oplus U \otimes M(U) \quad (6)$$

感受野大小扩张值 d 和注意力分支容量缩减比 r 是调控LLDBAM的两个关键超参数。经多次实验,在设定 $d=4$ 、 $r=16$ 时LLDBAM具有最佳效果,因此选定该参数进行网络的训练与测试。基于这一原理,LLDBAM与TODCN、MCRM相协同,在SGITG形成特征提取、梯度优化、注意力聚焦的递进计算链路,提升生成器对小目标、弱特征的捕捉能力,确保伪红外图像中目标细节的真实性与完整性。

2.2 轻量特征敏感对抗鉴别器

MIMRN的核心在于SGITG与LFAD之间的动态对抗博弈训练,使二者共同推动 I_r 生成质量的迭代提升。若生成器性能过优,鉴别器无法提供有效误差反馈,将导致生成器优化停滞;若鉴别器性能过强,则生成器因判别误差过大难以收敛^[18]。同时,考虑到无人武器端的低算力约束,鉴别器需在保证特征判别精度的前提下实现轻量化设计。因此,本文以平衡算力开销与弱特征敏感性为目标,通过集成轻量化双阶段卷积增强对伪红外目标的特征敏感性,并配合二元交叉熵对数损失函数调控博弈关系,实现判别性能与计算效率的最优平衡,为SGITG提供精准且高效的误差反馈。

2.2.1 轻量化双阶段卷积

在低照度环境下无人武器平台对地目标识别任务中,伪红外图像与真实红外图像的差异多表现为微弱热源纹理、低信噪比边缘等细粒度特征,且因机载平台算力有限,常规卷积网络存在以下两项缺陷:一是固定核权重缺乏对低光场景动态噪声、光影突变的适应性较差,难以捕捉微弱特征差异;二是深度网络为提升性能需增加参数量,导致计算开销激增、无法适配低算力条件,而轻量级网络若仅缩减核尺寸,又会因感受野不足,难以兼顾低光图像全局光影分布与目标局部细节的提取,最终影响鉴别器对真假红外图像的判别精度^[19]。

视觉模型中的Token作为连接图像原始像素与高层语义特征的关键中间单元,其关联聚合能力直接影响特征表达与计算效率的平衡^[20]。以此为切入点,本文借鉴人类视觉系统动态异尺度感知机制,设计了如

图5所示、采用大核感知配合小核聚合的轻量化双阶段卷积,在低计算开销下实现全局上下文与局部细节的精准提取。

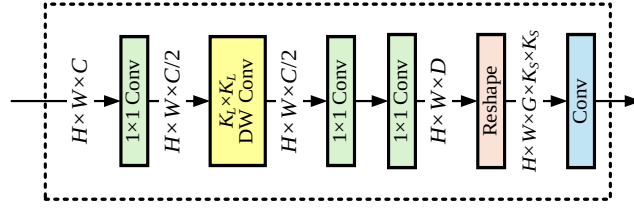


图5 轻量化双阶段卷积结构

Fig.5 Structure of LT Conv

输入特征图 $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ 首先进入LT Conv中大核感知模块,通过 1×1 逐点卷积将通道维度压缩至 $C/2$ 以降低计算开销,再通过核尺寸为 7×7 的大核深度卷积(DW Conv),在极低计算成本下高效扩展感受野,充分捕捉全局空间上下文关系,最后通过另一层逐点卷积生成编码全局Token关联信息的上下文自适应权重。之后,该权重传入小核聚合模块,先按默认组数 $G = C/8$ 的策略对输入特征图通道进行分组,每组包含 C/G 个通道以实现分组内权重共享,再将自适应权重重塑为 $G \times 3 \times 3$ 的动态核,基于该动态核对每个Token对应的 3×3 邻域特征执行卷积操作,完成局部特征的自适应精准聚合,最终输出融合全局上下文与局部细节的特征图。LT Conv通过大核感知全局光影分布并以动态小核聚合局部微弱特征,增强了对低光场景下真假红外图像差异的敏感性。同时其轻量化设计通过通道压缩、深度卷积与分组卷积的组合,在保障特征提取精度的前提下降低了参数量与计算开销。此外动态核机制能够灵活适应低照度环境下的噪声变化,相比固定核卷积层能更精准的捕捉真、伪红外图像间的机械差异,推动SGITG与LFAD快速达到纳什均衡。

2.2.2 二元交叉熵对数损失函数

在MIMRN的对抗训练框架中,SGITG与LFAD的博弈平衡直接决定 $\mathbf{I}_f$ 的生成质量,而损失函数作为二者性能优化的核心导向。考虑到无人武器需在全天候复杂环境中实现实时、高精度目标识别,损失函数不仅要能精准量化真假红外图像的差异,还需具备训练稳定性强、计算开销低、泛化性优的特点,以支撑网络的稳定、高效运行。

传统图像迁移任务中常使用L1范数损失函数,该方法虽对异常值具有一定鲁棒性,但在低照度环境下,可见光图像信噪比低、目标细节模糊,伪红外图像与真实红外图像的差异主要体现在高频边缘纹理和低频热源分布的细微差异上,而L1范数以绝对误差均值为优化目标,对这类细粒度差异的捕捉敏感性不足,易导致生成图像出现细节丢失、边缘模糊等问题。同时,L1损失在反向传播过程中易出现梯度稀疏现象,尤其在训练后期,生成器难以通过梯度更新进一步优化细节生成效果,影响模型收敛速度与最终泛化能力^[21]。

故针对上述问题,本本引入二元交叉熵对数损失函数(Binary Cross Entropy with Logits Loss, L_{BCEL})。 L_{BCEL} 直接将 sigmoid 激活函数与二元交叉熵损失集成,无需额外手动执行激活操作,既简化计算流程又避免梯度消失或梯度爆炸风险^[22],其表达式如式(7)所示。

$$L_{BCEL}(y, y^*) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\sigma(y_i^*)) + (1 - y_i) \log(1 - \sigma(y_i^*))] \quad (7)$$

其中 N 为训练批次中的样本数量; $y_i \in \{0, 1\}$ 为第 i 个样本的真实标签, $y_i = 1$ 为真实红外图像 $\mathbf{I}_r$, $y_i = 0$ 为伪红外图像 $\mathbf{I}_f$; y_i^* 为未经过 sigmoid 激活的 LFAD 线性输出; $\sigma(\cdot)$ 为 sigmoid 激活函数。 L_{BCEL} 通过逐样本计算真实标签与预测概率的对数误差并取平均,实现对鉴别器分类性能的精准量化。当为真实红外图像($y_i = 1$)时, L_{BCEL} 聚焦 $\sigma(y_i^*)$ 趋近于 1 的概率;当输入为伪红外图像($y_i = 0$)时, L_{BCEL} 则聚焦 $\sigma(y_i^*)$ 趋近于 0 的概率,从而引导 LFAD 精准区分真假红外图像的特征差异,同时为 SGITG 提供明确的误差反馈。

L_{BCEL} 通过直接量化对 $\mathbf{I}_f$ 与 $\mathbf{I}_r$ 的判别概率误差,能够精准引导 LFAD 强化对红外图像物理特征的判别敏感性,同时推动 SGITG 针对性优化 $\mathbf{I}_f$ 与 $\mathbf{I}_r$ 的差异,尤其在低照度环境下,可有效缩小 $\mathbf{I}_f$ 在热源纹理、目标边界等关键特征上与 $\mathbf{I}_r$ 的区别,为后续目标识别提供高质量的图像输入。此外, L_{BCEL} 将 sigmoid 激活与交叉熵损失集成的设计,既简化计算流程、降低算力开销,又避免了梯度消失或爆炸问题,保障训练稳定性与实时性。

同时, L_{BCEL} 对真假样本的细微差异敏感性强,能精准捕捉高频边缘与低频渐变特征,且梯度分布均匀,可缓解深层网络梯度消失,强化模型对低照度环境下小目标细节的还原能力与复杂场景的泛化鲁棒性。

综上,在MIMRN的整体优化目标中,总损失函数 L_{total} 由对抗损失 $LFAD_v$ 与 L_{BCEL} 加权构成, L_{total} 的表达式如式(8)所示。

$$L_{total} = L_{adv} + \lambda L_{BCEL} \quad (8)$$

其中超参数 λ 用于平衡对抗训练与细节优化的权重,经实验测试, $\lambda=1.2$ 时MIMRN获得最佳效果。该组合既通过对抗损失保障 I_f 的整体红外特征一致性,又借助 L_{BCEL} 的精准量化特性强化细节真实性与训练稳定性,最终实现SGITG与LFAD的高效纳什均衡,为后续可见光与伪红外图像的融合及目标识别提供拟真数据源。

2.3 语义感知融合网络

由于现有图像融合方法多聚焦表层视觉质量评估,难以满足低照度环境下空中无人武器在目标识别、跟踪等任务中对语义信息的需求,即便融合图像定量指标优异,仍可能因低光下小目标细节丢失、边界模糊导致后续检测失效^[23]。为解决这一问题,本文以平衡待检测目标的语义特征凝练引导与细节强化为目标,构建了网络结构如图6所示的梯度残差语义融合模块,以实现 V_i 与 I_f 的高效融合,为后续目标识别提供兼具高频纹理细节与低频热源语义的优质特征输入。

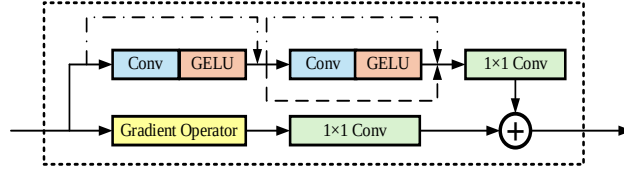


图6 梯度残差语义融合模块结构

Fig. 6 Structure of GRSFM

GRSFM采用双分支整体框架,依次完成特征提取、语义融合、图像重建流程。通过引入语义损失反馈机制,将目标识别任务所需的语义信息反向传递至融合网络,使融合过程不再局限于像素级视觉优化,而是针对性保留原图像中支撑目标判别的关键信息,确保融合图像与原图像特征等效。GRSFM的主路径采用密集连接结构,包含2个均搭配GELU激活函数的 3×3 卷积层与1个 1×1 卷积层,分别提升特征非线性表达能力与整合通道数,该路径通过特征密集复用强化低光环境下目标的深层语义表征,弥补暗光导致的特征辨识度不足问题。此外,GRSFM的残差路径引入梯度运算(Gradient Operator),通过计算特征梯度尺度并经 1×1 卷积消除通道差异,精准捕捉小目标边缘轮廓、纹理等细粒度细节,缓解背景像素干扰带来的细节丢失。

综上,SPFN通过GRSFM中融合主路径深度语义特征与残差路径细节梯度特征的结构,即提供了目标全局热源语义信息,保障多尺度目标的整体特征完整,又能够补充高频细节信息,解决伪红外图像清晰度不足、可见光图像低光下噪声干扰的问题。这一方法可使 F_o 既具备红外图像的热源辨识度,又保留可见光图像的纹理细节,有效提升了复杂环境下目标与背景的区分度,降低了后续目标识别网络的误检、漏检风险。

3 实验结果

为验证算法在伪红外图像生成与目标识别任务中的性能,本文设计多组对照实验,依次开展损失函数优化验证、真伪红外训练效果对比、各频段识别性能测试、多尺度目标识别验证及算法横向对比。

3.1 实验设置

为检验MIMRN的伪红外图像生成及其目标识别效果,本文以8864张自建数据集联合9233张FLIR双光数据集作为训练样本,对真实红外图像和伪红外图像中的行人与车辆目标进行标注,训练集和验证数据集比例为9:1,训练周期为160次,输入图片尺寸为 640×480 。所用计算机CPU为英特尔I5-8300H, GPU为NVIDIA GTX 3080。网络训练过程中优化器为SGD,动量参数设置为0.937, batch size设置为2,并在训练阶段应用Mosaic数据增强方法。算法所处系统环境为Ubuntu 20.04,网络编写语言为Python 3.11,配合

Pytorch作为辅助高级API。

3.2 损失函数优化验证

为筛选适配低光场景的最优损失函数,本文针对L1范数损失及KL散度损失、余弦嵌入损失、软间隔损失、均方差损失和二元交叉熵对数损失等五项替代损失函数展开对比分析,其训练损失函数曲线如图7所示。

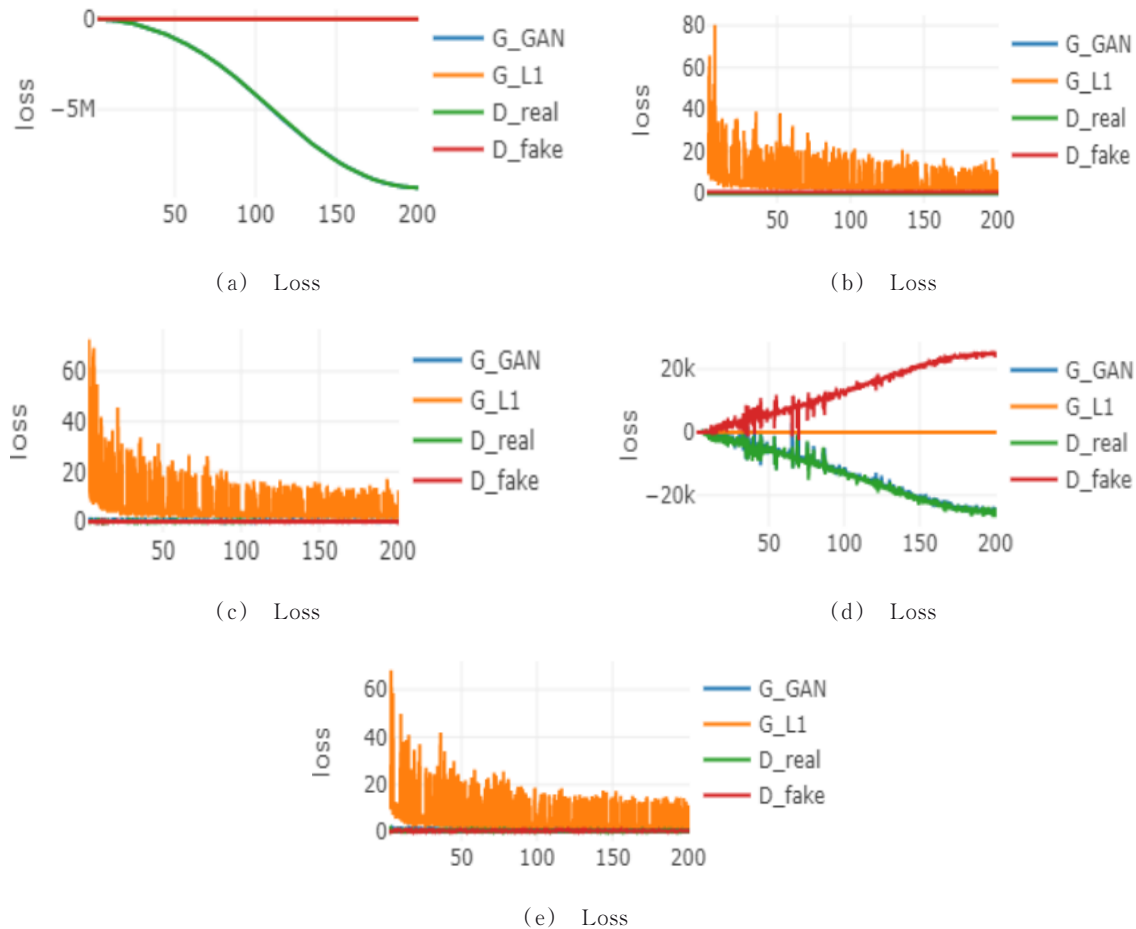


图7 不同损失函数对比训练曲线。(a)KL散度损失;(b)软间隔损失;(c)均方差损失;(d)余弦嵌入损失;(e)二元交叉熵对数损失;子图横坐标:训练迭代轮次,纵坐标:损失值。

Fig. 7 Comparison of training curves with different loss functions. (a) KL divergence loss; (b) Soft margin loss; (c) Mean square error loss; (d) Cosine embedding loss; (e) Binary cross-entropy logarithmic loss. Horizontal axis: training epoch; Vertical axis: loss value.

如图7所示,L1范数损失对异常值具备一定鲁棒性,但易导致模型最大损失处于较高水平;KL散度损失训练前期不稳定,鉴别器对可见光图像的判别损失需180个epoch后才趋于平稳;余弦嵌入损失下D_fake与D_real始终大小相等且符号相反,损失波动大,稳定周期长达170个epoch;软间隔损失与均方差损失虽将最大损失降至80左右,但稳定后仍存在小幅波动,难以进一步提升精度。而本文采用的二元交叉熵对数损失表现最优,模型最大损失降至60左右,仅需120个epoch即可趋于稳定,最终总损失降至15左右。这一结果表明,二元交叉熵对数损失函数不仅能够有效降低模型的最大损失,还能加快损失函数的收敛速度,减小了稳定时的总损失以及损失波动,使模型训练更为稳定与高效。

此外,为从客观角度评价不同损失函数对伪红外图像生成质量的影响,本文通过均方误差(Mean Square Error, MSE)、结构相似性(Structural Dissimilarity, SSIM)和峰值信噪比(Peak Signal to Noise Ratio, PSNR)来评价所生成图像的质量。实验以不同损失函数计算下生成的5张伪红外图像作为样本,分别计算上述评价指标的均值,所得到的不同损失函数图像生成质量评价结果如表1所示。

表 1 不同损失函数图像生成质量评价结果

Table 1 Evaluation results of image generation quality with different loss functions

Loss function	MSE	SSIM	PSNR
L1 norm loss	5.12	0.903	33.9
Mean squared error loss	4.87	0.911	34.4
KL divergence loss	4.88	0.908	34.3
Mean squared error loss	5.83	0.882	32.8
Cosine embedding loss	4.95	0.910	34.2
Binary cross-entropy log loss	4.57 ↓	0.920 ↑	34.9 ↑

实验结果显示,使用二元交叉熵对数损失后生成图像 MSE 显著下降,降幅达 10.7%;SSIM 和 PSNR 得到上升,分别提升 1.9% 和 2.9%。至此证明了 L_{BCEL} 能够从网络训练的角度,增强伪红外图像生成效率、质量与稳定性。

3.3 真伪红外训练效果对比

为验证伪红外图像替代真实红外图像训练的可行性,本文分别采用真红外图像与伪红外图像训练 MIMRN,不同训练图像下损失曲线与均值平均精度曲线分别如图 8、图 9 所示。

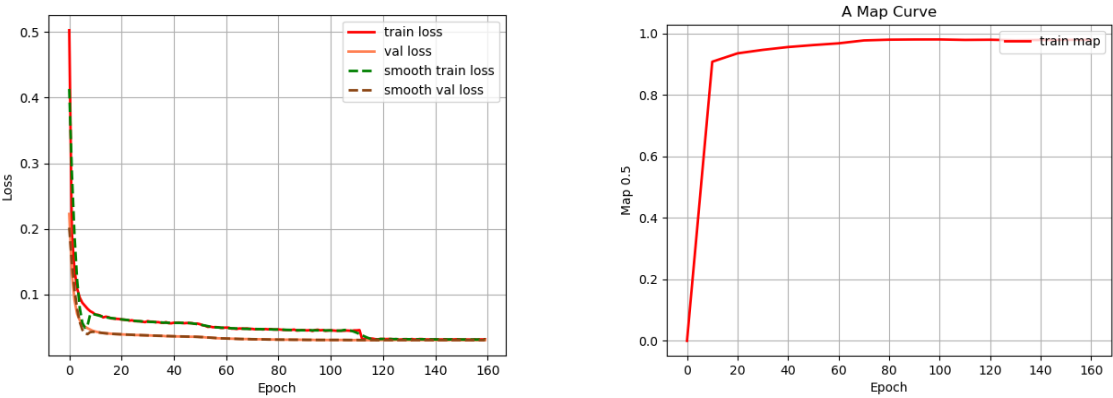


图 8 真红外训练损失与均值平均精度曲线。(a)损失函数变化曲线;(b)平均精度变化曲线

Fig. 8 Training loss and mean average precision curve for real infrared images.(a) Loss function curve; (b) Mean average precision curve.

(a)Loss (b)Loss

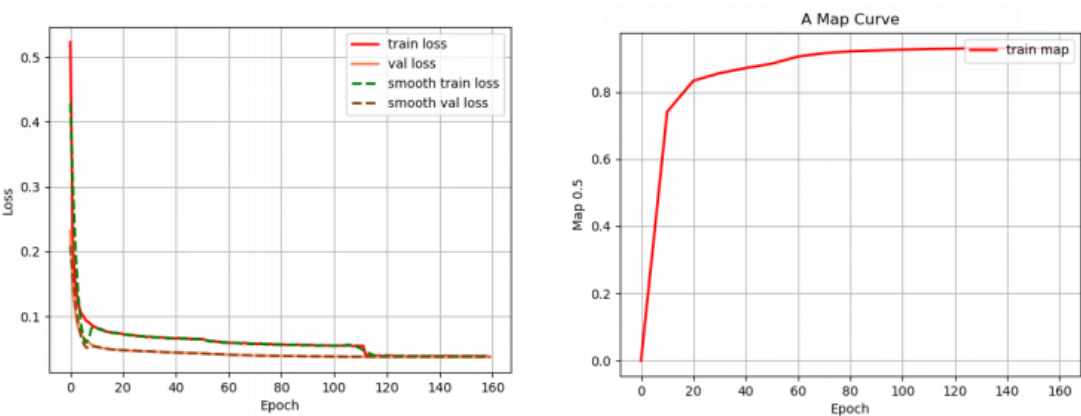


图 9 伪红外训练损失与均值平均精度曲线。(a)损失函数曲线;(b)平均精度曲线

Fig. 9 Training loss and mean average precision curve for fake infrared images.

(a)Loss (b)Loss(a) Loss function curve; (b) Mean average precision curve.

对比图 8 和图 9 可知,真红外图像训练初始训练损失 (train loss) 为 0.503,最终降至 0.031;验证损失 (val loss) 初始值 0.223,稳定于 0.030。伪红外图像训练初始 train loss 为 0.529,最终降至 0.038;val loss 初始值

0.231, 稳定于 0.037, 两类图像训练的损失值均实现有效收敛。识别精度方面, 在历经 160 个训练 epochs 后, 真红外图像训练后, 车辆目标检测准确率 (Precision) 为 94.63%、召回率 (Recall) 为 94.95%、平均准确率 (AP) 为 97.35%; 行人目标 Precision 为 97.73%、Recall 为 96.87%、AP 为 98.56%, mAP 达 97.96%。伪红外图像训练后, 车辆目标 Precision 为 94.09%、Recall 为 87.46%、AP 为 94.09%; 行人目标 Precision 为 96.46%、Recall 为 88.44%、AP 为 91.87%, mAP 为 92.98%。综上可知, 因真红外图像质量更高, 使其训练效果更优, 但伪红外图像训练仍能达到较高识别精度且与真红外图像差距较小, 验证了其替代真实红外图像的可行性。

3.4 各频段识别性能测试

为直观分析单一频段与融合后图像的识别效果, 利用 MIMRN 分别对可见光、真红外、伪红外及融合图像开展目标识别测试, 其不同频段图像识别结果如图 10 所示。

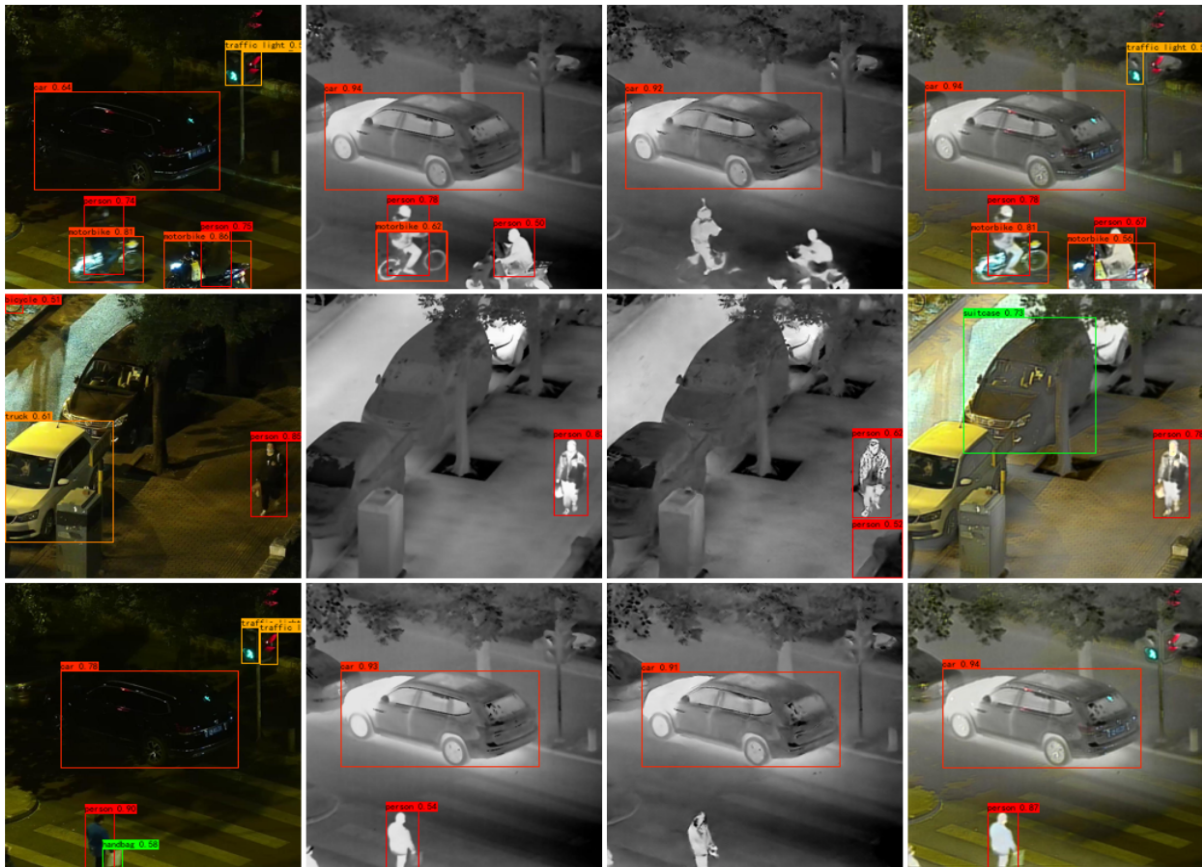


图 10 不同频段图像识别结果 (a) 可见光图像 (b) 真红外图像 (c) 伪红外图像 (d) 融合图像

Fig.10 Image recognition results of different frequency bands (a) Visible light image; (b) True infrared image; (c) Fake infrared image; (d) Fused image.

(a) (b) (c) (d)

从图 10 的测试结果可以看出, 在使用单一通道进行识别时, 图 10(a) 所示的可见光频段能够对部分光线较好的车辆与行人进行识别, 但无法识别特征不明显的行人与黑暗环境下的黑色汽车; 图 10(b) 所示的红外频段图像能够识别出车辆与行人, 但由于特征信息量较少、维度较低, 语义信息提取不够充分, 对未启动的低热源车辆的识别精度相对可见光频段更低。图 10(c) 所示的伪红外图像识别结果, 相较于图 10(b) 的真实红外图像稍差, 但具备基本纹理特征, 因清晰度不足使得部分目标难以精准识别; 图 10(d) 所示的融合识别策略借助双重频段的语义信息, 能够精准分割并标记目标区域与背景。这一过程有效减少了识别过程中预选区域判定的计算量, 不仅提升了识别精度, 还降低了环境干扰及融合后非目标特征增强对识别结果的影响。

3.5 多尺度目标识别验证

为了验证本文所提出算法在多尺度目标识别方面的效果,对MIMRN的最小可识别目标像素占比及其识别精度进行了测试。本文以FLIR数据集中的运动车辆视频流数据作为识别样本,针对同一目标在不同距离和像素占比的图像,分别采用红外、可见光以及伪红外可见光融合三种频段图像展开识别测试。测试过程从视频中目标车辆完整进入视场后开始,随着目标车辆持续远离,MIMRN会持续对车辆进行双频段融合自适应识别,同时记录识别精度以及车辆像素占比。当出现目标车辆连续3帧识别错误、连续10帧识别波动,或者目标识别精度达到最低识别阈值(60%)的情况时,便判定已达到MIMRN的最小目标识别极限。不同成像频段下目标多尺度识别结果如表2所示。

表2 不同成像频段下目标多尺度识别结果
Table 2 Multi-scale target recognition results of each frequency band

Image spectral band	Minimum recognizable target pixels	Proportion of minimum recognizable target pixels	Minimum target recognition threshold
Infrared	370×232 pixels	7.12%	64.2%
Visible light	452×292 pixels	11.1%	69.7%
Pseudo—infrared and visible light fusion	248×179 pixels	3.21%	55.4%

结合上述测试结果可知,MIMRN通过自适应融合的方式,在特征图层面上深化了目标在红外与可见光双频段上的语义信息,使识别的抗干扰能力更强,进一步提升了目标识别精确性。同时,具有目标兴趣区域标记能力的MIMRN能够有效地针对目标特征与所在位置进行精确识别,提高了算法对目标所在位置的锚定能力,降低了非目标区域对识别精度的影响。

3.6 算法横向对比

为验证MIMRN的综合性能,现将所提及的部分红外-可见光融合算法与MIMRN进行对比,以评判本文方法在双频融合领域的提升效果。由于其他参考算法仅涉及了融合或特征提取策略,尚未构成完整的端到端学习、训练与目标识别能力,故将该类算法的核心部分迁移至MINMRN的等效计算环节中,构建相同的数据计算流程与结构,则各算法与MIMRN的横向测试结果对比如表4所示。

表4 各算法测试结果对比
Table.4 Comparison of test results of each algorithm

Algorithm	Algorithm function and structural principle	Average recognition rate	Average computation speed
Reference[24]	Feature matching and fusion	71.3%	35fps
Reference[25]	Filter—based fusion	78.1%	37fps
Reference[26]	Filter—based fusion	76.2%	32fps
Reference[27]	Neural network fusion	79.8%	22fps
Reference[28]	Neural network fusion	81.1%	23fps
MIMRN	Feature extraction, fusion and recognition	81.9%	25fps

从表4的数据可以看出,MIMRN在整体结构上表现出较好的计算精度与稳定性。因采用了多源图像融合网络构架,MIMRN的平均计算速度相较于其他算法略低,但依旧满足实时解算的速度要求,且在目标准确率上表现更优。面对复杂环境下热源目标的识别任务时,MIMRN可借助自适应融合策略,获取目标在双频成像下的高维特征,同时利用融合迁移图像对纹理进行细化,有效减少噪声在识别过程中产生的干扰,提升了多尺度目标的识别效率。

由表5可知,在FID与LPIPS两项人眼视觉感知评价指标中,MIMRN的数值介于CycleGAN与Pix2Pix之间:其FID低于CycleGAN但高于Pix2Pix,LPIPS同样优于CycleGAN而劣于Pix2Pix。造成该现象的原因是MIMRN并未以人眼感知作为图像风格迁移相似度为评判标准,而是以人不在回路下无人设备线上端对目标识别的最终精度为网络训练收敛的标准。不同于传统图像生成网络面向人眼观感进行纹理还原与

表 5 成型网络模拟真实度对比
Table 5 Comparison of Simulation Fidelity of Generative Networks

Algorithm	FID	LPIPS
CycleGAN	88.24	0.16
Pix2Pix	72.12	0.07
MIMRN	78.5	0.11

灰度拟合, MIMRN 聚焦于复杂场景下行人与车辆的热源目标识别任务, 网络训练与特征融合过程更关注目标轮廓、辐射特性、判别性特征的保留与增强, 主动适度牺牲部分图像视觉平滑度以换取目标识别鲁棒性。实验结果表明, 尽管 MIMRN 在 FID 与 LPIPS 指标上有所降低, 但能够有效保留关键目标特征并抑制背景干扰, 提升双光融合驱动下的目标识别任务计算精度。

3.7 网络参数横向优化对比

为量化 LLDBAM 的参数, 对比其对网络整体轻量化的贡献, 在输入特征图尺寸为 $480 \times 480 \times 64$ 的条件下, 以 ResNet50 为基础网络架构, 将其与 SE-Net、CBAM 进行参数量对比, 结果如表 6 所示。

表 6 注意力机制子网络模块参数量对比
Table.6 Comparison of test results of each algorithm

Algorithm	Params
ResNet50+SE	28.09M
ResNet50+CBAM	28.10M
ResNet50+LLDBAM	25.96M

由表 5 可知, 相较于 SE-Net 与 CBAM, LLDBAM 的参数量最小。这一现象主要源于空间注意力分支中采用扩张卷积替代普通卷积扩大感受野, 同时通过优化缩减比 $r=16$ 减少了网络参数, 在保持注意力聚焦能力的同时, 更能满足微型计算终端的算力约束。

3.8 消融性实验

为验证 MIMRN 网络中各核心模块与关键组件的有效性, 本节设计消融性实验。在保持数据集、训练参数、测试环境完全一致的前提下, 通过依次移除生成器子模块、鉴别器、融合模块等关键结构, 定量分析各部分对模型最终识别性能的贡献, 实验结果如表 7 所示。

表 7 MIMRN 消融性实验结果
Table 7 Ablation experimental results of MIMRN

Network module setting	Average recognition rate	Average computation speed
w/o TODCN	76.1%	28fps
w/o LLDBAM	76.8%	27fps
w/o MCRM	75.3%	29fps
w/o LT Conv	75.5%	30fps
w/o LFAD	73.9%	31fps
w/o GRSFM	77.6%	26fps
w/o SPFN	77.1%	27fps
Full MIMRN	81.9%	25fps

由表 7 可知, 剔除任意一个所设计的功能模块后, 模型平均识别率均出现下降。其中去除 LFAD 轻量特征敏感对抗鉴别器后识别性能降幅最大, 表明对抗博弈约束能够有效约束伪红外图像生成质量, 提升跨域特征迁移效果; 生成器内部 TODCN、LLDBAM、MCRM 以及 LTConv 模块, 能够强化目标关键特征提取、抑制背景冗余信息; GRSFM 与 SPFN 融合结构可有效融合可见光高频纹理与红外低频热源特征, 进一步提升多尺度目标识别能力。

在计算速度层面, 删减模块后网络参数量与计算量减少, 推理帧率略有提升, 但识别精度损失明显; 完整的 MIMRN 模型虽帧率略低, 但仍满足线上无人装备端实时检测需求, 同时取得最高平均识别率 81.9%。

综合精度与速度结果表明,本文各功能模块设计合理、增益显著,整体网络在识别精度与实时性之间实现了良好平衡。

4 结论

本文提出了一种集成多模块协同优化的多源图像迁移识别网络,实现了基于可见光图像的伪红外图像生成及融合图像目标识别。采用目标偏移可变形卷积与轻量化双维度瓶颈注意力机制,通过灵活采样与双维度注意力加权,提取更具判别力的特征、抑制无关信息干扰,增强模型对图像变化的鲁棒性;设计轻量化双阶段卷积并采用二元交叉熵对数损失函数,优化训练过程中生成器与鉴别器的博弈平衡,提升高频边缘与低频渐变信息的捕捉效果。实验结果表明,该网络在降低低照度环境下目标识别成本的同时,强化了低成本武器平台的全天候目标侦察与探测能力,为无人武器装备的实战效能提升提供了有效技术支撑。

参考文献

- [1] LIU Zuoli, CUI Shoujun. Evolution of unmanned weapons and its implications for asymmetric security[J]. *Journal of International Security Studies*, 2023, 41(2): 23-48,157-158.
刘祚黎,崔守军. 无人武器变革及其对非对称安全关系的影响[J]. *国际安全研究*, 2023, 41(2): 23-48,157-158.
- [2] COCKBURN A. Kill chain: the rise of the high tech assassins[M]. New York: Henry Holt, 2015.
- [3] ZHANG F, FAN H, WANG K, et al. Research on intelligent target recognition integrated with knowledge[J]. *IEEE Access*, 2021, 9: 137107-137115.
- [4] KHAN S, MAZHAR T, SHAHZAD T, et al. A systematic literature review on the applications, models, limitations, and future directions of generative adversarial networks[J]. *IET Circuits, Devices & Systems*, 2025, 2025(1): 5384331. DOI: 10.1049/cdt2/5384331.
- [5] TU Hangyao, WANG Wanliang, CHEN Jiacheng, et al. A survey of image translation based on conditional generative adversarial networks[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2024, 36(1): 14-32.
屠杭垚,王万良,陈嘉诚,等. 基于条件生成对抗网络的图像翻译综述[J]. *计算机辅助设计与图形学学报*, 2024, 36(1): 14-32.
- [6] ZHOU X, ZHENG Y, YANG J. Bridging the metrics gap in image style transfer: a comprehensive survey of models and criteria[J]. *Neurocomputing*, 2025, 624: 129430.
- [7] WANG I, OH T. Design and experimental research of on-device style transfer models for mobile environments[J]. *Scientific Reports*, 2025, 15: 13724.
- [8] CHEN J, CHEN S, WEE L, et al. Deep learning based unpaired image-to-image translation applications for medical physics: a systematic review[J]. *Physics in Medicine & Biology*, 2023, 68(5): 05TR01.
- [9] ZHOU L, ZHANG T. AttCST: attention improves style transfer via contrastive learning[J]. *The Visual Computer*, 2023, 39(12): 6371-6383.
- [10] CHONG M J, FORSYTH D. JoJoGAN: one shot face stylization[C]//*Computer Vision-ECCV 2022*. Cham: Springer Nature Switzerland, 2022: 128-152.
- [11] MEN Y, YAO Y, CUI M, et al. DCT-Net: domain-calibrated translation for portrait stylization[J]. *ACM Transactions on Graphics*, 2022, 41(4): 1-9.
- [12] LI N, WANG H, ZHAO H, et al. Cross-modal visible-to-infrared image translation in remote sensing guided by thermal features[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-16.
- [13] ZHANG Xuhui. Infrared-visible image fusion and target detection method in complex sea surface scenes[D]. Guangzhou: Guangdong University of Technology, 2025.
张旭辉. 海面复杂场景红外-可见光图像融合及目标检测方法[D]. 广州: 广东工业大学, 2025.
- [14] HAN G, SHANG C, TAO Z, et al. Intelligent identification of micro-NPR bolt shear deformation based on modular convolutional neural network[J]. *Sensors*, 2025, 26(1): 184.
- [15] SU Yi, HU Boran, WEN Yan, et al. Visible-infrared image translation algorithm based on conditional generative adversarial networks[J]. *Infrared and Laser Engineering*, 2025, 54(12): 244-257.
苏毅,胡淳然,温焱,等. 基于条件生成对抗网络的可见光-红外图像转换算法[J]. *红外与激光工程*, 2025, 54(12): 244-257.
- [16] JinggangLYU, GAO Shuo, LI Yuzhi, et al. Facial expression recognition with channel attention guided global-local semantic collaboration[J]. *Computer Science*, 2026, 53(1): 195-205.
吕景刚,高硕,李玉芝,等. 通道注意力指导全局-局部语义协同的表情识别[J]. *计算机科学*, 2026, 53(1): 195-205.
- [17] LIU Xingke, YANG Qian, LUO Jianqiao, et al. Branch segmentation of pepper image based on spatial and channel attention[J]. *Computer Applications and Software*, 2025, 42(3): 202-208,243.

- 刘兴科, 杨前, 罗建桥, 等. 基于空间和通道注意力的花椒图像枝条分割[J]. 计算机应用与软件, 2025, 42(3): 202-208, 243.
- [18] ARORA S, GE R, LIANG Y, et al. Generalization and equilibrium in generative adversarial nets (GANs)[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney: PMLR, 2017: 224-232.
- [19] CHANG Y, MA D, DING Y, et al. Lightweight design for infrared dim and small target detection in complex environments[J]. Remote Sensing, 2024, 16(20): 3761. DOI: 10.3390/rs16203761.
- [20] HUANG Qidong. Research on adversarial robustness enhancement methods for vision foundation models[D]. Hefei: University of Science and Technology of China, 2025.
- 黄启栋. 面向视觉基础模型的对抗鲁棒性增强方法研究[D]. 合肥: 中国科学技术大学, 2025.
- [21] TANG Lei, PAN Bailin, LONG Bo, et al. Image domain translation algorithm based on generative adversarial networks[J]. Infrared Technology, 2025, 47(12): 1567-1575.
- 汤磊, 潘柏林, 龙波, 等. 基于生成对抗网络的图像域转换算法[J]. 红外技术, 2025, 47(12): 1567-1575.
- [22] PASZKE A, GROSS S, MASSA F, et al. PyTorch: an imperative style, high-performance deep learning library[C]//Advances in Neural Information Processing Systems 32. Red Hook: Curran Associates, 2019: 8024-8035.
- [23] CHEN Xiefa, LI Min, ZHAO Jingyu, et al. Low light target detection algorithm based on feature fusion enhancement[J]. Computer Engineering and Applications, 2025, 61(22): 235-244.
- 陈谢发, 李敏, 赵晶钰, 等. 基于特征融合增强的低光照目标检测算法[J]. 计算机工程与应用, 2025, 61(22): 235-244.
- [24] WANG Ting. Research and application of infrared and visible image fusion[D]. Xi'an: Xi'an University of Technology, 2019.
- 汪廷. 红外图像与可见光图像融合研究与应用[D]. 西安: 西安理工大学, 2019.
- [25] FENG Yufang, YIN Hong, LU Houqing, et al. Infrared and visible image fusion based on improved convolutional neural network[J]. Computer Engineering, 2020, 46(8): 243-249, 257.
- 冯玉芳, 殷宏, 卢厚清, 等. 基于改进卷积神经网络的红外与可见光图像融合[J]. 计算机工程, 2020, 46(08): 243-249+257.
- [26] JEE S, KANG M G. Sensitivity improvement of extremely low light scenes with RGB-NIR multispectral filter array sensor[J]. Sensors, 2019, 19(5): 1256.
- [27] JUNG C K, ZHOU K L, FENG J W. FusionNet: multispectral fusion of RGB and NIR images using two stage convolutional neural networks[J]. IEEE Access, 2020, 8: 23912-23919.
- [28] HUI T, TANG X, LOY C C. LiteFlowNet: a lightweight convolutional neural network for optical flow estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8981-8989.

Research on Multi-source Image Transfer Fusion and Target Recognition Algorithm Based on Generative Adversarial Network

CAO Zhaorui¹, Li Xinye², PAN Qian¹, YE Mingliang²

(1 School of Equipment Engineering, Shenyang Ligong University, Shenyang 110159, Liaoning, China)

(2 School of Mechanical Engineering, Shenyang Ligong University, Shenyang 110159, Liaoning, China)

Abstract: To address the limitations of existing infrared image migration algorithms—namely weak generalization, insufficient detail representation, and structural redundancy—this paper proposes a novel visible-to-infrared image migration method. The research is motivated by the need to extend the spectral detection domain of low-cost visible light systems for unmanned weapons and to enhance low-light visible imaging systems. The primary objective is to achieve robust infrared feature transfer from visible images, thereby enabling accurate multi-scale target recognition under complex all-weather conditions while reducing dependency on physical infrared sensors. A Multi-source Image Migration Recognition Network (MIMRN) is designed, integrating several innovative components. First, a strongly generalized infrared migration generator is constructed, replacing the conventional U-shaped architecture with a multi-level residual connection structure. This modification reduces structural redundancy, accelerates training convergence, and facilitates reaching the Nash equilibrium in adversarial learning. Second, a lightweight feature-sensitive adversarial discriminator is developed, where the number of convolutional layers in the

PatchGAN architecture is adjusted to enhance feature extraction and detail capture. The discriminator is trained using a binary cross-entropy log loss function instead of the traditional L1 norm loss, improving sensitivity to both high-frequency edge information and low-frequency gradient variations. Third, a deformable convolution and a bottleneck attention mechanism are introduced to enable more flexible sampling and to extract more discriminative features, thereby enhancing robustness against various image transformations. Fourth, a gradient residual semantic-aware fusion module is employed to merge the pseudo-infrared images with the original visible images, facilitating precise recognition of multi-scale targets across composite spectral bands. The entire network is optimized to balance the generator - discriminator game during training, improving structural similarity and detail fidelity in the generated pseudo-infrared images. Experimental results demonstrate that the proposed MIMRN effectively generates target-specific infrared physical characteristics from visible images, significantly reducing the perceptual and structural differences between pseudo-infrared and real infrared images. The method notably lowers the cost of target recognition in low-light environments. Under all-weather and complex environmental conditions, the network achieves high computational accuracy and robustness for multi-scale heat-source target recognition tasks. Quantitative evaluations show that training recognition models with pseudo-infrared images yields improved target identification accuracy compared to baseline methods. Furthermore, fusing visible images with pseudo-infrared images provides additional performance enhancement, enabling visible-light cameras to partially compensate for the absence of physical infrared sensors. The approach successfully strengthens the all-weather reconnaissance and detection capabilities of low-cost weapon platforms. This paper presents a multi-source image migration recognition network that integrates multiple collaborative optimization modules to achieve visible-to-pseudo-infrared image generation and fused-image target recognition. By incorporating deformable convolution for adaptive offset sampling and a lightweight two-dimensional bottleneck attention mechanism for discriminative feature weighting, the method extracts more robust features while suppressing irrelevant information interference, thereby enhancing robustness to image variations. The lightweight two-stage convolution design coupled with the binary cross-entropy log loss function effectively balances the adversarial training process between the generator and discriminator, improving the capture of both high-frequency edge details and low-frequency gradient structures. Experimental validation confirms that the proposed network reduces target recognition costs under low-light conditions while significantly enhancing the all-weather reconnaissance and detection capabilities of low-cost weapon platforms. This work provides effective technical support for improving the combat effectiveness of unmanned weapon systems in practical deployment scenarios.

Key words: image fusion; generative adversarial networks; multispectral imaging; target recognition

OCIS Codes: 100.4994; 100.2960; 100.2000; 110.3080

CSTR: 32255.14.gzxb20265508.0810003