

文章编号 1004-924X(2024)15-2439-15

## 用于夜间撒布地雷探测的可见光偏振图像融合

沈康宇<sup>1</sup>, 崔博伦<sup>2</sup>, 吕其峰<sup>1</sup>, 王 驰<sup>1\*</sup>

(1. 上海大学 机电工程与自动化学院, 上海 200444;

2. 北京空间机电研究所, 北京 100094)

**摘要:** 为了克服在夜间光线微弱情况下, 撒布地雷目标与周边地面背景的光谱强度差异较弱的难题, 研究一种端到端的无监督可见光偏振图像融合增强算法。融合图像利用撒布地雷的偏振特性在夜间增强地雷目标的同时, 尽可能地保留场景的细节纹理信息。融合算法网络由特征提取模块、特征融合模块和图像重构模块构成。在特征融合方面, 引入混合注意力机制以加强网络对特征张量中显著信息的提取能力。并通过设计基于像素内容分布的损失函数, 引导融合图像保留更多源图像中显著像素区域特征, 实现网络的端到端输出。针对夜间撒布地雷数据集, 同 7 种主流图像融合方法进行主客观评价, 并在 SSIM, VIF 等 8 项评价指标中表现为同类最优, YOLOv5 模型中经融合增强的图像在地雷检测任务中表现要优于强度图像 (mAP@0.5 领先 8%, mAP@0.5:0.95 领先 11%)。本模型具备先进性, 并对后续地雷目标检测任务有积极影响。

**关键词:** 图像融合; 撒布地雷; 偏振特性; 可见光; 低照度

中图分类号: TN209; O436.3 文献标识码: A doi: 10.37188/OPE.20243215.2439

## Visible-polarized image fusion for nighttime dispersal of mines

SHEN Kangyu<sup>1</sup>, CUI Bolun<sup>2</sup>, LYU Qifeng<sup>1</sup>, WANG Chi<sup>1\*</sup>

(1. School of Mechanical Engineering and Automation, Shanghai University, Shanghai 200444, China;

2. Beijing Institute of Space Mechanics & Electricity, Beijing 100094, China)

\* Corresponding author, E-mail: wangchi@shu.edu.cn

**Abstract:** To address the challenge of weak spectral intensity differences between dispersed mine targets and the surrounding ground in low light conditions at night, an end-to-end unsupervised visible-polarized image fusion enhancement algorithm is explored. This algorithm uses the polarization characteristics of scattered mines to enhance nighttime mine targets while preserving scene texture details. The fusion algorithm network consists of a feature extraction module, a feature fusion module, and an image reconstruction module. A hybrid attention mechanism is incorporated to improve the network's ability to extract significant information from the feature tensor. Additionally, a loss function based on pixel content distribution is designed to ensure the fused image retains prominent pixel features from the source image, enabling end-to-end network output. For the nighttime landmine scattering dataset, evaluations using seven mainstream image fusion methods showed superior performance across eight metrics, including SSIM and VIF.

收稿日期: 2024-05-17; 修订日期: 2024-07-02.

基金项目: 北京市航空智能遥感装备工程技术研究中心开放基金课题 (No. AIRSE20233); 国家重点研发计划资助项目 (No. 2023YFF0722902); 国家自然科学基金资助项目 (No. 62175144)

The fusion-enhanced image in the YOLOv5 model surpassed the intensity image in landmine detection tasks. This model is state-of-the-art and positively impacts subsequent mine detection missions.

**Key words:** image fusion; scattered mines; polarization characteristics; visible light; low illumination

## 1 引言

利用投射的方式将地雷撒布在预定区域,通常称为撒布地雷,具有速度快、灵活性强、机动性高等优点,是实现远距离、大面积布雷的有效手段<sup>[1]</sup>。在夜间低照度复杂环境下<sup>[2]</sup>,由于撒布地雷表面有伪装涂层,地雷与周边地面背景的光谱强度差异较弱,因此,如何有效克服夜间光线微弱情况影响,实现对近地面地雷等目标的增强显示,是地雷探测和伪装揭露技术的研究难点。偏振图像与可见光强度图像往往具有不同的特征属性<sup>[3]</sup>,强度图像通常包含丰富的场景和纹理信息,但成像的明暗容易受光照信息的影响。尤其是在夜间低照度情况下,撒布地雷目标在强度图像中表现为低像素值或者与周围环境像素差异较小的特点,难以被肉眼观察。偏振成像技术往往能检测到很多肉眼不可见的偏振信息,用来表征一个物体的偏振特征,与光强度无关。研究表明<sup>[3-5]</sup>,撒布地雷等人造目标具有不同于环境的偏振度特征,因此,将强度图像同偏振图像融合,能在低照度环境下增强撒布地雷目标图像的同时,尽可能地保留强度图像中的丰富场景信息,融合后的图像更利于目标探测。

传统的图像融合方法大多基于变换域和稀疏表示。基于变换域的融合方法会对源图像进行多尺度分解,按照一定的规则对相同尺度的系数进行融合<sup>[6]</sup>。常用的多尺度分解方法包含金字塔分解和小波变换<sup>[7-8]</sup>。然而,在面对高维信息表示时,传统的小波变换不能很好地表示图像中的边缘方向等信息<sup>[9]</sup>。为了克服这个问题,Candes和Donoho提出曲波变换(Curvelet transform, CVT)来保持边缘的几何特征<sup>[10]</sup>; Do和Vitterli将拉普拉斯金字塔变换同方向滤波器组结合,提出Contourlet变换<sup>[11]</sup>。Arthur<sup>[12]</sup>在Contourlet变换的基础上提出了NSCT(Nonsub-sampled Contourlet)变换,通过使用非下采样金字塔和非下采

样方向滤波器组来代替Contourlet变换中的拉普拉斯金字塔和方向滤波器组,改善了下采样操作引入的空间失配问题。基于稀疏表示(Sparse Representation, SR)的融合方法,旨在从大量训练源图像中学习得到一个过完备的字典,再用过完备字典对源图像进行稀疏编码,得到稀疏表示系数,最后通过对稀疏表示系数融合、解码得到最终的融合图像<sup>[13]</sup>。相比于多尺度分解,该方法具有更高的稳定性,但在信息替代的过程中也容易丢失图像的高频信息,往往结合多尺度变换联合使用<sup>[14]</sup>。

传统的图像融合方法需要人为制定融合规则,不同应用场景的融合效果受具体融合规则的影响,而深度学习因其强大的自学习和特征表示能力,在图像融合领域得以广泛应用。Ma<sup>[15]</sup>将生成对抗网络(Generative Adversarial Network, GAN)应用于图像融合领域,提出一种红外图像同可见光图像端到端融合的无监督网络模型FusionGAN。2020年,Zhang等<sup>[16]</sup>将卷积神经网络(Convolutional Neural Networks, CNN)应用于可见光偏振图像融合中,提出PFNet模型,但在特征融合方面,PFNet只对特征图像进行了简单拼接。Li<sup>[17]</sup>提出的DenseFuse融合模型则引入相加和 $L_1$ 范数两种融合策略。上述融合策略仍需要人为手工设置,不参与模型的训练学习,存在一定的局限性。本文提出一种用于夜间撒布地雷的可见光-偏振图像融合增强算法,该算法的网络模型从强度图像和偏振图像中学习融合图像的端到端映射,并采用无监督的方式进行训练,使得融合图像在夜间增强撒布地雷目标的同时,尽可能地保留图像场景的细节纹理信息。

## 2 原理

### 2.1 网络模型结构设计

算法整体框架结构如图1所示,网络主要由

特征提取模块、特征融合模块和图像重构模块 3 部分组成,能够实现单通道强度图像和偏振度图像的融合。特征提取模块从强度图像和偏振图像中提取图像的深层特征信息,为了提高特征提取的准确性和网络运算速度,两个特征提取模块共享权重参数。每个特征提取模块都包含卷积层(Conv)和梯度稠密块(Sobel Gradient Dense

Block, SGDB)。卷积层使用反射填充来保证特征矩阵大小不变,并使用 ReLU 作为激活函数来提高模型的非线性能力。为了更好地提取源图像中的高对比度区域特征,本文在传统致密块的基础上提出了基于 Sobel 算子的 SGDB,使得特征提取模块在保留浅层特征的同时尽可能地保留浅层特征的梯度信息。

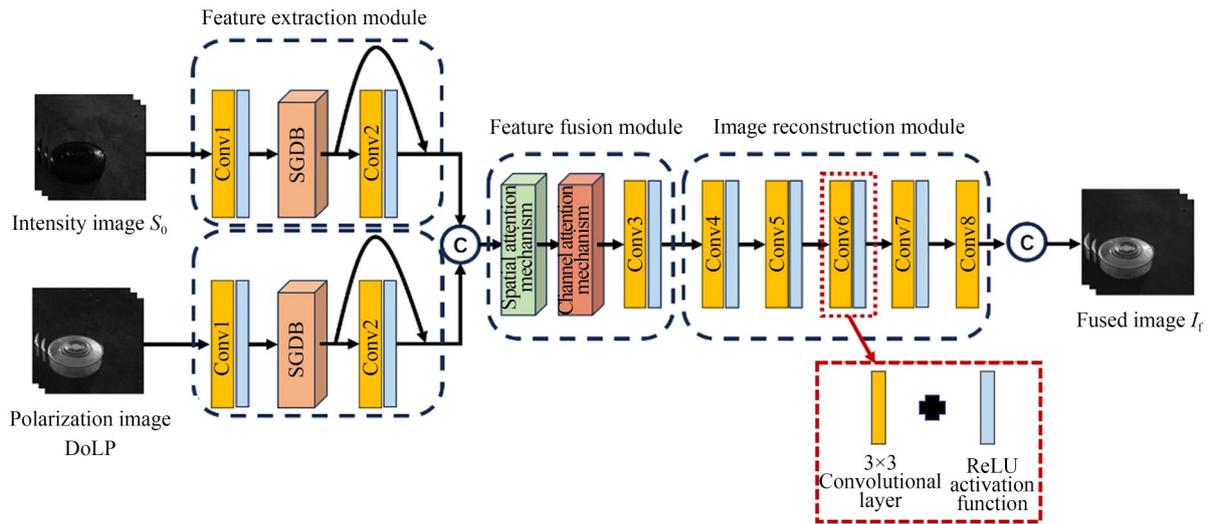


图 1 可见光-偏振图像融合算法的网络模型

Fig. 1 Network model of visible-polarized image fusion algorithm

SGDB 结构如图 2 所示,它由上下两条数据流组成,上部数据流为一个包含两个卷积层的标准致密块,每个卷积层的输出通道数为 16,并以 ReLU 作为卷积层的激活函数。每个卷积层的输出会直接作为输入馈送给下一层,以此尽可能多地保留浅层的特征信息,减少模型的过拟合。下部数据流为一个基于 Sobel 算子的梯度提取块,

用于提取源图像中的梯度信息并直接送入下一层同上层数据流的输出拼接,以保留源图像中高对比度部分的梯度信息。

特征融合模块对送入强度图像和偏振图像的特征张量进行特征过滤和特征融合,突出偏振图像中撒布地雷目标和强度图像中的纹理细节。这需要网络结构能够同时关注偏振图像和强度图像中的显著信息,但在两者图像中显著信息通常分布不一致。因此,本文在特征融合模块中引入了如图 3 所示的空间注意力机制和通道注意力机制<sup>[18]</sup>,来加强对特征张量中显著信息的表示,以提高融合特征中有效信息的利用率。

图像重构模块用于将融合的特征张量重构为最终融合图像。由 5 个卷积层组成,输出通道数分别为 128, 64, 32, 16, 1, 融合特征张量通过这 5 个卷积层正向传播实现通道维度的逐步压缩,最终输出一个具有突出高偏振度目标和清晰纹理细节场景信息的融合图像。

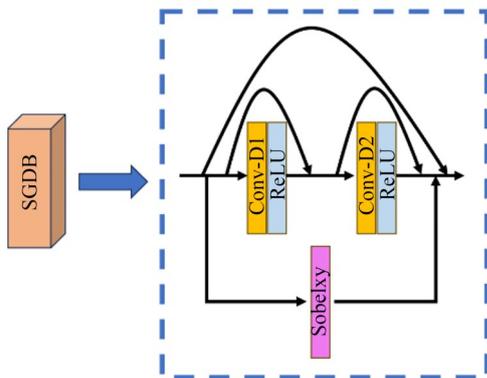
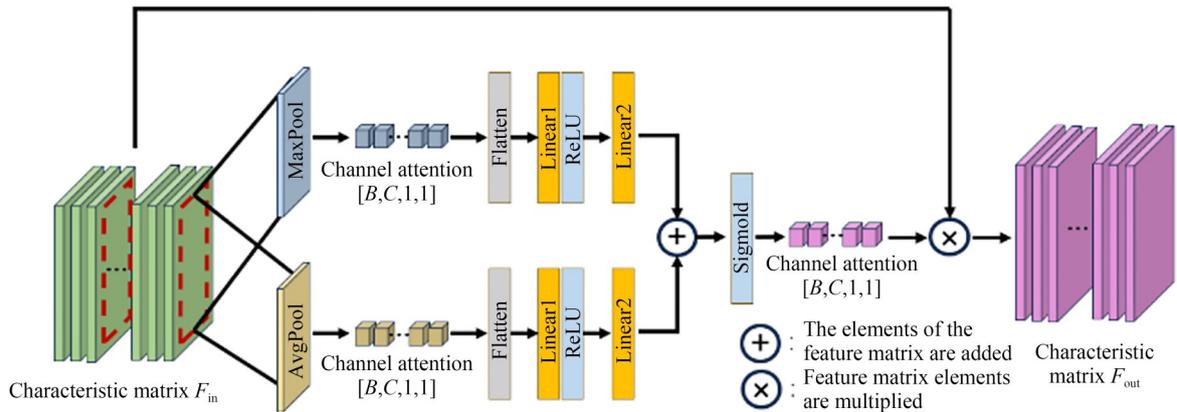
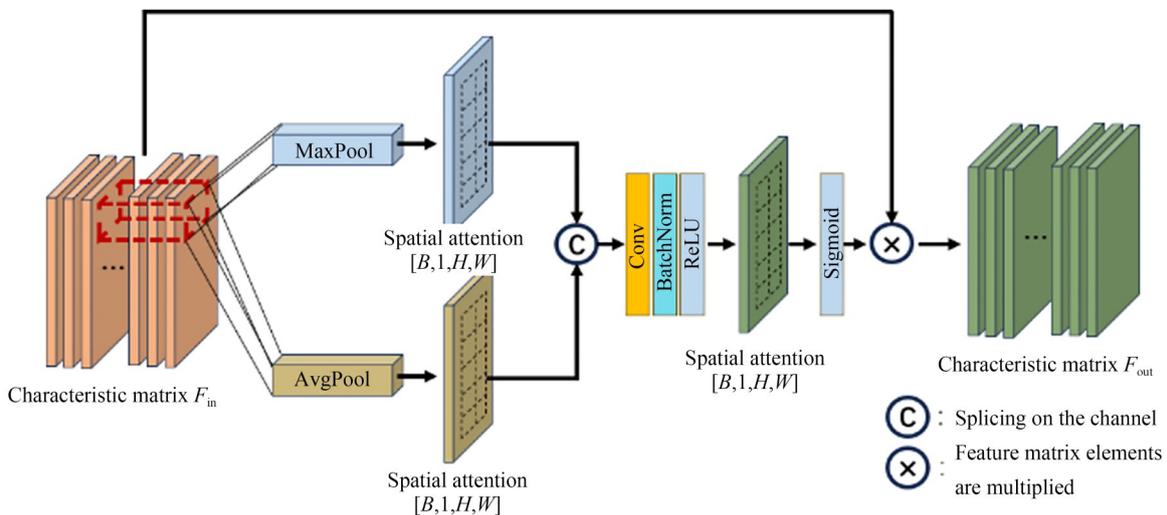


图 2 梯度稠密块网络结构示意图

Fig. 2 Structure of Sobel Gradient Dense Block (SGDB)



(a) 通道注意力机制  
(a) Channel attention mechanisms



(b) 空间注意力机制  
(b) Spatial attention mechanisms

图 3 注意力机制模型结构

Fig. 3 Structure of attention mechanism model

### 2.2 损失函数的设计

为了有效衡量融合图像与目标图像之间的差异,通过反向传播引导网络参数更新。本文从两个角度设计损失函数。一方面,为了使融合图像具有较好的结构特征,采用多尺度多权重结构相似性损失函数,来约束融合图像与源图像在多尺度上的结构相似性;另一方面,针对图像中高像素值区域和高像素梯度值区域,设计一种像素内容损失函数,使融合网络能够有针对性地提取和保留显著目标区域特征。本模型所采用的总

体损失函数定义如下:

$$Loss_{total} = \lambda Loss_{mswssim} + Loss_{pix}, \quad (1)$$

其中: $Loss_{total}$ 为总损失函数, $Loss_{mswssim}$ 为多尺度多权重结构相似性损失函数, $Loss_{pix}$ 为像素强度损失函数, $\lambda$ 是控制两个损失函数之间权重的超参数。

随着距离探测器的距离远近,撒布地雷目标在图像上呈现的目标尺寸也不同,为了提高模型处理不同尺度图像细节的能力,采用多尺度多权重的结构相似性指标(Mutil-scale and weighted SSIM, MSWSSIM)<sup>[19]</sup>作为相关的损失函数,其

定义如下:

$$Loss_{mswssim} = 1 - \frac{1}{5} \cdot \sum_{w \in \{3, 5, 7, 9, 11\}} (\gamma_w \cdot loss_{ssim}(I_{so}, I_f; w) + (1 - \gamma_w) \cdot loss_{ssim}(I_{DoLP}, I_f; w)), \quad (2)$$

其中: $\gamma_w$ 为权重系数; $loss_{ssim}(x, y; w)$ 表示在窗口尺寸为 $w$ 下 $x$ 图像和 $y$ 图像的局部结构相似性指标,具体计算公式定义如下:

$$Loss_{ssim}(x, y; w) = \frac{(2\bar{w}_x \bar{w}_y + C_1)(2\sigma_{w_x w_y} + C_2)}{(\bar{w}_x^2 + \bar{w}_y^2 + C_1)(\sigma_{w_x}^2 + \sigma_{w_y}^2 + C_2)}, \quad (3)$$

其中: $C_1$ 和 $C_2$ 为常数,分别为 $1 \times 10^{-4}$ 和 $9 \times 10^{-4}$ ; $w_x, w_y$ 分别是图像 $x$ 和图像 $y$ 在窗口 $w$ 内的区域, $\bar{w}_x$ 和 $\bar{w}_y$ 分别代表 $w_x$ 及 $w_y$ 区域内的均值; $\sigma_{w_x}^2$ 和 $\sigma_{w_y}^2$ 分别为 $w_x$ 和 $w_y$ 的方差; $\sigma_{w_x w_y}$ 为 $w_x$ 和 $w_y$ 的协方差。

像素内容损失用来约束融合图像对源图像中高像素值区域和高梯度值区域的保留,由强度损失和梯度损失两部分组成,定义如下:

$$Loss_{pix} = Loss_{int} + \alpha Loss_{grad}, \quad (4)$$

其中: $Loss_{int}$ 为强度损失, $Loss_{grad}$ 为梯度损失,超参数 $\alpha$ 用来平衡强度损失和梯度损失。

强度损失从单个像素点上的强度值出发,利用Frobenius范数<sup>[20]</sup>来计算融合图像和目标图像之间的像素差异,约束生成的融合图像同源图像之间的像素强度差异。强度损失的公式如下:

$$L_{int} = \frac{1}{HW} \left\| w_p (I_{DoLP} - I_f) \right\|_F^2 + \frac{1}{HW} \left\| w_1 (I_{S_0} - I_f) \right\|_F^2, \quad (5)$$

其中: $\|\cdot\|_F$ 为Frobenius范数,用于测量两个矩阵之间的距离; $H$ 和 $W$ 分别代表图像的高度和宽度; $I_{DoLP}$ 和 $I_{S_0}$ 分别为输入的偏振图像和强度图像, $I_f$ 为模型预测的融合图像; $w_p$ 和 $w_1$ 分别为两个Frobenius范数相加的权重参数,其值由两个源图像和融合图像之间的像素差异决定,计算公式如下:

$$w_p = \frac{\exp(I_{DoLP}/\tau)}{\exp(I_{DoLP}/\tau) + \exp(I_{S_0}/\tau)}, \quad (6)$$

$$w_1 = 1 - w_p, \quad (7)$$

其中: $\tau$ 作为缩放因子来缩放源图像的像素值之间的差,设置为0.1。当偏振图像的像素值比强度图像的像素值更大时,同像素位置的偏振信息

更显著,即权重系 $w_p$ 更大。

同理,梯度损失则是从像素点同周围的像素梯度差异出发,利用Frobenius范数来计算融合图像和不同源图像之间的像素梯度值差异,公式如下:

$$Loss_{grad} = \frac{1}{HW} \left\| w_{grad,p} (Grad_{DoLP} - Grad_f) \right\|_F^2 + \frac{1}{HW} \left\| w_{grad,1} (Grad_{S_0} - Grad_f) \right\|_F^2, \quad (8)$$

其中: $Grad_{DoLP}$ 和 $Grad_{S_0}$ 为偏振图像和强度图像分别经过Sobel算子后的梯度图像; $Grad_f$ 为模型输出融合图像的梯度图像; $w_{grad,p}$ 和 $w_{grad,1}$ 分别为两个佛罗贝尼乌斯范数相加的权重参数,计算同式(6)和式(7)。

### 3 实验及结果分析

#### 3.1 数据集的建立

为了更好地描述地雷表面偏振特性,引入偏振度(Degree of Linear Polarization, DoLP)参量<sup>[21]</sup>来表征偏振光强度在整体光强度中的比例。DoLP与 $I, Q, U$  3个参量有关,其定义如下:

$$DoLP = \frac{\sqrt{Q^2 + U^2}}{I}. \quad (9)$$

偏振度将光的偏振程度进行了量化,其值域在0~1之间。当 $DoLP=0$ 时,光为完全非偏振光;当 $DoLP=1$ 时,光为完全偏振光;当 $DoLP$ 在0~1之间时,光为部分偏振光。

如图4所示,实验所用撒布地雷为不同材料和种类的雷壳,包括防坦克金属雷壳、防坦克塑料雷壳、防步兵塑料雷壳和防步兵橡胶皮雷壳。偏振图像拍摄装置如图5所示,该装置以多角度测量架为主体,搭载偏振模块、相机模块、数据传输及处理模块,完成对被测目标的偏振光谱特征提取。

相机模块选用基于微光成像技术的PCO.edge4.2sCOMs微光相机,如图6所示,用于采集夜间撒布地雷的强度图像,通过分时成像方式计算图像的偏振度图像。采集时间为晚上19:00~23:00,持续3个月,这期间夜间环境平均照度介于0.1~0.9 Lux之间。采集不同的光照条件、种类、背景、距离和角度的撒布地雷图像共425对,

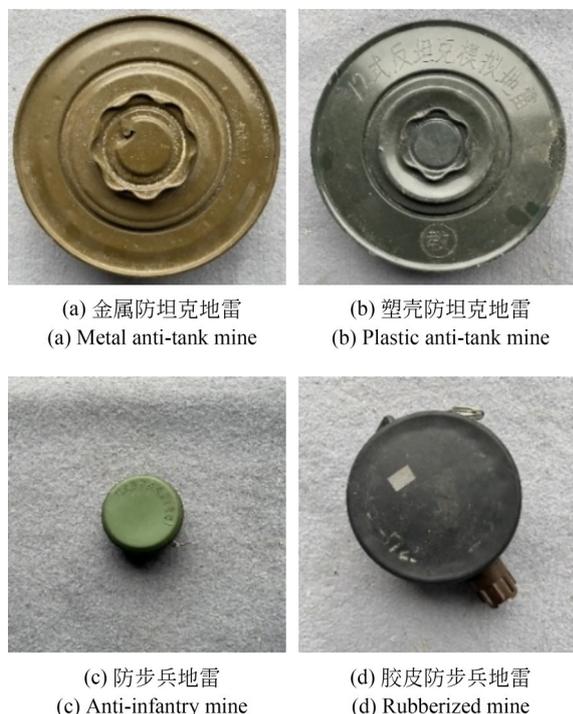


图 4 数据集所用地雷目标

Fig. 4 Targets for mines used in dataset

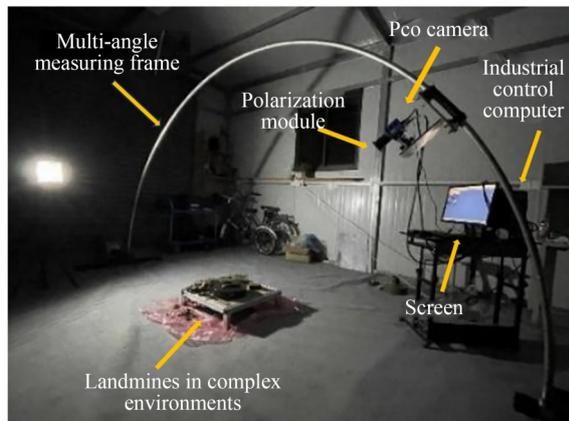


图 5 多角度偏振观测装置

Fig. 5 Multi-angle polarization observation devices

如图 7 所示,每组照片中左侧为强度图像,右侧为偏振图像。所涉及主要地面背景有草类植被、沙土、水泥地等复杂环境。数据集以 6:2:2 的比例划分为训练集、验证集和测试集。

### 3.2 实验设置及超参数选择

所有的强度图像和偏振图像在送入网络训练之前均进行归一化处理,数值位于 $[0,1]$ 区间。为了尽可能避免在反向传播时出现梯度消失或梯度爆炸的问题,网络中所有卷积层采用 Xavier



图 6 微光相机

Fig. 6 Low-light camera

正态分布方式初始化权重参数,偏置参数统一初始化为 0<sup>[22]</sup>;对于全连接层的权重参数以均值为 0、标准差为 0.01 的正态分布方式初始化,偏置参数初始化为 0。

模型批训练的大小设置为 2,训练迭代周期为 200 轮。使用 Adam 优化器来调整网络参数,学习率为 0.001,并在 100 个迭代周期内以 0.99 的速率呈指数衰减。平衡 SSIM 损失函数与像素损失的超参数  $\lambda$  设置为 0.1,平衡像素强度与像素梯度损失的超参数  $\alpha$  设置为 1。所有实验均在 NVIDIA GTX 3080Ti GPU 和 Intel Xeon (R) CPU E5-2686 v4 上进行训练。

### 3.3 对照实验和主观评价

为了验证本文模型的先进性,使用上述融合算法对夜间撒布地雷的偏振图像和可见光强度验证集图像进行融合,融合结果同目前 7 种主流的融合算法结果进行对比,所涉及的对照方法包含 2 种基于多尺度分解的图像融合方法 (CVT<sup>[23]</sup>, NSCT<sup>[24]</sup>)、1 种基于表示稀疏矩阵的融合方法 (SR<sup>[25]</sup>)、1 种综合多尺度和稀疏矩阵的融合方法 (NSCT-SR<sup>[26]</sup>) 以及 3 种基于深度学习的图像融合方法 (FusionGAN<sup>[15]</sup>, PFNet<sup>[16]</sup> 和 DenseFuse<sup>[17]</sup>)。为了保证公平性,这 7 种方法的参数设置均与原文保持一致,其中 DenseFuse 采用  $L_1$ -范数的融合策略。

图 8 为不同方法针对沙土背景下塑壳防坦克地雷目标的强度偏振融合图像。其中,  $S_0$  为目标的强度图像; DoLP 为目标的偏振度图像,即偏振图像。由图 8 可见,本文提出的方法能够使融合

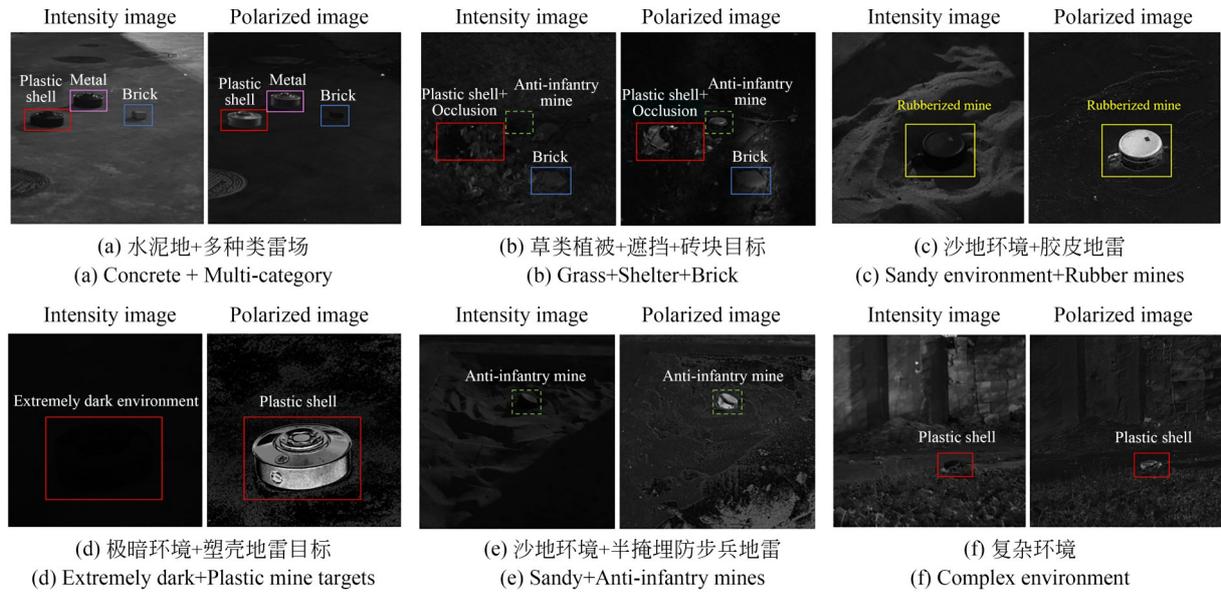


图 7 数据集中地雷目标图像  
Fig. 7 Targets for mines used in datasets

图像在最大程度凸显地雷目标的偏振特征,同时能够保留沙土背景的纹理信息,图像更加符合人眼视觉观测。CVT和NSCT方法在融合图像的过程中均会损失一部分地雷目标的偏振度信息,导致地雷目标成像不显著。SR以及NSCT-SR方法在融合成像过程中,地雷目标存在部分区域失真的情况。FusionGAN以及DenseFusion方法则在此场景下表现较好,能够在保留地雷目标偏振特征的同时,一定程度上增强沙土背景的显

示。PFNet成像则与本文提出方法相近,但是在地雷目标的环状侧面区域存在局部颗粒过曝的现象,融合图像对像素分布约束较差,尤其在极低照度情况下表现更为明显。

多落叶复杂环境下不同方法对塑壳防坦克地雷的强度偏振融合图像如图9所示,强度图像 $S_0$ 中包含丰富的场景纹理信息,能够清晰辨别图像左上侧电线杆上的条纹色彩信息,但是由于地雷目标表面的伪装特性,地雷目标在强度图像中

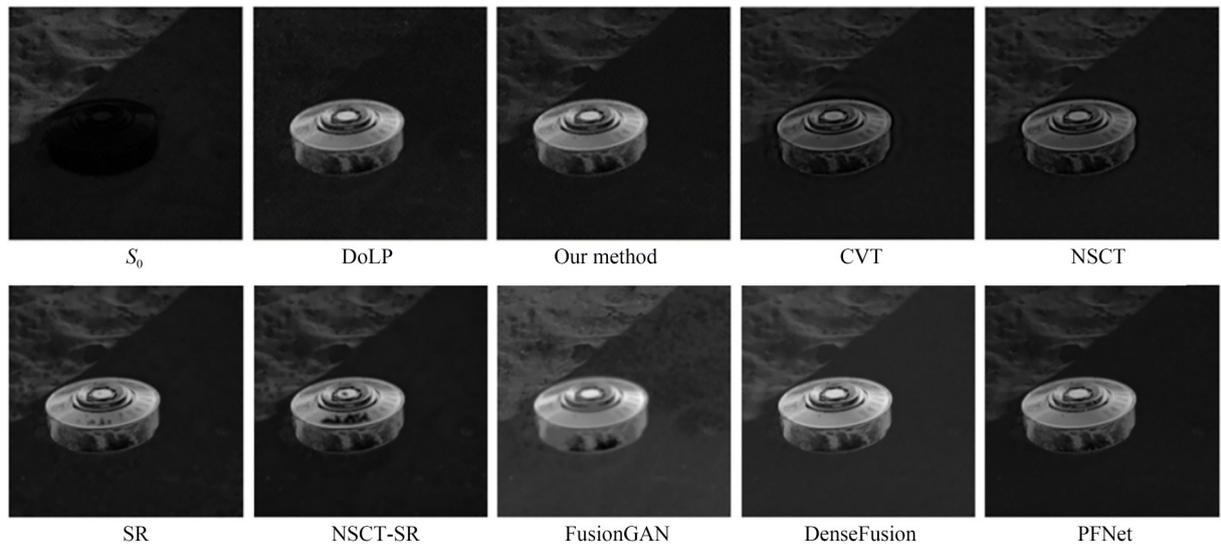


图 8 沙土背景下不同方法的融合图像  
Fig. 8 Fused images of different methods in sandy soil background

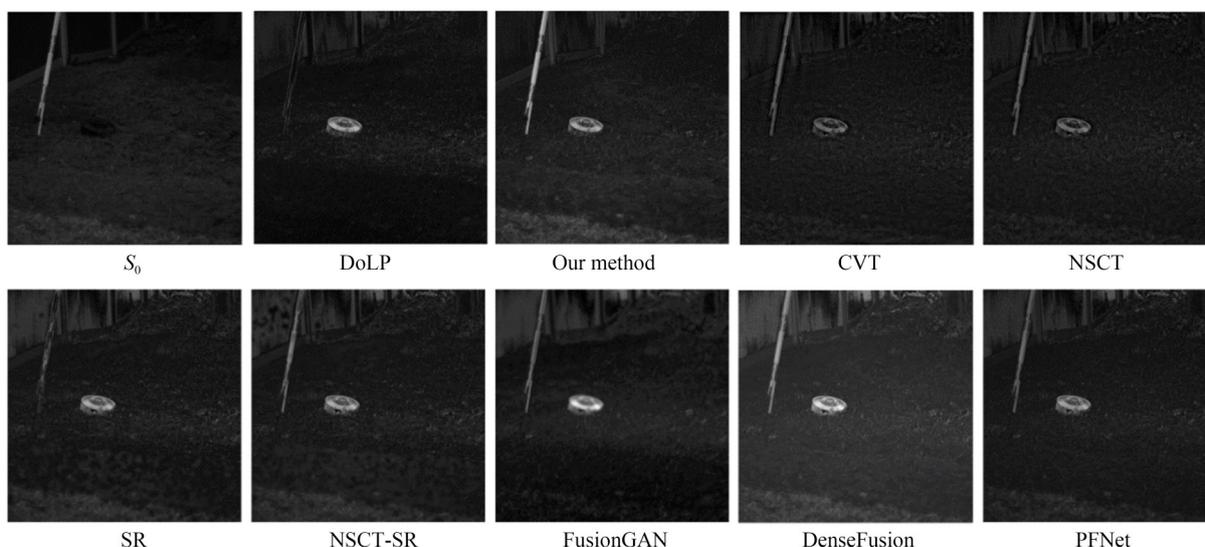


图9 多落叶复杂环境下不同方法的融合图像

Fig. 9 Fused images from different methods in complex environment with multiple deciduous leaves

表现不明显;偏振图像很好地表征了地雷目标表面的偏振特征,但是缺失了电线杆的色彩条纹信息,尤其是场景中的白色区域更为明显。强度图像同偏振图像具有互补特性,该场景的融合图像很好地说明了融合算法的合理性。因此,本文提出的算法能够在增强地雷目标显示的同时,很好地保留电线杆上的色彩条纹信息。从电线杆上的白色条纹区域来看,在该场景下 NSCT-SR 以及 PFNet 方法都在融合过程中以偏振图像为主导,忽略了强度图像中的场景色彩条纹信息。与本文提出的方法相比, DenseFusion 虽然在融合图像中增强了地雷目标,但是其右上侧的墙体背景存在一定程度的色彩失真。

综上,从主观评价角度出发,本文提出的融合方法能够较好保留偏振图像中高偏振度的地雷目标信息和强度图像中光照良好的场景纹理信息,即使在低照度等复杂环境下,融合图像也能在增强地雷目标的同时包含丰富的场景信息,使得融合图像更符合人眼观测。相较于其他图像融合方法,本文提出算法在散布地雷特性检测方面具备先进性。此外,本算法在树叶遮挡、多地雷多目标、沙土半掩埋以及草地中远距离等场景上也具备优良的融合效果。

### 3.4 客观评价

为了客观评价不同方法的融合效果,本文选用 12 种常用图像质量评价指标<sup>[27]</sup>来衡量融合图

像的质量,结果见表 1。其中, MSE 和 PSNR 属于误差敏感指标,侧重于反映融合图像像素点间的差异; SSIM 和 MS-SSIM 是结构相似性指标,从亮度、对比度和结构三方面出发来评价融合图像同源图像之间的相似程度<sup>[28]</sup>; VIF,  $Q_{abf}$  以及 SCD 侧重于人眼的视觉感知方面来评估融合图像<sup>[29]</sup>; MI, EN 用来表征图像信息的丰富程度<sup>[30]</sup>; 统计指标 SD, AG 和 SF 反映了图像像素的离散程度及内容的纹理变化。其中,除了 MSE 为逆向指标,即 MSE 越小表示融合图像质量越好,其余 11 项指标均为正向指标。结果显示,本文提出的方法在 12 项评价指标中 8 项最优 (SSIM, MS-SSIM, EN, AG, SCD, VIF,  $Q_{abf}$ , MI), 1 项次优 (SD)。根据 SSIM 和 MS-SSIM 指标,本方法能够在不同尺度上保持与源图像高度的结构一致性。最优的 VIF 进一步证明了本方法在模拟人眼视觉感知方面的有效性。基于梯度的融合性能指标  $Q_{abf}$  的优异表现揭示了本方法在保留边缘信息和纹理细节方面的显著能力。MI 指标高,表明融合图像与源图像间信息共享的丰富性, EN 指标则暗示融合图像在信息内容方面的丰富度, AG 表征了图像的纹理和边缘信息,本方法在这些指标上也取得了较好的效果。在 PSNR, MSE 上,本方法弱于 SR 类方法,但 SR 类方法在融合图像计算上需要较长的时间。

表 1 不同方法融合图像的客观评价结果

Tab. 1 Objective evaluation results of fused images by different methods

Methods	SSIM	MS-SSIM	PSNR	MSE	EN	SD	AG	SCD	VIF	$Q_{abf}$	MI	SF
Our method	<b>0.758 5</b>	<b>0.781 3</b>	27.84	53.21	<b>5.707</b>	19.68	<b>3.819</b>	<b>1.597</b>	<b>0.837 3</b>	<b>0.606 5</b>	<b>2.689</b>	10.69
CVT	0.737 4	0.779 3	25.97	68.19	5.205	13.44	3.803	1.536	0.709 6	0.545 8	1.444	10.46
NSCT	0.742 4	0.780 7	26.01	67.73	5.203	13.54	3.719	1.571	0.778 9	0.584 6	1.549	10.50
SR	0.723 7	0.739 2	28.92	46.26	5.556	19.56	3.337	1.360	0.710 9	0.473 0	2.801	10.05
NSCT-ST	0.731 2	0.742 6	<b>28.96</b>	<b>46.21</b>	5.605	19.57	3.593	1.411	0.739 2	0.512 6	2.811	10.49
FusionGAN	0.608 0	0.702 3	24.51	91.86	5.544	17.92	2.327	1.175	0.500 6	0.205 3	1.375	5.499
DenseFusion	0.677 2	0.746 3	24.98	86.82	5.703	<b>21.95</b>	3.804	1.461	0.819 3	0.485 2	2.623	11.14
PFNet	0.746 8	0.770 5	27.38	57.55	5.381	19.49	3.711	1.501	0.746 5	0.565 3	2.544	<b>14.55</b>

## 4 消融实验

### 4.1 梯度损失消融实验

像素内容损失  $Loss_{pix}$  的引入,使得融合图像更加注重像素强度和像素梯度的保留。为了进一步评估强度损失  $Loss_{int}$  同梯度损失  $Loss_{grad}$  之间的平衡性能,本文在不引入结构相似性损失  $Loss_{msssim}$  的前提下,仅靠像素内容损失对网络模型参数进行单独学习训练,即总体损失函数如下:

$$Loss_{total} = Loss_{pix} = Loss_{int} + \alpha Loss_{grad}. \quad (10)$$

实验通过改变平衡系数  $\alpha(0, 0.5, 1, 5, 10, 100, 1000)$  来观察梯度损失对于像素内容损失的影响,单次实验的训练迭代周期为 100 轮,以第 100 轮的训练权重参数生成融合图像,并以各项

客观评价指标为依据,对融合图像的质量进行评估,具体参数如表 2 所示,结果保留 4 位有效数字。

由表 2 显示,当梯度损失  $Loss_{grad}$  的平衡系数  $\alpha=1$  时,融合图像质量的 PSNR, MES, EN 及 SF 指标最佳, MS-SSIM, SD, AG, VIF 指标次佳,模型融合图像的综合性能趋于最优。当  $\alpha=100$  时,强度损失  $Loss_{int}$  同梯度损失的负面协同效益达到最大,各项评价指标都趋于最低。由于没有引入 SSIM 损失函数对融合图像的 SSIM 指标进行约束,从数值上看,随着梯度损失的比重上升 ( $\alpha$  变大), SSIM 以及 MS-SSIM 指标呈逐渐下降的趋势,这种情况会随着 SSIM 损失函数的引入而改变。

表 2 平衡系数  $\alpha$  对融合图像质量的影响Tab. 2 Effect of balancing coefficient  $\alpha$  on quality of fused images

$\alpha$	SSIM	MS-SSIM	PSNR	MSE	EN	SD	AG	SCD	VIF	$Q_{abf}$	MI	SF
0	<b>0.755 6</b>	<b>0.784 0</b>	27.18	52.67	5.317	18.44	3.346	1.573	0.799 8	0.545 7	<b>3.082</b>	9.868
0.5	0.746 6	0.775 2	27.86	48.62	5.503	<b>19.61</b>	3.854	1.610	0.910 1	0.645 0	3.004	10.71
1	0.745 1	0.775 4	<b>27.99</b>	<b>47.50</b>	<b>5.512</b>	19.49	3.858	1.640	0.910 4	0.645 7	2.948	<b>10.72</b>
5	0.743 9	0.775 3	27.60	50.59	5.495	19.04	3.838	<b>1.668</b>	<b>0.913 8</b>	0.646 6	3.061	10.71
10	0.739 7	0.774 5	26.78	57.95	5.465	19.03	3.829	1.660	0.909 5	<b>0.646 7</b>	3.003	10.68
100	0.083 9	0.477 6	23.61	109.3	1.810	2.470	0.812	0.410 4	0.030 1	0.066 3	0.493 2	2.579
1000	0.608 2	0.644 5	24.01	101.9	5.334	17.44	<b>3.956</b>	0.532 7	0.758 2	0.631 8	2.712	11.48

为了观察梯度损失对于融合图像的影响,图 10 为不同平衡系数  $\alpha$  下的融合图像。当  $\alpha=0$ , 0.5, 1, 10 时,融合图像整体仍以强度信息为主导;当  $\alpha=100$  时,强度信息和梯度信息的负面协

同效益达到最大,融合图像除了个别形状边缘信息,整体的灰度较低,融合效果较差;当  $\alpha=1000$  时,融合图像整体由源图像的梯度信息为主导,物体及环境的边缘形状在融合图像中表现最为

明显。可见,随着平衡系数 $\alpha$ 的不断增大,梯度损失占总体损失的比重不断增大,源图像的梯度信息对于融合图像的指导效果则越明显。

综上所述,当平衡系数 $\alpha=1$ 时,融合图像能

够在保留源图像的强度信息的同时凸显出源图像的梯度边缘信息,融合图像质量不管从融合指标或者图像成像效果的角度上都趋于最优,因此,本文提出的模型超参数 $\alpha$ 设置为1。

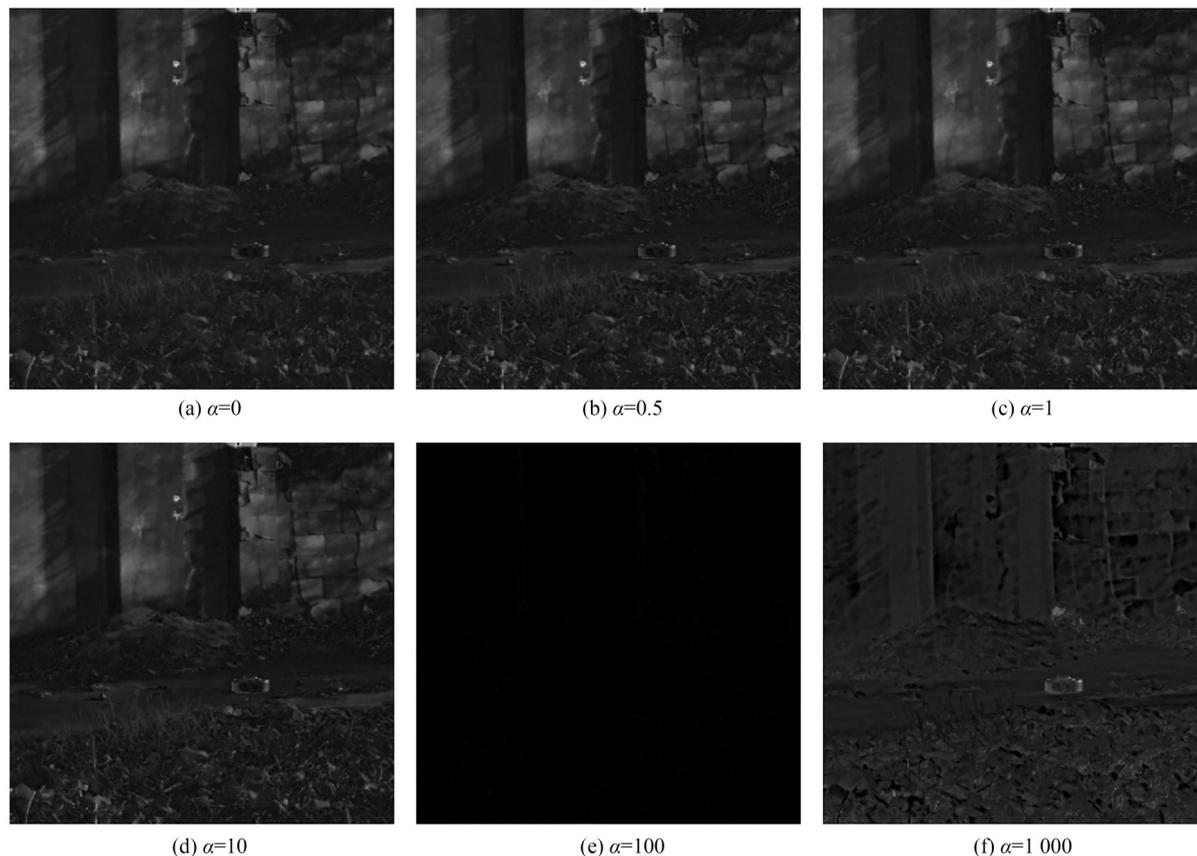


图 10 不同平衡系数 $\alpha$ 下的融合图像

Fig. 10 Fused images with different equilibrium coefficients  $\alpha$

#### 4.2 结构相似性损失消融实验

当平衡系数 $\alpha=1$ 时,融合图像在 PSNR, MSE 等统计评价指标的性能趋于最优,但是在结构相似性方面(SSIM 和 MS-SSIM)仍存在不足,所以在保持最佳像素内容损失 $Loss_{pix}$ 的前提下,通过引入多尺度多权重结构相似性损失 $Loss_{mswssim}$ ,来进一步提高融合图像同源图像在亮度、对比度以及结构方面的相似程度。平衡系数 $\lambda$ 分别为 0, 0.1, 0.5, 1, 10, 100, 1 000 时,模型产生融合图像的客观评价结果如表 3 所示。为了保证实验的一致性和可比性,每次实验的训练迭代周期为 100 轮,并以第 100 轮的训练权重参数生成融合图像。最后,以各项客观图像质量评价指

标作为依据,对融合图像进行评估,结果保留 4 位有效数字。

表 3 显示,随着平衡系数 $\lambda$ 的不断增大,多尺度多权重结构相似性损失函数 $Loss_{mswssim}$ 所占总体损失函数的比重越大,融合图像的 SSIM 表现就越好,但是随着 SSIM 主导模型生成融合图像的比重增加,PSNR, MSE, EN 等 10 项参数均呈下降趋势,与 SSIM, MS-SSIM 形成互斥关系。为了能够生成综合性能最优的融合图像,基于消融实验结果,本文选择平衡系数 $\lambda$ 为 0.1 构建总体损失函数,使得损失函数在原有像素内容分布损失的基础上,引入多尺度多权重结构相似性损失来改善融合图像在结构相似性方面的表现。

表 3 不同平衡系数  $\lambda$  对于融合图像质量的影响

Tab. 3 Effect of different balance coefficients  $\lambda$  on quality of fused images

$\lambda$	SSIM	MS-SSIM	PSNR	MSE	EN	SD	AG	SCD	VIF	$Q_{abf}$	MI	SF
0	0.745 1	0.775 2	27.99	47.50	5.512	19.49	3.858	1.640	0.910 4	0.644 7	2.948	10.71
0.1	0.758 5	0.779 9	27.82	53.55	5.448	19.67	3.812	1.592	0.844 4	0.600 3	2.689	10.72
0.5	0.759 5	0.781 3	27.60	56.74	5.375	19.19	3.770	1.611	0.819 0	0.585 1	2.561	10.63
1	0.759 1	0.779 1	27.75	56.14	5.389	19.66	3.758	1.554	0.813 1	0.576 7	2.515	10.64
10	0.760 4	0.781 5	27.70	57.41	5.329	18.87	3.712	1.574	0.793 1	0.568 7	2.440	10.56
100	<b>0.761 6</b>	<b>0.783 1</b>	27.54	58.34	5.302	18.67	3.695	1.586	0.788 4	0.565 9	2.431	10.51
1 000	0.761 1	0.781 5	27.73	57.32	5.335	19.09	3.707	1.574	0.791 9	0.566 4	2.444	10.54

### 4.3 注意力机制消融实验

为了验证注意力机制的效果,本文在空间注意力机制、通道注意力机制和无注意力机制 3 种情况下对模型进行了单独训练。每次训练的迭代周期为 100 轮,以第 100 轮的训练权重参数生成融合图像,以各项客观评价指标为依据,评估融合图像的质量,并与采用混合注意力机制(通道+空间)模型结果进行对比,具体结果如表 4 所示,结果保留 4 位有效数字。

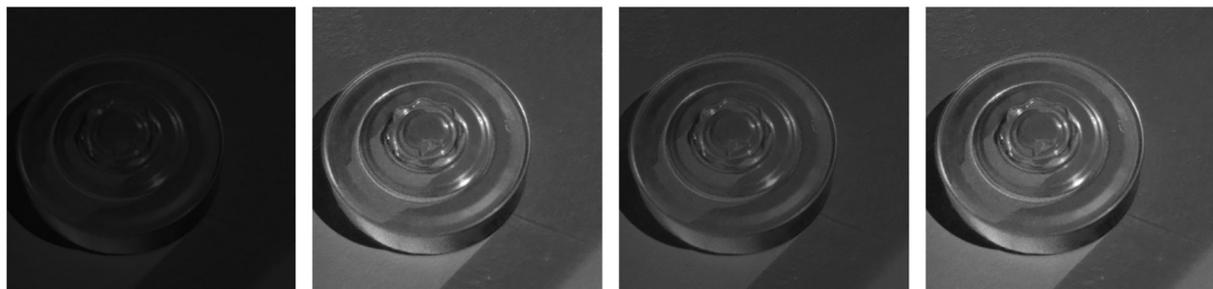
表 4 显示,从客观评价上分析,采用注意力机制的模型在各项指标上要比无注意力机制的模型呈现更高的数值,表明引入注意力机制能

够有效地提升融合图像的质量。其中,采用通道注意力机制的模型在 SSIM,PSNR 等 11 项参数上要远高于空间注意力机制,说明通道注意力机制对本模型的融合效果提升最为明显,这也与特征提取模块中所采用的 SGDB 结构相关,源图像中不同特征信息在通道维度进行拼接,通道注意力机制的引入使得模型能够自适应地调节通道权重,进而强化那些具有显著信息的特征通道。从主观评价角度出发,不同注意力机制下模型生成的融合图像如图 11 所示。结果显示,采用通道注意力和混合注意力机制的模型生成的融合图像具有更高的信息对比

表 4 不同注意力机制下模型生成融合图像的客观评价

Tab. 4 Objective evaluation of model-generated fused images under different attention mechanisms

Fusion module	SSIM	MS-SSIM	PSNR	MSE	EN	SD	AG	SCD	VIF	$Q_{abf}$	MI	SF
Lack of attention	0.516 2	0.738 6	24.43	95.86	3.104	4.154	0.662 4	0.965 2	0.181 0	0.025 1	1.392	2.067
Spatial attention	0.752 4	<b>0.798 6</b>	25.85	70.37	4.863	11.47	2.736 0	1.387 0	0.507 1	0.378 0	1.978	8.059
Channel attention	0.754 4	0.776 6	27.25	57.53	5.408	18.83	<b>3.819 0</b>	1.553 0	0.842 8	0.597 8	2.608	10.66
Mixed attention	<b>0.758 5</b>	0.779 9	<b>27.82</b>	<b>53.55</b>	<b>5.448</b>	<b>19.67</b>	3.812 0	<b>1.592 0</b>	<b>0.844 3</b>	<b>0.600 3</b>	<b>2.689</b>	<b>10.71</b>



(a) 无注意力机制  
(a) No attention

(b) 仅通道注意力  
(b) Channel attention

(c) 仅空间注意力  
(c) Spatial attention

(d) 混合注意力  
(d) Mixed attention

图 11 不同注意力机制下模型生成的融合图像

Fig. 11 Fused images generated by models with different attention mechanisms

度,能够有效地突出地雷这类高偏振特性目标,并在融合图像中高亮显示。

综合上述分析,结合使用空间注意力机制和通道注意力机制,可以让模型在特征提取时更加关注图像中的关键信息,并从多个维度提取更为丰富和有效的特征,进而提升融合图像的质量。

## 5 对夜间撒布地雷检测任务的影响

为进一步探究融合图像对于目标检测的影响,本文选用YOLOv5<sup>[31]</sup>来评估融合图像上的目标检测性能。同时,为了增强模型的泛化能力,对训练集进行随机裁剪、翻转、缩放、旋转等常规数据增广处理,并对训练集和验证集中的地雷目标进行手工标注。分别将未增强的可见光图像和使用本文方法生成的融合图像送入模型训练,迭代轮次设定为200,使用精确率(Precision,  $P$ )、召回率(Recall,  $R$ )和均值平均精度(mean Average Precision, mAP)指标<sup>[32]</sup>来评估检测性能。

未增强的可见光强度图像和增强地雷目标显示的融合图像的目标检测性能评价结果如图12所示, mAP@0.5表示交并比(Intersection over Union, IoU)阈值为0.5时的mAP值, mAP@0.5:0.95表示不同IoU阈值(起始值0.5,终值0.95,步长0.5)下所有mAP的均值。结果表明,在精确率方面,可见光图像与融合图像相差不大(约2%),但在不同IoU阈值下地雷目标的mAP值,融合图像都要优于可见光图像(mAP@0.5领先8%, mAP@0.5:0.95领先11%),说明经过偏振特征增强后的融合图像能够在夜间低照度环境下为检测器提供更丰富的地雷目标信息。在召回率方面,融合图像比可见光图像高12%,表明在检测精确率相近的情况下,基于融合图像训练的模型具有更小的地雷漏检性能。图13为部分测试集中的检测效果,每组图片中,左侧为可见光强度图像的检测结果,右侧为相应融合图像的检测效果。相较于可见光图像,基于融合图像的检测器具有更高的置信度,检测精度大幅度提高。

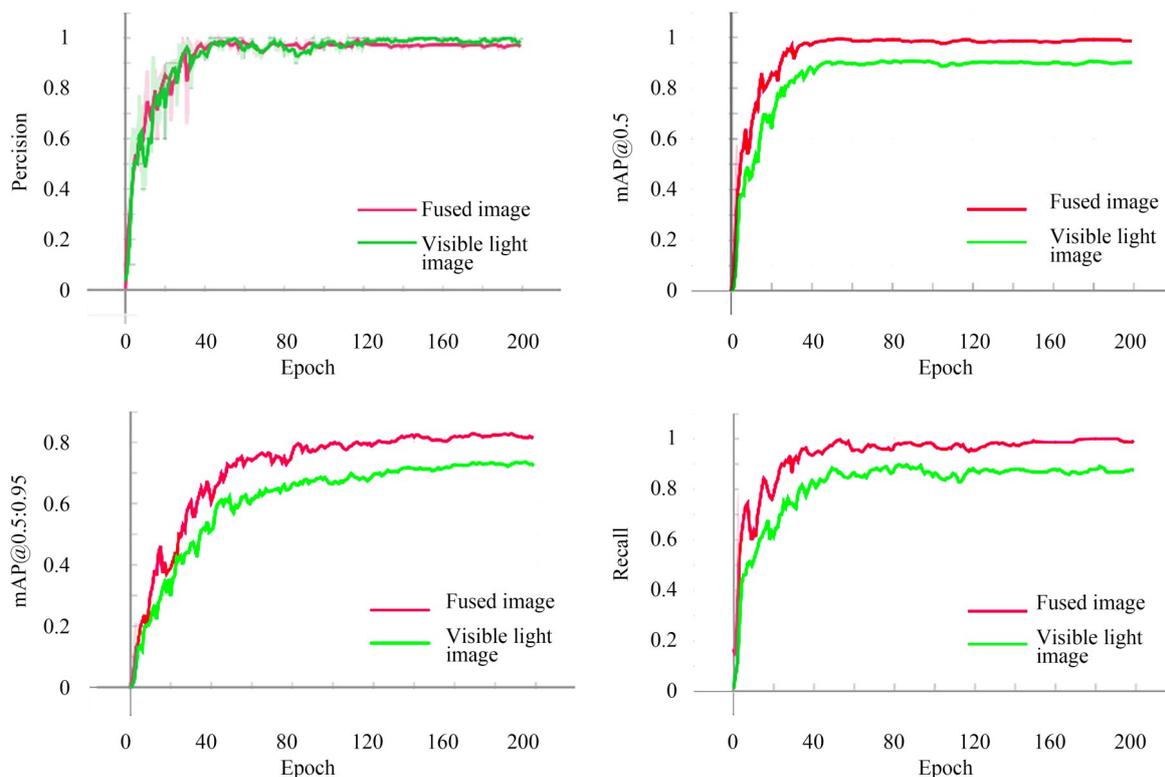


图12 可见光强度图像和融合图像对目标检测性能的影响

Fig. 12 Effect of visible intensity images and fused images on target detection performance

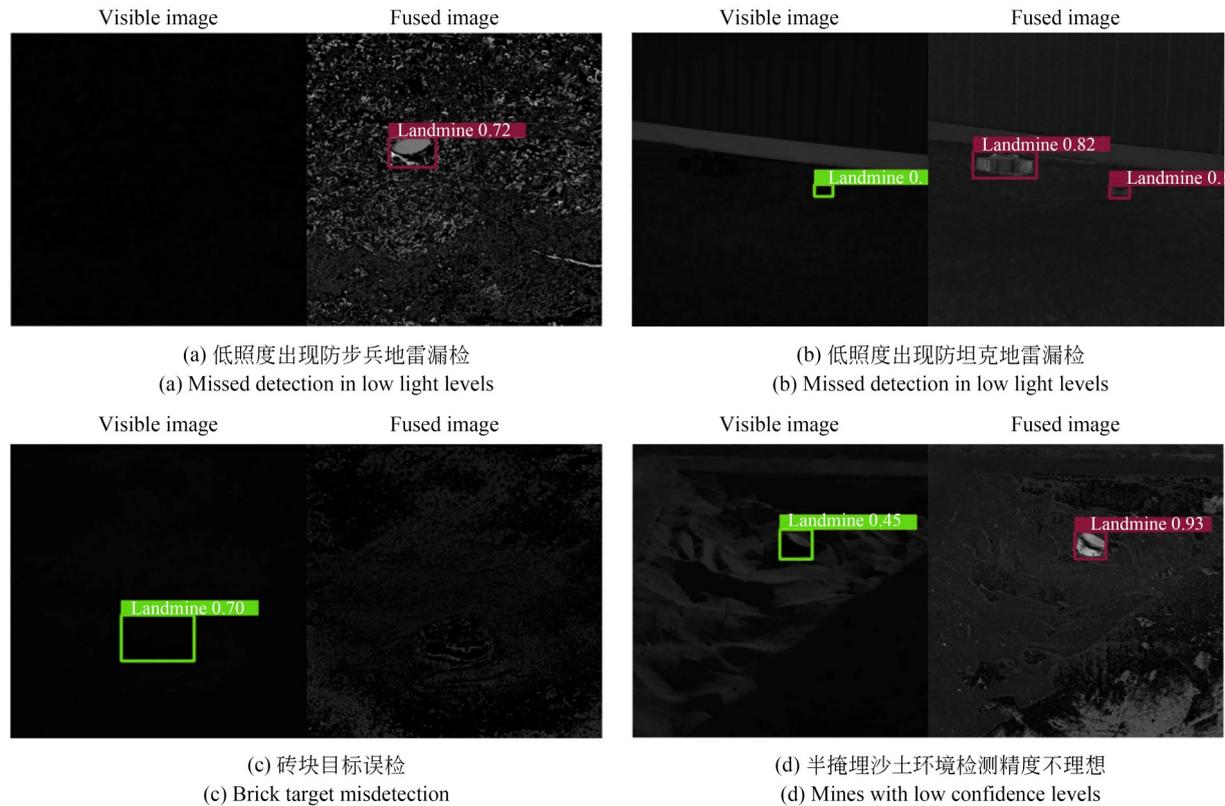


图 13 测试集部分可见光强度图像和融合图像的检测效果

Fig. 13 Detection effect of visible intensity images and fused images in part of the test set

## 6 结 论

本文提出一种端到端的无监督可见光偏振图像融合增强算法,使得融合图像在夜间增强撒布地雷的同时,尽可能地保留图像场景的细节纹理信息。在特征提取方面,基于传统致密块提出了基于 Sobel 算子的 SGDB,该特征提取模块在保留浅层特征的同时尽可能地保留浅层特征的梯度信息。在特征融合方面,本文引入混合注意力机制以加强网络对特征张量中显著信息的提取能力,提高融合特征中有效信息的利用率。此外,本文还设计了一种基于像素内容分布的损失函数,并引入多尺度多权重损失函数,以引导融

合图像保留更多源图像中的高像素值区域和高梯度值区域特征,实现网络的端到端输出。针对夜间低照度撒布地雷数据集,选用 7 种主流的融合算法与本文提出模型进行实验对比,并对各方法的融合图像进行主客观评价。结果表明,相较于其他方法,本文所提出的融合方法对夜间撒布地雷的偏振图像与可见光强度图像有更好的融合效果,能够较好保留偏振图像中高偏振度的地雷目标信息和强度图像中光照良好的场景纹理信息。最后,结合 YOLOv5 模型进一步探究了融合图像对目标检测任务的影响,证明融合图像能够利用地雷偏振特性来增强地雷目标的特征信息显示。

## 参考文献:

- [1] HOWE J D, MILLER M A, BLUMER R V, *et al.* Polarization sensing for target acquisition and mine detection [C]. *SPIE Proceedings, Polarization Analysis, Measurement, and Remote Sensing III*. San Diego, CA. SPIE, 2000: 202-213.
- [2] KYBA C C M, AUBÉ M, BARÁ S, *et al.* Multiple angle observations would benefit visible band remote sensing using night lights[J]. *Journal of Geophysical Research: Atmospheres*, 2022, 127 (12): e2021jd036382.
- [3] LIN Y H, HUANG H H, WANG Y J, *et al.* Im-

- age-based polarization detection and material recognition [J]. *Optics Express*, 2022, 30 (22) : 39234-39243.
- [4] ZENG L F, WANG B, XIN G X, *et al.* A fault location method based on polarization analysis for coal mine [J]. *Frontiers in Earth Science*, 2023, 11: 1097314.
- [5] YANG Y B, ZHANG G B, YAO C L, *et al.* Application of induced polarization method in mineral resource exploration [J]. *Sustainability*, 2023, 15 (4) : 3840.
- [6] WANG S F, MENG J, ZHOU Y, *et al.* Polarization image fusion algorithm using NSCT and CNN [J]. *Journal of Russian Laser Research*, 2021, 42 (4) : 443-452.
- [7] DEB MAJUMDER B, ROY J K, PADHEE S. Recent advances in multifunctional sensing technology on a perspective of multi-sensor system: a review [J]. *IEEE Sensors Journal*, 2019, 19 (4) : 1204-1214.
- [8] ZHAN L, ZHUANG Y, HUANG L. Infrared and visible images fusion method based on discrete wavelet transform [J]. *Journal of Computers*, 2017, 28 (2) : 57-71.
- [9] MA J Y, MA Y, LI C. Infrared and visible image fusion methods and applications: a survey [J]. *Information Fusion*, 2019, 45: 153-178.
- [10] CANDÈS AND DAVID L DONOHO B E J. *Curvelets: A Surprisingly Effective Nonadaptive Representation for Objects with Edges* [M]. Stanford, Calif. : Dept. of Statistics, Stanford University, 1999
- [11] DO M N, VETTERLI M. The contourlet transform: an efficient directional multiresolution image representation [J]. *IEEE Transactions on Image Processing*, 2005, 14(12) : 2091-2106.
- [12] CUNHA A LDA, ZHOU J P, DO M N. The nonsubsampling contourlet transform: theory, design, and applications [J]. *IEEE Transactions on Image Processing*, 2006, 15(10) : 3089-3101.
- [13] ZHANG Q, LIU Y, BLUM R S, *et al.* Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: a review [J]. *Information Fusion*, 2018, 40: 57-75.
- [14] 朱攀, 刘泽阳, 黄战华. 基于 DTCWT 和稀疏表示的红外偏振与光强图像融合 [J]. *光子学报*, 2017, 46(12) : 1210002.
- ZHU P, LIU Z Y, HUANG ZH H. Infrared polarization and intensity image fusion based on dual-tree complex wavelet transform and sparse representation [J]. *Acta Photonica Sinica*, 2017, 46 (12) : 1210002. (in Chinese)
- [15] MA J Y, YU W, LIANG P W, *et al.* Fusion-GAN: a generative adversarial network for infrared and visible image fusion [J]. *Information Fusion*, 2019, 48: 11-26.
- [16] ZHANG J C, SHAO J B, CHEN J L, *et al.* PF-Net: an unsupervised deep network for polarization image fusion [J]. *Optics Letters*, 2020, 45 (6) : 1507-1510.
- [17] LI H, WU X J. DenseFuse: a fusion approach to infrared and visible images [J]. *IEEE Transactions on Image Processing*, 2019, 28(5) : 2614-2623.
- [18] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional Block Attention Module [M]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 3-19.
- [19] ZHANG J C, SHAO J B, CHEN J L, *et al.* Polarization image fusion with self-learned fusion strategy [J]. *Pattern Recognition*, 2021, 118: 108045.
- [20] BÖTTCHER A, WENZEL D. The Frobenius norm and the commutator [J]. *Linear Algebra and Its Applications*, 2008, 429(8/9) : 1864-1885.
- [21] 巩文哲, 褚金奎, 成昊远, 等. 基于无监督学习和注意力机制的水下偏振图像融合 [J]. *光学精密工程*, 2023, 31(21) : 3212-3220.
- GONG W ZH, CHU J K, CHENG H Y, *et al.* Underwater polarization image fusion based on unsupervised learning and attention mechanisms [J]. *Opt. Precision Eng.*, 2023, 31(21) : 3212-3220. (in Chinese)
- [22] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [C]. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Italy: JMLR Workshop and Conference Proceedings, 2010: 249-256.
- [23] NENCINI F, GARZELLI A, BARONTI S, *et al.* Remote sensing image fusion using the curvelet transform [J]. *Information Fusion*, 2007, 8 (2) : 143-156.
- [24] ZHANG Q, GUO B L. Multifocus image fusion using the nonsubsampling contourlet transform [J]. *Signal Processing*, 2009, 89(7) : 1334-1346.

- [25] LIU Y, CHEN X, WARD R K, *et al.* Image fusion with convolutional sparse representation [J]. *IEEE Signal Processing Letters*, 2016, 23(12): 1882-1886.
- [26] LIU Y, LIU S P, WANG Z F. A general framework for image fusion based on multi-scale transform and sparse representation[J]. *Information Fusion*, 2015, 24: 147-164.
- [27] 杨艳春, 李娇, 王阳萍. 图像融合质量评价方法研究综述[J]. *计算机科学与探索*, 2018, 12(7): 1021-1035.
- YANG Y CH, LI J, WANG Y P. Review of image fusion quality evaluation methods[J]. *Journal of Frontiers of Computer Science and Technology*, 2018, 12(7): 1021-1035. (in Chinese)
- [28] WANG Z, BOVIK A C, SHEIKH H R, *et al.* Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [29] XYDEAS C S, PETROVIC V. Objective image fusion performance measure [J]. *Electronics Letters*, 2000, 36(4): 308.
- [30] MA J Y, YU W, CHEN C, *et al.* Pan-GAN: an unsupervised pan-sharpening method for remote sensing image fusion [J]. *Information Fusion*, 2020, 62: 110-120.
- [31] WU W T, LIU H, LI L L, *et al.* Application of local fully convolutional neural network combined with YOLOv5 algorithm in small target detection of remote sensing image[J]. *PLoS One*, 2021, 16(10): e0259283.
- [32] PARK I, KIM S. Performance indicator survey for object detection[C]. 2020 20th International Conference on Control, Automation and Systems (IC-CAS). October 13-16, 2020. Busan, Korea (South). IEEE, 2020: 284-288.

## 作者简介:



沈康宇(2001—),男,安徽滁州人,硕士研究生,主要从事图像融合的研究。  
E-mail: 595096493@qq.com

## 通讯作者:



王 驰(1982—),男,河南周口人,博士,教授,2009年于天津大学获得博士学位,主要从事精密测试及仪器等方面的研究。E-mail: wangchi@shu.edu.cn